Digital Media in Politics and Society

Lecture Script

Andreas Jungherr

2024-10-12

Table of contents

Welcome and housekeeping

Welcome to the lecture notes to *Digital Media in Politics and Society*. The lecture series is developed and written by Andreas Jungherr, Professor for *Political Science, especially Digital Transformation* at the University of Bamberg, Germany. The lectures are designed for the Master Degree programme in political science at the University of Bamberg and its focus areas *Computational Social Science* and *Governance of Innovative and Complex Technological Systems*.

This course follows the flipped-classroom approach. In class, we will discuss the topic of the respective session and any open questions you might have. In order to profit from these sessions, it is mandatory that you read the notes to the respective session and listen to the lectures. Both will be made available approximately one week before the respective topic is discussed in class. In the final session of the course, there will be an exam testing you on what you have taken away from the class. In preparation for the exam, make sure to study the review questions made available to you on this site.

You can find the script to this lecture on this website.

Copyright

The materials have been produced using Quarto.[1]

[1]To learn more about Quarto books visit https://quarto.org/docs/books.

1. Core terms and (ridiculously) brief history

> "Don't know where we're going, but there's no sense being late". – Quote from the movie *Quigley Down Under* (1990).

One of the most important challenges societies face today is how to make sense of digital media. Mechanisms, conditions, and opportunities of and for their uses remain unclear. The only thing obvious is that a lot of people have different expectations:

Writing in 1996, the early cyber-activist and lyricist for the *Grateful Dead* John Perry Barlow declared the promise of digital media:

> "We will create a civilization of the Mind in Cyberspace. May it be more humane and fair than the world your governments have made before."
>
> Barlow (1996)

In 2017 Facebook CEO Mark Zuckerberg spoke about the promise of connection through digital media:

> "For the past decade, we've focused on making the world more open and connected. We're not done yet and we will continue working to give people a voice and help people connect. But even as we make progress, our society is still divided. So now I believe we have a responsibility to do even more. It's not enough to simply connect the world; we must also work to bring the world closer together."
>
> Zuckerberg (2017)

But the mood has turned darker since those days. In 2021, speaking about the supposed role of digital media in spreading doubts about the vaccine against the novel Corona virus, the President of the United States of America Joe Biden declared:

> "They're killing people."
>
> Kanno-Young and Kang (2021)

And doing some performative thinking about artificial intelligence Tesla CEO Elon Musk mused:

> "I think we should be very careful about artificial intelligence. If I were to guess like what our biggest existential threat is, it's probably that. (…) With artificial intelligence we are summoning the demon."
>
> McFarland (2014)

So what is it? Are digital media a blessing or a course? Let's have a look at the ledger.

Clearly, digital media have brought some good. They have opened up discourses for new voices. They have allowed new parties to emerge and compete in elections. They also have allowed societies, organizations, and people to gain new insights and improve.

But clearly, digital media have also brought a lot of bad. They have allowed challengers to diverse and inclusive societies to compete and mount their attack on open societies. They have allowed people to publish and widely distribute false or misleading information about politics, society, and health. They have also opened up discourse to many discriminatory and downright hateful voices whose only goal seems to be to attack and denigrate those with different beliefs.

In short, digital media are neither uniformly good or bad for society or democracy. They are both. So, the question is, how to capitalize on the good digital media bring while mitigating the bad? This lecture series is here to assist you in this task. The goal of this lecture series is to help you to make sense of digital technology - the changes it brings, the opportunities it provides, and the challenges it presents.

> **Goal**
>
> The goal of this lecture series is to help you to make sense of of digital technology - the changes it brings, the opportunities it provides, and the challenges it presents.

In order to do so, we look at some of the biggest controversies about the uses of digital media in politics and society. We look beyond the headlines and see what scientific evidence is available, how this evidence is produced, and what it does tell us about the role of digital media in politics and society. This podcast will introduce you to the best available evidence on ongoing controversies, enable you to ask better questions on the role of digital media in politics and society, and show you the tools that allow you to answer them.

Here are the main topics, we will be talking about:

- Core terms
- Brief history
- Data
- Algorithms
- Public arena
- Challenge
- Artificial intelligence (AI) and democracy

- Computational social science (CSS)

Digital media are hear to stay. No matter how much some people might wish, there is no way back to a time and politics before. So, we'd better start figuring out how this works.

Like the man says: "I don't know where we're going, but there's no sense bein' late."[1]

Time to get a map!

1.1. Defining digital media

First, we need to be clear about how we are using the term *digital media* going forward. In talking about digital media we need to account both for the technology as well as the institutions and organizations that develop, provide, and maintain digital technology. In our book *Retooling Politics* Gonzalo Rivero, Daniel Gayo-Avello, and I defined the term digital media accordingly as:

> "We refer to institutions and infrastructures that produce and distribute information encoded in binary code. On the one hand, this anchors us with uses of a specific technology: the production, encoding, storing, distributing, decoding, and consumption of information in binary code. On the other hand, it allows us to broadly discuss institutions, organizations, and practices associated with the use of this specific technology."
>
> Jungherr et al. (2020), p. 7-8.

This definition foregrounds that in order to examine the impact of digital media on society, we need to move beyond digital media narrowly understood as technology and account for their societal embedding. At the same time, technological aspects are also important and cannot be neglected. Accordingly, digital media in politics and society can be studied on different levels: We can study digital media by focusing on technology, the affordances they provide to users and providers, user psychology, or societal structures digital media are embedded in, shape, and mutually are shaped by. Focusing on either of these levels alone won't tell the full story of how digital media work in politics and society. Instead, we need to combine findings from different analytical approaches in order to understand where different levels interact, opportunities emerge, or where there are limits on the impact of digital media on politics and society.

For example, it might be that people encounter misleading information in digital communication environments and on the basis of these information learn false facts or even be persuaded by a position misleadingly argued for.[2] This would be a psychological effect.

[1]Quote from the movie *Quigley Down Under* (1990). Director: Simon Wincer. Script by John Hill

[2]For more on disinformation in digital communication environments, see our session on *Targeting, manipulation, and disinformation* later this semester.

At the same time, people also encounter information on television, in newspapers, or in their social environment. As a consequence, they encounter many information directly contradicting the misleading information they encountered online. Accordingly, they can be expected to unlearn the information they picked up. Social structures - news media or social embeddedness - thus counteract the structure digital communication environments and limit the strength of effects only emerging there. To think about digital media in politics and society sensibly, we need to account for both these levels and their interaction. If we neglect this, we will misdiagnose their impact. Only focusing on the psychological effects of digital media, we would overestimate the effect of digital media on society. Only focusing on the structural embeddedness alternatively would mean underestimating their effect. Only the combination of both levels allows us to characterize the role of digital media in politics and society correctly. This will be a recurring motive going forward.

1.2. Characteristics of digital media

In the discussion of digital media's impact on politics and society, four characteristics of digital technology keep reappearing. They are:

- Digitization and digitalization,
- Lowered information costs,
- Interactivity, and
- Networks.

1.2.1. Digitization and Digitalization

The success of digital technology lies in the power of encoding wide varieties of information in an uniform format. In his history of digital technology the information scientists Robin Boast describes this process:

> "(...) what makes the digital, as we use it today, digital is that the combination of ons and offs, in very specific albeit complex ways, encodes information. Over the past 150 years these codes have encoded all types of information, including all of our media. Translating or encoding something, a mediation, into a code of ons and offs - this is digital, and this is the foundation of all digital technology."
>
> Boast (2017), p. 10.

Digitization is the process of transforming analogue information into digital bits and making them thereby subject to computational operations.[3] By digitizing information

[3]On the contrast between *digitization* and *digitalization* see also Brennen and Kreiss (2016).

or processes digital technology can lead to a shift within or between organizations. Early adopters can develop competitive advantages given their uses of digitized information. For example, by digitizing contact information of supporters, a political organization might develop a more efficient way to collect donations and thereby outperform competitors. Digital information might allow an actor or organization to pursue their goals more efficiently, but their goals and behavior remain the same as before.

In contrast, *digitalization* refers to shifts in the goals, behavior, or nature of actors and organizations in order to capitalize on opportunities provided by digital technology. For example, digitizing information and processes might allow a political organization to more efficiently collect donations from supporters. In turn, this might lead to existing organizations transforming in order to better capitalize on these opportunities or even new organizations emerging that are inherently optimized to the opportunities provided by digital media and distinct from previous organizations. Examples for transformations like these can be found in the Obama Presidential campaign 2008 and the activism organization MoveOn.org.[4]

Analytically, digitization allows us to focus on the process and the impact of information being digitized or measured in the first place. Digitalization on the other hand foregrounds the transformative processes within organizations, structures, or fields attempting to capitalize on these new opportunities. We will come back to both these processes throughout the course of this lecture series.

1.2.2. Lowered information costs

Another key feature of digital technology is the massively lowering in the cost of publishing, distributing, accessing, and archiving digital information.[5] By lowering the costs of information, nearly everyone can publish information, political analysis, or commentary online at little cost. Once published, anyone with an internet connection can access this information. As a result, we find ourselves in a situation of information abundance, where we can find seemingly endless content corresponding with our interests or needs.

At first, information abundance might seem like a good thing. More information provides people with the basis to make better decisions and allows them better to control elites and the government. But looking closely, things are more complicated. Information abundance weakens gatekeepers by allowing voices and positions to route around their selection decisions. This can broaden discourse but also threaten it once opponents of open and democratic societies gain a foothold in the public arena. By providing information alternatives to commercial information providers and providing attractive alternatives for ad display, digital media weaken the economic foundations of news production

[4]For examples for digitalization in action within political organizations see Kreiss (2012) for the Obama campaign 2008 and Karpf (2012a) for MoveOn.org.

[5]For a core text on the economic consequences of lowering information costs by digital technology see Shapiro and Varian (1999). For two key texts on the political effects of lowered information costs see Neuman (1991), Bimber (2003).

and news as a business. Finally, information abundance allows people to avoid political news much more effectively than in the past and thereby to opt out from the provision of political information altogether. Taken together, these unexpected consequences of information abundance might harm democracies by weakening the foundations of its underlying information environment. We will come back to these developments in the sessions on the *challenge to institutions* and *discursive power*.

1.2.3. Interactivity

Past media technologies have been characterized as one-to-many broadcast communication systems.[6] Examples include newspapers, radio, or television. A journalist can communicate through print or broadcast media with an audience. This makes it one-to-many communication. This set of media technologies allows a small set of people with access to mass media to communicate with an audience of many people. But these technologies do not allow the audience to speak back. As a consequence, one-to-many communication systems are seen to support existing political, economic, and social power structures.

Digital media provide an alternative to this one-directional form of communication. They provide interactivity by allowing the audience to speak back.[7] Digital media allow people to publish information, for example on websites or on social media profiles. This broadens the set of people able to communicate with large audiences from the comparatively small set of professional editors or journalists. But various features of digital media also allow the audience to speak back, be it through comments on sites directly available to anyone reading the original article or by speaking back on their own websites or social media profiles. Digital media allow many to communicate with many.

While digital media featured opportunities for many-to-many communication from its beginning, interactivity has been discussed most prominently in the context of the so-called Web 2.0. The label Web 2.0 refers to a set of technological developments, user-centric practices, and business models that shifted internet use from information publishing, finding, and consuming toward engagement between users and authors. This enabled realtime exchanges on websites and platforms, thereby opening up the range of active participation for internet users. This technological shift was accompanied with a burst of think-pieces on the new interactive nature of digital technology allowing for *true* dialogue between internet users and political, economic, or social elites. By allowing for greater interactivity, Web 2.0 technology was seen as a leveler to power inequalities between elites and the public.[8] Today interactivity is seen more critically, be it as a tool

[6] For classic theories of mass communication see Laswell (1948), E. Katz and Lazarsfeld (1955).

[7] On the concept of *interactivity* see Quiring (2016). For an optimistic account of what happens to the news once the audience speak back see Rosen (2006).

[8] For examples of the somewhat exuberant hopes regarding interactivity and Web 2.0 see Levine et al. (2000), O'Reilly (2005). For an early critical view of digitally enabled interactivity on websites by politicians see Stromer-Galley (2000).

for elites to merely pretend to be accessible or as a channel for unruly and uncivil user comments threatening to derail political discourse. We will come back to this when we discuss digital media as *challenge to institutions* and *the public arena*.

1.2.4. Networks

The final characteristic of digital media, that we need to discuss is the network.[9] Digital media connect people. This leads to the emergence of digitally connected social networks. Through these networks information travels quickly allowing for the quick and wide distribution of information. But these networks can also be used in order to coordinate people around causes or to spread the word about grievances or participatory opportunities.

The network characteristic of digital media provides alternatives to more established forms of political organization and has been seen by some as decisively supporting political activism. Accordingly, the role of digital media in international protests has been prominently discussed, examples include the events surrounding the Arab Spring of 2010, the Occupy protests of 2011, or the Black Lives Matter protests of 2013 and later. Others have pointed to the weaknesses of digital networks as an alternative to political organizations. While digital networks have been strong in channeling enthusiasm into demonstrations of support online or political action, they have proven weak in translating this energy into lasting political action or initiatives. Declaring the death of political organizations as we know them might thus be premature. We come back to these issues when we will be talking about the *challenge to institutions*.

1.3. A (ridiculously) brief history of digital media

When we talk about digital media today, we mainly talk about their negative effects on society. We lament the power of monopolistic platforms like Amazon, Facebook, or Google. We share scare stories about the supposed impact of Russian bots on US elections. We worry about the climate of political discourse in face of rough and uncivil exchanges online. Wherever we look, digital media appear as a threat. This was not always the case.[10]

In the heady days of the nineteen-nineties - when digital technology was in its fifties, the internet was 21, and the World Wide Web had just seen the light of day - people

[9]For a broad discussion of the role of networks in society see Easley and Kleinberg (2010). For a powerful account of digitally networked activism see Tufekci (2017). For more on the forms of digitally networked activism see Bennett and Segerberg (2013). For the limits of digitally networked activism see Gurri (2018).

[10]For a actor-focused attempt at a history of the digital age and key inventions see Isaacson (2014). For a broad history of Silicon Valley and the economic, political, and social contexts see O'Mara (2019). For an early history of the development of the internet see Hafner and Lyon (1996).

saw the internet predominantly as a medium of liberation and empowerment.[11] The ability to publish websites independently or to communicate freely and pseudonymously on internet fora was seen as an opportunity for people whose voices were previously unrecognized to find each other and get their voices heard. These expectations gained broad traction in society in the early two-thousands. The dot-com crash in March 2000 devaluated the stock of many companies in digital tech. In the attempt of creating new excitement about digital technology, digital entrepreneurs started to talk about the Web 2.0 and associated technological innovations that enabled greater degrees of user interactivity than the previous version of the world wide web. Interaction, connection, and user empowerment became powerful narratives of the time.

But Web 2.0 did not only empower users. If anything it empowered technology companies even more. While the power of the world wide web lay in the ability of users to publish websites without support of companies or institutions, Web 2.0s backbone was provided by a few large companies: Alphabet (formerly Google), Amazon, Apple, and Meta (formerly Facebook) emerged as powerful players achieving all but monopoly status in their chosen field. The power of these companies led to many accounts lionizing the achievements of their founders and CEOs.[12] At the same time, voices critiquing their power and increasing reach into ever more areas of society emerged.

Economists tend to emphasize the economic opportunities emerging from platform business models.[13] By allowing market participants to find each other, platform companies create business opportunities and value, where there was none before. Prototypical examples include Airbnb - a company connecting people looking for a place to stay and those willing to sublet apartments - or Uber - a company connecting people looking for a ride and those willing to provide taxi services.

In contrast, sociologists and media scholars tend to point to the dangers of emerging power imbalances between platform companies and market participants.[14] They warn against the associated powers of platform companies to shape the production conditions and structures of markets they come to dominate. Platform business have also been critiqued for bypassing regulation - such as Airbnb circumventing tourism restrictions and bed-limits set by city governments, or Uber by circumventing labor laws by treating drivers not as employees but instead as independent contractors. Outside Silicon Valley this critical view of platforms and the underlying business models has come to dominate public and regulatory discourse.

[11] For typical accounts of the expected empowering features of digital media see Rheingold (1993), Levine et al. (2000). For an ideational history see Turner (2006). On a sympathetic participant's view of the so-called Web 2.0's history see O'Reilly (2017).

[12] For an account of the major companies see Galloway (2017). On Google see Auletta (2009), Levy (2011). On Amazon see Stone (2013), Stone (2021). On Paypal and its aftershocks see Soni (2022). On Facebook see Levy (2020). On Instagram see Frier (2020). On YouTube see Bergen (2022).

[13] For optimistic accounts of the opportunities provided by platforms see Rochet and Tirole (2006), D. S. Evans and Schmalensee (2016), Parker et al. (2016). On some of the prototypical platform companies like Airbnb or Uber see Stone (2017).

[14] For critical accounts of platforms see van Dijck et al. (2018). For some of the challenges of regulation platform companies like Uber see Thelen (2018).

The critical view of digital media has become so widespread that some have diagnosed a "techlash" of late – a backlash against digital technology.[15] Beside the business practices of platform companies and their dominant market power, actual and perceived transgressions of technology companies or their founders have contributed to this. Apparent negligence by tech companies in reigning in malicious actors like the English consultancy *Cambridge Analytica*, or curbing hate speech, or stopping the spread of misinformation have all contributed to their image in public and among regulators to suffer severely. In fact, the dominant view at the moment seems to be that technology companies are a threat to democracy and healthy societies. Nevermind that the empirical evidence is indicating that the effects of malicious actors like *Cambridge Analytica* is all but neglectable and that the spread of misinformation online appears to be exaggerated. While digital media are clearly an arena where political conflicts of the day are staged, their causal role in the widely diagnosed deterioration of political competition and discourse is far from certain.

The bad news is that digital media probably never were as beneficial for society as the early proponents believed. The good news is, that the flip side to this argument is also likely to be true: Digital media are probably far from so detrimental as feared by their detractors today. While the good was never as good as hoped for, at least the bad is not as bad we sometimes fear.

The more discerning readers are probably just about to point out that this (very) brief history of digital media was actually a (very) brief history of US digital media. As always, the more discerning readers are correct. The discussion of digital media and its role in the world is still predominantly told based on cases from the USA. But this is limiting. While ten years ago one could reasonably make the case that the largest digital technology companies were based in the US, this is no longer the case. For example, in the meantime Chinese companies have grown into powerful providers of digital platforms used all over the world, think TikTok or WeChat. They even have developed into important providers of digital hardware providing the technological backbone for digital communication all over the world, think Huawei.[16] This means two things going forward. For one, the provision and regulation of digital media will become much more subject to geopolitical considerations and competition than it has in the past. But more broadly, it also means that academics need to widen their gaze from a predominant fixation on observations based on Western countries or the USA and account for the uses and effects of digital media in other international contexts.

[15]On the techlash see Hoffmann (2020). On the limited evidence for the negative direct impact of technology companies on the political ills of the day see Acerbi (2020), Mercier (2020), Jungherr et al. (2020), Jungherr and Schroeder (2021). On Theranos and the proverbial fall of grace of a founder see Carreyou (2018). On Facebook see Frenkel and Kang (2021). On Uber see Isaac (2019). On WeWork see E. Brown and Farrell (2021). On Peter Thiel see Chafkin (2021).

[16]On China's tech company ecosystem see Fannin (2019) Tse (2015). On Alibaba see Clark (2016). On AI in China see Lee (2018). On the international reach of digital companies from China see Hillman (2021), Segal (2021).

1.4. Cultures

Already this (very) brief history of digital media shows that many different groups of people have worked on the development, roll-out, application, and regulation of digital media over time. This includes the military and the security apparatus of states, scientists, coders, counter-culture figures, business people, and enthusiasts. The influence of these groups waxes and wanes over time depending on external events, new innovations, or shifts in societal, economic, or political power balances. Each of these groups shares specific interests, sensibilities, and concerns making it distinct from the others. As a consequence, they see digital media, its promises, and associated dangers differently from each other.

It is only natural for military or security specialists to emphasize security threats and demand better tools allowing for greater control of digital communication. It is just as natural for libertarians to shy away from control and instead emphasize opportunities emerging from free and open communication. Business entrepreneurs and economists see the good new business opportunities on digital media bring, while those critical of capitalism in general will chafe at what they perceive as exploitation. There is truth in many of these positions, while each one on its own is limiting and risks misrepresenting digital media and its role in politics and society.

One could tell the (very) brief history of digital media and the shift from enthusiasm to fear in public perception as a story of shifting influence between different societal groups and the respective prominence of their view of digital media. According to this reading, the diagnoses of the role and impact of digital media in society and politics did not change, only the influence of groups and accordingly the prominence of their view of digital media.

The theoretician of the early internet Manuel Castells identifies in a dated but illustrative section of his 2001 book *The Internet Galaxy* four cultures shaping the history of digital media:

> "(...) the techno-meritocratic culture, the hacker culture, the virtual communitarian culture, and the entrepreneurial culture. Together they contribute to an ideology of freedom that is widespread in the Internet world."
>
> Castells (2001), p. 37.

For Castells, internet culture takes it techno-meritocratic element from the engineering and academic roots of the developers and users of the early internet. This element brings the valuation of reputation, openness of arguments to challenge through public exchange, and credit given for originators of ideas or arguments. This culture is currently probably most visible in developer communities around the internet encyclopedia Wikipedia and open source software development.[17]

[17] For Wikipedia and open source culture see Raymond (1999), Benkler (2011).

To the techno-meritocratic element, he adds the culture of hacking. This refers to a culture of continuous tinkering and improving of software within a decentralized community of practice contributing to the development of technological standards and software.[18] It also refers to a culture of challenging authorities by finding, exploiting, or making public weaknesses within software, systems, or processes.

The communitarian influence adds another layer to these highly meritocratic but also competitive and sometimes harsh elements driven by academic, engineering, and hacking influences. The communitarian element is driven by the normative counter cultural influences that were so important among the early adopters of the internet outside of academia.[19] For Castells, this layer of internet culture is formed by the valuation of hierarchy-free, horizontal communication among users, free speech independent of potential censorship by governments or mass media, and the chance for free participation of users contributing to the shared community space.

In addition to these layers that sometimes overlap and sometimes are in conflict, there comes the entrepreneurial and commercial layer of internet culture. For Castells, this represents the expectation of participants in the development of digital technology to be financially rewarded. Here, technological development is not just about the free sharing of technology and insights, as for scientists or hackers, and not about enabling free speech, exchange, and participation for users, as for communitarians, it is about monetizing the development of tools and services.

At different points in time, the tensions between these different layers become evident. In the early naughts, enthusiasm around the social web, or the Web 2.0, seemed to offer a surprising harmony between these elements. Engineers and hackers were working hand in hand to develop new technology, services, and features for digital media. Those, increasing ease of use, and dropping costs of connectivity added new attractions to an ever increasing base of internet users which were drawn online. Entrepreneurs funded by venture capital and driven by the hopes of future riches, were happy to offer their tools and services apparently free of charge. At the same time, they rhetorically aligning themselves with the values of communitarians, emphasizing the importance of dialogue among equal users, conversations, and the participatory power of using digital tools and services in shaking up commercial, social, and political hierarchies. These declarations have been subsequently criticized as hollow gestures with the term *Californian Ideology*.[20] Yet, this critique largely remained academic until a series of shocks made the inherent contradictions among the layers of internet culture apparent to all.

The NSA spying scandal and the involvement of important players in the digital industry made painfully visible the ongoing interconnection between the military /espionage / industrial complex and the internet economy that was inherent from the start of the

[18]For hacker culture see Levy (2010), Thompson (2019).
[19]For virtual communitarianism see Rheingold (1993), Turner (2006).
[20]For a critique of *Californian Ideology* see Barbrook and Cameron (1995).

development of digital technology financed by the US military.[21] The widespread coverage of the scandal led to a sudden increase in public awareness of digital technology allowing for widespread mass surveillance. This raised a painful contradiction to the claims of personal empowerment and free speech routinely associated with the use of digital media.

In early 2018, there emerged another controversy that illustrated the tension between the commercial interests of digital entrepreneurs and the users of their services. The scandal surrounding the massive collection of information on users on Facebook by the consultancy company *Cambridge Analytica* made the public suddenly aware of the amounts of access online platforms routinely allowed third parties to the data of their users. While Facebook's business model had been no secret, the details surrounding *Cambridge Analytica* illustrated for many for the first time the breadth and scale of data third parties were able to access. The emerging controversy made painfully clear that the commercial interests of entrepreneurs did not necessarily align with the communitarian values of early user communities allowing for free expression and exchange. Although continuously present in public statements and press releases these values were apparently mattered little in the business decisions guiding the monetization of digital platforms and tools.

This goes to show that even in the early days, there was not only one way of looking at digital media. Instead there were multiple approaches, multiple groups of people, multiple goals, and multiple norms. Sometimes, these approaches aligned, sometimes they contradicted each other. At times some approaches dominated but then were replaced by others. The way of seeing digital media depends on conditions of the time and which group with their view of digital media currently dominates. More than likely, this also holds for the currently dominante negative view on digital media and its impact on society and politics.

1.5. Review questions

1. Please define "digital media" according to Jungherr et al. (2020).
2. Please list four characteristics of "digital media" discussed in the lecture.
3. Please list the four cultures of the internet as identified by Castells.
4. Please list and discuss the analytical levels by which we can assess the impact of digital media on society. Please present an example on how the interaction of these levels can change our assessment of this impact.
5. Castells lists four cultures of the internet. Please discuss whether you find these concepts still hold, if some are missing, whether and why time has rendered some obsolete.

[21] For background on the NSA spying scandal see E. J. Epstein (2017).

2. Computational social science (CSS)

2.1. What is computational social science?

2.1.1. The promise of computational social science

Reading accounts of computational social science, one cannot help but feel excitement about their transformational potential for the social sciences.

In one of the foundational articles sketching the outlines and potentials of the emerging subfield of computational social science, David Lazer and colleagues wrote in 2009:

> "(…) a computational social science is emerging that leverages the capacity to collect and analyze data with an unprecedented breadth and depth and scale. (…) These vast, emerging data sets on how people interact surely offer qualitatively new perspectives on collective human behavior (…)."
>
> Lazer et al. (2009), p. 722.

More recently, Anastasia Buyalskaya and colleagues reiterate this early optimism:

> "Social science is entering a golden age, marked by the confluence of explosive growth in new data and analytic methods, interdisciplinary approaches, and a recognition that these ingredients are necessary to solve the more challenging problems facing our world."
>
> Buyalskaya et al. (2021), p. 1.

Marc Keuschnigg and colleagues expect that:

> "(…) CSS has the potential to accomplish for sociology what the introduction of econometrics did for economics in the past half century, i.e., to provide the relevant analytical tools and data needed to rigorously address the core questions of the discipline. (…) The new CSS-related data sources and analytical tools provide an excellent fit with a sociological tradition interested primarily in the explanation of networked social systems and their dynamics."
>
> Keuschnigg et al. (2018), p. 8.

These are just three examples but many accounts of computational social science voice similar optimism regarding the promise they expect CSS to hold for the study of societies and human behavior. These promises usually come in two forms: The first promise focuses on the increased coverage of social phenomena and human behavior through digital trace data and digital sensors, the second goes even further and expects a transformation of the nature of the social sciences.

On the most fundamental level, CSS can be understood as a response to the growing availability of data. The ever more intensive and diverse use of digital technology creates a constantly growing reservoir of data that documents individual behavior and social life in ever higher resolution. Lazer and colleagues describe this potential already in 2009:

> "Each of these transactions leaves digital traces that can be compiled into comprehensive pictures of both individual and group behavior, with the potential to transform our understanding of our lives, organizations, and societies."
>
> Lazer et al. (2009), p. 721.

Digital technology provides new types of data, offers new and broader approaches for the measurement of the world and social life through sensors and devices, and through digitization makes available and computable vast amounts of previously collected data that up until now could only be analyzed within the limits of their analogue form. The enthusiasm seems therefor more than justified.

First, any interaction of users with online services creates data traces, hence the term *digital trace data*.[1] In principle, digital trace data provide a comprehensive account of user behavior with, and mediated by, digital services. This makes these data highly promising for social scientists, since they promise to provide a comprehensive account of those behavioral and social phenomena that happen on digital services or are mediated by them. Examples for phenomena like these are interaction patterns in political talk online, public interactions with news on digital media, or digital political activism. Additionally, behavioral and social phenomena not primarily associated with digital media but connected to them become also visible in digital trace data. Examples include the analysis of suspected trends in political polarization or extremism based on political talk online or interaction patterns between users on digital media, the mapping of cultural trends based on content on digital media, or discursive power in digital and traditional media.

But, careful! The operative words in the paragraph above are *in principle*. In practice, most digital traces remain out of reach of most researchers. While digital media companies have access to vast troves of digital traces emerging from the uses of their digital services, researchers have only access to highly limited slices of these data that companies choose to make available to them. This can either happen through dedicated

[1] For more on digital trace data see Howison et al. (2011); Golder and Macy (2014); Jungherr (2015); Salganik (2018).

programming interfaces, API, or through exclusive agreements between companies and select researchers for privileged access. This limits the realization of the promise of digital trace data in the social sciences considerably and raises severe practical and ethical concerns in the use of these data. We will come back to this later.

Second, digital technology also extends the number and reach of sensors measuring the world. This could be data emerging as a byproduct of another service, like satellite imagery, or the output of sensors specifically designed by researchers.[2] In principle, this data type is only bound to increase with the availability and wide distribution of Internet of Things devices. Yet, this expected wealth of data reinforces important questions of people's privacy in a world of all-seeing, all-sensing digital devices and the legitimacy of data access for academics, researchers, and industry.

Finally, digital technology also provides new perspectives and opportunities for the work with data available in analogue form. By digitizing existing data sets, researchers can deploy new approaches and methods to already existing data sets.[3] This promises new perspectives to old questions by making these analogue data sets available to analytical approaches provided by computation.

This massive increase in the number and diversity of available data sources extends the reach of social scientists. We can expect to cover more social phenomena and more of human behavior in greater detail and wider breadth. This can offer us a window to new questions and phenomena, as well as enabling us to examine well-known phenomena from a different vantage point. This might also allow social scientists to get a better systems-level view of society and human behavior. This has led some to expect computational social science to contribute to a transformation of the social sciences in general.

For some scholars, the availability of vast data sets documenting human behavior has inspired the hope that the social sciences might transcend their status of a "soft" science into an "actual" scientific discipline.[4] In other words, a discipline with models allowing for the confident prediction of the future. In this view, more data do not only mean an increase of the coverage of social processes or human behavior but actually would allow for a "measurement revolution" (Watts, 2011) in the social sciences. Thus, social science might transcend its current state of after-the-fact explanation and evolve into a science with true predictive power. This hope rests on a view of society as being shaped by underlying context-independent laws that have mostly remained invisible to scientists due to the lack of opportunities to acquire data that can now be accessed. As with most ambitious dreams, the realization of the transformation of the social sciences seems far off.[5]

[2] For more on sensor data see Pentland (2008); Stopczynski et al. (2014).

[3] For more on the potential of digitized data corpora see Piper (2018); Underwood (2019); Cirone and Spirling (2021).

[4] For accounts of coming transformations see Watts (2011); Hofman et al. (2017); González-Bailón (2017).

[5] For why context-dependency is not a bug but a feature of the social sciences see Flyvbjerk (2001); Elster (2007/2015); Gerring (2001/2012).

We can find many studies that illustrate the first promise of computational social science. Increases in data documenting social phenomena and human behavior are significantly extending the tool-box available to social scientists. Here, CSS is proving to be a success and to become ever more important as access to data and knowledge about computational methods increase and diffuse among social scientists. The second, expectation (which you can either take as a promise or a thread depending on your faith-based affiliations) of a transformation of social science into a more strictly predictive science remains unfulfilled as of yet. While the faithful might be tempted to treat this as an indicator that we simply need even more data, it might be more plausible that the nature of the social sciences resists this sort of transformation. The subject of social science is the examination of context-dependent phenomena. This makes prediction in the social sciences an instrument of theory-testing and not an instrument of planning and design, as for example in engineering or physics. While CSS might increase the reach and grasp of social scientists, it does not necessarily make us into socio-physicists, nor is it a tragedy if it won't.

But what is computational social science, besides it providing social scientists with new data?

2.1.2. Computational social science: A definition

While it is true, that digitally induced data riches were a decisive factor in the establishment of computational social science, CSS is more than the computational analysis of digital data. Sure early work in CSS might have spend more time and enthusiasm in the counting of digital metrics and the charting of new data sets than strictly necessary. Also, this somewhat limited activity combined with the hardly contained exuberance of some early proponents of CSS might have given rise to the caricature of CSS as a somewhat complicated effort at counting social media data. More generally, it is limiting to focus definitions of CSS on specific topical subfields. It is true that much early work in CSS focused on digital communication environments. But this is more an artefact of early availability and accessibility of data sets documenting user behavior on social media – especially Facebook and Twitter – than a constitutive feature of CSS. Instead, CSS is the scientific examination of society with digital data sets and computational methods. This can extend to the examination of digitally enabled phenomena but does not have to stop there.

For one, far more and more diverse data sets are now available than in the early days of computational social science, ten years ago. As a result, current research in CSS no longer works primarily with social media data, but instead uses far more diverse datasets. Examples include large text corpora documenting news reporting or literature, historical and current parliamentary speeches, as well as image or video data. At the same time, historical data records are increasingly being made digitally accessible and provide rich opportunities in the social sciences. Also, there is growing awareness among practitioners of computational social science for the need of providing stronger connections between

CSS studies and social science theory. This holds for connections to established theories as well as the development of new theoretical accounts.[6]

In order to characterize computational social science, exclusively data- and method-centric definitions of CSS are therefore too one-sided and consequently outdated. In 2021 Yannis Theocharis and I suggested a definition of CSS, taking current developments into account while also foregrounding what differentiates CSS from other approaches in the social sciences.

> **Defintion: Computational social science**
>
> "We define computational social science as an interdisciplinary scientific field in which contributions develop and test theories or provide systematic descriptions of human, organizational, and institutional behavior through the use of computational methods and practices. On the most basic level, this can mean the use of standardized computational methods on well-structured datasets (e.g., applying an off-the-shelf dictionary to calculate how often specific words are used in hundreds of political speeches), or at more advanced levels the development or extensive modification of specific software solutions dedicated to solving analytically intensive problems (e.g., from developing dedicated software solutions for the automated collection and preparation of large unstructured datasets to writing code for performing simulations)." (Theocharis & Jungherr, 2021, p. 4).

In this definition, the specific properties of new data sets take a backseat. Instead, the definition foregrounds *theory-driven* work with computational methods in the social sciences. At the same time, it recognizes the importance of descriptively oriented work. This is important not least because CSS opens up new types of behavior and phenomena that only arise as a result of digitization or which were previously beyond the grasp of social scientists. Accordingly, there must be room in CSS for first systematically recording and describing new behaviors or phenomena without forcing them hastily into the limits of well-known but possibly unsuited theories.[7]

The definition also foregrounds an important point of tension in precisely differentiating CSS from other fields in the social sciences. Nearly all contemporary work in the social sciences relies on computational methods and digital or digitized data. This includes the storage and processing of digital data (such as digital text, image, or audio files), computationally assisted data analysis (such as regression analyses), or data collection through digital sensors (such as eye tracking or internet of things enabled devices). In this work, computation is often a necessary precondition. For example, while it is pos-

[6] For more examples for the analysis of news articles see Barberá et al. (2021). For literature corpora see Piper (2018); Underwood (2019). For television news casts see Jürgens et al. (2022). For political ads see Schmøkel and Bossetta (2022). For parliamentary speeches see Rauh and Schwalbach (2020). For digitized historical data see Cirone and Spirling (2021). For the theory-CSS disconnect see Jungherr and Theocharis (2017); Jungherr (2019).

[7] The following three paragraphs follow Theocharis and Jungherr (2021), p. 4 f.

sible to run multiple regressions with pen and paper, the success of this method in the social sciences depends on the digital representation of the underlying data sets and computational resources available to process the data. In the most general reading of the provided definition the use of any computational method in data handling and analysis would qualify as computational social science. One could thus argue that nearly any form of contemporary social science would constitute computational social science. Obviously, this is not helpful in identifying constituting elements of the field and subsequent potentials and challenges.

In talking about CSS specifically, it might be helpful to focus more on studies and research projects in which computational methods and practices are not used as plug-and-play solutions but instead demand for varying degrees of customization with regard to data collection, preparation, analysis, or presentation. Again, this is best thought of as a distinction in degree. On one end of the scale, we find projects that require some coding with regard to the sequential calling of pre-existing or slightly modified functions or data management. On the other end of the scale, we find research projects that demand the development of dedicated software solutions, for example in automated and continuous data collection, preparation and structuring of large unstructured raw data, or the development of dedicated non-standardized analysis procedures. Projects at different ends of this scale share issues arising from their focus on social behavior, systems, or phenomena but they vary significantly with regard to their computational demands. Projects that use standardized computational methods might thus be basically indistinguishable from other areas in empirical social science research. On the other hand, projects at the other end of the scale are likely to face challenges indistinguishable from software development in computer science.

Any conceptualization of computational social science should thus not be tied to a specific set of methods, data sets, or research interests. Instead, the constituting element of CSS differentiating it from other approaches in the social sciences, is the degree to which research projects demand for the inclusion and development of computational methods and practices over the course of a project. At the same time CSS is a specific subfield in computer research in that it focuses on social systems and phenomena. Consequently, approaches and methods have to account for the specific conditions of this research area.

Computational social science occupies a bridging position between the social sciences, computer science, and related disciplines. This enables researchers to conduct interdisciplinary research into both new and already known social phenomena by combining social science theories and methods as well as concepts and methods from computer science. In this bridging function, CSS gives the social sciences access to advanced computational approaches and methods, while opening up subjects of study in the social sciences to computer science and related disciplines. In the dialogue between the disciplines, CSS contributes to the institutionalized transfer of knowledge and practices and helps at overcoming historically grown barriers between fields. If successful, computational social science does more than just transfer knowledge or methods. It combines theoretical and methodological approaches from related disciplines into viable concepts and research designs and applies them in order to establish scientific knowledge on social phenomena.

2.2. The computational social science project pipeline

Our discussion of computational social science and its promises and challenges has remained rather abstract. It is time to turn to CSS as a practice. For this, let's have a look at the typical CSS project pipeline. While CSS projects come in a stunning variety of data sets used, methods employed, and questions asked, more often than not, these projects share a pipeline of tasks, problems, and decisions that is typical for CSS. Examining this pipeline allows us to think about engaging in CSS as a practice, while at the same time providing you with a blueprint for potential research projects that might lie in your future.

The typical pipeline for computational social science consists of the following steps:

- research design,
- data collection,
- data preparation,
- linking signals in data to phenomena of interest,
- data analysis, and
- presentation.

Let's have a look at each of these steps in detail.

2.2.1. Research design

As with any research project in the social sciences, projects in computational social science should start with a research design.[8] Researchers must ask themselves how to go about in answering a specific question in a reliable, transparent, and inter-subjective way. This can include questions testing a theoretically expected causal mechanism between two phenomena, explorative questions of new phenomena for which no plausible prior theoretical expectations exist, or the systematic description of phenomena or behavior. The nature of the question then dictates the choice of data, method, and process.

To date, some of the greatest successes of computational social science lie in the description of social phenomena and characteristics of groups and individuals.[9] The best of these studies showcase the impressive measurement opportunities of CSS - such as the

[8]For a helpful introduction to doing research in computational social science see Salganik (2018).

[9]For the estimation of environmental conditions based on satellite images see Brandt et al. (2020). For the connection between online communication and news coverage see Jungherr (2014); Wells et al. (2016). For the reaction to external events on digital media see Zhang et al. (2019).

estimation of environmental conditions based on satellite images in hard to reach areas or the interconnection between online communication and outside factors, such as media coverage or external events. CSS as a field has been less interested in connecting findings systematically to theoretical frameworks in the social sciences, providing explanations or causal mechanisms for patterns identified, or even connecting digital signals robustly to concepts or phenomena of interest. Currently, CSS has been less successful in connecting their findings to theories in the social sciences or advancing new systematic theories. This gap offers interesting new perspectives for new research designs.

CSS has to transition from its early stage of producing predominantly isolated empirical findings to a more mature stage in which studies are more consciously connected with theoretical frameworks, allowing the field to speak more actively to the debates in the broader social sciences trying to make sense of underlying phenomena. This might mean treating predominantly diagnostic efforts as only a first step and focusing researchers' attention more actively on connecting digital signals to meaningful concepts and starting to work on explaining patterns found in data based on causal mechanisms. This might also mean extending concepts and theories currently in use among social scientists for the conditions found in online communication spaces while at the same time remaining mindful of relevant research interests and frameworks in traditional social science.

While most work in computational social science follows predominantly descriptive empirical approaches, such as the analysis of text, image, or behavioral data, there are other approaches that offer different types of insights. One example for this are experiments.[10] By deliberately manipulating actual or simulated digital communication environments, researchers can identify causal effects of specific design decisions or targeted interventions. Even if this approach is effortful in terms of design and implementation, it offers great potential for knowledge.

The need for experimental research designs has been recently illustrated by a study by Burton et al. (2021). The authors show in their paper that in data rich contexts, such as those found in the work with digital trace data, many different explanatory models fit data. Some that conceivably might be true, others that are obviously meaningless. This raises the danger that by using purely correlative research designs in CSS, researchers might fool themselves in believing patterns support their theory of interest while in fact falling for spurious correlations emerging from large and rich data sets. The presence of large data sets makes careful research designs more important not less.

Another alternative method is theory-driven simulation or modeling of social systems or individual behavior.[11] This approach has lost relative influence in the course of the rapidly increasing data availability through social media services. Nevertheless, the

[10] For experimental evidence on causal effects of specific design decisions in digital environments see Salganik et al. (2006); Salganik and Watts (2009). For evidence on causal effects of interventions see Bail et al. (2018); Munger (2017). For dangers of purely correlative designs in data rich environments see Burton et al. (2021). For simulation see J. M. Epstein (2006); Macy and Willer (2002); Miller and Page (2007).

[11] For more on simulations see Macy and Willer (2002); J. M. Epstein (2006); Miller and Page (2007).

strongly theory-driven background of this approach offers a promising alternative to the often predominantly data-driven exercises of research based on social media data.

2.2.2. Data collection

After settling on a research design and choosing the appropriate data to answer your question, the fun of getting data starts. It is no accident that the discussion about the alleged wealth of digital data in the social sciences often elegantly skips over questions of whether and how these data can be collected, processed, managed, and made available. In fact, data collection and processing are often the most time-consuming, complicated, and at the same time least visible and most thankless tasks in computational social science.[12]

Data collection in computational social science has become more complicated over time. This is due to digital media becoming more difficult to collect and increasing scientific standards in working with said data. In the early phases of CSS, the topic of data collection often took a backseat and the procurement of social media data was often enabled by companies running digital services. Over time, however, social media services have become significantly more restrictive in terms of the data access they allow outsiders. At the same time, there was growing awareness in academia that even the generous provision of social media data via official interfaces only provided a fraction of the data necessary for answering demanding research questions. Additionally, CSS practitioners found themselves challenged that their focus on a few well-researched platforms, such as Twitter, only would allow for limited statements about digital communication, human behavior, or societies. This raised the call for more cross-platform research, again raising the demands for data collection and preparation. While true, one cannot help but note that these challenges are often raised by people skeptical of computational work to begin with, if not quantitative methods in general.

Overall, this means that data collection and processing for CSS projects has become significantly more complicated. Different data sources often have to be monitored continuously over long periods of time. Some of these can be queried via official interfaces, so-called Application Programming Interfaces (API) (e.g. Facebook, Twitter, or Wikipedia), while access to some data sources (e.g. individual websites) demand specially adapted software solutions. Both approaches are complicated and prone to different types of errors. With long-term data collections, there is a risk, among other things, that API or non-standardized data sources can change unnoticed.[13] Accordingly continuous quality assurance must be ensured which can demand for significant investment of resources and time. Overall, the increasing demands on the breadth, scale, and quality for data collection increasingly require the development of research software adapted to the respective project and can no longer only be mapped with relatively little programming effort and access to isolated API.

[12] For more on data and data quality in CSS see Posegga (2023).
[13] For an account of shifting rules of API see van der Vlist et al. (2022).

2.2.3. Data preparation

Even less well discussed than data collection are issues for computational social science projects arising from the preparation of data for analysis. While API provide clearly structured data, unstructured data from less standardized sources must first be structured after collection. This usually requires the transfer of raw data into database structures developed for the research project. Most research projects also require semi- or fully automated labeling steps in which individual data points are supplemented with meta data (e.g. by coding text according to interpretative categories). In the case of extensive projects, these must be secured and stored together with the originally collected data and made available for further analysis. The use of different software in various steps of data preparation, such as collection, structuring and annotation, complicates this aspect. The design of database structures and work processes, ensuring a consistent and high-performance infrastructure for the analysis of complex data sets, is not trivial and often requires more than rudimentary knowledge in modeling the corresponding database structures.[14] Additional knowledge of software development using various libraries and technologies is often required as well.

2.2.4. Linking signals in data to phenomena of interest

The next step in computational social science projects follows the research design and runs parallel to data collection and preparation for analysis. This is providing the connection between signals visible in data and phenomena of interest. Examples for this might be specific interaction patterns between Reddit users as expression of political polarization in society, or mentions of politicians on Twitter as being indicative of their subsequent electoral fortunes. It is important for researchers to critically interrogate their data on whether these signals are actually connected with the phenomenon of interest.

Data emerge based on different data generating processes.[15] For example, publicly available Twitter messages are the result of a complicated filtering process leading a user to post a tweet referring to specific topics or persons. Twitter is a performative medium documenting objects of attention or opinions the specific subset of people active on Twitter want to publicly be seen as interacting with or referring to. This makes Twitter a powerful tool to understand dynamics of public attention of politically vocal Twitter users but probably not a tool to understand public opinion in society overall. Data collected on other digital media come with different data generating processes that need to be reflected in the interpretation of identified patterns.

[14] For more on data management for social scientists see Weidmann (2023).

[15] For a discussion of different data generating processes of different sources of digital media see Jungherr and Jürgens (2013). For a detailed discussion of Twitter's data generating process and its consequences for work based on Twitter-data see Jungherr et al. (2017).

This also means CSS needs to get serious about indicator validation.[16] Today, much of CSS relies on face validity. If a digital signal seems to reasonably reflect a phenomenon of interest, no systematic validation tests are undertaken. This allows the quick production of seemingly meaningful findings, speaking to contemporary concerns in public debate. Yet, there is the serious danger of mistaking digital signals for phenomena they are not actually documenting, as for example mistaking signs of attention to politics for political support or predicting the flu by looking for signs of winter. Measurement in the social sciences often means searching for evidence of latent variables that have no direct objectively measurable expression. This demands for the active reflecting, theorizing, and testing of whether identifiable signals can be reasonably expected to express a concept of interest. This makes reliance on face validity in the social sciences dangerous and prone to error.

2.2.5. Data analysis

The next step in typical computational social science projects, data analysis, is much better documented and well discussed than the previous stages. There is a wide variety of methods available in CSS. The use of methods naturally follows choices in research design and the demands and opportunities connected with the available data. Later, we will be examining some typical analytical approaches within CSS in greater detail, so here I will only briefly mention some analytical approaches and choices available to you.

One typical approach is automated or semi-automated content analysis of different digital corpora. For example, the computationally-assisted analysis of text in the social sciences is already very well established and is increasingly complemented by the use of more advanced machine learning methods and (semi-)automated analysis of image data.[17] These analyses can be very rudimentary, for example by identifying and counting the occurrences of specific words in text. Or they can be more demanding, for example by looking for expressions of a latent concepts (such as political ideology) in speech or interaction patterns. Analyses can be performed by human coders or automatically. Still, independent of the choice for simple or demanding analytical target or automated versus human coding, these studies have to address fundamental questions of coding validity and reliability that have been well established in the literature on content analysis.

Another approach closely associated with CSS is network analysis.[18] Social network analysis is a long-established research practice in social science with a rich body of theories

[16] For the need of indicator validation in computational science and the need to theorize the connection between digital signals and phenomena of interest see Howison et al. (2011); Jungherr et al. (2016); Jungherr (2019). For mistakes driven by face-validity see Lazer et al. (2014); Jungherr et al. (2017); Burton et al. (2021).

[17] For more on the analysis of text in CSS see Benoit (2020); Gentzkow et al. (2019); Grimmer et al. (2022). For the analysis of images see N. W. Williams et al. (2020); Schwemmer et al. (2023).

[18] For more on network analysis see Wasserman and Faust (1994); Easley and Kleinberg (2010); Howison et al. (2011); DellaPosta et al. (2015); Granovetter (2017).

and methods. Methods of network analysis allow the investigation of different relational structures between human (or non-human) actors, often with the aim of understanding the meaning and effects of these structures in different application areas. Instead of an "atomistic" research perspective that sees people primarily as isolated individuals, network analysis pursues a "relational" perspective that takes people's relationship structures seriously and by mapping them tries to identify their impact. Network analysis is a very prominent approach in CSS. On the one hand, this is due to the fact that digital communication as such is fundamentally closely linked to the concept of networking and networks. Since corresponding societal processes and individual usage behavior are responsible for much of the digital trace data used in CSS, it is no wonder that network analysis is an obvious choice in the analysis of these data. However, this seemingly intuitive proximity often obscures the necessary interpretative steps involved in analyzing networks based on digital trace data.

Increasingly, there are also studies that connect different data types.[19] For example, some studies connect people's survey responses to their digital traces (such as web tracking data). The benefit of studies following this approach is the opportunity to offset some of the limitations of using only one data type. For example, simply relying on people's survey responses on what type of news media they claim to have consumed is prone to error. People forget, misremember, or might not admit to consuming specific media. On the other hand, inferring people's political leaning or opinions simply based on digital traces is also fraught. Combining both data types might in principle provide a broader picture of their behavior and effects of their online behavior or information exposure. Other studies combine data collected on different digital media platforms, following a similar research logic. But while offering a broader view into some questions, these combined approaches bring other drawbacks that need to be critically reflected and accounted for.

2.2.6. Presentation: Ensuring transparency and replicability

The final step of any computational social science project is the presentation of its findings. I will not bore you with generalities about the writing and publication process. Instead, let us focus on one crucial element in finalizing a project: providing transparency about your choices and making sure it is replicable by other researchers.

In many social sciences we find important movements that push for the development and institutional adoption of more transparent research practices allowing for a more reliable interrogation of research findings by third parties while at the same time limiting primary researchers' degrees of freedom in adjusting research questions and designs after knowing the outcomes of data analyses. Proposed remedies include systematically providing public access to data sets, the publication of code underlying data preparation and

[19] For more on the potential of combined data sets of different types see Stier et al. (2020). For challenges and limits to this approach see Jürgens et al. (2020).

analysis, and pre-registration of planned research designs and analytical protocols.[20] While the importance of this program is recognized in fields such as economics, political science, or psychology, it is largely lacking within CSS.

There are two areas systematically introducing opaqueness into computational social science:

- data underlying research projects, and
- transparency with regard to the robustness and inner workings of advanced methods.

One of the central selling propositions of CSS is its use of large and rich data sets. These data sets often stem from commercial online platforms. Accordingly, they come with various concerns regarding the privacy of users whose behavior is documented in them and intellectual property rights of the companies providing researchers with access to them. This brings two challenges: First, how do we ensure access to relevant data for researchers in the first place; and second, once access has been granted, how can researchers provide others access to said data to double check their findings. In these cases, rules set by platforms governing access to proprietary data can serve as cloaking device, rendering data underlying highly visible CSS research intransparent. Here, the field has to become more invested in developing data transparency standards and processes. This might mean pushing back against some of the often arbitrary rules and standards of data access set by platform providers. Those are often designed with commercial uses in mind and serve primarily to protect the business interests of platforms and their public image instead of serving the interests of their users or society at large by enabling reliable and valid scientific work.

Another area of opaqueness in CSS arises from the use of advanced computational methods in an interdisciplinary context. The different disciplines at the intersection of CSS come with different strengths and sensibilities. While typically, there is high comfortableness and skill among computer scientists in software development and the use of quantitative methods, social scientists typically are more interested in addressing actual social instead of predominantly technical questions. This brings the danger of scientists primarily driven by interests and sensibilities in social problems uncritically using analytical tools provided by computationally minded colleagues without critically reflecting on these tools' inner workings and boundary conditions. In the worst case, this can lead to social scientists misdiagnosing social phenomena based on an uncritical and unreflected use of computational tools and quantitative methods.

At the same time, the development of robust methods in CSS is hampered by a prototype-publication culture. Researchers are incentivized to publish innovative methods which once published are treated as proven by the field. Critical testing of methods and their implementations in code across varying contexts is currently not encouraged by

[20] On replication and open science see Angrist and Pischke (2010); Open Science Collaboration (2015); Christensen et al. (2019); Wuttke (2019)

publication practices of the leading conferences and journals in the field. This inhibits the development of a robust collective validation effort of methods and measures.

Already this brief sketch of the typical CSS project pipeline shows the diversity and richness of computational social science. The field is neither defined by specific data types or analytical methods. Rather, CSS is a broad research approach embracing different methods and perspectives. Individual researchers or even most mono-disciplinary teams cannot convincingly represent this diversity. The future of CSS lies in the interdisciplinary merger of the various social sciences, computer science, and natural sciences. This is easier said than done. As anyone who has tried it will tell you, interdisciplinary research is easy to talk about but difficult to practice. To get better at this, it is important to collect and document specific experiences of different projects or research teams. Some documentations are starting to be published.[21] But this can only be the beginning of a systematic reflection.

2.3. Text analysis: Typical approaches in computational social science 1

2.3.1. Text analysis in political science

One approach typically associated with computational social science that is very prominently used in political science is text analysis.[22] One reason for this is the long tradition of using text corpora in various subfields of political science going back thirty years or more. Examples include the measurement of positions and preferences of parties based on manifestos, the analysis of positions and tactics of parties and politicians based on parliamentary debates, the automated identification of events and actors in text corpora for the development of event- and actor-databases in international relations and conflict studies, or the analysis of news coverage in agenda research and discourse studies. These subfields identified early on the potential of computational approaches in the pursuit of their research goals using large text corpora. Accordingly, we find rich traditions and practices in these subfields going back thirty years or more in working with text. This includes standardized approaches in the collection, preparation, analysis, reporting and the provision of large text corpora, starting with human centered approaches and today reflecting deeply on the uses of and standards for the work with computational methods.

[21] On the practice of computational social science and the establishment of research teams see G. King (2011); Gilardi et al. (2022); Windsor (2021).

[22] For general accounts of computationally-assisted text analysis in the political sciences see Benoit (2020); Grimmer et al. (2022). For the collection and analysis of parliamentary debates see Schwalbach and Rauh (2021). For text analysis and event detection see Beieler et al. (2016). For text analysis of news coverage Barberá et al. (2021). For text analysis in literature or philosophy corpora see Piper (2018); Underwood (2019).

Text is a rich medium reflecting cultural and political concepts, ideas, agendas, tactics, events, and power structures of the time. Large text corpora open windows and allow comparisons across countries, cultures, and time. Different types of text contain representations of different slices of culture, politics, and social life. They therefor are of interest to various subfields in the social sciences and humanities.

For example, political texts (such as parliamentary speeches or party manifestos) express the political ideas of the time and chosen tactics of politicians or parties.[23] At the same time, they are also expressions of public performances of these ideas and tactics. They are therefor product of choices by political actors how they want to be seen by journalists, their constituents, party members, or the public at large. This makes them not direct expressions of their true nature or positions but instead expressions mediated through conventions associated with the type of text under examination. These and other potentially relevant mediating factors need to be accounted for in subsequent analyses and the interpretation of any identified patterns. While remaining aware of these mediating factors the analyses of corpora of political texts between countries and over time allows for the identification of various interesting political phenomena, this includes shifts in the positions of parties, the introduction of new ideas into political discourse, agenda shifts, or tactical choices in language or rhetoric. This and their relatively codified and structured format makes political texts popular in political science.

Another important text source in political science is news text.[24] News contain a variety of signals that are of great interest to social scientists. This includes the reporting of key political events and figures in politics and society. News cover the what, when, where, who, and sometimes even why of important political or societal events. Extracting these features from journalistic accounts allows the establishment of standardized, large-scale databases of international events and actors. Approaches like these have been successfully used in conflict studies. News texts are also a prominent basis for the analysis of political agenda setting and agenda shifts. Identifying the frequency and time of the coverage of selected topics, researchers can identify the relative importance events have in press coverage and compare that with their importance in political speech, public opinion surveys, or digital communication environments. Finally, the analysis of news coverage also allows for the analysis of discourse dynamics over time. How are current important topics discussed in the media, what are the aspects different sides emphasize, what are the arguments, and who are prominent speakers given voice to in the media? These questions can be answered based on the analysis of news coverage and

[23] For a large collection of party manifestos across countries and time Merz et al. (2016). For collections of parliamentary debates see Rauh and Schwalbach (2020).

[24] Since news content is protected under copyright law, establishing big publicly available text corpora of news coverage is difficult. Luckily, various news organizations start providing standardized access to their archives, which enables the reliable collection and analysis of their respective coverage. See for example New York Times. Content of other news organizations can be accessed through third-party providers or specific licensing deals. For event data see https://www.gdeltproject.org. For an overview of agenda setting research see McCombs and Valenzuela (2004/2021). For examples of discourse analyses based on news coverage see Ferree et al. (2002b); Entman (2004); Baumgartner et al. (2008); Benson:2014aa.

that provide important insights into the way societies negotiate contested topics, such as foreign policy, immigration, or reproductive rights.

By collecting and preparing for analysis large text corpora scholars can therefor access and make available vast troves of knowledge on various questions and in different subfields. The tremendous collective efforts in digitizing and making available text corpora are a massive accelerating factor in this effort.

To get a better sense of computer-assisted text analysis in action, let's have a look at three recent studies, using different text types and different methods in answering their respective questions.[25]

2.3.2. Making sense of party competition during the 2015 refugee crisis with a bag of words

In their 2021 article *How the refugee crisis and radical right parties shape party competition on immigration*[26] Theresa Gessler and Sophia Hunger study a corpus of 120,000 press releases by parties from Austria, Germany, and Switzerland. The authors are interested in whether parties changed the emphasis of the topic immigration and their position on immigration in their press releases between 2013 and 2017. The authors ask whether the attention of parties in their press releases to the topic immigration followed a long-term trend in politicizing the topic driven by the emergence of far-right parties or instead whether attention shifts were driven by the heightened levels of public attention on immigration during the events of 2015. With their study, they contribute to the scientific debate about party competition and agenda setting and position their findings with regard to theories in both areas. At the same time, they provide an instructive example of how to anchor an empirical study within theory, creatively establish and justify new data sources, and use an intuitive and comparatively accessible computational method in the analysis of text.

In order to answer their questions, the authors introduce a new data source: monthly press releases by parties. In their research design section, they justify this choice. The predominant data source for work in comparable fields are party manifestos. Those have proved valuable in the study of political competition and agenda shifts but their characteristics limit their applicability for the authors' purposes. Due to their sparse publication rhythm, following the electoral calendar, they do lend themselves for the analysis of long-term trends but not for the identification of short-term shifts driven by current events and sudden shifts in public opinion. Press releases, due to their higher frequency and their connection to current events are more promising in this regard. This reasoning opens the intriguing possibility that theorizing about long-term trends within

[25] For an overview of computer-assisted text analysis in political science see Grimmer et al. (2022). For practical advice on how to do computational text analysis in R see Silge and Robinson (2017); Hvitfeldt (2022), for quanteda an R package of high popularity with political scientists see Benoit et al. (2018), for Python see Bengfort et al. (2018); Lane and Dyshel (2022).

[26] Gessler and Hunger (2022).

party positioning is less about the subject being necessarily primarily shaped by long-term trends but an artefact from the availability of data sources allowing only for the analysis of this type of question. Mobilizing new data sources thereby potentially opens up new aspects of the phenomenon that remained invisible before.

The authors collected 120,000 press releases from major parties in Austria, Germany, and Switzerland published between 2013 and 2018. To identify press releases referring to immigration, the authors developed a dictionary containing words referring to immigration and integration. To evaluate the performance of their dictionary, the authors hand-coded 750 randomly-selected press releases and tested the quality of different dictionary approaches and a specifically trained support vector machine classifier (SVM). Their dictionary outperformed others in the identification of immigration-related press releases and performed in similar quality as the SVM. Accordingly, they chose their computationally less demanding and at the same time interpretatively more accessible dictionary to classify the remaining press releases over the SVM. The proportion of thus identified press releases referring to immigration of all press releases during a given month allows the authors to identify the relative salience of the topic and temporal shifts over time in comparatively high temporal resolution.[27]

In order to identify the relative position on immigration of parties, the authors use Wordscores.[28] First developed by Michael Laver, Kenneth Benoit, and John Garry, Wordscores try to identify relative topical positions of parties based on the similarities and distinctiveness of words they use in text. The more similar the words, the more similar the positions. The more distinctive, the further they are. Simplifying their approach somewhat, Wordscores allow Gessler and Hunger to identify whether parties meaningfully diverge from their original word use regarding immigration and whether over time they converge or diverge with words used by parties of the radical right in their sample. They take this as proxy for position shifts of parties with regard to immigration in either accommodating or confronting positions of the radical right.

Using these approaches, Gessler and Hunger find that controlling for other factors it does seem that mainstream parties during the refugee crisis reacted to the greater attention paid to immigration by radical right parties by increasing their own attention to the topic in their press releases. But after the crisis subsided, they decreased their attention to the topic back to their original levels. In contrast, regarding the positions nearly all parties did not converge toward positions taken by the far right.

With their study, Gessler and Hunger not only provide compelling evidence on political competition between European mainstream and radical right parties during the 2015 refugee crisis. They also show how mobilizing a new data source can provide new evidence, allowing new perspectives in the analysis of scientifically long established subfields. By capitalizing on the greater temporal resolution provided by press releases,

[27] To get a better sense of how to evaluate the quality of competing text analysis methods see the online appendix of Gessler and Hunger (2022).

[28] For more on Wordscores see Laver et al. (2003); Lowe (2008).

the authors open a window into short term patterns in political competition, otherwise invisible to researchers depending on established data sources only available in much lower frequency. The study is also an interesting case in attempting to identify shifts in two latent concepts (i.e. relative topic salience and position of parties) based on the analysis of text.

2.3.3. Who lives in the past, the present, or the future? A supervised learning approach

In his 2022 article *The Temporal Focus of Campaign Communication*[29] Stefan Müller analyses the degree to which parties refer to the past, the present, and the future in their party manifestos before upcoming elections. He anchors this question with voting behavior theory, which considers retrospective and prospective considerations of voters. Thus temporal considerations matter to voters in elections, but do they also matter for campaign communication by parties. To answer this question, Müller analyses 621 party manifestos published between 1949 and 2017 in nine countries. Other than for Gessler and Hunger (2022), this time the data source is clearly up to the task. Party manifestos are directly connected with elections and should reflect tactical considerations by parties regarding their self-presentation toward partisans, constituents, journalists, and the public at large.

To answer his question, Müller collected all machine-readable manifestos in English or German from the Manifesto Corpus, leaving with 621 manifestos from nine countries.[30] He then had human coders label sentences as referring to the past, present, or future. Either by directly labelling or by using pre-labeled data sets Müller ended up with an annotated sample of 5,858 English and 12,084 German sentences. This allowed him to train and validate several different computational approaches for the classification of the remaining sentences in the data set. He trained and validated a Support Vector Machine (SVM), a Multilayer Perceptron Network, and a Naive Bayes Classifier. Since all classifiers performed all comparatively well, the author chose the SVM since it provided the best trade-off between performance and computational efficiency.

Through this approach, Müller finds that 54% of sentences refer to the future, 37% the present, and 9% the past. But there is some variation between countries. In general, though, it appears like incumbent parties focus somewhat more on the past than opposition parties. This makes sense considering the different roles of incumbent and opposition parties in political competition. Incumbents run at least in part on their supposedly positive record of the past and opposition parties naturally challenge said record.

[29] See S. Müller (2022).

[30] For the Manifesto Corpus see Merz et al. (2016). For detailed information of the validation and comparison between the different classification approaches see the online appendix to S. Müller (2022).

To get a better sense of how parties refer to the past, present, or future, the author uses German and English versions of the Linguistic Inquiry and Word Count (LIWC) sentiment dictionary.[31] The dictionary lists terms that for test-subjects carried positive or negative emotional associations. By calculating the emotional loading of words used in sentences referring to the past, present, or future, Müller infers whether parties spoke positively or negatively about different temporal targets.

Using this approach, Müller finds that opposition parties tend to speak more negatively about the past than incumbents. Again, this finding is in line with the different roles of incumbent and opposition parties in political competition.

By using a pre-existing data set, the Manifesto Corpus, and further annotating it, Müller can show that parties indeed use temporal references differently according to their structural roles in political competition. The paper offers an interesting example for the use and evaluation of various supervised computational classifiers in the analysis of large data sets, enabling classification efforts for data sets whose size would make manual classification infeasible.

2.3.4. Political innovation in the French Revolution

And now, let's have a little bit of fun.

In their 2018 article *Individuals, institutions, and innovation in the debates of the French Revolution*[32] Alexander T. J. Barron, Jenny Huang, Rebecca L. Spang, and Simon DeDeo present an analysis of speeches held during the first parliament of the French Revolution, the National Constituent Assembly (NCA), sitting from July 1789 to September 1791. They have access to a corpus provided by the French Revolution Digital Archive containing 40,000 speeches by roughly a thousand speakers. The speeches held during this time frame are of great interest not only to historians but also parliamentary and democracy scholars, since they open a window into the process of epistemic and political sense making and innovation processes within one of the first modern parliamentary bodies that provided the template for many subsequent parliamentary institutions and democratic discourse in general.

Barron and colleagues approach the text corpus through the lens of information theory. They are interested in determinants for the emergence of new ideas in parliamentary discourse and their persistence. For this they identify distinct word combinations through latent Dirichlet allocation (LDA).[33] LDA is a popular automated approach for reducing the dimensionality of text to a set of automatically identified topics that are characterized by the frequent clustered occurence of words. Barron and colleagues use LDA to identify

[31] For more on the Linguistic Inquiry and Word Count (LIWC) sentiment dictionary see Tausczik and Pennebaker (2010).

[32] See Barron et al. (2018). For the corpus see The French Revolution Digital Archive: https://frda.stanford.edu.

[33] For latent latent Dirichlet allocation (LDA) see Blei et al. (2003).

clusters of co-occurring words, topics, and assign them two metrics they calculate based on ideas from information theory: novelty and transcience.

With *novelty*, the authors refer to the degree to which a distinct word pattern in a speech as identified through LDA differs from patterns in prior speeches. With *transcience*, the authors refer to how the same word pattern differs from those in future speeches.[34] The higher the transcience, the higher said difference. To measure these metrics for each distinct word pattern, they calculate a measure called Kullback-Leibler Divergence (KLD). Doing so allows the authors to quantify two important but interpretatively demanding concepts: How frequently are new ideas introduced in parliamentary discourse and how long do they persist? Once you have quantified these features of distinct word packages, you can start looking for determinants of either outcome. Who is responsible for the introduction of new ideas? And what are the contextual conditions for these ideas to survive and thrive?

Barron and colleagues show that in general the National Constituent Assembly (NCA) was a parliamentary environment clearly open for the introduction of novel ideas, but many of these new ideas did not persist long. Still, many speeches were at once highly novel and only weakly transient, this condition the author call *resonance*. Looking closely, the authors show that individuals differ with regard to their tendency to introduce new and resonant ideas. In fact, among the top 40 orators high-novelty speakers are usually associated with the political left and the bourgeoisie. In contrast, low-novelty speakers are on the political right and belong to the nobility. Going into detail even further, the authors show that high-profile individuals can deviate from these patterns, such as the left-wing radicals Maximilien Robespierre and Jérôme Pétion de Villeneuve whose speeches showed exceptionally high values of novelty and resonance. They consistently introduced new ideas in their speeches that were picked up by others and persisted over time. On the other side of the political spectrum, speakers like Jean-Siffrein Maury and Jacques de Cazalès exhibited low novelty and high resonance. The authors take this as evidence for their role in keeping the conversation in parliament coherent (low novelty) while at the same time being able to influence its future course (high resonance).

The authors point out that their results correspond with previous findings by historians taking more traditional analytical routes. But by translating meaningful and interpretatively demanding concepts into a small set of elegant quantitative metrics, Barron and colleagues provide a systems-level view of innovation and persistence in parliamentary debate. This further allows them to quantitatively identify the impact of features on individual, structural, and institutional levels for the introduction and subsequent fate of new ideas.

The study by Barron and colleagues shows powerfully how the use of computational methods combined with innovative theoretical concepts can open up new insights not only in our present and digital life but instead provide new perspectives to the past.

[34] For details on the operationalization of *novelty* and *transcience* see the online appendix to Barron et al. (2018). For details on the Kullback-Leibler Divergence (KLD) see Kullback and Leibler (1951).

Studies like these are bound to grow both in frequency and resonance with the continued digitalization of ever more historical data sets and archives and offer promising perspectives for interdisciplinary research.

Why not start by checking which other historical parliamentary records are available to you?

As with any set of examples, I could have chosen different studies that would have been just as interesting or would have provided insights into different approaches to text analysis. But already these three brief examples illustrate the breadth in available data sets, methods, and questions computational text analysis can be employed to answer. It is no wonder then to find this approach to be a highly prominent pursuit in computational social science and beyond.

In reading studies like these, or in CSS in general, always make sure to check for online appendices. Often, there you find the actual information about the ins and outs of doing the analysis, sometime get instructions of replicating reported findings, and get a much better sense in general of how a specific method was implemented.

2.4. Digital trace data: Typical approaches in computational social science 2

2.4.1. Digital trace data

As we have seen, in computational social science there are great hopes and enthusiasms connected with the availability of new data sources. Particularly one new data source features in these accounts: digital trace data.[35]

Once people interact with digital devices (such as smart phones and smart devices) and services (such as Facebook or Twitter), their digitally mediated interactions leave traces on devices and services. Some of those are discarded, some are stored. Some are available only to the device maker or service provider, some are available to researchers. This last category of digital trace data, those that are stored and available to researchers, has spawned a lot of research activity and enthusiasm over a new measurement revolution in the social sciences. But somewhat more than ten years into this "revolution", the limits of digital trace data for social science research are becoming just as clear as their promises. Before we look at studies using digital trace data, it is therefore necessary that we look a little more closely at what they are, what characteristics they share, and how this impacts scientific work with them.

[35] For research in digital communication environments in general see Salganik (2018). For digital trace data in particular see Howison et al. (2011); Golder and Macy (2014); Jungherr (2015).

In their 2011 article *Validity Issues in the Use of Social Network Analysis with Digital Trace Data* James Howison, Andrea Wiggins, and Kevin Crowston define digital trace data as:

> "(...) records of activity (trace data) undertaken through an online information system (thus, digital). A trace is a mark left as a sign of passage; it is recorded evidence that something has occurred in the past. For trace data, the system acts as a data collection tool, providing both advantages and limitations. The task for using this evidence in network analysis is to turn these recorded traces of activity into measures of theoretically interesting constructs."
>
> Howison et al. (2011), p. 769.

Replace the term *network analysis* in the last sentence with *social science* and you have the crucial task in working with digital trace data before you. This translation of "traces" into "theoretically intersting constructs" demands for accounting for specific characteristics of digital trace data. Back to Howison and colleagues:

> "1) it is found data (rather than produced for research), 2) it is event-based data (rather than summary data), and 3) as events occur over a period of time, it is longitudinal data. In each aspect, such data contrasts with data traditionally collected through social network surveys and interviews."
>
> Howison et al. (2011), p. 769.

Again, replace *social network* with *social science* and you are good to go. It is probably best you go, find the article, and read this section of the text yourself, but let me briefly explicate the concerns expressed by Howison and colleagues as they relate to social science more generally.

First, digital trace data are *found* data. This makes them different from data specifically designed for research purposes. Usually, social scientists approach a question through a research design specifically developed to answer it. You are interested in what people think about a politician? You ask them. You are interested in how a news article shifts opinions? You design an experiment in which you expose some people to the article but not others and later ask both groups for their opinion on the issue discussed in the article. With digital trace data, you do not have that luxury. Instead, you often start from the data and try to connect it back to your interests. What do people think about a politician? Well, maybe look on Twitter and count her mentions. If you really get fancy, maybe run the messages through a sentiment detection method and count the mentions identified as positive and those identified as negative. Want to identify the effects of a news article? Check if people's traces change after exposure. Already these two examples show that found data can be used in interesting ways. At the same time, you often have to compromise in working with them. Thinking purely from the perspectives of research design and identification approach, digital trace data will often leave you frustrated as they simply might not cover what you need. On the other hand, once you accommodate yourself with what signals are available to you in found data, you might land on new questions and insights that following a purely deductive approach, you might have missed. This is especially true for questions regarding the behavior of users in digital communication environments or the inner workings of said environments.

Second, digital trace data are *event data*. They document interactions and behavior in collections of single instances. Like a Facebook page, write and @-message on Twitter, comment on a post on Reddit, edit a Wikipedia page, click on a site, on and on. These events can carry a lot of information. For example, measuring the impact of an add through clicks of featured links is a perfectly good approach, as far as this goes. But in social science, we often are interested not only in specific interaction or behavior events. Instead, we ask for the reasons underlying these events, such as attitudes or psychological traits. To get at these, we need to understand how users interpret their action that led to a data trace. Take one of the examples from earlier: If we want to understand public opinion on a politician or a political topic, researchers often look for mentions of an actor or topic in digital trace data. But the motives for mentioning a politician or topic on a digital service vary. One could express support, critique, neutrally point to a topically connected event or quote, or one could try and be funny in front of friends and imagined audiences. Some of these motives might be identifiable by linguistic features around the term, others might not. In connecting the event visible in digital trace data, mentions of actors or topics, with the concept of interest, attitudes toward them, means taking into account the data generating process linking the documented event to the concept of interest. This step is crucial in the work with digital trace data but often neglected in favor of a naive positivism, de facto positing that digital traces do not lie and speak to everything we happen to be interested in.[36]

Third, Howison et al. (2011) point to digital traces demanding for a *longitudinal perspective*. Data documenting singular events need to be aggregated in order to speak to a larger phenomenon of interest. But what aggregation rule is appropriate or not? That is open to question. For example, many sociologists are interested in the effects of friendship relations between people. Friendship is an interpretatively demanding concept. This raises many measurement problems. Some people interact with people they consider friends often. Others interact with people they consider friends only seldom but hold deep affection and trust. Just looking at interactions in person, on the phone, or online will therefore not necessarily tell us, who the people are our subjects consider friends. Traditionally, sociologists would survey people to identify people they themselves identify as friends. They therefore have access to the result of the personal calculus of respondents over all interactions and experiences with a person resulting in their assessment of the person as friend or not. Simply looking at digital trace data, as for example email exchanges, public interactions on Twitter, or co-presence in space measured by mobile phones or sensors only provides us with single slices of this calculus.

[36] For more on accounting for data generating processes linking available signals to concepts of interest see Jungherr and Jürgens (2013); Jungherr et al. (2016); Jungherr (2015).

Leaving us to guess the aggregation rule translating single events visible in digital trace data into the latent concept of interest. This is true for social relationships, as friendship, but also for other concepts of interests, such as attitudes or traits. Researchers need to be very careful and transparent in how they choose to aggregate the events visible to them in digital trace data and take them as expression of their concept of interest, especially if its an interpretatively demanding concept.

Finally, Howison et al. (2011) emphasize that digital trace data are "both produced through and stored by an information system" (p. 770). This is important to remember. It means that both the recording of the data as well as access to it depend on the workings of said information system and the organization running it. For one, this means that it is important to differentiate between social and individual factors contributing to an event documented in a data trace and features of the underlying information system. An example for an individual factor leading to a data trace could be my support for a given candidate that makes me favorite a tweet posted on her Twitter feed. Alternatively, a system-level feature could be an algorithm showing me a tweet by said candidate prominently in my Twitter-feed in reaction to which I favorite it in order to read it later. The are two different data generating processes driven by different motives but not discernable by simply looking at the digital trace of the event.

The second consequence of the prominent mediating role of the information system is our dependence of its internal reasons for recording and providing access to data traces. Researchers depend on information systems and their providers in providing them with access and setting access rules. This can be comparatively rich, as currently is the case of Twitter, or comparatively sparse, as is currently the case with Facebook. In any case, shifts in access possibilities and rules are always possible and do not need to follow coherent strategies. This makes research highly dependent on the organizations collecting and providing access to data and introduces a highly troubling set of concerns regarding ethics of data access, conflicts of interests between researchers, organizations, and users, and the transparency and replicability of research findings.

While these challenges persist in the work with digital trace data, this new data type has found a prominent place in social and political science. Unfortunately, the degree to which these challenges are reflected in actual research varies considerably. Nevertheless, let's take a closer look.

2.4.2. Digital trace data in political science

Digital media have led to far-reaching changes in social life and human behavior.[37] These new phenomena lead to new research questions, which digital trace data have been used to address. Areas in political science, that have seen the strongest impact of digital

[37]For accounts of how digital media impact politics see Jungherr et al. (2020), the public arena see Jungherr and Schroeder (2022), for their impact on discourses see Jungherr, Posegga, and An (2019).

media and research using digital trace data include the practice of politics, political communication, structures and dynamics in the public arena, and discourses.

Various studies using digital trace data focusing on politics and political communication have addressed the behavior of political elites, partisans, and the public.[38] This includes the use of digital media in political campaigns, protests, or the coordination of citizens and civil society.

Other studies examine structures of the public arena, media use, and discourses in digital communication environments. This can be, for example, investigating how people use the media. While the traditional approach would exclusively ask people in surveys for their media usage patterns, digital trace data offer powerful additional perspectives through greater reliability and greater resolution.[39] An example for this are web tracking data that track and document website visits by respondents. In addition to new perspectives on the use of news sources, digital trace data also offer new perspectives on the influence of different media types and sources. Here, examining agendas of the most important issues in different digital and traditional media and their mutual influence dynamics are promising research areas.

More specific to digital media still are studies focusing on usage dynamics and behavioral patterns of people in their use of digital services. This focus area lends itself especially well for the study through digital trace data.[40] Examples include studies focusing on Facebook, Reddit, Twitter, or YouTube. Beyond the study of commercial services, digital trace data have also been successfully been used in studying the behavior of people on e-government services, such as online petitions.

Other studies are using digital trace data to examine the ways governments react to the challenge of digital media.[41] In particular, authoritarian states see themselves increasingly challenged in their control of the public by digital media and the associated new possibilities for information and coordination of their population. Digital trace data have provided researchers with promising instruments for documenting and examining digital media provision and attempts at government control in different countries.

Beyond the study of phenomena directly given rise by the use of digital devices and services, researchers are also trying to infer general phenomena based on signals in

[38] For examples of the study of campaigns with digital trace data see Jungherr (2015); Jungherr et al. (2022), protest see González-Bailón et al. (2011); Jungherr and Jürgens (2014); Theocharis et al. (2015), for civil society see Theocharis et al. (2017).

[39] For examples the integration of survey and digital trace data see Subhayan Mukerjee (2018); Jürgens et al. (2020); Scharkow et al. (2020), for agenda dynamics see Neuman et al. (2014); Posegga and Jungherr (2019); Gilardi et al. (2021).

[40] For Facebook see Stier et al. (2017), for Instagram see Kargar and Rauchfleisch (2019), for Reddit see An et al. (2019), for Sina Weibo see Chen et al. (2013), for Twitter see Gaisbauer et al. (2021), for WeChat see Knockel et al. (2020), for YouTube see Rauchfleisch and Kaiser (2020), for online petitions see Jungherr and Jürgens (2010).

[41] For examples for studies examining government control with digital trace data see Chen et al. (2013); Lutscher et al. (2020); Tai and Fu (2020); Lu and Pan (2021).

digital trace data.[42] Examples include the estimation of political alignments of social media users or the prediction of public opinion or election results. While often original and sophisticated in the use of methods, the validity of resulting findings are contested since they often risk misattributing meaning to spurious correlations between digital traces and larger societal phenomena.

Now, let's look at two studies a little more closely to get a better sense of how work with digital trace data actually looks.

2.4.3. Making sense of online censorship decisions

The most direct way to use digital trace data is to learn about the digital communication environment they were collected in. But before your eyes glaze over now in expectation of another starry eyed discussion of hashtag trends on Twitter, there is more to this than the good, the bad, and the ugly of social media. For example, looking at what happens on digital media can tell us a lot about how states regulated speech or try to control their public. One important example for this is China.

By now, there are a number of highly creative and instructive studies available that use data collected on digital media to understand the degree, intensity, and determinants of Chinese censorship activity. In their 2020 paper *Specificity, Conflict, and Focal Point: A Systematic Investigation into Social Media Censorship in China*[43] the authors Yun Tai and King-Wa Fu examine censorship mechanisms on WeChat.

WeChat is an important social media platform in China, which in December 2019, was reported to have more than 1.1 billion monthly active users. It is an umbrella application that bundles many different functions for which Western users would have to use different applications. For example, WeChat allows, among other functions, blogging, private messaging, group chat, or e-payment. Users and companies can publish dedicated pages on which they can post messages and interact with others.

WeChat provides no standardized access to its data through API. So the authors developed a dedicated software to crawl the app, which they termed WeChatscope. The software uses a set of dummy accounts that subscribe to WeChat pages of interest. New URLs posted on these pages are saved and then visited and scraped hourly continuously for 48 hours. At each visit by the crawler, the sites pointed to by the URLs are scraped and meta-data and media content downloaded and saved in a database. If a page disappears, the software saves the official reason for removal given by the platform. The reason "content violation" is given in cases were content is deemed a violation of related law and regulation.

[42] For estimating political alignment of social media users based on digital trace data see Barberá (2015). For the use of digital trace data to infer public opinion and prediction elections see Beauchamp (2017). For critical accounts see Jungherr et al. (2016); Jungherr et al. (2017); Rivero (2019).
[43] See Tai and Fu (2020).

For their study, Yun Tai and King-Wa Fu collected 818,393 public articles on WeChat that were published between 1 March and 31 October 2018. These articles were posted by 2,560 public accounts. Of those 2,345 articles were removed for “content violation”. These articles are what the authors are interested in. More precisely, they are interested in how these articles censored by Chinese regulators differed from others. In order to do so, they first decided to pair each censored article with a non-censored article published on the same account that topically was as similar as possible to the censored article. To identify those pairs of censored/not-censored articles, the authors ran correlated topic models (CTM). This left them with 2,280 pairs of articles published on 751 accounts. To identify the potentially minute difference between articles that led to censorship, the authors used a random forest model. The approach is well suited for identifying meaningful signals in large numbers of input variables.[44]

Using textual terms to predict censorship decisions, the authors identified “perilous” words, terms whose appearance was more frequent in censored articles than remaining articles. They further differentiated between general terms and those that were unique identifiers of entities (such as place names or organizations), times, or quantities. They found that these “specific” terms were especially perilous, increasing the probability for the censorship of articles considerably.

The authors go on and add to this analysis in further steps. But for our purposes, we have seen enough. So, let’s stop here.

The authors connect their very specific findings to literature on conflict, multi-party games, and coordination. Based on the considerations from these theoretical literatures, they conclude that Chinese censorship reacts strongly and negatively to “specific terms” as those might serve as focal points for subsequent coordination or mobilization of users. Thus not only ideas are suppressed but also the linguistic signifiers allowing for coordination of people around specific causes or places.

The study of government censorship is a continuously moving target, as it remains the hare and hedgehog race between the censor and the censored. The study by Yun Tai and King-Wa Fu is therefor surely not the final word on internet censorship or Chinese censorship. Still, their study provides an important puzzle piece in this debate. More important for us, their study provides a highly creative and instructive example of how to use data collected on digital media for the study of speech and government control. Further, it is also an interesting case of how to connect the highly specific and often abstract results from computational text analysis with more general theoretical debates in the social sciences.

All the more reason for you to read the study yourself.

[44]For correlated topic models see Blei and Lafferty (2007). For random forests see Breiman (2001).

2.4.4. It's attention, not support!

Sometimes, we do not only want to learn about what happens in digital communication environments. Sometimes, we want to learn about the world beyond. Digital trace data can help us also in the pursuit these questions. But we need to be a little more careful in reading them, in order not to being misled. A look at studies using digital trace data trying to learn about public opinion is instructive.

Many people use social media to comment publicly about their views about politics or comment on current events. Taking these messages and trying to learn about public opinion could be a good idea. Right from the start, there are some obvious concerns.

First, not everybody uses social media and those who do, differ from the public at large. Also, not everybody who uses social media posts publicly about politics or news.[45] So we are left with an even smaller potentially even more skewed section of the public whose public messages we base our estimate of public opinion on.

Second, public posts on politics and the news are *public*. This might seem like a truism but is a problem for studying public opinion. Many people hold opinions about politics and the news but only a very politically active and dedicated person will posts those publicly, especially in the case of political controversy. By publicly commenting about politics on social media, people demonstrate their political allegiances and convictions for all the world to see. This includes family, friends, colleagues, competitors, and political opponents. Everyone can see what they think about politics and are invited to comment, silently judge, or screenshot.

Third, not only academics are turning to social media to learn about public opinion. For example, journalists, politicians, and campaign organizations are all watching trends and dynamics on social media closely to get a better sense of what the public thinks or is worried about.[46] We might worry about the power they give social media to influence their actions or thinking, given the biases listed above, but this will not make them stop doing it. So anyone publicly posting about politics might not be doing this to express their honest and true opinion. Instead, they might be doing it tactically to influence the way journalists or politicians see the world and evaluate which topics to emphasize or which positions to give up.

As a consequence, studying public opinion based on public social media messages means studying comments, links, and interactions by a highly involved, potentially partisan, non-representative group of people, a portion of whom might be posting tactically, in order to influence news coverage or power dynamics within political factions.

Anyone looking at these obstacles and still thinking, social media posts might be a good way to learn about political opinion truly must be fearless. But as it turns out, there are many who try to do just that. Could it be that they are right?

[45] See Auxier and Anderson (2021).
[46] See Anstead and O'Loughlin (2015); McGregor (2019); McGregor (2020).

In joint work Harald Schoen, Oliver Posegga, Pascal Jürgens and I decided to check what public Twitter messages can tell us about public opinion and what they can't. In the 2017 paper *Digital Trace Data in the Study of Public Opinion: An Indicator of Attention Toward Politics Rather Than Political Support*[47] we compared Twitter-based metrics with results from public opinion polls.

To get access to relevant Twitter messages, we worked with the social media data vendor *Gnip*. We bought access to all public Twitter messages posted during a time span of three months preceding Germany's federal election of 2013 containing mentions of eight prominent parties. Given Twitter's current data access policy, we probably would have simply used the official Twitter API to identify and access relevant messages. Back then, this was not possible. As we were only interested in public opinion in Germany, we only considered messages by users who had chosen German as interface language in interacting with Twitter. This choice might underestimate the total number of messages referring to the parties in question but the resulting error should not systematically bias our findings.

We then calculated a number of Twitter-based metrics for each party following prior choices by authors trying to infer public opinion during election campaigns based on Twitter. This included the mention count of parties in keywords and hashtags, the number of users posting about parties in keyword or hashtag mentions, the number of positive and negative mentions, and the number of users posting positively or negatively about a party. To identify the sentiment of messages mentioning a party, we used a Twitter convention prevalent in Germany at the time. Users identified a message as being in support or opposition to a party by using its name in a hashtag followed by a + or - sign (e.g. #cdu+ or #csu-). This choice might not be replicable across other countries or time, but it is a robust approach to check whether for our case explicitly positively or negatively tagged messages are more strongly connected with public opinion than normal mentions.

This variety of considered metrics reflects the challenge in working with digital trace data discussed by Howison et al. (2011) under the term *event data*. We might have objective ways to count the events in which users mention political parties in a specific way but we have to interpret what this event tells us about their intentions or attitudes.

Continuing, we then compare our Twitter-based metrics with the vote share of each party on election day and compared the resulting error with that of opinion polls. All chosen Twitter-based metrics perform massively worse than estimations based either on the results of the previous federal election in 2009 or polling results. Calculating metrics per day instead of aggregating them over the whole time period shows this error fluctuating strongly over the course of the campaign, with no apparent improvement.

Again, this finding connects back to the concerns raised by Howison et al. (2011) with the term *longitudinal data*. Mentions of parties happen over a long time period before an election. There are no fixed rules by which to decide over which time period to

[47]See Jungherr et al. (2017).

aggregate the mentions to calculate a metric to predict election outcomes. Any choices in this regard are therefore arbitrary.

We get a better sense of the nature of party mentions on Twitter by looking at their temporal distribution. The number of party mentions spiked in reaction to highly publicized campaign events. People commented on politics in reaction to campaign activities, candidate statements, and media coverage. Twitter mentions were therefore indicative of attention paid to politics by politically vocal Twitter users but did not allow to infer the public's attitudes or voting intention.

The paper thus shows that there is indeed something we can learn about the world by looking at social media, it just might not be everything we tend to be interested in. In the case of Twitter and public opinion, it looks like Twitter data can show us what politically vocal people on Twitter were paying attention to and were reacting to. But it did not allow for the inference of public opinion at large or the prediction of parties' electoral fortunes. Getting this distinction right between what we wish digital trace data to tell us (or what other people claim they tell us) and what they actually can tell us giving underlying data generating processes and their inherent characteristics is important. Only by addressing this challenge explicitly and transparently will work based on digital trace data mature and achieve greater recognition in the social science mainstream.

The two studies chosen here offer only a small spotlight on the possibilities in working with digital trace data. As was the case with the examples chosen to illustrate work based on text analysis, there are many other interesting and instructive studies out there. So do please not stop with the studies discussed here, but read broadly to get a sense of the variety of approaches and opportunities open to you in working with digital trace data.

2.5. Learning about the world with computational social science

Computational social science offers us new ways to learn about the world. New data sources emerge, either made available through digitization or bursting into existences through digitalization. New computational methods become available to social scientists. And people coming from different interdisciplinary backgrounds develop new interests in social phenomena and human behavior. This makes CSS into a promising interdisciplinary area. A space of crossroads for people who share interests in social phenomena and human behavior and where people from different scientific backgrounds meet.

Thus it comes as no surprise that CSS comes in many names and has many relations: Some think of it in topical subfields, such as Computational Communication Science (CCS), data science, or sociophysics. Others think of it not as a field but primarily see it through associated methods, such as agent based modeling, network analysis, or computational text analysis. For others still, it is simply a subfield of applied computer science concerned with social systems and phenomena. Each of these perspectives comes with specific insights, strengths, and contributions. Too many in fact to be given appropriate space in this and the preceding episodes.

Still, there are shared concerns among researchers "who develop and test theories or provide systematic descriptions of human, organizational, and institutional behavior through the use of computational methods and practices" (Theocharis & Jungherr, 2021, p. 4). Those include facing the conceptual challenges of translating social science theories and interests into computational concepts and operationalizations, connecting new data sources with established theories or interests while remaining open to new phenomena and behavioral patterns, and the integration of practices and workflows from different disciplines in order to capitalize on the new opportunities emerging from interdisciplinary efforts. Every scholar and every interdisciplinary team across the multitude of CSS subfields face these concerns and challenges to different degrees. There is value in remaining aware of the shared roots, concerns, and challenges of computational social science instead of splintering too early into subfields driven by topical interests or methods. This splintering would risk diluting the collective attention and effort to discussing and working through the challenges of CSS.

Computational social science in all its facettes offers scientists new ways to learn about the world and the impact of digital media on politics and society in particular. True, the emotional response to the incessant interdisciplinary efforts in shaping, contesting, and improving goals, methods and practices of CSS can suddenly shift from being exhilarating to exhausting. For anyone, working in CSS means moving out of their comfort zone of field-specific theories, methods, practices, and workflows. Truly lived scientific interdisciplinarity is challenging. But at the same time also exciting and promising. As in every new area of research, work in CSS is characterized by uncertainties. At the same time, however, it is an area that, precisely because of its freshness and the open questions associated with it, brings unbelievable dynamics thematically, theoretically and methodologically. This makes it, without question, one of the most exciting and rewarding areas of social science today.

2.6. Review questions

1. Please define computational social science following Theocharis and Jungherr (2021).
2. Please discuss critically the promise and challenges of computational social science.
3. Please sketch the typical steps of the project pipeline in computational social science and discuss critically some of the decisions researchers face along the way.

4. Please discuss the data generating process of Twitter and its consequences for what type of question Twitter data is suited and what type it is not.
5. Please discuss why we can expect to find insight in social and political phenomena by the computationally assisted analysis of text? What are potentials? What are limitations?
6. Please define the term digital trace data following Howison et al. (2011).
7. Please discuss along the lines of Howison et al. (2011) for what type of research project digital trace data are suited and for what type they are not?
8. What questions do you have to answer to ensure your work with digital trace data is valid according to Howison et al. (2011)?

3. Data

Data are crucial for the discussion of politics and digital media. Understanding the core concepts and issues arising from the quantification of social and political life and the resulting data is important for engaging in many of the subsequent controversies of the uses of digital media in politics. Digital media, devices, and sensors collect data documenting the world, society, and human behavior. This has been seen by some as a measurement revolution, providing many new avenues for the social sciences as well as new business opportunities in the economy. Perceived potentials and dangers in the increases in the volume and breadth of coverage of digital data are broadly discussed, but it is also important to examine how these new data sources relate to the social or behavioral phenomena they supposedly cover. New data riches have to be translated into meaningful measures of phenomena of interest and society.

Digital data are the newest step in the quantification of reality and social life. This leads to three important questions, that need to be addressed in research: How do we turn the world into numbers? What do these numbers enable us to do? And how do we as a society structure this process and define rules and regulation of what is and is not allowed? These are powerful questions framing research projects and they will help us structure this chapter.

This chapter will start by presenting core issues arising from the quantification of social and political life. Following this, it will introduce readers to hopes and limits connected with the term *big data*. This will be followed by the discussion of fundamental questions of scientific measurement. Once these foundational concepts and questions are discussed, the chapter will turn to how political organizations and other actors are translating the world into data and are using data to better understand their environment, the effects of their actions, control their members, and improve their work. This then will form the basis of discussing the trade-offs between increasing the capacity of organizations and states through data and the considerable privacy concerns of people whose data are collected, processed, and potentially shared with others.

The chapter will introduce students to core issues to data collection, use, and governance through digital media and devices. They will learn core concepts in the discussion about data and measurement and will encounter key examples and trade-offs that need to be considered in the examination and discussion of data uses by political organizations and actors and the governance of data collection and data use through regulators and the state.

3.1. Data and quantification

Computation and digital media have reinforced interest in the opportunities and challenges of data and quantification in various societal fields. Data promise the objective observation of the world, insights into hidden patterns and causes, and foresight into future developments or the outcomes of specific choices or interventions. Data are crucial for making the world legible and changing it through targeted interventions. In this sense, data provide the basis for the modern world and its scientific understanding.

Digital media and computation have increased the availability of data in ever more fields. They also have extended analytical opportunities. It is no surprise then to find digital data to be a topic of both enthusiasms as well as fears. But before we can take a closer look at both hopes and fears, we first have to be clear about what exactly we mean by the term data and what the preconditions are for data providing a true representation of the world.

> **i** Defintion: Data
>
> Data are symbolic representations of entities in the world and their relationships. They are the result of some measurement process that maps entities' properties to numerical values of a variable. The numerical relationship between variables represents the relationship between entities in the real world.[1]

This definition points to important features of data that make them useful but also constitute limits to their use. Data provide reduced symbolic representations of specific characteristics of entities of interest. Numeric symbolic representation allows for the documentation of entities, events, or behaviors. Mathematical calculation allows the identification of causal or correlative connections between recorded entities. Models developed based on these connections allow the prediction of likely outcomes given specific inputs. Data are therefor an important feature and allow for a deeper understanding of the past, control of the present, and even provide a glance of the future. But to do so, we first must translate entities of interest in the world into reduced symbolic, numerical representations. This process is called *quantification*.

> **i** Defintion: Quantification
>
> Quantification refers to the process of translating entities, or selected characteristics of entities, into numbers.[2]

Quantification, the reduction of entities to numbers, is powerful. It allows to document

[1]This definition is based on Hand (2007), p. 7 and Hand et al. (2001), p. 25.

[2]For a popular account of how quantification allows to learn about the world see Hand (2007). For a popular account of quantification and related issues from a sociological perspective, see Mau (2017/2019). For an academic review of associated sociological research see Mennicken and Espeland (2019).

past and present, as well as to plan for the future without having to account for the richness of entities in the world. The translation of entities of interest into numbers provides a model of entities, their relationships, and potentially the world. It also allows for the performance of mathematical operations. These operations do not only speak to the numeric representation, the model, but also allow for inferences on the underlying entities and phenomena, potentially uncovering otherwise hidden patterns and causal relations, or the prediction of future developments or effects of interventions. Data not only represent the world but also allow for targeted interventions.

While quantification offers impressive new opportunities, it also has limits. In reducing entities in the world to numbers, quantification reduces the world to countable signals. This means disregarding much of the world. This can be useful, but this can also be dangerous. In the process of quantification unimportant features of entities might be counted, while important might be missed or disregarded. Or in the process of reducing entities to numbers, their essence might be lost. Accordingly, anyone working with data will have to consider the underlying data generating process and if necessary critically interrogate it. We will return to this, when we will be looking at measurement.

Also, quantification not only represents the world, in some sense it is recreating it. Quantification assigns entities, or selected features, to categories. These categories then become the structuring devices people use to learn about the world and shape it through interventions. The definition of categories and the interpretative mapping of entities and features become important constitutive features of data-driven reasoning. While the analysis of data follows objective rules of probability and mathematics, the process of defining categories and assigning entities is constructive and interpretative. Accordingly, these aspects of quantification need to be subject to critical interrogation.[3]

Quantification and resulting data promise insights about past, present, and future. But to distill insight from data, we must not only analyze data and their inherent relationships. We also need to actively account for limits of quantification and measurement approaches in relation to our interests. Otherwise even the most impressive efforts in quantification and quantitative analysis will lead to mistaken views about the world.

3.2. Big data

Digital media and digital technology more broadly have provided society and researchers with new data sources, that appear to be cheap and ubiquitous. In the past, collecting data was difficult and expensive. People had to actively go out, count, and record their objects of interest. In social science, they had to run expensive surveys in which people visited or called respondents and had them answer questionnaires. This made data expensive to collect.

[3]For the classic development of this argument see Foucault (1966/1994). For an account of the history of quantification see Porter (1995/2020). For a sociology of classification and its consequences see Bowker and Star (1999).

Digital technology changes this. Existing, analogue data sets are digitized and thereby made broadly available and accessible for computational analysis. At the same time, digital technology collects original data continuously. Examples include digitally enabled sensors that collect information about their environment and translate signals into data. Sensors like this can be found in cars, mobile phones, and many internet enabled smart devices. Additionally, people's interactions with digital services create data traces, documenting their behavior.

These data are collected and used by device makers and service providers and can serve as basis for the improvement of existing devices and services or the creation of new ones. While these data sources are primarily available to commercial actors, academics can gain access to selective slices of these data through standardized open protocols – such as application programming interfaces (APIs) defining rules and limits of public data access. Alternatively, academics can pursue privileged collaboration with companies. Digital technology thus promises a new data abundance for business, society, and academia. Data are suddenly cheap. This created a lot of enthusiasm among business consultants, journalists, and some scientists about the supposed potentials of unknown data riches. The term *big data* has become the focus point of these enthusiasms.[4]

Originally, the term big data referred to large data sets that technically could not be held or processed in one database or on one machine. But soon it gained popularity as a term covering the new data riches provided through digital technology and the associated economic benefits. One of the earlier characterizations of big data was proposed by the business consultancy Meta Group and – after its subsequent acquisition by the consultancy Gartner – has become known as the Gartner definition.

> **Definition: Big data**
>
> " "Big data" is high-volume, -velocity and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making." (Laney, 2001).

Since the original formulation of this definition in 2001, many additional *Vs* have been suggested in covering new or neglected characteristics of big data. But the original three Vs – volume, velocity, and variety – should suffice for our purposes. They illustrate the suspected promises of big data very clearly, while also pointing to one of the crucial shortcomings of subsequent efforts and debates. The definition points to its origin, being interested in the technical issues arising from handling data sets made available through digital technology. Data come in great volume, surpassing the capabilities of standard computational set-ups and statistical methods. Data come in great velocity, on the one hand allowing for real-time analysis of unfolding phenomena, but at the same time also providing challenges by shifting features within the data and data generating processes.

[4] For a measured discussion of big data see Holmes (2017). For the uses of big data in the social sciences see Schroeder (2014).

Finally, data come in great variety, such as text, image, video, audio, or meta-data. As a consequence, data need to be structured in order to allow for subsequent analysis.

The definition puts its focus clearly on the technical features of big data, and not their characteristics as representation of objects in the world. This can be forgiven, given the definition's origin with a technical consultancy trying to prepare its clients for future opportunities or challenges. Unfortunately, the same focus on technical features, simplistic fascination with size and volume, and an overwhelming disregard for issues arising from the translation of entities in the world in symbolic representation have dominated the subsequent debate and use of the term. This is less forgivable.

Nearly all discussions of big data treat them as true representations of whatever happens to be of interest to the speaker. This could be buying intentions, psychological traits, or political affiliations. Whatever happens to be the desire of the inquirer, big data shall provide. If we would follow the big data boosters, we could expect to find everything and all in big data, no matter what we are looking for.

This is of course not the case. What goes for other types of data also holds for big data. The translation of entities into symbols entails the construction of their representation. Quantifying the world depends on creatively translating objects of interest into measurable signals and translating those into data. These important interpretative steps are just as important in the age of big data as before. Arguably they are even more important now than ever, since people now have access to data not primarily collected with their analytical goals in mind. To realize the potential of these found data for research, demands for very active and creative steps in quantification. Lets take a look at one of the most prominent and talked about categories of big data: digital trace data.

> **Definition: Digital trace data**
>
> "(…) records of activity (trace data) undertaken through an online information system (thus, digital). A trace is a mark left as a sign of passage; it is recorded evidence that something has occurred in the past. For trace data, the system acts as a data collection tool, providing both advantages and limitations." (Howison et al., 2011, p. 769)

Examples for digital trace data include messages and metadata documenting contributions, behavior, and interactions of users of social media services, like X, Reddit, or YouTube. Due to their comparatively easy accessibility through application programming interfaces (APIs), data like these have become prominent in the work of social and computer scientists. Additionally, they provide the basis for many consultancy and media analysis services. Digital trace data therefor have come to provide representations of social systems and human behavior for academic work as well as professionals in media and other businesses.

But there are specific challenges to symbolic representations of the world based on digital trace data. While early enthusiasts expected digital data sources like these to provide allencompassing insights into human behavior and social systems, time has surfaced severe limitations and sources of bias. Two sources of bias are especially relevant in working with big data built from digital trace data: biased coverage and biased behavior.

> **Definition: Bias in data sets**
>
> A bias in a data set refers to a systematic error or deviation from a representative sample. It occurs when a data set doesn't accurately represent the population or domain it claims to, leading to skewed results. For example, a dataset claiming to represent the general population would be biased if it primarily consisted of individuals between the ages of 18 and 36.

Data collected by digital services, while high in volume, are still limited by who is using these services to begin with. This means that a service's active user base determines the share of social life that is covered by data collected on it and that can meaningfully quantified and modeled. The active user bases of even the biggest social media services are highly concentrated within specific demographics and systematically exclude specific socio-demographic groups. This is true for age, by systematically underrepresenting older people, but potentially also for political leaning, with different services catering to partisans of different political stripes. Some of theses biases in the coverage of these data sets can in principle be identified, other remain hidden be it for lack of information about users or be it for unobserved shifts in the composition of the user base contributing to a given data source.

There is also the risk of measuring biased behavior through digital trace data. Digital trace data only offer us access to behavior of people mediated through the affordances, interfaces, and code of the providers of digital services. For example, we might be interested in the attitudes of X users. But the only thing we get is their public posts. These might be true expressions of opinions, thoughts, and reactions. But they might just as well be the results of a strategic public performance, in where users play a public role and post fitting messages. Or they might be the result of behavioral incentives provided by the services themselves, such as algorithmic content selection.

These are just a few examples for how the information system recording and providing data is also shaping data. So in the end, we cannot be sure, that we measure human behavior beyond the digital environment it was collected in. For the purposes of using the data for inferences beyond the confines of the system generating them, the data therefor are clearly biased and unsuited.[5]

These limits of digital trace data, and big data more generally, are beginning to feature more clearly in the debate. Early on, these concerns were largely ignored in a spirit of daring can-do. But over time, the limits of work based on data like these became more

[5] For a foundational discussions of how to interpret digital data traces see Howison et al. (2011). For a discussion of the mediation process translating phenomena, behavior, and attitudes into digital data traces see Jungherr et al. (2016).

apparent. Also, the more these data came to matter beyond the confines of academia, the more the accuracy of diagnoses and prediction based on these data came to matter. An important factor in this was the growing use of large data sets in applications and services enabled through artificial intelligence (AI).[6]

The richness and volume of newly available digital data are a core contributor to recent advances in artificial intelligence (AI). With the growing awareness of AI-enabled opportunities as well as associated risks, biases in data sets have received increasing attention. At the core of associated concerns are fears of biased data sets leading to biased outcomes of decision making based on them. Especially automated algorithmic learning and decision making have received much attention in this discussion.[7]

Questions of who and what gets counted in big data are thus increasingly getting more attention. Still, there remains much to do to overcome the current naive positivism prevalent in the work with big data. One core issue in the work with big data, and in quantification in general, is how the entities in the world get translated into symbols. This brings us to the question of measurement.

3.3. Measurement

By translating observations into numbers, quantification provides opportunities for new and important insights about the world and objects of interest. But as with any translation, making entities and phenomena countable means also losing some of their features. Quantification makes some things visible, while hiding others. To better understand this process, we have to examine how things become numbers, we have to examine *measurement*.[8]

> **Definition: Measurement**
>
> "The assignment of numbers to represent the magnitude of attributes of a system we are studying or which we wish to describe." (Hand, 2004, p. 3)

Today, measurement is pervasive and underlies much of contemporary life. The economy runs on data, models, and prognoses allowing for the optimization of production, the planning for inventories, or the setting of prices. States use them for policy planning, designing interventions, and the allocation of state resources. Scientists depend on them for the explanation of the world and the prediction of trends. We ourself, carry personal trackers that measure our movements and heart beat to determine our fitness and track

[6] For discussions of research with big data see Salganik (2018). For a discussion of how to account for the limits of digital trace data in the social sciences see Jungherr (2019).

[7] For critical discussions of data sets in artificial intelligence and algorithmic decision making see Barocas and Selbst (2016), Bolukbasi et al. (2016), Buolamwini and Gebru (2018), Mayson (2019).

[8] For comprehensive discussion of measurement in various scientific fields see Hand (2004).

our training progress. This pervasiveness of measurement makes it difficult to imagine that this might have been different.

The historical roots of measurement and data are very prosaic. They lie in the late Middle Ages and start with accounting. International merchants started to rely on numbers to keep track of orders and inventory, allowing them to run intricate and far-reaching international trade networks. Soon states started to adapt to these innovations in order to increase their ability to collect taxes and raise armies. The foundations of measurement are therefor very practical and lie at the heart of the processes that gave rise to the modern society.[9]

Looking more closely at measurement, we can differentiate between two measurement approaches: *representational measurement* and *pragmatic measurement*.

On the most fundamental level, measurement can be the mapping of empirical relationships between distinct objects by quantifying specific observable attributes. The resulting numerical relationships represent the empirical relationships between the objects. This is *representational measurement*.[10]

For example, we can examine the relative strength of a protest movement over time by counting participants in protest events at different points in time. If we count more participants at events over time, we can conclude that the movement gains strength. If we count fewer participants over time, we can infer the opposite. This type of measurement is straightforward. Numerical values assigned during the measurement process are constrained by the relationship of empirically observable characteristics. But as we will quickly find out, most phenomena of interest – especially in the social sciences – do not lend themselves to this direct measurement approach.

> **Definition: Representational measurement**
>
> The assignment of numerical values to variables, in such a manner that the numerical relationships between variables corresponds to empirically observed relationships between measured entities. This assignment is not arbitrary; instead, it's directly informed and constrained by the patterns and properties observable in the actual entities being studied.

Things become a little more difficult in the measurement of phenomena that do not lend themselves to direct observation but that still merit quantification. The state of the economy, collective happiness, public opinion, people's psychological traits or attitudes: none of these concepts can be observed directly but all are subject of measurement. This requires *pragmatic measurement*.[11]

[9] For two informative accounts about the historical origins of measurement see Crosby (1996), Deringer (2018).

[10] See Hand (2004), p. 27–52.

[11] See Hand (2004), p. 52–60.

Concepts like these do not directly map to empirically observable and directly comparable characteristics. So to measure them scientists have to first decide on how the concept of interest manifests in measurable signals. They have to construct the measurement, which means deciding on what the target concept is and how it should manifest indirectly in observable objects. While representative measurement is directly constrained by empirically observable properties of entities, pragmatic measurement is constrained by shared conventions among those doing the measuring about what constitutes valid measurement approaches and their practicality in use.

Definition: Pragmatic measurement

The assignment of numerical values to variables to approximate non-directly observable concepts. This is achieved through the theoretical mapping of these concepts to empirically observable indicators. While these measurements are grounded in shared conventions and are valued for their practical utility, their validity is contingent upon the robustness of the link between the theoretical concepts and their empirical proxies. Such connections, and the inferences drawn from them, remain subject to interpretation and critique.

One example for pragmatic measurement from political psychology is the measurement of latent attitudes. For example, *libertarianism* is an abstract concept, capturing a set of ideas about protecting the rights of individuals against the state.[12] If we want to find out whether people agree or disagree with libertarianism, we cannot observe this directly. Libertarians share no physical characteristics making them different from others, egalitarians say, that we could empirically observe. Instead, scientists have proposed a set of statements, each of which corresponding with some aspect of libertarian ideas or convictions.[13] Those survey respondents who agree with these statements more strongly or consistently than others, we can label libertarians. Of course, the statements used to measure libertarianism are open to interpretation and critique, whether they truly capture the concept of interest or not. Also, we could use other ways to construct a pragmatic measurement of libertarianism. For example, we could look at digital media posts and classify them as being in accordance or conflict with libertarian ideas. Doing so, would allow us to classify users as libertarians based on their propensity of posting content expressing connected statements or ideas.

Pragmatic measurement is a powerful approach to examine and structure the world, especially with regard to concepts or phenomena that do not lend themselves easily to direct empirical observation. At the same time, pragmatic measurement does not only represent reality, it also constructs it by defining ways on how to measure concepts and phenomena which cannot be directly observed empirically. Accordingly, the process of defining these measures warrants close observation and critical interrogation. Especially

[12] For a history of ideas of libertarianism see Zwolinski and Tomasi (2023).

[13] For a proposition of how to measure individuals' alignment with libertarianism through survey responses see Iyer et al. (2012).

with regard to whether the proposed measures correctly and fairly represent the supposed target concepts or wether they unfairly represent institutional, economic, or political interests.

Example: Public opinion

The quantification of the relative strength of factions in political competition offers examples for both representational and pragmatic measurement. The most direct way to determine the relative strength of political factions is counting votes after an election. This is a case of representational measurement. The vote count is a direct representation of the empirically determined votes in favor of either party.

But although this measure has a clear and easily identifiable empirical counterpart, the interpretations of what these votes *mean* diverge strongly. Some will claim that the winning party has a clear mandate of exercising their platform. If not, why would people have voted for it otherwise? Others will claim that being successful at the polls does not provide any indicator that people actually support the positions of a party. They could merely have voted for it out of protest given their opposition toward the political status quo. Still others might point out that a vote for a party does not speak of policy support or protest. It might simply be a tactical choice to create conditions under which parties might be forced into a coalition government. These are just three possible interpretations among many others. These examples are not meant to be exhaustive. They simply illustrate that even for data that are the result of a direct and simple representative approach, interpretations of their meaning diverge widely, always depending on the interests and intentions of those doing the interpreting. This is even more so with measures of public opinion.

In between elections political parties' fortunes might shift. But without votes to count, their relative strengths are hard to quantify. It is here, that public opinion research comes in. Public opinion is an abstract concept with no direct empirical expression. This makes public opinion an object for pragmatic measurement. To determine public opinion at any time people are surveyed, be it for their support of parties, assessment of politicians, or opinions on policies or political topics. While answers in these surveys can be counted, the reading of these numbers is open to interpretation and construction. Accordingly, public opinion surveys serve interested parties as instruments in the pursuit of their goals. The specifics in the measurement of public opinion – the samples drawn, questions asked, weighing procedures applied, and interpretations – are subject for critical reflection and interrogation. The same goes for the impact of the quantification of public opinion and political will, in general.[14]

It is important to keep the distinction between representational and pragmatic measurement in mind. Both representational and pragmatic measurement translate phenomena into seemingly objective numbers. But while in the case of representational measure-

[14]For more on public opinion, its history, and uses see Herbst (1993); S. Igo (2007).

ment these numbers are constrained by directly empirically observable characteristics, pragmatic measurement defines what empirically observable signals are seen as an expression of underlying empirically unobservable concepts or phenomena. In the best case, this makes invisible but important aspects of the world visible and opens them up for documentation, tracking over time, prediction, and targeted interventions. In the worst case, pragmatic measurement hides its interpretative and constructive characteristics behind the seeming numerical objectivity borrowed from the more narrowly empirically constrained representative measurement approach. In interpreting and critically interrogating data, we therefor need to take into account by which approach they were measured.

Additionally, measurement might fail on a practical level. On a very foundational level, there is the question of a measure's *validity*.[15] A measure is valid if it captures the phenomenon of interest correctly. For example, asking people to state the frequency of media use in the recent past and treating respective answers as objective would represent an invalid measurement. [16] Personal recollections of something as ephemeral as media use cannot serve as documentation of actual media use.

A measure also needs to be *reliable*.[17] It needs to produce accurate quantified representations of the empirical features of an entity. A simple example of an unreliable measure is a scale that returns wildly fluctuating weight assessments at repeated weighing of the same object. An example for the reliability of a more complicated measurement approach would be if two people tasked with content analysis return independently from another the same content classification. Conversely, if they would fail to do so, the measure would be not reliable.

There is also the issue of *bias* of measures. [18] Measures are biased if their results contain systematic errors. To stay with the example of a scale, a scale would be biased if it systematically adds or subtracts a specific value to each measurement. For an example for a more elaborate measurement approach we can turn to public opinion surveys. A biased sampling process might lead to specific parts of the population being underrepresented among survey respondents. Accordingly, their preferences cannot be registered, which in turn leads to the return of a biased measure of public opinion. Questions of biased datasets and results of subsequent analyses have gained high relevance in the discussions about artificial intelligence and algorithmic decision making. By relying on datasets that systematically over- or underrepresent marginal or historically discriminated groups in society, risks the continuation of discriminatory or marginalizing practices into present and future by reliance the results of seemingly objective data analyses. We will return to this issue later.

Finally, we also need to consider that in some cases the measured objects *react to measurements* and change their characteristics. This is especially true for the social sciences.

[15] For a discussion of measurement validity see Hand (2004), p. 131–134.
[16] For more on the limits of self-reports in media research see Prior (2009).
[17] For a discussion of measurement reliability see Hand (2004), p. 134–145.
[18] For a definition of bias see Hand (2004), p. 133–134.

Publishing prognoses on expected economic developments will lead to people adjusting their behavior in order to profit from these developments, thereby potentially reinforcing the predicted trends or rendering them mute. Similarly, publishing data on the prevalence of hate speech on digital platforms can lead to the platforms adjusting their policies. This might mean filtering potential hate speech at the point of publication, deleting following moderation after publication, or shadow banning it making it invisible to all users but the author. These adjustments change the prevalence of hate speech but are also invalidating the previously used measurement approach. Quantifying entities that have the ability to react to measurement approaches and their results can therefor limit the long-term use of these approaches and present non-trivial challenges to the validity of identified patterns or predictions over time.

These limitations and challenges are well understood for traditional measurement approaches with clearly defined measurement targets and instruments. Here, we have clearly defined quality control procedures that allow to assess potential biases or reliability issues. These issues are less well understood – or even acknowledged – for newer measurement approaches, as in the work with big data.

In light of this discussion about measurement, its characteristics, limitations, and challenges, we can conclude that quantification, the symbolic representation of the world in data, does not necessarily lead to an objective and correct model of the world. Instead, measurement can both on a conceptual and on a practical level introduce errors and biases leading data to misrepresent the world and the relationship between objects of interest within it. The importance of data in contemporary societies means that we need to actively account for this by actively and transparently constructing measures as well as critically interrogating those presented by others. This is especially important in the work with new and innovative measurement approaches and data types, as for example those found on and produced by digital media.

3.4. Observing the world through data

Data provide a reduced representation of the world and promise to uncover hidden patterns and connections between entities that remain invisible in their full, unreduced manifestation in the world. This allows people and organizations to make sense of the world, and increase their level of control. This includes the state, political parties, or companies. These actors need to understand unfolding developments, outcomes of interest, contributing factors, and have an understanding of how they are able to shape outcomes in their interest. For this, they need data. By reducing the complexity of society and the world, data make the world legible and thereby provide the opportunity for intervention. This includes rationalizing and standardizing society and the world by creating terms, concepts, and categories.

Data structure entities in the world according to standardized categories, allowing actors to make sense of reality, assess problems, intervene as they see fit, and evaluate their success. But this process of making the world legible means also engaging in what Hand called pragmatic measurement.[19] Actors have to actively and creatively translate entities in the world into standardized and countable objects. This holds either for entities that are directly of interest to them or those they take as signals of underlying phenomena they want to track or influence. This is a process of translation and abstraction in which meaning and details are lost. This risks overlooking or misreading important elements. Still, some sort of reduction is necessary to read and interact with society and the world. The sociologist James C. Scott illustrates this in the discussion of the work of state officials:

> "Officials of the modern state (…) assess the life of their society by a series of typifications that are always some distance from the full reality these abstractions are meant to capture. (…) The functionary of any large organization"sees" the human activity that is of interest to him largely through the simplified approximations of documents and statistics (…). These typifications are indispensable to statecraft. State simplifications such as maps, censuses, cadastral lists, and standard units of measurement represent techniques for grasping a large and complex reality; in order for officials to be able to comprehend aspects of the ensemble, that complex reality must be reduced to schematic categories. The only way to accomplish this is to reduce an infinite array of detail to a set of categories that will facilitate summary descriptions, comparisons, and aggregation."
>
> J. C. Scott (1998), p. 76–77.

But for this to work, the translation of society and the world into categories and numbers, quantification must validly account for the actual objects of interest and contributing elements. This makes this not an effort in simple *representational measurement* but instead *pragmatic measurement*, foregrounding the challenges associated with the later. This goes double for the questions of whether targeted interventions are successful or not. A correct reading depends strongly on the fitness of the pragmatic measurement process underlying the translation of the world in data. If the map of the territory is wrong or unfit, navigating by it will not lead you to your destination.

Example: Campaign dashboards

Many parties in Western democracies support their local chapters and campaigners with digital information systems providing standardized solutions for voter outreach and fundraising. Services like these were first popularized by Barack Obama's campaign for the 2008 US-Presidential race. The campaign provided its local chapters with a centralized web-enabled service. This allowed the campaign to coordinate voter outreach by creating call- and walk-lists for volunteers targeting likely voters, that the central campaign office had identified as promising. In the subsequent

[19] See Hand (2004), p. 52–60.

2012 Presidential race, the functionality was extended by mobile apps by which individual campaigners could log voter contacts, making the visible to campaign headquarters.
Obama's success and the perceived contribution of data-enabled mobilization efforts popularized these services. In the following years, the functionality of the original services was extended and various new services were developed by vendors affiliated with other parties and in different countries. So by now services like these are a common feature in international campaigns in democracies, all be it with different functionalities and varying importance to the campaign depending on local data-privacy laws, resources spent on development, and the reliance of campaigns on local voter outreach.
What these services share is that they provide campaign headquarters with the opportunity to observe and shape local campaign activity more effectively than before. When in earlier times, headquarters depended on local chapters reporting back to them on how campaigning was going, these digital services allow headquarters to track local activities directly and to coordinate activities. By having volunteers tracking their activities, headquarters get a view in aggregate and detail of how well the campaign is going and the energy in the field. By providing local chapters with outreach priorities, either targeted directly at addresses or more broadly at the street level, campaign leadership can coordinate activities according to centrally decided strategy.
Digital campaign technology therefore clearly supports central campaign bureaucracies in collecting and aggregating data that make local campaign activities visible and allow central bureaucracies to shape them in turn. But the successful application and deployment of these services depend on more than mere technology. What are the legal conditions for campaigns to collected and use data about prospective voters and volunteers? How much resources can an organization spend on development and training? How strongly does a campaign depend on local voter outreach? How high is data-literacy within the organization allowing for meaningful quantification, analysis, and subsequent action? These are some of the contextual conditions that need to be considered in trying to understand these services' contribution to campaign strategy, operations, and ultimately to electoral victory or defeat.[20]

One important way for organizations to observe the world and shape it is metrics-based management.

Definition: Metrics

Metrics are quantified measures that allow organizations the tracking, status assessment, and evaluation of predefined processes.

[20]For more on the practice of digitally assisted door to door campaigning see R. K. Nielsen (2012). For more on the development and maintenance of digital campaign technology see Kreiss (2012) and Kreiss (2016).

Organizations use metrics to increase efficiency and productivity. They break down their processes in small identifiable steps and track the achievement of these steps together with inputs and outputs. If done well, this allows for the detailed monitoring of important processes, identification of inefficiencies, the design of interventions to improve efficiency or productivity, and central control.

The idea of metric-based management is an offspring of the Taylorist approach to "scientific management".[21] In the early nineteen hundreds, the engineer Frederick Winslow Taylor, advocated the calculation of standard levels of outputs for each job contributing to factory outputs. Workers who hit these metrics or outperformed them were financially rewarded, while those who fell behind were payed less. These ideas became very popular. They were ported by US defense secretary Robert McNamara to quantify and track progress during the Vietnam War, they were the inspiration to metric-based management, moved into the public service sector under the term "new public management", and recently experienced a revival in Silicon Valley tech companies under the term "objectives and key results (OKRs)".

The idea behind these schemes is usually the same: define a set of key steps important in the pursuit of an organizations' goals and track their achievement over time. By tracking inputs – such as raw material or working hours – and outputs – such as units produced – managers can supervise the process, incentivize the behavior of workers, or look for hidden inefficiencies. At least that's the promise.

In practice, management by metrics depends on the suitability of metrics covering the relevant inputs and outputs. This sounds trivial, but often what is easy to measure is not necessarily relevant, while what is relevant is not necessarily easy to measure. By failing in the pragmatic measurement process, by misconstructing measured signals, metrics-based management can lead an organization to focus on achieving metrics instead of pursuing what is necessary to produce their outputs.[22] The fundamental problem in the measurement or quantification of society and the world therefore also hold in these cases.

Through the greater availability and pervasiveness of data, enabled by digital technology, the reach of metrics has extended. This is true for established businesses that now start to look to digital metrics to assess their success of failure, such as news media assessing the success of articles or journalists by the number of views or interactions they generate in digital communication environments. Or this could be actors who up until now had no access to data documenting their success or failures in real-time, such

[21] For Taylorist "scientific management" see F. W. Taylor (1911). For metrics in the Vietnam War see Halberstam (1972). For metrics based management see Wooldridge (2011). For "new public management" see Pollitt and Bouckaert (2017). For "objectives and key results (OKRs)" see Doerr (2018).

[22] For a spirited critique of metrics based management see Muller (2018). For reactivity to metrics by those measured see Espeland and Sauder (2007)

as politicians or micro-celebrities.[23] These actors also now find many digital metrics available to them, supposedly tracking their fates while providing them opportunities for interventions targeted at improving it.

By a multitude of publicly visible or private metrics, digital technology provides many actors new opportunities for making legible their environment in which they pursue their goals. Unfortunately, through their simplicity and instant availability, these metrics are often not critically interrogated as to whether they actually speak to the goals of actors using them.

For example, while a politician can easily track likes on Facebook, it is much harder to determine whether these likes actually correspond with preferences of her constituents. So, in the worst case, she might optimize her positions and rhetoric for the audience of her Facebook posts while losing from sight, the preferences of her actual voters.

If not used consciously and critically, metrics can easily create a quantified cage for actors relying on them. By making it easy to optimize toward them, metrics can provide the illusion of control to central authorities using them to govern an organization. But while an organization might do well according to the abstract quantification of the world – metrics – its actual fate might be much less benign. As J. C. Scott (1998) points out, while an abstract map is necessary for the centralized control of complex systems, the same map can mislead governing units and navigating through it can lead to failures, small and large. This is just as true for metrics in the age of big data, as for those coming before.

3.5. Shaping the future through data

Symbolic representations of the world do not only allow people and organizations to see the world, they also allow them to shape the future. Building mathematical models of the world, provides new ways to form expectations about relevant future developments as well as effects of specific actions and interventions. Having access to better models about the world allows people and organizations to position themselves with greater foresight in the world regarding future developments and effects of their actions. This allows them to outperform the competition and shape the future according to their interests.

> **i** Definition: Model
>
> In the given context, a model is defined as the formal expression of structure within data. While a model might represent structures between real-world entities, such

[23] For the use of metrics by news media see Christin (2020). For the use of metrics in policing see Ferguson (2017). For the use of metrics in political activism see Karpf (2016). These studies are not only helpful in learning about these specific cases. They are also helpful by providing templates of how to approach the use of metrics in organizations and associated effects scientifically.

a correspondence is not guaranteed. The fidelity of this representation depends on the accurate portrayal of these entities within the dataset in question.
Models can be categorized as:
Descriptive: These models articulate a structure found in data without making claims about causality or explaining relationships.
Mechanistic: These models formally denote relationships between variables based on theoretical expectations. In doing so, they translate real-world theories into formal, numerical representations.

Models rely on the numeric representation of the world. This allows for the mathematical identification of structures between sets of symbols. As far as these symbols provide an accurate representation of the world, structures identified through mathematics should also account for structures between the represented entities in the world. This includes the identification of relationships between variables and the entities they represent, such as their systematic co-occurrence or the presence of one entity causing that of another.

For example large online retailers can analyze buying patters and build statistical models of which products tend to be bought together or subsequently. Of course, more interesting are models successfully identifying events or actions that cause users to buy an item. By having information about exposure to ads, clickstreams, or buying history the identification of causal models might be possible as well. More generally, the same logic applies to recommender systems, that point users to content or products given their own prior browsing or buying history or that of other users that resemble them in for the model relevant features.[24] These models allow online services – be they retailers, news sites, or content providers – to shape the information users see in ways that in the past generated outcomes of interest – be they sales, clicks, or interactions. Technology companies shape the future by structuring the digital environments users find themselves in according to insights from models.

Given enough resources, data access, and dedicated personell organizations can capitalize on the opportunities of data analysis. This includes governments, parties, or collective action organizations. Still, while the promises of data-driven practices are often proclaimed and discussed, in fact the realization of their potential carry considerable demands in resources, analytical capabilities within an organization, quantifiable and quantified aspects of interest, and an organizational culture open to integrate data-driven insights and adjust tactics and strategy accordingly. These preconditions are much less often found or even discussed, than broad speculations about data-driven promises. Accordingly the potential of the uses of data and models are often overestimated and need to be critically interrogated for specific cases and contexts.

An important feature of having models about the workings of the world – or at least that part of the world one is interested in – is that models allow to predict future developments or the effects of actions and interventions. As long as they are accurate,

[24] On the workings and uses of recommender systems see Narayanan (2023).

of course. Having accurate models about the world allows for the prediction of the future. This includes future large-scale developments – as in models of the economy or the climate – as well as future behavior of people – as in buying decisions or consumption patterns. As David J. Hand puts it:

> "The mathematical model represents an understanding of how things work, and this understanding can then be used in other situations where data collection is unfeasible. (...) using mathematical models, based on data which have been collected in simpler situations, we can produce predictions (...)."
>
> Hand (2007), p. 68

Model-based prediction formalizes knowledge gained on the basis of quantified representations of the world. They allow people, organizations, and societies to plan for the future, adjust their behaviors and actions, and potentially even shape it to their benefit. Of course, this does not mean that these predictions will always be correct. But if used diligently and in knowledge of their limitations, they can help.

Example: Prediction in campaigning

Campaign organizations and political parties are also experimenting with the use of data-enabled practices and model-based predictions. Again, the Obama campaigns of 2008 and 2012 are often quoted examples that serve as template for expectations for data in politics. But – as always – it pays to look closely.

Much has been made out of the Obama campaign's supposed ability to identify likely voters by predicting people's vote choices based on available data. But a close study of the precision of these targeting models by Hersh (2015) showed that the predictive power of the campaign depended strongly on the data available to it. The campaign relied on official voter files. In some states these voter files contained information on whether a voter registered as Democrat, Republican, or Independent. Hersh showed that in states where this data was available the Obama campaign was able to target their outreach much more precisely than in states where these data were not available. The quality of predicting vote choice thus clearly depended on the campaign knowing the self-declared party affiliation of prospective voters.

Other prediction tasks performed better though. In their account of the Obama campaigns' data-driven practices Nickerson and Rogers (2014) show that the campaigns used experiments to generate data that allowed them to predict the best e-mail wording to maximize donations from their supporters.

Arguably, this practice allowed the Democratic Party to get too good at eliciting money from their supporters. Key consultants and commentators have been arguing that this reduction of political supporters to their likelihood to donate have led the party organization to lose sight of the more meaningful aspects of political organizing and over time weakened the ability of the party to mobilize and count on the support of their volunteers.[25] Despite considerable but isolated successes quantification, modeling, and prediction might thus have come to weaken the organization in the mid- to long-term.[26]

An even more elaborate use of quantified representations and models is the use of simulations. Here, models are used to simulate systems, such as societies, traffic, or human movements in confined spaces. Simulations lie at the core of product development, architecture of public spaces, and design. They are also used in the social sciences in order to test dynamics, interaction patterns, and adaptation within complex social systems.[27]

Taking stock, model-based predictions allow for the development, testing, and use of new medicine and vaccines. They allow for the planting, raising, and harvesting of crop. They allow people and organizations to plan ahead and try to shape conditions according to their interests. In science they allow the formalization, communication, and testing of theories about the world and over time contribute to improved understanding and cumulative knowledge by exposing inaccurate or wrong ideas. Models, and quantification more broadly, are thus the basis of modern science, business, and life and contribute to our lives being less solitary, poor, nasty, brutish, and short. They allow the shaping of the future not just mere exposure to it.

3.6. Privacy

The natural corollary to quantification and data-driven insight and capabilities is privacy. The simple formula is the more data, the greater the potential analytical insights, the greater the capacity of organizations, companies, or states to make profits, shape people's option spaces, or the future. People's interests and rights in keeping aspects of their lives, character traits, interests, and behavior private can appear as an annoying speed-bump on the road to greater capacity and profits. This is especially relevant in the collection and use of data documenting people's uses of digital devices or services.

New capabilities in data collection, retention, and their uses raise hopes and desires within companies and governments for access to ever more data on ever more users on ever more aspects of the world. But here, interests between companies, governments, and users do not necessarily align and might in fact diverge. Greater capacity of companies and governments might run counter to the interests and rights of people who find themselves documented in data. While state regulators might seem a bastion of

[26]See Sifry (2023) for a critical reflection on the long-term effects of quantified campaign tactics by the Democratic Party.

[26]For more on data driven predictions in the Obama campaigns see Hersh (2015) and Nickerson and Rogers (2014). For a discussion of data-driven campaigning in Germany see Jungherr (2016a).

[27]For introductions to the use of simulation within the social sciences see Miller and Page (2007) and J. M. Epstein (2006).

protection of people's rights, it is not clear that their interests necessarily align with that of the people.

For example, Hersh (2015) has shown that in the US for a long time parliamentarians have worked continuously toward extending the data government agencies were obliged to collect in voter files and officially provide campaigns with access to. Also governments have an interest in having access to more data on people, be it to the benefit of crime prosecution and prevention, service provision, or public health. The flip side of greater government access to data is of course greater control over citizens and the potential for repression, especially in autocratic regimes. But trade-offs between government access to data and people's privacy should never be treated lightly, also in democracies. In short, government regulators do not necessarily have to be natural allies of people in the protection of their rights to privacy.[28]

Privacy is an inherently contested concept. But the legal scholar Alan F. Westin provides a definition that focuses on the control of information. This definition provides a helpful basis for the discussion of tensions between privacy and availability of digital data.

> **Definition: Privacy**
>
> "Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others." (Westin, 1967, p. 7)

Digital technology provides unprecedented opportunities for collecting, aggregating, processing, and disseminating information. But these opportunities mainly hold for companies collecting and disseminating data, not with the people who they document. In fact, these opportunities often directly negatively impact the ability of people to control the collection, use, and dissemination of information about them. The legal scholar Daniel J. Solove has proposed a taxonomy that is helpful in mapping these threats.[29]

Solove points to a set of potential privacy threats connected with the collection, processing, dissemination of information, and invasion of privacy. All of which feature strongly in the discussion of digital data. Most people are unaware of the data that are collected about them, their behavior, and their contacts when using digital devices or services or by being wittingly or unwittingly documented by digitally-enabled sensors. They also have little idea, never mind control, over how the collected data are processed by data collectors or owners. Data processing can include aggregation of data on people across different sources, the identification of users or their traits by probabilistic inference and associated labeling, or subsequent secondary uses either by the companies collecting the

[28] On the larger history of privacy and increased in the US see S. E. Igo (2018). For a discussion of how politicians actively expand the information collected in official voter files and the access of campaign organizations' to them see Hersh (2015). For a critical discussion of how privacy rights have over time be framed as limits to innovation and state capacity see J. E. Cohen (2013).

[29] Solove (2008).

data or those buying access to them. People also have no control over who companies provide with access to their data, be it intentionally or unintentionally. Also, collecting data on people and the data-driven inference of traits and interests might allow companies to influence decision making by people using devices or services to further their own interests over that of their users. The collection, processing, and retention of data with digital technology provides significant threats to the control anyone can hope to have over information pertaining to them.

Example: Cambridge Analytica

The by now infamous case of *Cambridge Analytica* provides a perfect example for the threats sketched by Solove. In the run-up to the 2016 US Presidential election the English campaign consultancy Cambridge Analytica claimed to be able to identify voters' psychological traits based on their Facebook activity and be able to target them with content optimized to get them to act according to a campaign's goals. After Donald Trump's win in 2016, the company claimed this as proof for the success of their approach. By now it seems pretty clear that Cambridge Analytica was not central to the Trump campaign[30] and that the approach advocated by them was highly unlikely to bring the claimed effects.[31] But the story of Cambridge Analytica is still instructive as a case in which nearly all of Solove's threats to information privacy come true.[32]

First, there is the question of data collection. Who collects what data? In the case of Cambridge Analytica, this starts in 2014. A researcher affiliated with a research lab at the University of Cambridge had developed a Facebook App allowing people to test their personality traits. But the researcher did not only collect the responses to the quiz. He used the app to pull additional information through the Facebook application programming interface (API) on the participants and their Facebook contacts. He then connected and stored these information to the responses of the personality quiz. At that point people had lost control over the information collected about them.

The second threat comes from information processing. Again, this case is instructive. The researcher linked signals from the information collected on Facebook to the responses of participants to his personality quiz. He and the people at Cambridge Analytica were convinced that they now had a model allowing them to predict personality traits based on Facebook data. Again, these claims are dubious and highly contested. But for the violation of privacy, the actual validity of the results of data processing does not matter, it only matters that the data are used in ways out of the control of people documented by them.

The third threat is information dissemination. In this case, this happened when the Cambridge researcher provided the company Cambridge Analytica with access to the data collected by his app and the models based on them.

Finally, by trying to influence people in their voting behavior based on models developed on data collected for another purpose Cambridge Analytica also hits the fourth threat: invasion. By actively shaping information environments of people based on data documenting traits and interests, the company interferes with their decision making.

By using digital information systems, like Facebook, Google, TikTok, YouTube or X, users cannot expect to be in control of the information posted in these systems and produced by their use. Instead, the companies running these systems can and will decide how to use these information and which third parties they provide access to the data or its processed derivates. Digital information systems thereby inherently challenge the notion of privacy referenced above.

People, their contributions, behavior and interactions, are tracked in digital information systems and the resulting data are stored and later used. This could be for commercial, political, security, or scientific purposes. Most people will be unaware of these subsequent uses when using digital media. Add to this that many successful digital business models depend on the targeted provision of ads based on users' traits and interests. While much is unknown about the mechanics behind these models, they rely on broad data collection and processing without necessarily securing informed consent of people documented by that data.[33] There is an inherent tension between the interests of companies providing digital services and the rights to privacy of their customers and users.

Accordingly, designing meaningful privacy for these data is difficult. This has inspired much debate and controversy. In response to these issues, the European Union (EU) implemented in 2018 *The General Data Protection Regulation (GDPR)*. The GDPR advances strict limits to the uses of data collected by companies and has influenced other international data regulation. But privacy concerns remain.[34]

Even if data are used appropriately and according to the wishes of the people they are documenting, privacy remains an issue. For one, there is the question of data ownership. This has come up repeatedly in the context of US campaigns. Imagine you support a candidate during the primaries of your party but ultimately the candidate has to drop out. What happens to the data the candidate's campaign collected on its supporters, such as e-mail addresses, donation behavior, issue preferences, or contact information? Does the candidate or the contractor delete the data once the run is through? Or do they retain the data? If so, for how long? Or do they decide to throw their lot in with another candidate and transfer the data to them? These are questions that came to

[32] For more on the Cambridge Analytica scandal see Kroll (2018).

[32] For more on the limited effects of psychometric targeting advocated by Cambridge Analytica scandal see Hersh (2018).

[32] The following account is based on Kroll (2018).

[33] For more the digital ad business see Auletta (2018); Crain (2021). For a critical take on the efficiency of these models see Hwang (2020).

[34] For more on the GDPR and its international influence see Bradford (2020). For more on the mutual influences and dependencies between US and EU in data regulation see Farrell and Newman (2019a).

matter for the Democratic Party once it had to be decided who could gain access the extensive data the Obama campaigns collected after his last successful race.[35]

Questions of data privacy do not become easier through an inherent tension between privacy and analytical capacity. While it is easy to prioritize people's privacy over companies' increased capabilities to roll out ads, this becomes harder once the trade-offs are between users' privacy and more important analytical goals. This includes crime prevention or public health. For example during the recent Corona pandemic there were various attempts to use digital media, apps, and trace data to track the spread of the epidemic and infection chains. While promising from the perspective of increasing state-capacity to fight the pandemic, pursuing these options to the full would severely impact people's privacy by for example tracking their movement and contacts. These concerns needed to be accounted for in the development and deployment of tracking apps and the official use of other available data sources, potentially leading to a loss in functionality and thereby weakened support in the fight against the pandemic.[36]

Of course states are not only interested in access to data during times of public health crises. Famously, the NSA-spying scandal, following the revelations of US military contractor Edward Snowden in 2013, showed the degree to which the US government was accessing data collected by companies providing digital information systems to users all over the world.[37] Even more extensive are the efforts of the Chinese government to collect information about its people and to shape and sanction their behavior. The Chinese *Social Credit System* is much discussed, although its workings and effects are very difficult to assess confidently. This makes it hard to differentiate between public facing claims and actual workings of social control through quantification and data, leading to a likely overestimation of the capabilities of the system.[38]

The perceived power of data for surveillance and social control has also given rise to fears of data-enable espionage and foreign influence. This can include new forms of signal intelligence, were states try an capture digital communication of other states or foreign nationals. More contentious are suspicions of espionage and foreign influence enabled by the provision of digital infrastructure through foreign firms. This could recently been seen in concerns voiced in the US of the emerging dependencies of allied states on communication infrastructure provided by the Chinese company *Huawei* and the growing popularity of the Chinese social media platform *TikTok* in the West. Reliance on these structures is seen to provide the Chinese state with a backdoor to the information flows and public discussions in other countries. With mounting geopolitical tensions, this is

[35] For more on the subsequent uses of Obama's data see Meckler (2012) and Timberg and Gardner (2012).

[36] For a general discussion of privacy and medical data see Price and Cohen (2019). For privacy in Covid-tracking approaches specifically see I. G. Cohen et al. (2020).

[37] For more the NSA scandal see Gellman (2020).

[38] For more on China's Social Credit System see Brussee (2023); Creemers (2018); Knight and Creemers (2021).

seen to provide data-driven risks.[39]

The opportunities for quantification and data-enabled insights provided by digital technology are shadowed by making people's control of information about them increasingly untenable. Accordingly, various scholars such as Spiros Simitis, Helen Nissenbaum, and Julie E. Cohen have argued in favor of extending the discussion of privacy away from the individual and toward the design of information systems people are embedded in and covered by.[40] This perspective will come to matter again, in the discussion of platform business models and companies and artificial intelligence. The scientific study and design of data collection and use practices and protocols clearly also has to consider privacy concerns on both the individual and the systems level.

3.7. The promises and limits of quantification and data in the study of the social and political world

Data and more generally quantification hold promises in understanding and shaping the world as well as social and political life. But to realize this promise there are preconditions that have to be met. On a foundational level, quantification needs to be sensible. Without a careful translation of entities in the world into signals and data, resulting analyses will be void and potentially misleading. Here, it is ironic that in a time when digital technology provides unknown data riches, the effort in translating the world into data receives very little attention. We need to be careful, that the society with the largest amount of data at its disposable, is not at the same time the society with the most naive understanding of what these data do and do not represent. Realizing the potential of quantification means not only capitalizing on analytical opportunities found in quantified representations of the world, it also means remaining aware of the limits of representation.

The relevance of data in the age of digital technology means also that the processes of quantification, the use of data, their regulation, and the subsequent impact on society are important questions for the social sciences. This includes questions of how organizations and actors approach quantification. How do they try and capture meaningful aspects of the world? How do they translate these signals into data? How aware are they about the dangers of quantification in losing aspects of the world or are they only focused on the upsides? Another set of questions focuses on the uses of data once measured. How do organizations and actors use the quantified symbolic representation of the world? How are they incorporating data-driven insights within decision making or are analytics only window dressing? Do they actively account for the limits of data or are they treating them as unmediated insight? Finally, the social sciences also need to address questions of

[39] For a background on classic and current forms of foreign influence see Rid (2020). For more on tensions regarding reliance on Huawei see Segal (2021).

[40] See J. E. Cohen (2012); J. E. Cohen (2019); Nissenbaum (2009); Simitis (1987); Simitis (1995).

how quantification and the uses of data are governed and the impact that these practices have on society.

Let's take a step back for a moment. Looking at the sociological literature on quantification, it is hard to miss the uneasiness that underlies much it. There is a persistent implicit or explicit sense of loss. By reducing the world to numbers, some essence of worldness or humanity appears to be always under threat. By continuously making new societal fields and aspects of human behavior legible, data open them up for colonization by market forces and centralized control.

There is much to commend this view. For one, we have seen in the section on measurement how important it is to decide which aspect of an entity or phenomenon to quantify and which to leave out. We have also seen that naive interpretations of data without accounting for the specifics of what aspects a given measurement approach makes visible and which it hides leads to a misreading of the world. But at the same time quantification and data lie at the heart of modern life and have led to crucial improvements in health, quality of life, and prosperity. The issue therefore does not lie with quantification or data as such, it lies with their responsible and informed or irresponsible and naive uses.

In its extreme the unease toward the reduced representation of the world found in data brings to mind a short story from Jorge Luis Borges.[41] In *Del rigor en la ciencia* Borges tells the tale of a group of cartographers so enamored with their skills and the beauty of realistic representation, they start an effort to map their empire and its features point by point, ending up with an exact replica in size and detail. While beautiful to its originators, the map was useless to those who followed and the work was disregarded. This story might serve as a cautionary tale to those warning of the dangers of data, quantification, or reduction in general instead of pointing to specific flaws inherent to these approaches or contingent to their uses.

For the other extreme, we also find a cautionary tale in literature. In Hermann Hesse's novel *Das Glasperlenspiel*,[42] we meet an order of scholars dedicated to the study of a universal symbolic system of science and culture, the glass bead game. But in their total dedication to symbolic representation they lose sight of the world as such and the social and political conditions for their pursuit. In fact, once the Master of their order gives up his study to be in the world, he perishes.

This illustrates the risks with either rejecting data and quantification in favor of an unspecified essentialism not available to symbolic representation and the purely positivistic reliance on symbolic representations of the world with no specific interest in the world as such or the mechanisms by which the translation of the world in symbols leads us to misread the world in its representation. We will encounter this issue again when we will be talking about artificial intelligence.

[41] Borges (1946/1975).
[42] Hesse (1943/2002).

Quantification, data, and models are important building blocks to modern societies. But realizing their promises demands for informed uses. This means accounting for both their potential and limits. For social scientists and practitioners alike this means not only improving their analytical skills but also to develop a sophisticated understanding of the relationship between symbolic representations and the world as such. Without a solid understanding of how representations map and do not map onto the world as such, any mathematically identified structures on the representational level do not have to carry meaning beyond the realm of symbols. And these are preconditions for the use and assessments of algorithms to shape human behavior and the world.

3.8. Further reading

For an instructive popular account of using data to learn about the world see Hand (2007).

For a helpful overview over different sociological approaches to quantification and its consequences see Mennicken and Espeland (2019).

For a sociological account of how the state and other central authorities make the world legible and increase their level of control over it see J. C. Scott (1998).

For a broad account of the role of models in the social sciences see Page (2018).

For more on the statistics models are based on see McElreath (2020).

For more on privacy see Solove (2008).

3.9. Review questions

1. Provide definitions of the terms:

- Data
- Quantification
- Big Data
- Digital trace data
- Bias in datat sets
- Measurement
- Representational measurement
- Pragmatic measurement
- Public opinion
- Metrics
- Model
- Privacy

2. Please discuss how big data can contribute to the quantification of social and political life. Discuss the potential and limitation of at least two sources of big data specifically.
3. Please select a target concept of interest asking for pragmatic measurement. Then choose a source of digital trace data and discuss how the available signals allow – or do not allow – for the measurement of the target concept.
4. Discuss how models allow organizations and people to shape the future.
5. Using the framework provided by Solove (2008) discuss the different ways digital technology constitute threats to people's privacy.

4. Algorithms

Making sense of data, acting on them, and using them to predict the future asks for algorithms. Algorithms are sets of steps by which to solve pre-defined tasks. Originally, the term *algorithm* refers to a predefined set of steps by which people would solve a given mathematical problem. But today the term appears predominantly in connection with computing. Algorithms allow computers to perform tasks. They are crucial in the advances of computer-enabled analysis and automation.

The uses of algorithms vary widely. Algorithms can identify patterns in data, surfacing hidden connections between phenomena or actions. For example, expressing interest in a given brand on social media might be linked to the support of a specific party. Or being informed about a supposed political crisis might make people donate money to a political party. Potential patterns abound and algorithms help in identifying them.

Similarly, algorithms can also be used to automate actions. Based on rules learned from past data, algorithms can automatically present people with specific options. For example, if in the past people who expressed a liking for a given brand would also go on and support a given party, the algorithm could automatically present them with content from said party without having them first look for it. Or, if in the past specific people reacted to a crisis prompt in an email with a donation, while others reacted to a personal request by the candidate's spouse, the algorithm can automatically decide which recipient to present with which information in an email blast in order to drive donations.

Algorithms thus help us understand the world better and interact with it more efficiently. But algorithms are also a source of worry. How can we be sure that patterns identified from data documenting the past should be replicated in the present or future? How can we know what the algorithm learns from data? And how can we be sure that the actions taken by algorithms conform with the goals, with which we have designed and deployed them?

Studying the use and impact of algorithms in society means being aware of their technical workings but also their uses in different societal fields. How are algorithms designed? What tasks are their deployed to solve? And how are their uses and outcomes monitored and interrogated, not only by those using them but also those who are affected by them. The study of algorithms therefor has mathematical, computational, and social components. This chapter will provide the reader with an overview of the technical as well as the social components in the study of algorithms.

4.1. What are algorithms?

We live in a world shaped by algorithms. In digital communication environments, algorithms decide what information we see, which music or films are suggested to us, or which people we are invited to connect and interact with. Algorithms decide about which contributions to digital communication environments are banned and which are allowed. Algorithms are also shaping our world beyond the merely digital part. They assign people individual risks, be it credit default, fraudulent behavior, abuse, or recidivism. And algorithms act. Algorithms buy and sell stock and they decide whether a self-driving car should stop or accelerate. Algorithms shape the information we see, options we have, and confront us with actions we have to content with. No wonder, then, algorithms their uses, workings, and effects face increasing scrutiny and contention.

The term *algorithm* has become a general catch-all term for the risks and fears associated with the powers computers hold over individual lives and in societal fields. The term has become associated with opaque and incontestable mechanisms that allow governments and companies to control people or – worse yet – that through runaway technological change have evolved beyond the control of even those who develop or deploy them. But these far-reaching fears and generalizations tend to obscure the actual nature, workings, and effects of the use of algorithms. Instead of inspiring and enabling critical reflection, evaluation, and improvement of algorithms these accounts risk blanket rejection and creating a sense of helplessness. To turn widespread interest or concern into something more productive means looking at algorithms more closely and analytically.

First, we need to be clear about the term and its origins. We have come to associate the term *algorithm* with computers. But in fact, the term goes back far beyond the age of computation. Algorithm is the latinized version of the name of an important mathematician from the middle ages, Muḥammad ibn Mūsā al-Khwārizmī. Al-Khwārizmī was a ninth century mathematician from a region of central Asia, south of the Aral Sea in what today is Uzbekistan and Turkmenistan. He was the author of the earliest known book on algebra. Once the book was translated into Latin in the 12th century, the name of its author became synonymous with the practice of mathematical calculation following routine arithmetic procedures. Accordingly, in mathematics the term algorithm refers broadly to “any process of systematic calculation, that is a process that could be carried out automatically” (Chabert, 1994/1999, p. 2). This of course includes processes that were defined well before there was the term algorithm, such as *Euclid’s algorithm* that described a routine to identify the greatest common divisor between two numbers, and those that came after, such as Euler’s method for solving differential equations. More fundamentally even, we can also think of basic math routines performing long division by hand as algorithms.[1]

But we can think even more broadly about the term. The term *algorithm* can cover

[1] For a historical account of different types of algorithms from Mesopotamia to computation see Chabert (1994/1999).

all procedures that provide a standardized set of steps solving of a specific repeatedly encountered problem. In this sense, any recipe in a cookbook is an algorithm, allowing for the standardized step-by-step approach to solving a specific problem, such as preparing a given dish. Algorithms, in this sense, lie at the heart of many standardized practices of modern life. Be it cooking, sports, or the pursuit of hobbies. They are an essential part of teaching trades, skills, and professions. And they are a crucial feature of organizations, guaranteeing their proper and standardized workings. Examples include internal administrative guidelines within government bureaucracies, police departments, or large companies for addressing specific repeatedly encountered tasks. Algorithms, in this sense, are a core feature of modernity, standardization, and professionalization.[2]

Currently, we tend to associate the term *algorithm* most prominently with computers and the automation of processes or decision making. Fundamentally, when used in the context of computing, the term algorithm means the same as in other areas but the series of steps to solve a task must be computable by a machine.

Defintion: Algorithm

"The modern meaning for algorithm is quite similar to that of *recipe*, *process*, *method*, *technique*, *procedure*, *routine*, *rigmarole*, except that the word"algorithm" connotes something just a little different. Besides merely being a finite set of rules that gives a sequence of operations for solving a specific type of problem, an algorithm has five important features" (Knuth, 1968/1997, p. 4).
For Knuth (1968/1997) these features are:

1. Finiteness;
2. Definiteness;
3. Input;
4. Output;
5. Effectiveness.[3]

According to this definition, an algorithm must be finite. In other words, it must complete after a series of steps. These steps need to be clearly and unambiguously specified, they need to be definite. An algorithm works on inputs and returns outputs. Those stand in a specified relation to the inputs. Finally, following Knuth (1968/1997), algorithms should be effective in that each of the steps specified to solve a task should in principle be doable by a person using pen and paper. These sets of features can be used in the analysis of algorithms to compare the performance of algorithms in translating inputs into outputs. An important element here is the number of steps necessary to perform the task, with a smaller number of steps being more efficient and demanding for less computing time.

Algorithms lie at the heart of computation. They are the basis for any computational

[2] For a broad discussion of algorithms in social life see Daston (2022).
[3] See Knuth (1968/1997), p. 4–6.

operation and their discussion can rightly focus on the math, expression in code, and efficiency.[4] But algorithms are increasingly used to solve tasks that are not predominantly restricted to technical contexts but also have social consequences.[5] Algorithms shape digital systems that directly impact, or interact with people. Accordingly, in their analysis we have to move beyond merely technical questions to social, psychological, and structural questions. For this, we have to turn to the fields algorithms are used in, tasks they are set to solve, direct effects on people, and indirect effects on the societal fields and social contexts they are deployed in.

Computational algorithms are the key in translating data into action and insight. In Chapter 3 we saw how data make the world readable, offer insights about regular and potentially causal connections between entities, and allow for actors to plan and adapt for the future. Algorithms can build on these potentials and realize them. In fact, given the massive scale of newly available data, the automation of analysis and action through algorithms becomes a condition to do so. It is no surprise, then, to find computational algorithms to be used in a wide variety of context in which they shape and impact social and individual life. This variety is too broad to cover it here comprehensively. However, we can group their uses in three categories based on different goals, workings, and concerns. These are the use of algorithms for insight, decision support, and action.

4.2. Algorithmic insight

Algorithms are widely used to provide insight. Data do not speak for themselves, data-enabled insight depends on analysis. This is done with algorithms designed to process and analyze data to uncover patterns, trends, and hidden relationships. Examples include data visualization tools, data mining techniques, and certain machine learning models used in exploratory data analysis. Algorithms like these are essential in sectors like research, business intelligence, and academic studies, where raw data must be transformed into meaningful insight.

While there are many different kinds of algorithms for data analysis, it is helpful to differentiate between two approaches:

Many algorithms transfer mathematical calculation procedures into computer-readable routines. Programmers provide computers with a pre-defined series of steps for solving tasks in data analysis. This allows computers to automatically perform data analysis on vast datasets and to provide insights. This includes data preprocessing, sorting, searching, clustering, and pattern recognition. Since all steps and analytical procedures are pre-defined, these approaches are reliable and comparatively easy to explain but depend on the quality of the pre-defined steps and rules.

[4]For technical introductions see Cormen et al. (1990/2022), Kleinberg and Tardos (2005).

[5]For an introduction to algorithms focusing on their uses to solve tasks in the larger world see Louridas (2017).

Example: Identifying campaign donors, K-Nearest Neighbors

Let's assume a political campaign has a dataset with information about people. It is interested in gaining insight into which of the people the campaign comes into contact with are likely donors. One easy procedure that allows this is the K-Nearest Neighbors algorithm (KNN).[6] Roughly speaking, KNN does the following:
The campaign has a data set with information about people. Such as their age, income, the number of times they have voted in the past, and whether they have recently donated. Now, for each new person entered into the campaign database, KNN calculates whether they are likely to do so in the future by looking at a small set of people in the dataset (let's say three) that resemble them with regard to characteristics that the campaign has information about; in our case this would be age, income, and the times they have donated in the past. If most of these similar people have donated, there is a likelihood that the person of interest will be open to donate in the future. If not, not.

Alternatively, one could use machine learning algorithms.[7] Machine learning algorithms also operate based on a set of pre-defined steps, but many introduce stochastic and probabilistic elements during their execution. This incorporation of randomness means they may not always yield identical outputs when executed multiple times with the same input, making them less deterministic in nature. Machine learning algorithms tend to be more data intensive than other approaches but also are more responsive to unforeseen patterns in data and temporal shifts. On the other hand, due to their stochastic and probabilistic elements, they can be hard to interpret and their returns difficult to explain or assess, leaving some uncertainty with regard to their use in real-world scenarios.

Example: Identifying campaign donors, machine learning

Let's stay with our previous example but now look at how it could be solved through a machine learning algorithm. Let's say a Random Forest algorithm.[8] The campaign begins by collecting a comprehensive dataset on past donor behavior. After ensuring that the data is cleaned and processed, relevant features are selected and processed appropriately. The campaign randomly divides the dataset into a training set (e.g., 80% of all records) and a test set (e.g., 20% of all records). Using the training set, a Random Forest model is trained to recognize patterns and characteristics that are indicative of donation behavior. The model's feature importance ranking is analyzed to understand which donor characteristics are most predictive of donation behavior. The trained model's performance is evaluated on the test dataset using appropriate metrics to ensure it predicts accurately and generalizes well to unseen data. Once satisfied with the model's performance, the

[6] For K-Nearest Neighbors (KNN) see Hastie et al. (2001/2009), p. 459–484.
[7] For a general overview of machine learning see Alpaydin (2016/2021). For a technical introduction see Hastie et al. (2001/2009).

> campaign can then use it to predict the likelihood of new entries to the dataset being potential donors based on their characteristics.

Both traditional and machine learning algorithms offer valuable insights, yet they possess distinct strengths and limitations. Traditional algorithms are particularly helpful for data sets with a limited number of features per observation, or low dimensionality. In such scenarios, the design and application of classical statistical models are straightforward, and the simplicity of these algorithms often suffices. The limited variables mean there's minimal benefit from the complexity of machine learning algorithms.

On the other hand, for high-dimensional datasets with numerous features per observation, traditional algorithms quickly reach their capacity. It's in these complex scenarios that machine learning algorithms shine, leveraging their ability to handle vast amounts of data.[9] However, these algorithms are resource-intensive to train. Their often stochastic and probabilistic nature can make them challenging to interpret, leading to potential skepticism and criticism from the public. For tasks requiring high reliability and transparency, the complexities and opacities of some machine learning models may be less suitable.

Actors in various fields rely on algorithms to provide them with insight. Examples include campaigning, in which parties use data and algorithms to identify likely voters, contributors, and model likely voting behavior.[10] Another example is the use of data by news rooms to identify trending topics and features of successful stories.[11] In these and other cases, the use of algorithms to generate insight not only changes the behavior and opportunities of people using algorithms, it also can lead to a reconfiguration of structures, organizations, and institutions in the field around new opportunities and demands associated with their realization.

> **Example: Algorithms, Obama, and the aftershocks**
>
> Famously the US presidential campaigns by Barack Obama in 2008 and 2012 relied heavily on algorithms to generate data-enabled insight about voters, donors, and volunteers. These insights were so important to the campaign that it changed the traditional organizational structure of political campaigns and put digital and analysis specialists at the core of the decision making group.[12]
> Key to these adjustments has been the power of data- and algorithm-enabled practices for campaigns to generate insight about potential donors and to provide ways of maximizing their contributions to the campaign. These successes generated a whole industry of donor analytics and became crucial to campaign practices within the Democratic Party across the ticket.[13] But heavy reliance on donation

[8] For Random Forests see Hastie et al. (2001/2009), p. 587–604.
[9] For more on the modeling of high dimensional data see J. Wright and Ma (2022).
[10] For the use of data-enabled insights by campaigns see Nickerson and Rogers (2014).
[11] For the use of algorithm-enabled insights in the news see Christin (2020).

requests in campaign communications with their supporters might over time have contributed to a shift in the relationship between party and supporters. Sifry (2023) argues that this practice has led to a largely transactional and commodified relationship between the party and its supporters, severely reducing the resonance of its progressive claims and over time contributing to a donation fatigue.
This is an example that over the course of a few campaign cycles, innovative early uses of algorithms for insight provided a competitive advantage for early adopters. The public perception of the contribution of algorithmically gained insights to electoral success created a large set of imitators who adapted the same techniques and approaches. This then led to adaption and changes in the behavior of people rendering formerly successful practices mute. If we want to understand the use and impact of algorithms, we therefor must turn beyond the narrow analysis of the workings of algorithms and their uses and take a broader look at societal fields and action and adaption over time.

Of course, just because an algorithm does provide a result, that does not make the result true or useful. In the interpretation of results provided by algorithms, people need to remain aware of potential limitations and sources of error. This includes translating the general limitations of quantification introduced in Chapter 3 to the specific contexts algorithms are used in. More specifically, it also includes critically interrogating whether the steps in the analysis reflect both the characteristics of the underlying data and the problem they are supposed to solve.

4.3. Algorithmic support

Algorithms are also used to provide suggestions and advice. This includes algorithms that suggest courses of action or scenarios about future developments based on regularities identified in past data. Examples include algorithms that recommend users of digital media platforms content of potential interest, algorithms that advice police forces on where to expect a concentration of criminal activity, or algorithms that advice doctors on whether specific symptoms indicate a specific disease. Algorithms play a crucial role in decision-making processes across various sectors, helping individuals and organizations to make informed choices by offering data-driven suggestions.

In discussing these algorithms, their uses, effects, and evaluation, it is helpful to differentiate between algorithms providing advice and suggestions to experts and those providing advice and suggestions for lay people and users of digital communication environments and services.

[13]See Kreiss (2012).
[13]See Kreiss (2016).

4.3.1. Algorithmic support for experts

In a growing set of professional fields, algorithms support experts by providing assessments, prognoses, and scenarios. Algorithms promise to offset some of the vagaries and inefficiencies of human decision making and analysis of evidence.

Ideally, decision making in professional and institutional contexts aims for fairness to those concerned, consistency across cases, comprehensive accounting for the available evidence, and explainability. But even trained experts can fall short of these expectations. For example, psychological biases can unduly influence decisions of trained decision makers. Cognitive shortcuts, such as heuristics, can fail decision makers in specific contexts and lead them astray. Also, people can struggle with keeping aware of all the evidence relevant to a decision.[14] These are some of the limitations of human decision making that can can lead to inefficient, unfair, or wrong decisions. Here, algorithmic and data-enabled decision support can help.

Algorithms can support decision making by predicting expected outcomes based on available data. Using rules provided by modelers or rules based on autonomously identified regularities in past data, algorithms model patterns, processes, behaviors, and outcomes of interest. They then can predict the likelihood of events happening or entities falling in a given category based on these models. This makes algorithms useful in various areas.

In medicine, computational algorithms, such as IBM's *Watson for Oncology* assist physicians by predicting the likelihood of a patient suffering from a disease based on given symptoms or test results and by suggesting treatment plans with greater likelihood of success than alternatives.[15] In finance, insurers use credit scoring models to assess the creditworthiness of individuals or businesses.[16] In engineering, computer-aided design systems (CAD) offer suggestions on design optimization or error correction.[17] In biological research, algorithms propose molecular structures for new potential drugs or materials or offer insights into genetic variants and their potential implications.[18] In climate research, algorithms help in the development of climate models and the downscaling of general expectations to specific geographical areas.[19]

In these cases experts are supported by algorithmic support systems that suggest courses of action, diagnoses, or scenarios for given cases or moments. Algorithms provide experts with scores indicating whether people or other entities fall in a relevant category, for

[14] For an overview of the principles and psychology of judgment and decision making see Baron (1988/2023). For a foundational text on biases in human decision making see Tversky and Kahneman (1974). For a critique of putting to emphasis on psychological biases in explaining people's decision making see Gigerenzer (2018). For cognitive heuristics see Gigerenzer and Gaissmaier (2011). For limits to expert forecasting see Tetlock (2005/2017).

[15] See Somashekhar et al. (2018).

[16] See Citron and Pasquale (2014).

[17] See Bi and Wang (2020)

[18] See Jumper et al. (2021).

[19] See Kotamarthi:2021aa.

example whether they are likely voters or likely to default on a credit. Or they provide experts with scenarios of likely future developments or outcomes. Sometimes, these algorithmic assessments rely on one model. Sometimes they rely on multiple models and provide experts with different outcomes to compare and choose from.

In general, the goal of algorithmic decision support for experts is to increase efficiency and speed of decision making especially in time-sensitive or resource-starved environments, such as medical diagnoses or in the criminal justice system. In other cases, algorithmic support is more about providing access to a greater evidence base, providing access to different models and different outcomes, and for handling large amounts of data. This is not so much about efficiency but about providing practitioners with better access to insight based on greater troves of information made accessible through models.

The positive impact of algorithmic support systems has a counterweight in a set of concerns. If algorithms are trained on biased data, they will produce biased outcomes. For instance, if a hiring algorithm is trained on historical company data that favors one gender over another, it might reproduce that bias in its recommendations.

Many advanced algorithms, especially deep learning models, are often treated as *black boxes* because their decision-making processes are not easily interpretable. This lack of transparency can make it challenging to trust or validate the advice provided. Experts might become overly dependent on algorithmic advice and neglect human intuition or expertise. This is particularly concerning in fields where human judgment is crucial, like medicine or law.

Many algorithmic systems rely on vast amounts of data, which can raise concerns about user privacy and data security, especially when algorithmic support is provided on a software-as-a-service basis by external companies. Without proper safeguards, sensitive information could be at risk of being accessed by unauthorized parties.

Also, over time the suggestions of algorithms might turn into de facto algorithmic cages for experts and professionals. After all, if one deviates from an algorithmic recommendation and this turns out badly, the potential penalty is higher than simply following the algorithm even if that recommendation turns out to be wrong.

Example: Predictive policing

An area where the use of algorithms directly touches on democracy and the power of the state over its subjects is the criminal justice system. For example, algorithmic support systems are increasingly used by various police departments in the USA.[20] Based on data on previous criminal activity, algorithms provide assessments of which areas are expected to feature heightened criminal activity so police can preemptively patrol them. Algorithmic assessments can also extend to individuals, such as assessments whether selected people have a heightened risk of criminal behavior and therefor merit higher police attention.[21]

Algorithmic support systems promise police departments a better way to allocate scarce resources more effectively to where they are most needed. This is clearly an important contribution. At the same time, there are important limits to data-driven and probabilistic approaches to policing, that have given rise to strong critique.
This includes shifts within systems relying on algorithmic support to underlying probabilistic approaches, abandoning less quantifiable approaches that nevertheless might be more suited to the task at hand.
Just as importantly, by relying on data that for reasons of historical discriminatory practices over-represent specific demographic groups in crime-statistics, algorithms guiding contemporary police activity risks reproducing these discriminatory practices. Over time these patterns might be reinforced through a vicious feedback loop of historical over-representation in data, heightened police attention akin to racial and social profiling, and subsequent overrepresentation in arrest reports and new data.
Over reliance on data-driven algorithmic support might thus contain a hidden trade off between efficiency and fairness within the criminal justice system.[22] Of course this general critique can be also applied to other areas relying on algorithmic support systems.

4.3.2. Algorithmic support for non-experts

Algorithms can also provide support for people in their daily lives. These algorithmic support systems have great reach and are already present for many different types of uses and practices.[23] Very prominently this includes algorithmic recommendation systems on digital services, such as news-feeds on social media sites, video and music streaming, or on online shopping sites. In navigation and travel algorithms provide suggestions for routes and travel itineraries. In more specialized contexts algorithmic support systems provide advice in personal life decisions in dedicated services or apps, such as food and diet suggestions, workout plans, or dating.

These algorithms are important in helping people navigate choice-rich and fragmented information environments and option spaces. Algorithmic recommendation systems are important and can help people navigate overwhelming and confusing information environments by suggesting information. They can also help people to make better choices regarding health, nutrition, or personal finance by shaping option spaces on the basis of data-driven analysis and prediction. At the same time, the growing reliance on algorithmic support systems in ever more areas of public and personal life can turn into a problem for politics and society, if these systems are implemented and handled without reflection, evaluation, and critique.

[22] For a sociological account of predictive policing in practice see Brayne (2021).
[22] For an overview of predictive policing see Ferguson (2017).
[22] For a foundational critique of probabilistic methods in the criminal justice system see Harcourt (2006).
[23] See Narayanan (2023).

For algorithmic support for non-experts, similar concerns exist as for algorithmic support directed at experts. But beyond bias, transparency, de-skilling and over reliance, and privacy concerns there are also other risks to consider. This includes a homogenization of choices. If everyone relies on the same set of algorithms for recommendations, diversity in consumption patterns will necessarily be reduced. This in turn will lead to a homogenization of tastes and culture. For instance, if everyone watches the same suggested movies or reads the same recommended books, it could limit cultural diversity and novelty.[24]

This algorithmically driven cultural homogenization could lead to different outcomes. For one, it is possible, that international consumption and production patterns would shift toward generalized US cultural products, increasing the already felt cultural hegemony of the USA. This would also contribute to a spread of US cultural and political concerns to other countries, irrespective of the specifics of local contexts and cultures. Alternatively, algorithmic recommendations might shift patterns toward a more generalized international taste, due to large audiences for cultural products in Asia. Market forces would thus shift cultural production and consumption not toward US-based patterns but toward those of the largest consumer group of international consumers. Similar dynamics already emerged with American media and sports companies reaching out to Asian markets and in turn adjusting products and public statements to the sensibilities, concerns, and interests of Asian audiences and governments.[25]

Additionally, greater reliance on algorithms for content discovery and distribution puts companies providing these algorithms into a powerful position opposite creators who come to depend on producing content recognized and distributed by algorithms. When algorithms determining the visibility and discoverability of content shift, so can the relative prominence of those who create content for algorithmically shaped information systems.[26]

A recent case illustrating this dependency is Instagram's shift from an algorithm predominantly relying on signals from the social graph in content recommendation decisions to one predominantly relying on signals within content items, which led to strong pushback by Instagram influencers with large followings.[27]

Example: Algorithmic recommendation

Digital information environments are often shaped by algorithmic recommendation systems. Based on predefined or automatically identified rules, users are shown information that is likely of relevance to them or that is likely to have them engage with it. On a social networking site like Facebook, this could be a post by a close friend or family member, on an entertainment app like TikTok this could be a

[24]See M. Frey (2021).
[25]See Schwartzel (2022), Zhu (2022).
[26]See Cotter (2019), Duff and Meisner (2023).
[27]See Mignano (2022).

funny video of a dancing hamster, on a microblogging service like X this could be a news item on current events.

Rules determining recommendations vary and can shift over time. For example, generally speaking there are recommendation algorithms that show people content prominently that others who they are connected with have posted or have interacted with. This would be a logic based on users' social graph, or more broadly, their network.

Alternatively, machine learning algorithms can identify signals within the content itself that at any given time or for any given subgroup of users indicate likely engagement. Based on these automatically identified patterns within content, algorithmic recommender systems would decide about whom to display what information. Other logics for rules, or mixtures of logics, can also apply.[28]

In most digital information environments, some sort of algorithmic recommendation is necessary for people to navigate these information and choice-rich environments. At the same time, companies running these services use algorithmic recommendation systems in support of their commercial interests. By keeping people engaged with the service, companies are in a position to show them more ads, making them a more valuable partner for companies trying to reach people. These uses stand in obvious tension, especially when algorithmic recommendation shape people's political and news exposure. For example, content with strong informational quality might be beneficial to people but not necessarily lead to interactions. Accordingly, it might not be recommended by algorithms. In contrast, other content creating controversy or arousing strong emotions might create interactions and therefore be recommended algorithmically, but over time might lead to a deterioration of the information environment. In short, purely commercial display and distribution logics are likely to be in conflict with logics focused on the quality of information or democracy.

Algorithmic support systems clearly hold great potential for experts in increasing their capabilities, information processing ability, and their efficiency. This has obvious consequences for the (rhetorical) power of experts and expert advice in democracies. Algorithmic advice systems aimed at the broad public can also be very helpful. They provide crucial assistance in choice-rich information environments and option spaces that without algorithmic support might be prohibitively difficult to navigate due to fragmentation or information overload.

While helpful, there clearly also are risks associated with these systems. These includes risks for people subject to algorithmic suggestions, the correctness of the suggestions provided, fields and systems coming to rely predominantly on algorithmic advice systems, and an increase in the power of companies providing algorithmic support systems in ever more political and social fields. Accordingly, the analysis of algorithmic support systems must always include their mechanics, uses, results, effects, and structural

[28] For different mechanisms behind algorithmic recommendation see Narayanan (2023).

embeddedness.

4.4. Algorithmic action

Algorithms need not stop at providing insight or advice, some also can take automated action. These algorithms are designed to process data, make decisions, and execute actions without requiring human intervention at every step. Examples come from different areas.

Algorithms can allow machines to move and act. This provides the basis for self-driving cars where algorithms evaluate data from multiple sensors, such as cameras, LIDAR, and radar, to navigate, decide when to accelerate, brake, or turn, and to react to unforeseen situations.[29] The same goes for their use in agricultural machinery, such as automated tractors where algorithms decide on when and where to plant, water, or harvest crops.[30] Another more controversial example is their use in military and civilian drones where algorithms allow drones to fly specific routes, adjust to environmental conditions, and even decide on targets.[31]

Algorithms are also used to autonomously shape and act in digital information environments or markets. We already encountered algorithmic recommendation systems. These recommendations can provide choices for users. Alternatively, recommended content can automatically be displayed, for example through autoplay on Spotify, YouTube, or TikTok. In these cases algorithms come to autonomously decide on what content to display next and only leave the user to cancel the playback instead of actively choosing between suggestions; thereby reducing user agency. Additionally, algorithms are used to monitor and moderate digital communication environments by autonomously detecting and sometimes removing or hiding content that violates content policies.[32]

Algorithms can also analyze and act in markets. A prominent example is online advertising, where algorithms decide which ads to show to which users, and when. These display decisions are based on user profiles and user behavior on the one side and ad auctions on the other, determining a demand-driven price for advertisers to reach users with specific profiles.[33] Another prominent example, comes from the use of algorithms for high-frequency trading (HFT) where algorithms can decide to buy or sell stocks in fractions of a second based on real-time market data.[34]

Automated algorithms are also used to manage large interconnected systems. This includes smart grids, where algorithms are used to monitor and balance energy consumption, dynamically adjusting energy distribution based on demand. On the level of the

[29] For algorithms in self-driving cars see Badue et al. (2021).
[30] For autonomous algorithm in agriculture see Bechar and Vigneault (2016).
[31] For the use of algorithms in autonomous drones see Floreano and Wood (2015).
[32] See Gorwa et al. (2020), Douek (2021).
[33] See Singh (2020), MacKenzie (2022), MacKenzie et al. (2023).
[34] See Menkveld (2016).

individual household, this translates into the use of smart thermostats where algorithms help devices learn user preferences over time and adjust heating or cooling settings autonomously to optimize for comfort and energy efficiency and smart homes more general, where algorithmically enabled systems can decide when to turn on lights, lock doors, or activate security systems based on user behavior and external factors.[35]

For these and other cases like them, algorithms promise to increase the efficiency of systems, manage complex interdependencies between subsystems and devices, allow the management and monitoring of large-scale systems that are beyond the control of individuals, and the automation of tedious or dangerous work. These are powerful promises that can increase people's productivity using algorithmically enabled devices. The automated, data-driven management of large-scale systems promises greater coordination as well as fewer waste.

Clearly, there are many opportunities for algorithmic action. At the same time, the fully autonomous action of algorithms raises concerns. Some are the same concerns that we have encountered with algorithmic insight or support. But we also have additional concerns that are connected to the question of scale, responsibility, and interrogability.

Algorithms can act at vast scale and speed, much beyond human capabilities. This brings opportunities, as with the automated management of smart grids, connected internet of things devices, or smart cities. This promises efficiency gains in resource expenditure, as for example energy, and timeliness of decisions. On the other hand, scale and speed can contribute to wrong decisions or those with unintended consequences to be executed automatically at a scale and where errors are difficult to detect, monitor, or control in time. Algorithmic damages might thus accrue without the opportunity for timely control or rollback.

There is also the question of responsibility of automated algorithmic action. If an algorithm makes a mistake, who is responsible? The company providing the algorithm? The company providing the data to train the algorithm? The company implementing the algorithm in their product? Or, is it the user? The most obvious example of the responsibility conundrum are self-driving cars, where the question of liability in case of algorithmically caused accidents it unclear.

Finally, there remains the question of interrogability. Scale and speed in action as well as their integration in machine-human assemblages makes it in cases of error difficult to assess how and why decisions were made and what the degree of the error lies with the underlying algorithm or other elements of the assemblage.

[35] For smart grids from a policy perspective see M. A. Brown and Zhou (2019). For smart homes see Alam et al. (2012). On energy savings in smart cities see H. Kim et al. (2021).

Example: Algorithmic trading

Finance is an example for an industry heavily shaped by algorithms.[36] In fact, hopes for the replacement of humans and the automation of financial markets go at least back to the early nineteen seventies.[37] Today algorithms support trading in various areas of financial markets. Most prominent here is one type of algorithmic trading, high frequency trading (HFT). In his review of the economic literature on algorithmic trading Albert J. Menkveld defines algorithmic traders as:

> "(...) all traders who use computers to automatically make trade decisions. An example (...) is one who aims to minimize the price impact of a large order that has to be executed. Such an order is typically executed through a series of smaller child orders that are sent to the market sequentially."
>
> Menkveld (2016), p. 8.

High-frequency traders are a particular subgroup of algorithmic traders. HFT run on "extremely fast computers running algorithms coded by traders who trade for their own account."[38] This category of algorithmic trader has received much attention and press. Back to Menkveld:

> "(...) the key distinguishing feature of HFTs is their relative speed advantage for trading a particular security. One reason for such an advantage is information technology that enables them to quickly generate signal from the massive amount of (public) information that reaches the market every (milli)second. Examples are press releases by companies, governments, and central banks; analyst reports; and trade information, including order-book updates not only for the security of interest but also for correlated securities."
>
> Menkveld (2016), p. 4.

The case of algorithmic trading is interesting as here different concerns meet. For one, algorithmic trading is often associated with public fears about loss of control about algorithms and markets or unintended consequences of uncontrolled algorithms running wild. If trading algorithms act in scale and speed beyond human control or intervention markets can crash, firms can get damaged, and private fortunes big and small can get lost. Algorithmic crashes are not merely academic thought examples. There are many examples for algorithmic crashes large and small, some of them very dramatic. Often forensics after the events show that algorithms acted too fast, had unknown or overlooked errors, and contributed to effects that were difficult to identify or disentangle even after the event. At the same time, algorithmic trading can also serve investors considerably by removing friction from trading.[39] Algorithmic trading thus shows both the opportunities as well as the risks of automated algorithmic action.

Algorithmic action thus clearly holds vast potential for a more efficient management of systems and markets. At the same time, the more systems rely on algorithms, the greater the dangers of runaway mistakes that might be consequential but difficult to identify and roll back.

4.5. Risks and fears

As we have seen, computer algorithms are used in ever more societal areas. This raises broad concerns. While in principle algorithms provide a set of clearly defined steps to solve a given problem, their current uses have raised the question whether this is still the case or whether current uses hide the steps contributing to a decision and making it difficult to understand or contest. We have already briefly encountered some of these concerns but some merit deeper discussion. In the following sections, we will focus on concerns about fairness, trapping people in algorithmically constructed bubbles and loops, the alignment problem, and the opaqueness of algorithmic decision making and its consequences.

4.5.1. Fairness

Once algorithms start shaping people's option spaces, the question of fairness emerges. Algorithms make, or at least support, decisions about people spanning various areas of their life: they assign people's credit ratings, they evaluate their job applications, they assess the likelihood of them engaging in criminal activity, or administer welfare benefits. These algorithmic assessments and decisions matter for the choices people have and the way they are treated by institutions of authority. This makes it important that algorithms are treating people fairly.

> **i** Defintion: Fairness
>
> A decision can be called *fair* if people who resemble each other regarding the decision task at hand consistently are treated the same. If people with specific characteristics not relevant to the task are consistently treated differently from those they otherwise resemble, the decision can be called unfair or biased.
> Take credit scores for example. If people with a steady job and high income consistently get good credit scores the underlying process can be called fair. A steady

[39] For a review of the scientific literature on algorithmic trading in economic see Cardella et al. (2014). For high frequency trading see Menkveld (2016). For a sociological discussion of high frequency trading as an economic sub field see MacKenzie (2021). For an agenda-setting journalistic account see Lewis (2014).
[39] See Black (1971a); Black (1971b).
[39] See Menkveld (2016), p. 2
[39] See Menkveld (2016), p. 19–20.

> job and high income are clearly directly relevant to the task of assessing whether people will be able to repay a loan. But if people with a steady job and high income who happen to be women get a lower credit rating, the decision process would be unfair. Clearly gender should not be a variable directly relevant to the ability of a debtor to repay a loan.
> Generally, different outcomes are not necessarily a sign of an unfair process. But decision making becomes unfair once differences emerge along characteristics not directly relevant to the decision task at hand.[40]

As we have seen before, algorithmic decision support for experts promises to improve on some of the limitations of human decision making. With regard to fairness, psychological biases and cognitive heuristics stand out as potentially skewing human decision making. Both might render decisions unfair by relying on not directly relevant factors. Recruiters might be looking at the school applicants graduated from as a heuristic for how to interpret their transcripts and infer the likelihood of them succeeding in a firm. Algorithms can take more information into account and produce replicable predictions for the likelihood of candidates succeeding in a firm. These can be rules-based algorithms that consider many different factors prior identified as being relevant, or these can be machine learning algorithms that in data-rich contexts can identify signals predicting success that modelers were not aware of beforehand.

Algorithms can support decision makers in making decisions fairer across various domains; either by applying complicated models consistently across contexts or by identifying new rules and models based on data. They also allow decision makers access to large amounts of evidence by incorporating it in models and their output. Additionally, they allow for the systematic auditing of decision making processes they are modeling, this potentially makes it easier to identify hidden biases and address them. But algorithmic decision support also carries specific risks to fairness that need to be accounted for.

The prominence of algorithmic decision making and decision support has led many academics, commentators, and practitioners to put special attention on the question of algorithmic fairness.[41] Are algorithms contributing to fairer and more explicable decisions or do they continue, or even worsen, discriminatory practices? Are algorithms treating all people alike or are they treating people with specific characteristics, behaviors, or group affiliations worse than others? In the examination of algorithmic fairness two important dimensions emerge. One potential driver of algorithmic unfairness stems from configurations of institutional or organizational uses of algorithmic decision systems and the associated policy goals. The other stems from the data algorithms use to build

[40]See Chapter 4 in Barocas et al. (2023).

[41]For a textbook account see Barocas et al. (2023). For two early influential accounts raising the question of algorithmic fairness in society see O'Neil (2016); Eubanks (2018). For an early account of biases in computing see Friedman and Nissenbaum (1996).

models and base their assessments on.[42]

On a foundational level, the fairness of algorithmic decision support systems depends on the structural configuration of their use and the goals of the organizations and institutions using them. Importantly, organizations can use algorithmic decision support to enable them to form *better* decisions. What *better decisions* means in this context is of course open to interpretation. Some, will try and pursue the best possible decision for a given task and very actively consider the consequences for and welfare of people subject to algorithmic decisions. Others, will interpret *better* simply as more efficient, cheaper, or better for them without necessarily considering broader implications. For those in the first category, algorithms that produce unfair results would be an issue to address, monitor, and solve. For those in the second category, unfair outcomes do not matter much as long as the algorithm achieves its primary goals for the organization. Unfairness resulting from these uses of algorithms would thus not be primarily a technological problem with a technological fix but a result from the organizational goals of algorithm use.[43]

This is of course not to say that unfairness cannot result from the use of algorithms by organizations genuinely trying to achieve fair outcomes. But in these cases, organizations can set up dedicated auditing units, processes, and provide transparency of uses and outcomes for outsiders. In the best case, this could lead to unfair results being temporary and subject to improvement.

Unfairness can also result from data and modeling choices. Importantly, patterns of past discrimination can manifest in data algorithms are trained on.[44] For example, if police officers stop Black motorists and pedestrians routinely with greater likelihood than Whites, they are more likely to pick up on otherwise undetected offenses by Blacks than Whites, for example carrying illegal drugs. This does not necessarily mean that Whites are less likely to carry illegal drugs, they are simply less likely to be caught during routine stops. But over time, police records will contain a greater number of cases of Blacks who carried illegal drugs than Whites. An algorithm predicting the likelihood of a person carrying drugs could thus easily use race as a predictor. Policing decisions based on the output of such an algorithm would thus continue and over time reinforce a police department's history of discriminatory practices.

This is just one intuitive example, but data can contain other more hidden forms of bias or discrimination as well. This includes differences between gender roles and job prospects in societies expressed in text corpora,[45] racial discrimination in medical data,[46], or discriminatory patterns in grading.[47] The identification of bias within data sets and

[42]See S. Mitchell et al. (2021).
[43]See S. Mitchell et al. (2021).
[44]See Barocas and Selbst (2016).
[45]See Caliskan et al. (2017).
[46]See Obermeyer et al. (2019).
[47]See Hanna and Linden (2012), and Sprietsma (2013).

the avoidance of biases in algorithmic assessments are very rich and promising areas of computer science and interdisciplinary research.[48]

The question of fairness within algorithmic decision making is a core question for both computer science and applied fields developing and evaluating algorithmic support systems in various contexts. As shown questions include aspects of the specific set up and constellation of algorithmic decision making and support in organizational and institutional contexts as well as technical questions associated with coverage and biases within data sets and specific modeling choices. This is an important area that is bound to grow in prominence with the growing use of algorithms in society and more pervasive awareness of them and their consequences.

4.5.2. Bubbles and loops

There are widely perceived risks associated with algorithmic recommendation systems for the public. Some fears focus on the risks of society fracturing into silos of shared tastes, interests, and partisanship. These fears react to suspected mechanisms behind algorithmic recommendation systems. Fears of fragmentation attach themselves to algorithms providing people with information and cultural products in accordance to their prior beliefs and interests.

One of the most pronounced public fears associated with algorithmic recommendation systems is the fear of *filter bubbles*. The reasoning behind the filter bubble is very intuitive. Digital platforms like Facebook, Instagram, TikTok, X, or YouTube all depend on algorithms to structure information environments for people. The companies running these platforms aim to increase the time users spend on them to achieve greater opportunities for ad display. They do so by shaping information people see according to their likelihood of interacting with it. This is where algorithms come in. Algorithms select content that has a greater calculated likelihood of people interacting with it than random content.

One potential mechanism to select content like this is to check content that people interacted with in the past and suggest to them similar content in the future. This is of course only one potential mechanism. Alternatively, the algorithm could suggest content that other people in a user's social graph – say their friend and follower network – interacted with. Or an algorithm could suggest content that a user's digital twins – other users on a platform that share demographic and behavioral patterns without necessarily being connected or known to each other – interacted with.[49]

The exact mechanisms might vary, but the expected effects stay the same. By selecting content similar to that which people interacted with in the past, or content that others that they resemble had interacted with, algorithms supposedly continue to show people

[48]See for overview of underlying issues and challenges Barocas et al. (2023).
[49]For a discussion of various mechanisms behind social media recommendation algorithms see Narayanan (2023).

information likely to correspond with prior interests or beliefs. Over time, this would reduce people's exposure on the platform to serendipitous information outside their revealed interests or exposure to information contradicting their expressed beliefs. Over time, algorithmic information filters would trap people in information bubbles of their interests and beliefs, without providing them with views from the outside.

Especially regarding politics, this mechanism was seen as a great social threat. In 2011, the political activist Eli Pariser expressed these fears in his book *The filter bubble*.[50] By showing people only information with a political stance, they already agree with, Facebook would trap people in bubbles of politically homogeneous information, leading people to lose sight of opposing political opinions and the reasons for them. Over time, Pariser argued, this would lead to political polarization and a break down of political empathy across partisan lines.

The filter bubble is probably one of the widest known ideas about the supposed effect of digital media and algorithms on politics. Luckily, its empirical foundation is very thin. Almost since its publication in 2011 various empirical studies have shown that people tend to encounter cross-cutting political information, even in heavily algorithmically shaped information environments.[51] Sometimes, they even seem to encounter more cross-cutting information in digital communication environments, than in personal exchanges.[52] Still, comparative studies show, that there is variation across digital services in the degree to which algorithms shape homogeneous or heterogeneous communication environments.[53] Specific implementations of algorithmic recommendation systems, their change over time, and specific usage patterns of different platforms clearly matter for the kind of political information people encounter there. Accordingly, it might be a little early to declare an all-clear for algorithmically shaped information environments.

The available evidence indicates that algorithmically shaped information environments do not necessarily lead to people predominantly encountering politically homogeneous opinions in accordance with their prior held beliefs. Accordingly, Pariser's suspected mechanism of how algorithms increase political polarization through filter bubbles does not hold. But this does not mean that algorithmic recommendation systems do not pose risks.

Importantly, most available evidence tends to look at general tendencies among the the general population. While it is true, that on average people do not tend to predominantly encounter political information in line with their beliefs, it might very well be that algorithms provide small, special interest, or radicalized groups with enough congruent information as to help them form and reinforce fringe, deviant, or radical beliefs and encourage them to action. In this scenario, the algorithm has not primarily the effect of isolating people in belief-bubbles, instead it functions more as a discovery and reinforcement device of fringe interests and beliefs.

[50]See Pariser (2011).
[51]See Flaxman et al. (2016); Scharkow et al. (2020); T. Yang et al. (2020).
[52]See Gentzkow and Shapiro (2011).
[53]See Kitchens et al. (2020).

Take the example of political radicalism. A user starts out by searching on a video sharing site for a fringe music group that happens to be openly or clandestinely aligned with right wing extremism. The algorithm recognizes this interest and after the first video ends suggests other content by the same band or connected to the political movement associated with it. Step by step, the algorithm pulls the user further into a communication environment with content of increasingly radical ideas. The algorithm starts a reinforcing loop. It uses a specific fringe interest and then provides suggestions for a user to start a journey further down the rabbit hole, potentially leading to political conversion and radicalization.[54] Different radical groups have been shown to exploit this mechanism for information dissemination and mobilization, including the far- and extreme-right, as well as Islamist groups.[55]

Under specific conditions, this pattern can emerge even for the general public. In situations where very little information is available algorithms can struggle to find content to recommend. This includes breaking news, specialized and coded terminology, or topics of fringe interests. Strategic actors can exploit these *data voids* and publish misleading or radicalizing content.[56] Recommendation algorithms on search engines or digital platforms will then point people using related search terms to this content, simply because information from other more balanced sources is not, or not yet, available. People starting out on their journey with content from these dubious sources can then be pulled algorithmically to other content from these sources or others like them.

To be sure, once alerted digital platforms can react, either by deleting or shadow banning illegal content or by stopping its algorithmic recommendation. But this does not necessarily solve the underlying mechanism. Importantly, this depends on the willingness of the platform to intervene and stop radicalizing loops. This might be the case for Western platforms interrupting feedback loops of illegal content, known foreign influence operations, or known militant or terrorist recruitment attempts. Radicalization on the frontiers of fringe but accepted opinions and radical domestic beliefs is much harder to identify and police through platform companies. Also, in some cases companies might have no interest to interfere, or might even tip the scale and accelerate loops.

Here TikTok's role in shaping the information environment during the Israeli intervention in Gaza in reaction to the October 7, 2023 Hamas terror attack on Israel is instructive. While a full scientific study is still not available, journalists commented on how TikTok gave a biased view of the conflict. While videos showing Palestinian suffering under the Israeli military intervention circulated widely on TikTok, videos documenting Israeli suffering during the preceding terrorist attack were all but invisible.[57] Here, questions emerge as to the influence the Chinese state has over TikTok, an internationally influential information structure provided by a Chinese company. And whether the Chinese state directly or indirectly influences algorithmic recommendation and distribution

[54] See Kaiser and Rauchfleisch (2020).
[55] See Perry and DeDeo (2021).
[56] See Golebiewski and boyd (2019).
[57] See Harwell and Lorenz (2023).

to further its geopolitical interests by shaping public opinion abroad. Especially during unfolding geopolitical crises, these questions loom large.

The presence and impact of algorithmic feedback loops is harder to show empirically, than testing for filter bubbles. It is difficult to identify in surveys and it does not hold for the general population but for small groups of people at risk of radicalization. Additionally, any empirical identification of specific loops depends on specific constellations of content providers and platforms and might be temporary.[58] Research activity is therefor lower on this specific risk than the search for elusive population-wide filter bubbles. But its impact should not be neglected in the discussion of potential harms of algorithmic recommendation systems shaping digital communication spaces.

4.5.3. Alignment

There are also considerable concerns with algorithmic systems about the alignment between the goal of the programmer and the goals pursued by the algorithm. In his book *The alignment problem*, the science writer Brian Christian charts a specific challenge underlying any design of a rule-based system (Christian, 2020). How can the designers ensure that the rules they develop and implement in a system, when followed by the letter, lead to the results they wanted to achieve:

> "(...) we conjure a force, autonomous but totally compliant, give it a set of instructions, then scramble like mad to stop it once we realize our instructions are imprecise or incomplete – lest we get, in some clever, horrible way, precisely what we asked for."
>
> Christian (2020), p. 12f.

This is what Christian calls *the alignment problem*:

> "How to prevent such a catastrophic divergence – how to ensure that these models capture our norms and values, understand what we mean or intend, and, above all, do what we want – has emerged as one of the most central and most urgent scientific questions in the field of computer science."
>
> Christian (2020), p. 13.

The alignment problem reinforces some issues, we already have encountered, such as the need for transparency regarding the uses, mechanisms, and effects of algorithms. But it also points us to a further – potentially more troubling – issue. What if we have a rule-based system which is transparent and on the face of it seems successful in achieving its goals but unobserved by the programmers and the auditors is misaligned to some unspoken but crucial values or goals of people running it. Oversights like these can mean that algorithms seem successful in pursuit of their goals but in that pursuit violate

[58] On the challenges of doing research on this question see Kaiser and Rauchfleisch (2019).

values or principles unknown to them. In artificial intelligence research, this problem is called *value alignment*.

The alignment question matters for any rule-based system but it matters especially once algorithms are deployed in scale. One strength of algorithms is their ability to provide automated and standardized decisions at scale. Once implemented, it does not matter to an algorithmic decision system how many cases it handles. Algorithms diagnosing the likelihood of cancer based on medical imagery do not care whether they work on 100 or 10.000 cases. Similarly, algorithms recommending music tracks, news items, or movies for your early evening break, do not care for how many people they provide recommendations. But the scale by which algorithms are deployed matters for auditors in identifying outcomes pointing toward an underlying alignment problem. Scale also matters for support staff fielding calls by people subject to these decisions trying to complain or to understand how their option spaces were shaped – and potentially restricted – by algorithms.

The question of scale also matters with regard to potential but unintended consequences of the application of algorithms in complex socio-technical systems. Algorithms are a technological intervention in social systems. While sometimes this happens in clearly delineated areas with little chance of spill-over effects, running algorithms at scale in only weakly delineated fields brings serious risks of unintended effects, for example through spillover or through reinforcing feedback loops. Examples include algorithms designed to increase engagement in digital communication environments. Should the algorithmic recommendation of entertainment and political content follow the same principles? Or do we need different rules for different content types? What are the potential unexpected and unintended side effects when recommending political content following rules that work for content designed primarily for entertainment? And how do we identify them? Introducing technological interventions at scale in complex social systems can carry large unintended consequences and needs to be approached and monitored with caution.

4.5.4. Opaqueness

It might come as a surprise that a concept standing for the sequential working through a list of predefined steps to solve a given problem has become a byword for opaqueness. But this is exactly what has happened to algorithms. Their uses in politics and society are opaque – be it by design or accident – and so are the specific rules they follow in their approach at solving tasks.

The legal scholar Frank Pasquale has found an evocative image for these concerns, the black box. In his book *The Black Box Society*, he discusses the opaqueness surrounding the uses of data and algorithms in assessing people's reputation through their data shadows, shaping their option spaces by providing answers to their searches, and providing or denying them financial opportunities through financial service providers (Pasquale, 2015). Pasquale finds that various actors work actively on keeping their uses of algorithms and their inner workings secret. The most troubling of which is the denial of transparency by companies providing algorithmic decision-support systems employed to provide or deny people opportunities, be it in the provision of credit or in granting or denying them parol, supposedly to protect trade secrets.

Beyond active attempts at keeping the uses of algorithms in society opaque for security, commercial, or legal reasons, opaqueness can also emerge as a byproduct of the use of machine learning algorithms. Here, algorithms are set to solve a problem. But they do so by identifying the most promising sequence of steps to their solution themselves. While clearly successful in many cases, the uncertainty just about how the algorithm goes about in solving the task creates opaqueness and uncertainty.[59]

Once algorithms shape people's option spaces across different societal fields – such as information spaces, finance, or the legal system – meaningful transparency about their uses, inner workings, and effects becomes crucial. Meaningful transparency encompasses information about in which fields, which actors use algorithms to achieve which goals. This also means that these actors are explicit about the rules, the sequence of steps, they programmed the algorithm to pursue. This goes double if they outsource these uses to vendors providing them with algorithmically-supported systems or solutions. It is also important to know what the outcomes of the algorithms are. Do they conform with expectations or are they creating biased or unintended results. Finally, people subject to algorithmic shaping need places to appeal algorithmic decisions and sources where they can find out why the decisions went the way they did. Only through regular and transparent audits and ongoing critical debate, can the sense of opaqueness created by the use of algorithms be countered and the potential of algorithms to improve society be realized.

This demands for transparency about algorithms, the data they use, their workings, the ways they are deployed, and evaluated. Without such transparency, algorithms threaten to create an algorithmic cage whose circumference shifts for unknown reasons and by invisible forces. Without meaningful transparency, this threatens to become a digital version of the dark comedies and nightmares about unresponsive, fickle, and opaque but merciless structures charted by Franz Kafka in his novels *The trial*[60] and *The castle*.[61] Like the structures charted by Kafka, an opaque algorithmic cage would suddenly shut close on its inhabitants without offering reasons or meaningful options of appeal. Clearly, this cannot be the template for the application of algorithms in politics and society.

[59] For an account of how people adjust to opaque algorithms affecting their option space see Rahman (2021).
[60] Kafka (1925/1999).
[61] Kafka (1926/1998).

4.6. What can be done?

These worries need to be taken seriously and addressed. Not the least because computationally automated decisions based on algorithms are now prevalent in various areas of society. Accordingly, covering the uses, effects, and regulation of algorithms is an important field of scientific activity going forward. Not the least because popular accounts might have little to do with actual uses or effects of algorithms.

At the same time, worries should not distract us and blind us with fear. The use of algorithms in society is not optional. So these issues should point us to improving algorithms and their use, not mislead us to entertain the illusion of somehow avoiding their use. This includes conscious and informed approaches to mechanism design underlying algorithms and algorithm auditing and forensics.

Mechanism design is an area in economics, interested in the design of rules to achieve desired outcomes in strategic environments.[62] This is crucial in algorithm development, where rules must be carefully crafted to align with societal objectives. For instance, when algorithms are used to distribute resources, such as in traffic flow optimization, mechanism design ensures that equity and efficiency are considered. The success of these algorithms is predicated on their ability to adapt to human needs and the nuances of social behavior. Mechanism design does of course not ensure this success, but at least it provides a systematic and contestable approach in the development of rules underlying algorithms.

While mechanism design is one way to work on better algorithms by designing and interrogating rules and processes, algorithmic auditing and forensics focus on the outcome of algorithms. Auditing focuses on the continuous analysis of algorithm outputs and checking whether outcomes are fair, unbiased, and transparent.[63] Algorithm forensics focuses on the analysis of an algorithm's actual decision making process, in case audits find results to deviate from the goal of algorithm deployment. Of course, algorithm auditing is not a one-time process but a continuous one, ensuring ongoing accountability and alignment with evolving societal values.

Evaluating algorithms is a difficult task. It requires not only technical assessments of performance and accuracy but also evaluations against societal impacts and ethical standards. Transparency is a key component of this evaluation, providing insight into the algorithm's function and fostering trust. For instance, an algorithm used in loan approvals should be transparent enough that applicants can understand why they were or were not granted a loan. This clarity is essential for building trust between users and algorithmic systems.

[62] See Roughgarden (2016).
[63] See Metaxa et al. (2021).

Example: Interrogating algorithms

Interrogating algorithms can take different approaches. But the following set of questions provides an impression of the potential scope and variety of these audits. Let's start with the basics:

- "What problem is the algorithm intended to solve?" This foundational inquiry sets the stage for our audit by clarifying the algorithm's purpose. It prompts us to consider if the problem was clearly defined and if an algorithmic solution is indeed appropriate.
- "How was this problem solved previously?" This historical context is helpful, offering insights into the evolution from past methods to the current algorithmic approach, highlighting improvements or differences, and the limitations of previous solutions.

Next, we navigate the rule development process:

- "Through what process were the rules developed?" Here, we scrutinize the governance of the algorithm. Were experts consulted? Were people subject to the algorithmic decisions and actions consulted? How were ethical considerations and potential biases addressed during this phase? The integrity of the algorithm is often rooted in the inclusivity and rigor of its development process.

Following this, the evaluation process, transparency, security, and regulatory compliance of an algorithmically enabled system need to be made explicit:

- "How is the algorithm evaluated?" An algorithm must not only be built on solid ground but also must continually prove its worth. What metrics or criteria are established for its evaluation? Is there room for independent review, and does the algorithm demonstrate transparency in its decision-making process?
- "Is the algorithm's operation transparent?" We demand clear explanations of how the algorithm processes inputs to produce outputs. Transparency isn't a luxury; it's a necessity for accountability and trust.
- "How does the algorithm protect personal data?" and "What safeguards are in place to maintain data integrity?" It's not enough for an algorithm to be effective; it must also be a guardian of user privacy and data security.
- "Does the algorithm adhere to pertinent regulations, and how does it remain current with legal standards?" An algorithm must follow rules and regulations within the digital realm and continue to evolve with the regulatory landscape.

An audit could now turn to the outcome of an algorithmically enabled system. First, we focus on the outcomes on the individual then on the societal level:

- "How does the algorithm account for fairness, and what measures are in place to combat biases?" This leads us to consider whether diverse datasets were employed to ensure equitable performance across different demographics.
- "What are the societal implications of deploying this algorithm?" From ethical dilemmas to potential job displacement, we must be prepared to confront and manage the ramifications of our technological advancements.

These questions are of course not exhaustive. But they can provide a first framework to illustrate the reach and different perspectives algorithm audits can take. It also illustrates the demands that systematic and continuous audits of algorithmic systems put before us.

Current practices in various sectors illustrate the application of mechanism design, auditing, and transparency. For example, in healthcare, algorithms assist in diagnosis and treatment recommendations. Here, mechanism design is crucial to ensure that the algorithms prioritize patient outcomes and ethical considerations.[64] But just as important as design is the audit of algorithms used in medicine.[65] Auditing these systems is an interdisciplinary effort, involving not only medical professionals and computer scientists but also data scientists evaluating whether the recommendations are accurate and free of bias.

These examples illustrate different approaches how the rules underlying algorithms can be consciously designed and their outcomes and workings examined. Of course, the ease of this endeavor varies. In dealing with algorithms deployed in clearly specified and limited contexts where processes are well understood and outcome distributions lend themselves to easy comparison with ideal results, this process is comparatively easy. In more complex environments, with many interacting features, and no clear benchmark for outcome distributions to be compared to, ensuring the correct and fair working of algorithms is much more difficult. Nevertheless, this section shows that algorithms and their results need not be a black box. Instead, algorithms and algorithmically enabled systems can be designed and evaluated. But society needs to choose to do so and demand of developers and deployers of algorithmically enabled systems the ability to act on this. As algorithms and algorithmically enabled systems shape the fabric of society more extensively, our focus must remain on developing and refining mechanisms that ensure their benefits are maximized while their risks are minimized.

[64] For an instructive examples of the challenges of identifying flaws in an algorithms in the field of organ donation see Murgia (2023).

[65] See Liu et al. (2022).

4.7. The promises and the risks of automation

Clearly, the combination of data and algorithms is very powerful. It promises not only new insights about the world, it also promises the ability to act on these insights and thus shape the world and future. Algorithms learn from data by identifying regularities that they then use to predict the future. But algorithms also use data as inputs to initiate action. The realization of the potential of data and algorithms are mutually dependent. The quality of algorithms and their output depend on data and their quality. At the same time, the realization of the potential within data, that we discussed in Chapter 3, depends on algorithms.

Algorithms automate insight and action. This holds great potential. Societies encounter a continuously growing set of complex challenges with no clear solution, be it climate change, migration, international conflicts, aging society. These challenges are difficult to navigate. Algorithms with their capabilities of synthesizing and creating insight from data can be of great use in tackling these challenges.

Individuals face ever increasing choices in life, consumption, and information environments, while at the same time often facing tightening economic and temporal resources. Algorithmically shaped choice environments and advice can help people in navigating otherwise potentially overwhelming options and to make better choices in face of growing constraints.

By automating tasks at work, algorithms can help workers to be more productive and business to automate processes. This is often discussed as a threat to workers and prosperity. But in times of severe labor shortages in Western democracies, this can also be an opportunity to ensure prosperity in the context of a rapidly aging population, either by substituting labor or by making people at work more productive.

Realizing these potentials is not certain, though. For one, it is unclear how far the combination of data and algorithms can extend. Does data-enabled insight hold for all walks of life or are there limits? Can automation extend beyond the mere digital, or digitally mediated? Questions that we will reencounter when we will discuss artificial intelligence later in Chapter 7. In discussing algorithms it is easy to get seduced by the potential, when in fact there are very clear technical limits to their applications. In recent years, these limits have continuously been pushed back through technological innovation. Nevertheless, limits persist and need to be accounted for in any serious discussion.

More specifically, algorithmically enabled systems come with additional challenges, like ensuring fairness, avoiding bubbles and loops, providing alignment, or making what was opaque transparent. These are non-trivial challenges as they often point to underlying problems within societies that either shaped data or the ways algorithmically enabled systems are developed and deployed. They do not void the potentials of algorithms for society, but these challenges have to be addressed for this potential to be realized broadly and not to only benefit the few at the cost of the many.

Making algorithms work for society, is an interdisciplinary effort. Computer scientists, data scientists, social scientists, and practitioners from various fields must work hand in hand to develop algorithmically enabled systems and deploy them for different tasks in different contexts. Only through and open dialogue about opportunities, risks, and failures, can there be improvement of systems and over time trust in algorithmically enabled systems. The coming challenges are too big as that we voluntarily good give up on the potentials algorithms hold. At the same time, inherent risks and limitations need to be accounted for and constructively addressed.

4.8. Further Reading

For a non-technical introduction to algorithms in computing see Louridas (2017).

For fairness in algorithmic decision making see Barocas et al. (2023).

For an account of how algorithmic decision making can increase inequality see Eubanks (2018).

For a popular account of the problems of aligning algorithms to the goals of their deployers and larger social norms see Christian (2020).

For an account of the dangers of opaque algorithms and automation in society see Pasquale (2015).

For an introduction to an economic perspective on how to design better algorithms see Roughgarden (2016).

4.9. Review questions

1. Please provide a definition for the following terms:

- algorithm following Knuth (1968/1997);
- fairness in the context of algorithms;
- bias;
- alignment problem.

2. Please discuss the ways that *bias* can negatively impact the way that algorithms gain insight from data? What can we done to mitigate associated risks?

3. How can we assess the *fairness* of algorithmic support systems? How can we improve on this?

4. Discuss the threat of the *alignment problem* for algorithms that automate action. How can we account for this problem?

5. Assess the threat of *filter bubbles* for society on the basis of the available empirical evidence. Where could the threat lie? Where not?

6. Please discuss how you would approach an algorithm audit for and algorithm recommending content to people on a video sharing site like *YouTube*. Sketch the relevant question that need to be asked and discuss what data you would access or analyses you would run under ideal access conditions.

5. Challenge

All over the world, we see people using digital media to question and challenge authorities, organizations, norms and behaviors they perceive as dysfunctional or unjust. Digital media are therefore an important element in the challenge of established social institutions, sometimes even enabling these challenges in the first place.

Some of these challenges are aimed at expanding social representation and strengthening democratic participation. We find examples of this in the use of digital media by social movements, such as Black Lives Matter in the US. Other challenges aim to restrict representation and participation, as the example of the use of digital media by various right-wing populist movements and parties shows. Digital media can therefore contribute to strengthening societies and democracies as well as to weakening them.

In this chapter, we will discuss this role of digital media in politics and society in detail.

5.1. Digital media and the challenge to institutions

Three examples show how digital media enable the challenge of institutions. Journalists in the United States are responding to their experiences during Donald Trump's presidency by calling for a new normative foundation of journalism. Instead of an objective and neutral journalism, they demand a journalism that takes a clear position on social, moral, and political issues, a journalism based on *moral clarity*. They use digital media to advertise their normative reorientation and to show their broad social acceptance.[1]

In Italy, against the backdrop of the Berlusconi years and various technocratic governments, political activists question whether representative democracy actually allows the population to control their elected representatives and to ensure that they act in the interests of voters. They use digital media to form a new party that works according to the principles of direct democracy and binds its elected officials very closely to the will of the party members – *Movimento 5 Stelle*, the five-star movement, is born.[2]

The belief that prevailing standards and conventions in science are no longer suitable for adequate quality assurance of findings is spreading among young scientists. Widespread conventions in statistical analysis, public availability of data, and the documentation of analysis steps are put to the test and the approaches of scientific luminaries and

[1] For *moral clarity* in journalism, see Gessen (2020); Lowery (2020); Wiedeman (2020).
[2] For more on the five-star movement see Deseriis (2020).

entire subfields are publicly criticized. This primarily does not happen in the channels of established journals or the conferences of umbrella organizations, which are intended for this purpose in science. But this happens via digital channels. The logic of authority is replaced by the logic of public verifiability. On the one hand, digital media serve as a public forum for discussion and criticism. At the same time, however, they also serve as an infrastructure that enables the establishment of new standards for analysis and documentation. Digital media provide the basis for *open science*.[3]

In all three cases, important social institutions - journalism, party democracy, and science - are publicly challenged with the help of digital media. The challenge comes from young people who are either already part of these institutions or who look at them from outside. Contemporary practices or functioning of these institutions are publicly criticized if they deviate from their normative legitimacy and alternatives to their normative bases or their real-existing working methods and practices are formulated.

The global accumulation of these challenges in very different countries and institutional fields is clearly related to digital media. But before we examine this mechanism in more detail, we must first ask ourselves what institutions are, what function they have for societies, and the conditions under which they are challenged or lose their legitimacy.

5.1.1. What are institutions and what do they do?

Institutions are an important and widely used concept in the social sciences. However, importance and popularity bring the disadvantage that the term is used in a variety of ways and with different meanings. So let's start with a definition of the term *institution* to make sure we are actually talking about the same thing.

In his essay *Political Institutions and Social Power*, the sociologist and political scientist Claus Offe defines institutions as:

> "(…) systems of *rules* that apply to the future behavior of actors. They constitute actors and pro-/prescribe their scope and mode of action. These rules can be sanctioned through mechanisms that are specified in the charter, or legal specification, of an institution. These rules are, consciously or habitually, observed and complied with by actors who are aware not only of the rules but also of the fact that these rules are being enforced and deviant courses of action sanctioned. Institutions often impose severe constraints on what actors are permitted to do."
>
> Offe (2006), p. 10.

In other words, institutions are systems of rules that influence people's behavior. They have different levels of control and sanctioning options that punish misconduct. Since

[3] For open science see Christensen et al. (2019); M. Nielsen (2012), Wuttke (2019).

they have a normative character, they can be criticized and challenged for their legitimacy and effectiveness.

Of course, this is only one possible definition. However, Offe's focus on institutions as societies' building blocks, their normative character and possible and legitimate challenges posed by individuals and groups is very suitable for examining the impact of digital media in society and politics. Other conceptualizations and definitions of what constitutes institutions may be more suitable for other research questions and projects and are of course completely valid but shall be ignored here.

The institutions from our previous examples correspond to Offe's definition. Take journalism as an example. Journalism can be understood as a binding set of rules for organizations and people belonging to the institution. These rules can relate, for example, to the selection of news items, the representation of political factions in a society or the quality assurance of reports. Violations of these rules are punished either within newsrooms and media houses, by professional associations or in public. Violations of rules can lead to the exclusion from newsrooms for individual journalists or to punishment by professional associations or the public for media organizations. In extreme cases, particularly intense violations of the rules and regulations can lead to a general population-wide loss of trust, which can also be associated with a general loss of legitimacy for the institution *journalism*.[4]

Institutions are not neutral elements of social or political competition. Instead, they are the expression and instrument of existing power relations in society. In the words of the sociologist Manuel Castells:

> "[Societies] are contradictory social structures enacted in conflicts and negotiations among diverse and often opposing social actors. Conflicts never end; they simply pause through temporary agreements and unstable contracts that are transformed into institutions of domination by those social actors who achieve an advantageous position in the power struggle, albeit at the cost of allowing some degree of institutional representation for the plurality of interests and values that remain subordinated. So, the institutions of the state and, beyond the state, the institutions, organizations, and discourses that frame and regulate social life are (...) crystallized power relationships (...) that enable actors to exercise power *over* other social actors in order to have power *to* accomplish their goals."
>
> Castells (2009/2013), p. 14.

This perspective emphasizes the role of institutions as the outcome of earlier social conflicts and expressions of earlier power relations and norms.

Back to Offe:

[4]For journalism as an institution, see Kiefer (2010).

> "Social power manifests itself in a mode of action that has the effect of setting parameters for the action of other social actors, be it in favorable or unfavorable ways, as seen by those others. In either case, the exercise of power is conflictual, controversial, and contested. In this conflict, some legitimating norm of (political, social, economic) justice is invoked and appealed to. The exercise of power affects others in ways that are perceived by them to be justice-relevant, either fulfilling or violating standards of justice. Given the controversial and essentially contested nature of these standards (…) any institution can be criticized for failing to live up to some version of justice."
>
> Offe (2006), p. 20.

For Offe, this institutional challenge can be triggered either by a *crisis*:

> "One way in which institutional failure may happen is through a more or less accidental change of conditions in the external world that undermines the viability of institutional patterns or limits their ability to function. If that happens, rules and institutionalized goals and power relations are rendered untenable, whether because of some emerging discrepancy between an institutional complex and its economic, demographic, or technological environment or because of an evolving lack of fit between institutional complexes (…). In either of these cases, actors who have so far complied with institutional practices will start a process of (potentially self-accelerating) defection. (…) institutions may lose their"fit" with the external context conditions on which they depend, and hence their viability."
>
> Offe (2006), p. 18–19.

Alternatively, a general loss of legitimacy by institutions can also lead to their challenge. This is what Offe calls *conflict*:

> "Other cases of institutional breakdown grow out of the failure, or loss of moral plausibility, of the implicit theory of a just social order that comes with any institution. Institutions can implode because of a shortage of the moral resources and loyalties that are needed for their support."
>
> Offe (2006), p. 19.

Here, Offe presents two mechanisms through which institutions can be weakened and publicly challenged.

In a *crisis* the institution loses its fit with current social events or concerns. This can happen through external shocks - for example technological change and subsequently changed social conditions. The lost fit of an institution leads to a deterioration in its functions and effects. Consequently, people no longer adhere to the institutional set of rules and the institution loses its power to sanction deviancy.

In a *conflict* an institution is ignored or publicly challenged if it has lost its moral legitimacy for parts of society. This can either be due to a degeneration of the institution itself – i.e. an internal decline that leads to parts of the institution no longer adhering to their own set of rules or its normative basis. However, it can also be due to a social change in values outside the institution. Over time, the values a society shares change. Challenges to institutions might arise once values that form the basis of its normative rules at the time of its founding become seen by parts of society as outdated. Finally, it may also be that there has been a shift in the social balance of power and that the institution is seen as an expression of old, outdated power relations and accordingly rejected.

Both of these mechanisms of institutional decline can be triggered or reinforced by digital media. At the same time, digital channels create space for articulation and coordination for people and groups who are dissatisfied with current institutional configurations. Of course, this alone does not say anything about the legitimacy, direction, or impact of such challenges. But more on that later.

5.1.2. Digital media as staging area for challenges to institutions

What is the role of digital media in the challenge to institutions? The sociologist Manuel Castells offers some helpful considerations in his book *Communication Power*.

In the preface to the second edition of *Communication Power*, Castells describes the role of digital media in the various challenges to established institutions that we encounter all over the world.

For Castells, the control of the means of communication by the state and social elites is an important element in their exercise of power. By controlling the flow of information and means of communicative coordination, social elites can prevent alternatives to their rule or the institutions that support their rule from being developed and coordinated. For Castells, new communication technology leads to the weakening of this central control and thereby enables challenges to established power relations and structures to manifest:

> "Any new technology of communication, such as the printing press, has challenged authority, because the seeds of revolt existing in most individuals who are embedded in perennial unjust forms of social organization can only grow and blossom when they are connected to other individuals, breaking the barriers of individual experience to become social mobilization and alternative projects of social organization."
>
> Castells (2009/2013), p. xxi–xxii.

This is where digital media come in. The classic model of mass-media communication largely followed a predictable one-to-many pattern. You had one medium – be it a newspaper, a television program, or a radio program – and this medium communicated to an audience of many people. One sender – many recipients. Now, digital media shifted this model to a many-to-many pattern by allowing people to become communicators themselves. In the early days of digital technology, this process was cumbersome. People had to learn to code or host a website to be able to put their voice online. Today, this has become much easier. People can simply open up a social media account and communicate their views in text, audio, or video. These views potentially can be seen by large numbers of other internet users. Communication runs from many senders to many recipients. This makes communication much harder to control. With mass media, it is comparatively easy to control what is said by whom to which audience. With many-to-many communication, this becomes much harder.[5]

Back to Castells:

> "Thus, multiple messages emerge and multiple meanings can be constructed by the actors, who at times agree on meaning, and at other times disagree over the construction of the meaning, but who are nonetheless largely independent of the agenda-setting strategy of the deciders in the mass communication paradigm."
>
> Castells (2009/2013), p. xxii.

For Castells many-to-many communication, such as digital media, allows for the emergence of a *communication sphere* that allows for those disappointed, discriminated against, or not represented by established institutions to voice their disappointments, find each other, and propose alternatives.

> "Established institutions, in every domain of life, are challenged by those who feel dominated, devalued, exploited, humiliated, and misrepresented. These challenges need to confront the coercive capacity of institutions as well as the persuasive ability of the dominant mindset that legitimizes existing forms of power relationships. (…) Unlike the deliberative institutional sphere that is systematically biased toward existing domination, the communication sphere is shaped by the multiple inputs it receives from a diversity of sources, as well as by their interaction. The larger and broader these inputs are, and the faster the speed of their interaction, the more the communication sphere becomes a driver of social change."
>
> Castells (2009/2013), p. xxii.

Digital media, in so far as they open up communication for many-to-many and lessen the degree of control of established institutions over who says what to whom, provide a space in which the perceived and objective failings of institutions can be exposed, discussed, and alternatives formulated. The growing prevalence and ease of use of digital media

[5] For an early discussion of different characteristics of one-to-many and many-to-manny communication environments and their consequence for strategic communication see Hoffman and Novak (1996).

thus renders institutions vulnerable to challenges. At the same time, this dynamic makes digital media into important elements to the widespread challenges to institutions we witness.

Still, it is important to voice two caveats at this point. First, challenging institutions can have good as well as bad consequences for societies. Challenging institutions in order for them to achieve a higher degree of inclusivity and representativeness strengthens societies. Challenging institutions in order to restrict access to others or silence voices on the other hand is much more troublesome. Challenges with either goal are prevalent and digital media plays a role in both. So digital media by themselves are neither a force for beneficial or detrimental change in societies, they simply allow for the challenge of the institutions that support the powers that be. How that challenge is phrased and in which direction change is supposed to happen lies with the challenger. We will get back to this.

The second caveat concerns that expectation that digital media are necessarily lessening the degree of control over communication. In the early days of digital media they were perceived as providing a largely horizontal communication environment in which voices and content became visible based on its merit and the interests of users. This view has been challenged importantly by researchers like Matthew Hindman who in his book *The Myth of Digital Democracy* pointed out that actually for information to reach a lot of people in digital environments, it had to be picked up by one of a few central information hubs that garnered a lot of attention. What had looked to early commentators, such as Castells, like a horizontal communication environment was actually dominated by a few central actors who decided where the attention of most internet users were focused on at any given time. Digital media might allow the publication of alternative voices but did not guarantee their visibility. In the words of Hindman:

> "It may be easy to speak in cyberspace, but it remains difficult to be heard."
>
> Hindman (2009), p. 142.

Since the early days, the potential for control of digital media has only increased with the move from publishing information on individual websites to a few central platforms – such as Alphabet/Google, Amazon, Meta/Facebook, or Twitter. Those platforms provide the backbone for content publication, distribution, and access to digital media. This potentially constitutes a return to the centralized control of communication to which Castells saw digital media providing an alternative. In the light of this, the support digital media provides to the challenge of institutions might turn out to be less of a constitutive element of digital media itself but a byproduct of a specific configuration in its implementation.

5.2. How do digital media drive the challenge of institutions?

5.2.1. Intermediary institutions and the flow of information in democracies

Democracies depend on structures that connect governments, political elites, and the public. They facilitate information flows between different actors and different societal levels in democracies. Institutions like political parties, interest groups, and the news media make publics visible to elites, elites visible to publics, and publics visible to each other. They enable information flows making visible or allowing for the social construction of concerns, grievances, and interests of publics to elites and governments, while making elites and governments visible and – within bounds – transparent to the public. In this function, they provide, aggregate, and filter information.

In political science, political parties, interest groups, and news media feature to varying degrees. But, especially with regard to the first two institutions, their role as information intermediaries is often neglected. This is somewhat surprising as their function in this role determines the sense of representation in democracies. In fact, the political scientist Jan-Werner Müller has recently characterized these intermediary institutions as *critical infrastructures* of democracy J.-W. Müller (2021), p. 89–137. In this, he emphasizes their crucial role and focuses attention on the normative principles that should guide their work and the analysis of their actual practices and impact.

By surfacing and aggregating voices, perspectives, and grievances intermediary institutions, like political parties or the news media, provide representation for groups in society. They aggregate information. They also filter voices, perspectives, and grievances that fall outside a shared democratic framework or violate shared democratic or discursive norms. These could be extremist voices on the political left or right advocating the exclusion of others from the body politic or the restriction of their rights. Or this could be hateful or discriminatory voices trying to poison discourse and exclude or denigrate others. In this, intermediary institutions filter information.

But representation is not just aggregation - simply counting individuals or mirroring groups and attitudes according to their relative strength. Instead, representation mediated through institutions like parties, interest groups, or news media is also about the construction of identities, agendas, allegiances, and conflicts. Channeling Bourdieu (1990), Jan-Werner Müller observes:

> "Here representation is not conceived as substantively or descriptively reproducing something that already exists. It is not a matter of mechanical reproduction. Rather, it is a process in which individuals offer to a possible constituency an image of themselves based on so far unrecognized ideas, interests, or aspects of their identities. As a result, citizens might perceive themselves and the politics they need in a novel light. A constituency is not so much reproduced, or even revealed, as talked into existence and, as a result, uses its political freedoms in novel ways."

J.-W. Müller (2021), p. 79–80.

For Bourdieu, this is a crucial feature of politics:

> "The power of imposing a vision of divisions, that is, the power of making visible and explicit social divisions that are implicit, is the political power *par excellence*: it is the power to make groups, to manipulate the objective structure of society. As with constellations, the performative power of designation, of nomination, brings into existence in an instituted, constituted form (...), what existed up until then only as (...) a collection of multiple persons, a purely additive series of merely juxtaposed individuals."
>
> Bourdieu (1990), p. 138.

Examples for this construction of identities or the structuring of political conflict through intermediary institutions include on a large scale the forming and reproduction of national identity - what the political scientist Benedict Anderson has called "imagined communities" Anderson (2016). On a smaller level, this also includes the formation of group identities around focusing events, surfaced grievances, or campaign slogans and the structuring of political conflict along these new lines as witnessed recently on Twitter around hashtags as #blacklivesmatter, #metoo, or #fridaysforfuture, practices that Jackson et al. (2020) discuss in their recent book *#HashtagActivism*. The nature, inner workings, and influence of intermediary institutions structuring information flows and representation in democracies matter therefor, as they structure politics and political conflict.

Now we know that and why intermediary institutions – like parties, interest groups, or the news media – matter in democracies. But how is that connected to digital media and the current technological and political changes we are witnessing? To answer this question, we have to turn to one of the few political scientists who explicitly thought about the impact of technology on these intermediary institutions and provided a theory on how they are impact through technological change.

In his 2003 book *Information and American Democracy: Technology in the Evolution of Political Power*, Bruce Bimber analyzes shifts in communication technology and their effects on political organizations in their function as intermediary institutions enabling the flow of information in democracies. In his analysis, Bimber clearly shows that intermediary institutions in democracies are closely connected with the communication technology of their day.

> "[...] information regimes exist in American political history as periods of stable relationships among information, organizations, and democratic structure. The features of an information regime are: (1) a set of dominant properties of political information, such as high cost; (2) a set of opportunities and constraints on the management of political information that these properties

> create; and (3) the appearance of characteristic political organizations and structures adapted to those opportunities and constraints."
>
> Bimber (2003), p. 18.

But information regimes are not necessarily stable. In fact, technological change can lead to their disruption. This is what Bimber calls an information revolution.[6] Bimber provides examples for these information revolutions in the US: such as the growing availability of the postal service and newspaper during the 19th century; and subsequently, the emerging mass audience through the growing availability of first the radio and then television. These information revolutions change politics by shifting the option space available to political actors, thereby potentially shifting the balance of power. Bimber writes:

> "An information revolution disrupts a prior information regime by creating new opportunities for political communication and the organization of collective action. These changes create advantages for some forms of organization and structure and disadvantages for others, leading to adaptations and change in the world of political organizations and intermediaries. This is to say that democratic power tends to be biased toward those with the best command of political information at any particular stage in history."
>
> Bimber (2003), p. 18.

Shifts in technology thereby can lead to shifts in the nature and inner workings of intermediary institutions shaping the flow of information in democracies. They provide new opportunities for new actors and limit opportunities of established institutions, adapted to an earlier stage of technological development.[7] This is exactly what we are currently witnessing with the impact of digital technology on politics and society. Digital technologies have shifted important features of the information environment democracies face, this leads to the challenge of established intermediary institutions – such as parties, interest groups, and the news media – and the emergence of new actors and practices. The driving factors behind these challenges are digital media. They do so in three ways, by:

- lowering the costs of information production, distribution, and access,
- lowering the costs and opportunities in the coordination of people, and
- providing the communication backbone for an increasingly dense network of interconnections between people all over the world.

We will now discuss the first two of these factors and their consequence for politics and democracy in greater detail.

[6] For a detailed discussion of these information revolutions, the underlying technological changes, and their subsequent effect on democratic institutions see Bimber (2003), p. 34–88.
[7] For more on this reasoning see Jungherr, Schroeder, and Stier (2019).

5.2.2. Lowered costs of information production, distribution, and access

In times when information production, distribution, and access are comparatively expensive people depend on intermediary institutions that have the necessary means to provide these tasks. In times when information is comparatively cheap, people do not have to rely anymore on intermediary institutions for their information. Institutions therefor lose their monopoly on information for a specific group and are potentially weakened through this loss of function.[8]

In the past people belonging to a specific political faction depended on a central institution – a party – to provide them with information about their local chapter, the goings on in the capital, and how the leadership of the faction positioned itself regarding current events. Typical communication channels available to the party were regular communiques – such as letters or member magazines – or regular meetings in which the leadership informed local party members about current events. At the same time, the party also made specific members, practices, or concerns visible to the party elite and the broader base – for example by featuring them in their communications. This made the organization and its attitudes and opinions visible to itself. The cost of collecting, producing, and publishing that information gave the intermediary institution power. The institution could decide which information to publish, which elite-position to communicate to the base, and what section of the base to feature. The institution had an important gatekeeper function for the faction. This gave it power over the base.

Today, members of political factions do not have to rely on party institutions for information anymore. Instead, they can take to politically aligned but not institutionally controlled digital media to learn about current events. Or they can take to Twitter or their messenger group of choice to learn what other politically likeminded people are saying. By making information production, distribution, and access cheap, digital media have provided new informational options beside those of traditional intermediary institutions. This weakens intermediary institutions in at least two ways:

First, and most obvious, by offering alternative voices the opportunity to reach people without the explicit or at least tacit agreement of an intermediary institution, digital media enable the challenge of the power of intermediary institutions over their fields.[9] As we have seen Castells diagnose, central power over information means central power over people. This central power over information is challenged. While in the past, the party leadership could use their institutional power over information flows to set the agenda or to feature or silence specific groups within the faction, today this is increasingly difficult. Many alternative media outlets are available to activists that inform them about alternative ways of seeing the world, different from those party orthodoxy or leadership favor and Twitter quickly surfaces dissident groups within factions who share a view different from the leadership.

[8] For a broader discussion of lowered information costs and new communication environments see Jungherr et al. (2020), p. 69–131.

[9] See for this reasoning Jungherr, Schroeder, and Stier (2019).

This makes intermediary institutions more difficult to lead. Internal differences can surface quickly and gain traction through the instant visibility on digital media. Examples include, fringe positions in political groups gaining visibility on social media and in turn being amplified through media coverage. This can hamper the opportunities for parties to establish comprise with other political groups or even empower extremists allowing them to publicly speak for a faction instead of more moderate representatives.

At the same time, this opens intermediary institutions up to outside influence. For example, by funding politically but not institutionally aligned digital news sites, outside interests, such as business or interest groups, can challenge the power of central intermediary institutions. By establishing cheap alternatives to institutionally vetted information sources, outsiders can shift the agenda within factions and feature selected groups or individuals that challenge the central authority of the institutions.[10]

Now, these challenges of intermediary institutions can be empowering. We have seen above, how institutions can deteriorate with regard to their central functions. By enabling challenges, digital media might thus point the way to helpful reform by empowering alternative approaches to gain visibility and power. At the same time, digital media can also be used by challengers who are not interested in a broadening of access to intermediary institutions but instead try to narrow it to benefit their own specific political creed. These challenges are also enabled through digital media.

5.2.3. Lowered coordination costs

Digital media have also lowered the costs for people to coordinate.[11] In the past, this was another function that intermediary institutions held for political factions or social movements. In technology regimes where coordination is costly, people need intermediary institutions to carry the cost and run the logistics. This could mean collecting, maintaining, and updating lists with contact information with organization members and sympathizers. Or this could be the collection of funds to tide over members who faced income loss through strike or protest action, or to cover legal costs members face due to actions for the group.

In technology regimes with high costs of coordination, political activists need to go through existing intermediary institutions – existing groups, non-governmental organizations, or parties – for collective action or to get their voice heard. This gave these institutions power over which type of position or group to support and which to ignore. This had beneficial effects for discourse and politics by filtering out radical positions or groups and being able to coordinate around central causes and focus public attention and participatory energies. On the other hand, if intermediary institutions fail in

[10]See Popkin (2021).

[11]For a broader discussion of the opportunities of digital media for political coordination see Jungherr et al. (2020), p. 132–157.

their selection function, this can deteriorate into exclusionism, allowing insiders of the institutions to pursue their interests and topics and exclude those by outsiders.

The lowering of coordination costs through digital media allows the challenge of these institutions. By providing low costs digital tools that allow interested people to find each other, organize, and provide tactical support during collective action digital media offer alternatives for activism and participation to established intermediary institutions, such a parties, activist groups, or non-governmental organizations.

This has both good and bad consequences. For one by using digital tools new groups of people with shared but unrecognized grievances can find each other and coordinate collective action. They do not depend on existing intermediary institutions validating and picking up their cause in order to achieve visibility in the public arena or get access to the resources necessary to organize large-scale collective action events. Instead, they can find supporters and sympathizers on existing digital media infrastructures, there they can also get the word out and document their cause and actions, as well as collecting the resources necessary to pursue them.

Free services like Google Mail or Google Docs provide any group with the resources to run a powerful distributed co-working environment without any costs. Meta/Facebook, Instagram, Twitter, YouTube, and TikTok allow to get the word out and to document activities. Crowdfunding platforms allow to collect financial resources from a broad base of supporters. And petition platforms allow groups to quickly visualize the supposed size of supporters behind their cause.

These opportunities enable groups not recognized or represented by established intermediary organizations to coordinate around their cause and gain public visibility. These opportunities lie behind public successes of groups like Occupy Wallstreet, Me Too, Black Lives Matter, and Fridays for Future.[12] In enabling these groups, digital media provide opportunities for the empowering of unrecognized groups and causes in society and discourse and thereby are strengthening democracy.

But there is no guarantee that digital media will be used by challengers of intermediary institutions to empower democracy by extending participatory rights or representation. Instead, they can also be used by challengers bent on restricting participatory rights or denying representation to others. Examples for this abound among the uses of digital media by extremist and far right groups in Western democracies. Enabling a challenge of intermediary institutions therefor by itself is neither good nor bad for democracy, instead each challenge needs to be interrogated on its own merits or demerits.

There is another detrimental effect of this challenge, we should discuss. The new opportunities for the coordination of people around causes has raised the expectation of some, that political organizations would lose their general necessity for activists and thereby over time be displaced by temporal cause-based alliances of people that once causes were achieved would dissolve with people free to find new alliances around new

[12]See Jackson et al. (2020).

causes at another day. While at the face of it promising, after all who would not want to do away with the often stuffy and slow structures of political organizations, this stance ignores the importance of political organizations beyond information and coordination functions. Political organizations also provide the link between political activism and political structures, such as governments or parliaments. This link is important to allow for grievances, concerns, and collective energies to constructively shape policy instead of remaining on the street or social media. Various digitally powered – or at least enabled – movements suffered from exactly this missing transmission. Think of Occupy Wallstreet or Fridays for Future. Organizations therefor remain necessary for political activism. But by losing their monopolies on both information and coordination, they may end up weakened in others ways. This overall weakening of political organizations might also end up limiting their ability to function as transmission belts between political activism and political institutions – such as parties, governments, or parliaments. This would weaken democracy.[13]

5.3. Crisis, conflict, and the digital challenge to parties

We see parties being challenged all over Western democracies. The analytical tools discussed above help us in charting and analyzing these challenges. These tools are:

- crisis and conflict in intermediary institutions, and
- emergence of new parties or factions within parties adapted to the new opportunities provided by lowered information and coordination costs.

So, let's first start by looking for signs of what Offe (2006) calls *crisis* and *conflict* within institutions.

5.3.1. Signs of crisis and conflict of parties as institutions

Let's quickly recap. As we have seen, Offe (2006) identifies two reason for the decline and public challenge of institutions: *crisis* and *conflict*. With crisis Offe describes an institution losing step with external conditions it operates within and through this faces decline. This leads to defections and challenges. With conflict Offe describes a loss of moral fit. If the social order, sense of justice, or morality shifts around and institution. That institution might find itself in conflict with society. Again, defections and challenges may result from this development. Now, do these diagnoses apply to political parties as institutions?

In most Western democracies, the discussion about the decline of parties goes back decades. We find public critiques in the popular press, in academic work, but also in the

[13]For a popular argument of *organizing without organization* see Shirky (2008). For a rejoinder best characterized *organizing with different organizations* see Karpf (2012a). For the dangers of political nihilism see Gurri (2018).

declarations of political challengers of the status quo. Diagnoses of decline vary somewhat between different countries and party systems but they tend to share a common set of claims.

Various studies show that on a foundational level parties lose their representative function for broad sections of populations in Western democracies.[14] Party membership declines, while members grow older, and large sections of the population find no – or choose not to seek – access to parties. Similarly, the socio-economic backgrounds of party members do not tend to represent the whole population anymore but only specific subgroups. Specific findings vary between countries, but the central tendency seems to hold. On a fundamental level, this tendency might lead to a loss of representation for specific groups in society who might end up with no spokespersons in parties or parliament anymore. At least, it points to a weakening of the function of parties as information transmitter between publics and political elites in order to make them visible to each other. A function for which J.-W. Müller (2021) has labeled parties intermediary institutions in democracy.

Going further, some have diagnosed for parties in Western democracies a loss touch with their core constituency. As this is seen to be true for all major parties, those established parties stop actually competing for power but share it in a form of cartel, with government power shifting between members of said cartel. This makes public control of peoples' representatives, political parties, and governments untenable and – in the words of Peter Mair – hollows out democracy.[15]

Others have diagnosed rot that reaches even further to the core of democracy. In countries where parties and their representatives depend on large donations to run competitive elections, such as the US, parties are seen as more beholden to their financial benefactors than their constituents. This raises the challenge that parties simply represent the rich donor class and large companies that finance campaigns, instead of the public.[16]

As said before, the details of these diagnoses and their severity vary between countries and party systems. But we encounter variations of them across a wide set of Western democracies. Now, can we connect these diagnoses to the terms *crisis* and *conflict* as understood by Offe? For this, we have to examine potential drivers and consequences of these phenomena.

Parties are location-based membership organizations. They connect members who share given places. Their structure starts with local chapters, moves up to counties, from the on to states, and ends on the national level. Local party members have local, regional, and state representatives for whom they vote and who are, at least in principle, answerable to them. Like all location-based membership organizations, parties thrived in a time when people were comparatively immobile through their live and changed jobs and places

[14]For more on the loss of loss of representation in politics see A. Schäfer and Zürn (2021).
[15]For more on the cartelization of parties see Mair (2013); R. S. Katz and Mair (2018).
[16]For more the loss of central power of party organizations to outside donors see Popkin (2021).

where they lived comparatively seldom. With greater mobility between jobs and places parties, like other location-based membership organizations, suffered. Many people tend to be less connected to the places they live, since they either recently moved there or are on the jump to move somewhere else. Why spend time and money on a location-based membership organization, when there are other opportunities to get rid of both? This might explain the attractiveness of cause- instead of party-based political participation for many young and highly educated people.

In this sense, parties as institutions might indeed face a crisis. The institution party started in the context of given societal structures – rather immobile people who were open to participate in location-based membership organizations – and developed institutional structures that were optimized for this context – local representation and party hierarchies that followed geographical units of greater magnitude – local, regional, state, national. Once these underlying social structures shifted – many people being much more mobile over the course of their lives – parties struggled to adapt and lost access to those people.[17] The often diagnosed loss of parties' representational abilities and functions could thus be driven – at least partially – from an increasing mismatch between institutional structure and societal context.

But we also clearly see signs of conflict. The role of parties as intermediary institutions is to connect publics with political elites and the government. This depends on them – as a set – being able to reach all groups and sections of society and provide them with access points to the political system. Once this is seen to fail, or even objectively is failing, parties lose the moral justification for their role and the associated privileges. Lack of actual, or perceived, representation of sections of the public through parties thereby provides the basis for the challenge of their legitimacy. This then provides the seeds either for the emergence of new challenger parties with greater claims of representing the supposedly unrepresented or is channeled into attempts at political mobilization and participation outside the established party-political system. This is only acerbated once the role of outside money starts to feature strongly in accounts of whose requests parties tend to respond to.

Already this quick and cursory look has shown that parties as institutions face challenges connected with both, their fit to current societal conditions and structures and their moral justification. This provides the basis for the challenge of parties as intermediary institutions in democracies. Now, how do digital media feature in these challenges?

5.3.2. Challenge through information

The sinking costs of information production, publication, and distribution through digital media provides an important tool for the challenge of parties, be it from the outside or inside.

[17] Of course, not everybody is moving around. For an emergent divide between people on the move with little ties to a given place and those deeply rooted see Goodhart (2017).

In their role as information providers and conduits parties interact with news media. Parties provide news media with speakers, topics, and positions for them to cover and feature. In this they are subject to selection decisions of editorial desks or news organizations. Accordingly, parties adjust topics, positions, and speakers to what plays in news media. In party systems with strong centrally controlled parties, this might be a conscious choice by leadership. In systems with weak parties, this is probably best understood as an emergent phenomenon. This provides the media power over which speakers and positions feature often in their coverage and in turn come to dominate the public image of a party or faction.[18]

In the past, this interdependence has been criticized strongly. For one, this collusion between gatekeepers to the public – news media – and powerful institutions of the status quo – political parties – was seen as a feature to restrict access to the general public for challengers to the status quo and alternative view points. Also, the adjustment to selection criteria of the media by politicians and parties was seen as a detriment leading to a dumbing down of political debate, sensationalization, and politics as entertainment.[19]

With the digital extension of communication environments by new channels and new sources, we can see challenges to parties manifest that undercut this mechanism. This can lead to a healthy and more pluralistic extension of the types of information and opinions represented in the public area but also to the challenge of the status quo through extremist or exclusionary actors, which previously would have found themselves excluded from the public arena by mass media exercising their gatekeeping power.[20]

An example for this is the way Donald Trump used his Twitter account to drive media coverage to his primary bid in 2015 for the Republican nomination as Presidential candidate in the 2016 US US presidential election. When reality TV personality Donald Trump announced his candidacy in the Republican primaries on June 16, 2015 it was difficult for commentators and the party establishment to know what to make of this. With Trump, there were 12 primary candidates. Some of whom were heavy weights within the party, with five former governors and four current air former Senators in the race. Accordingly, it was easy to chalk Trump's candidacy off as a publicity stunt to rekindle the embers of his celebrity.

But the size of the field and the number of candidates well connected to the party establishment turned out to support Trump's bid. Since the party establishment could not settle on one candidate to throw their full weight behind, no obvious front runner emerged early on with media coverage rotating from one to the other. The only fixture

[18]For news selection see Shoemaker and Reese (2014). For accounts of the media-institution nexus see Bennett (1983/2016).

[19]For an influential critique of the close connection between established institutions and the media see Bennett (1990). For critiques of the adjustment of political parties and politicians to media see Bennett (2005).

[20]For a discussion of shifts in information environments through digital media see Jungherr et al. (2020), p. 30–68. For adjustments of parties to these shifts see Jungherr et al. (2020), p. 69–102. For how this enables challengers see Jungherr, Schroeder, and Stier (2019).

in the media coverage of the primaries was Trump, who managed to scandalize and enrage the public regularly by tweeting a constant barrage of provocations destined to keep him in the media limelight consistently. This and his consistent challenge of the Republican establishment help him establish consistent coverage and over time support among the Republican primary voters. Trump used Twitter to give the media a story too good to refuse, after all outrage and scandal bring clicks and viewers. His use of digital media allowed him to circumvent the gatekeepers of the Republican Party and mount a challenge to established party creeds and personell. In fact, this challenge turned out to remodel the Republican Party well beyond his candidacy and Presidency.[21]

While Trump's success in this regard is still unique, many fringe groups or politicians within parties use digital media regularly to gain media and public attention and position themselves as de-facto spokespersons for the supposedly disaffected party base. This makes parties harder to govern and contributes to their image in public appearing as fractured and torn between extremes. This loss of control of what constitutes the party line is acerbated by parties also losing the monopoly on making the party visible to itself. By checking their social media or messenger groups, party members are nor longer dependent on the party in order to find out what other members like them are thinking or doing. Again, this makes it harder for the leadership to establish a coherent line on controversial topics and control challenges or insurrections from within.

The same mechanism is also in play when we look not at the challenge of parties from within but from without. As shown above, challenges not only address specific parties but the established party system as a whole. This challenge does not attack one party as having lost its way or left its supporters behind, this challenge extends these claims to all established parties in a given system. In the past, these system challengers would have had to rely on friendly news media to feature their challenge and give it viability for voters. Today, these system challengers can rely on alternative digital news sources to feature them and use digital media as distribution channel to reach interested people and to connect supporters. The informational opportunities provided by digital media allow them to – at least in part – substitute for exclusion or moderation through established gatekeepers. These system level challenges are put forward by movements on both the far left and the far right. But recently, the use of digital media to circumvent institutional gatekeepers to pose a system-wide challenge has featured most prominently with far-right populism.[22]

5.3.3. Challenge through coordination

The sinking costs of coordination provide another mechanism through which digital media enable the challenge of political parties. Digital media and the organizational

[21]For Trump and the digital media news media connection see Boczkowski and Papacharissi (2018); Jungherr, Schroeder, and Stier (2019); Schroeder (2018); Carlson et al. (2021). For the weakening of Republican party structures see Popkin (2021).

[22]For populism and digital media see Schroeder (2019); Schroeder (2021).

opportunities they provide are an important element in the work of party organizations, established and new.[23]

Established parties integrate digital media in their existing structures be it on the national or the local level. These are important developments that receive ample academic attention and that can teach us a lot about the adaptability of parties and innovation within political organizations. But for our purposes it is more important to examine the way that digital media enable new political parties or organizations challenging political parties in their political function.[24]

Already early in the political uses of digital media, scholars pointed to the opportunities of digital media in the creation of alternative forms of political organization – sometimes called cyber-, digital-, or platform-parties. Digital media provide people and movements with tools allowing them to coordinate, elicit feedback, mobilize, organize, and create public visibility allowing new organizations to emerge.[25]

By now there are many cases of new political parties emerging in multi-party systems. Examples include parties on the political left, such as Pirate Parties in multiple European countries, the Spanish Unidas Podemos, or the Italian Movimento 5 Stelle, the 5 Star Movement. More recently, a growing count of new parties and movements on the political right rely on digital media as well, such as Germany's Alternative für Deutschland (AfD). These cases warrant further study. Especially since the uses of digital media vary strongly between different parties: from using existing publicly available digital tools for office work and administrative duties of an organization to the development of dedicated software for internal coordination and voting, as with tools like Liquid Feedback or Rousseau.[26]

In political systems that make it harder for small parties to enter parliament, digital media are used by challenger groups within established parties to form and coordinate. This includes the use of digital media in the US by insurgent campaigns competing for the Presidential nomination of their party, such as Howard Dean, Barack Obama, or Bernie Sanders. But this also includes uses of digital media by factions within parties, trying to shift the balance of power, such as Momentum within the UK Labor party.[27]

The uses of digital media for these organizations vary. They start with support in mundane but important office or organizational work, such as keeping in touch through

[23] For a broader discussion of the uses of digital media in political organizations see Jungherr et al. (2020), p. 158–178. For campaigning see Jungherr (2023b).

[24] For use of digital media by German parties see Jungherr (2016b). For UK see Dommett et al. (2021); Dommett (2020). For US see Kreiss (2012); Kreiss (2016); Stromer-Galley (2014/2019); Pearlman (2012).

[25] For cyber parties see Margetts (2001); Margetts (2006). For digital parties see Gerbaudo (2019).

[26] For the European Pirate Parties see Almqvist (2016); Deseriis (2020). For the Spanish Unidas Podemos see Casero-Ripollés et al. (2016). For the Italian Movimento 5 Stelle see Deseriis (2020); Natale and Ballatore (2014).

[27] For the use of digital media by Howard Dean see Trippi (2004); Hindman (2005). For Obama see Kreiss (2012); Kreiss (2016). For Bernie Sanders see Bond and Exley (2016); Kreiss (2019). For Momentum see Dennis (2020).

email or keeping supporter lists current through publicly available digital tools. Digital media have also come to play a very important role in parties in political campaigns as channels for the collection of political donations. This is especially crucial in the US. Beyond this, digital media are also a very important element in the mobilization and coordination of volunteers. Various U.S. insurgency campaigns from Howard Dean, through Barack Obama, to Bernie Sanders have used digital media actively to generate and channel enthusiasm among supporters into the party organization and translate it into volunteer work, such as telephone or door-to-door canvassing. At the same time, channeling the enthusiasm of activists into a more structured campaign organization can prove to be challenging as activists and volunteers often can pursue politically more pure or extreme goals than the campaigns feels comfortable in pursuing with the general electorate in mind.[28]

What we see here then is that digital media provide challengers to the status quo with tools that serve fundamental coordination functions out of the box. They provide them with a communication hub for their organization, tools for identifying, binding, and coordinating volunteers, collecting and distributing resources, and internal distributed decision making and voting in the form of dedicated software solutions. To achieve either of these functions in the past, challengers would have to go through established political organizations and marshal their resources. This in turn would have forced them to adjust their challenge in ways to make it palatable to the organization. Today this is not necessary anymore. As we have seen above, this mechanism can support challenges to the status quo aimed at making the system more representative and more responsive. But it can also enable challenges that try to achieve the opposite.

We played through this analysis of digital media's role in challenges to institutions with the case of political parties. But other institutions might have provided similarly instructive cases: think of the challenge to news media through digital-born and alternative media, think of the challenge to governments – democratic and authoritarian – through digitally enabled activism, or the digitally enabled open science movement for more transparency and replicability in scientific studies.

Diagnoses of *crises* or *conflicts* of specific institutions might vary, so might the specific digitally enabled challenges to the status quo, and our normative assessment of their justification. But thinking in terms of opportunities provided to challengers through lowered costs of information and coordination, offers a promising lens through which to analyze and discuss these challenges as well.

Go ahead and give it a try! Choose a case along your lines of interest and see how far this framework takes you, or where it stops being useful.

[28] For mundane uses see R. K. Nielsen (2011); B. Epstein and Broxmeyer (2020). For donations see Hindman (2005); Kreiss (2012); Kreiss (2016). For volunteer work and coordination see Han (2014); McKenna and Han (2014); Cogburn and Espinoza-Vasquez (2011); R. K. Nielsen (2012). For the difficulty of keeping volunteers ideologically aligned see Enos and Hersh (2015); Jungherr (2012); Trippi (2004).

5.4. Challenges reexamined

We have now talked extensively about how digital media enable challenges to institutions. But of course this alone does not guarantee their success. Yes, digital media might help identify and document the weaknesses of institutions, their crises and conflicts. They might help challenges to form, being brought forward, and reach a broad public. But success of challenges is of course not determined by these opportunities alone. Instead, success relies on various context conditions and their specific structural embeddedness.

The other question, we repeatedly ran into was how to assess challenges. As we have seen, some challenges might be judged as strengthening democracy and empowering people, while others might achieve the opposite. Now, how can we as researchers address this question? Without of course attributing the challenges we happen to sympathize with strengthening effects and those we dislike detrimental effects.

We will focus on these two questions in the final part of this chapter.

5.4.1. How challenges fail

Discussing the different ways in which digital media enable challenges to established institutions can one leave with the not uncertain feeling that these challenges are bound to succeed and that the days of established institutions are over. But just cast a glance left and right and you will see established institutions apparently doing just fine. How do these observations go together.

While digital media have enabled challengers, they also have provided new opportunities for social control. We see this most clearly with digital challenges to autocracies. One of the strongest tales about the transformative potential of digital media is connected to their role during the so called *Arab Spring*. From December 2010 to 2012, a number of Arab countries were swept by local revolts of citizens demanding for more democracy and greater accountability of their rulers. Digital media featured strongly in these revolts as tools for coordination, tactical support, documentation of the events, and establishment of connections with the international community. The early success of these protests and their public uses of digital media contributed to the public impression of digital media as powerful technologies with the potential to destabilize even strong authoritarian regimes.[29] Subsequent events indicate that both the hopes for sustained change through these revolts as well the diagnosis of the contribution of digital media to their short lived success were overblown.

But not only the West was watching, authoritarian regimes all over the world payed attention to the events and determined not to go down the same route. As a consequence, they started to increase their efforts to control digital communication infrastructures and companies based in their countries. While these attempts at first seemed to Western

[29] For the mutual dance between challenge and control see Tucker et al. (2017).

leaders and commentators futile, subsequent efforts by Chinese and Russian governments showed that state control of digital communication was possible. Especially the Chinese efforts stand out in this, with the Great Firewall cutting off Chinese users from Western digital media and allowing the government to censor content. Going beyond access to information, the Chinese Social Credit System is an even more powerful attempt at monitoring, incentivizing, and punishing behavior of citizens through the state. While its actual workings and effects are still ill documented, it stands as powerful marker of what degrees of digital control are available to authoritarian governments. Under these conditions, digital media provide only limited – if any – opportunities for challenges to the status quo to form through information or coordination.[30]

Let's look at events closer to home. Since the early days of digital media, digital structures have become continuously more centralized. When in the past challengers of the status quo could publish a website on their own server today most people use central hubs – like Facebook, Twitter, or YouTube – to publish or access information. The companies running these platforms therefor have potentially the opportunity to elevate or hide information that they see in conflict with their business. Recently, voices have been raised asking for greater responsibility of companies running these services to identify and censor harmful and incendiary content around politics. The most visible of those was Twitter's decision to suspend the account of then US President Donald Trump after him inciting rioters at the US Capitol on January 6, 2021. While the decision was highly contested at the time, after losing access to Twitter and other social media platforms, Donald Trump never achieved the same power to dominate the media agenda that he wielded during his time on Twitter. While, arguable the discourse on Twitter did not lose much through the forced exit of Donald Trump, this episode shows the power companies wield that run central digital infrastructures.[31] By deciding whom and what topics to allow access to their users, these companies are central hubs with the ability to control public discourse, if they exercise this power or not.

Beyond increased opportunities for central control, challenges to the status quo also have to content with the persistence of structures. Even if challengers are enabled through new opportunities provided by digital media, their challenge still has to engage existing structures.[32] Overturning, or changing, the status quo depends on more than putting out sharply worded tweets or coordinating protests. Structures are deeply embedded in society and many people support the status quo. This can be a relief – if we think of challenges intend on weakening democracy or attacking the rights of others – or this can be depressing – if we think of challenges trying to deepen democracy and increase true representation.

In our expectations for the impact of digitally enabled challenges, we always have to

[30]For China see Creemers (2017); Jiang and Fu (2018); Roberts (2018); Mozur et al. (2022).

[31]For the deplatforming of Donald Trump see@Douek:2021tz; Holmberg (2021). For the effects of deplattforming see Rauchfleisch and Kaiser (2021).

[32]For the limits of digitally enabled challenges see Gurri (2018).

consider the contextual conditions these challenges are facing.[33] Sometimes, the opportunities provided by digital media will shift the balance toward challengers. Sometimes they won't. But in any case, digital media will increase the repertoire of available actions to challengers of the status quo.

5.4.2. How to judge the legitimacy of challenges?

Over the course of this chapter, we have encounter a wide variety of digitally enabled challenges. We have encountered challenges intent on shifting the political agenda in order to address the interests and needs of future generations, such as Fridays for Future. We also have encountered challenges intend on foregrounding the ways societies fail many of their citizens by disregarding their rights and bodily safety, such as Black Lives Matters or Me Too. But we have also encountered challenges intend on limiting the rights of political participation for many groups in society, such as far right extremism or far right populism.

Looking at the diversity of goals digitally media is used to achieve, it is clearly not possible to state that digital media is either exclusively enabling those aiming to strengthen democracy or those bend on weakening it. In the early days, writers like Castells seemed to assume that any challenge of the status quo was a good one and that digital media therefore decisively would weaken hegemonic and patriarchal structures and empower citizens. In light of todays challenges by populist far right actors, this seems premature and maybe even wishful thinking. Instead of cheering on any challenge to the status quo, we need to look closely and be explicit in our normative evaluation in why we assess it either as a positive or negative challenge. For this, let us quickly recap what democracy is all about and based on that develop categories that help us in assessing digitally enabled challenges with regard to their impact on democracy.

In the chapter about artificial intelligence, we already have talked about constitutive features of democracy and how they can be touched by technological change. These features also allow us to assess the impact of challenges to the status quo and come to normative assessments of their justification.

As we already have encountered, the discussion of and literature on democracy is vast and multifaceted. Discussions range from the philosophical foundations, normative ideals, historical expressions, empirical variations, to legal and procedural questions. But three important characteristics of democracy consistently feature in theses discussions:[34]

- Free and fair elections as a process to establish the basis for collectively binding decision making;
- Belief in the ability of people to make decisions for themselves about societal and political questions;

[33] For the need for context-aware analysis see Jungherr et al. (2020).

[34] For foundational discussions of democracy see Calhoun et al. (2022); Dahl (1998); Guttman (2007); Tilly (2007); Landemore (2012); Przeworski (2018).

- Equality of people with regard to representation and rights.

Each of these characteristics allows us to assess challenges to the status quo. First, how does the challenge impact the representation and rights of groups. Is the goal the extension of political representation and rights of marginalized groups, or is the goal restricting representation and rights to groups somehow conceived as the true electorate. While any hard and fast rule is bound to fail, right of the bat, challenges aiming to extend representation and citizen rights have a better case for strengthening democracy than those aiming to restrict representation and citizen rights.

Second, are we dealing with challenges that accept the results of free and fair elections or do they, for whatever reason, reject the binding power of elections. The rejection of free and fair elections as a binding process of collective decision making is a crucial warning sign for anti-democratic tendencies in challenges of the status quo. This being said, if challenges can bring forward specific and credible evidence for how elections in practice are neither free nor fair but do not, in principle, reject election results, this can be seen as supporting and strengthening democracy.

Finally, do challenges attack the ability of people to make decisions for themselves or do they aim to empower them broadly. We currently experience a series of challenges of democratic decision making in favor of expert rule. Often these challenges are based on the implicit or explicit claim that people would be unable to decide on important issues. Either they are not rational enough, lack foresight, or only act in their own narrow self interest. These challenges often suggest some sort of expert rule or to remove selected important topics from democratic decision making. In this, these challenges are inherently weakening democracy, even if they might be well-intentioned on moving society along on selected progressive paths.

Of course, challenges encountered in the wild might not fit these questions neatly. There might be dominant and radical factions deviating in the extend of their challenge. Or challenges might vary in their impact on democracy across these questions. Some might fight for the extension of citizen rights but feel that this is not an issue the public should decide democratically and elections not recognizing an extension of the electorate are accordingly not binding. Or there might be challenges aiming to restrict the rights of selected groups in society but otherwise adhere to the bindingness of elections. As with any framework, these questions have to be assessed carefully and diligently in each examined case and results have to be weighted carefully.

But if in this process one question would have to suffice, then my money would be on the question whether a challenge aims to extend representation and rights or whether it aims to restrict them. After all, democracy is about the self-rule of empowered citizens, without equal rights and representation across groups within a society, this remains an elusive goal. Challenges moving societies along a path toward greater equality of rights and representation therefor have a crucial function in strengthening and improving democracy.

5.5. Understanding the role of digital media in the challenge to institutions

In this chapter, we have discussed the ways digital media enable challenges to institutions. We have also seen that this enabling does not guarantee success. It is also important to note that digital media enable both challenges that arguably can be seen as aiming to strengthen democracy as well as those aimed at weakening it.

These observations point to a set of important take aways for the scientific work on digital media. First, we need to improve our understanding of how digital media are actually used in support of challenges of the status quo. This includes work focusing on the uses of digital media in organizations and associated adaption process. But this also includes work that focuses on digital artifacts of challenges, such as content posted on social media. Especially the second type of work invites contributions using methods from computational social science.[35]

Going further, the varying fates of digitally enabled challenges to the status quo also show that we need to increase our understanding of the contextual conditions for successful challenges. Some of those might be connected to digital media directly, others might not. Still, getting a better grasp of the necessary conditions for successful challenges of the status quo will also allow us to get a better understanding of the contribution of digital media to these successes.

This future work will also have to extend our analyses to include long-term fates of challenges. For example, for a few months it seemed like the events of the Arab Spring were successful in sustainably changing the balance of power in the countries concerned. A few years later, though, these changes proved to be short lived and often illusory. Work on the political effects of digital media can sometimes appear rather short of breath, with a strong focus on the short term and select cases of apparently successful uses. These studies can be highly instructive but often fail to fully account for the long-term effects of the use of digital media described. From these studies then it is easy to get the impression of the power of digital media while a longer term view might indicate their limits.

Given the multiplicity of challenges to the status quo, we currently experience, we need to deploy a clear normative framework in their discussion. Some challenges to the status quo are justified and their success might arguably end up strengthening democracy or even the very institution they challenge. Others, might end up weakening institutions and democracy. In discussing these challenges, our normative assessments must follow clear criteria in order for us not to end up celebrating challenges we happen to support and critiquing those we do not. In this chapter, I introduced one such framework.

[35] For research opportunities for computational methods to map action repertoire of challengers see Teo and Fu (2021).

The intermediary institutions democracies rely on are far from perfect. For years, the social sciences have pointed out the various ways these institutions fall short, fail to represent all people, or perpetuate injustices. To find these institutions challenged comes therefor as no surprise. Clearly, digital media play a role in enabling the current wave of challenges. Especially, the current prevalence of digitally enabled reactionary challenges from the extreme or populist right makes it tempting to write of digital media as a tool for political expression or coordination. But this would be a mistake. While some digitally enabled challenges are aimed at rolling back citizen representation and rights, so are many challenges that try to increase both. So instead of damning digital media in support of the first, or celebrating it for their support of the second type of challenge, we need to develop a better understanding of the underlying mechanisms and subsequent conditions for success. Only this will ensure that we are able to use digital media to strengthen democracies and better societies instead of falling prey to those using them to weaken both. Here, there clearly remains much to do for social science.

5.6. Review questions

1. Please define the term *institution* following Offe (2006).
2. Please define the term *crisis* in the context of institutions following Offe (2006).
3. Please define the term *conflict* in the context of institutions following Offe (2006).
4. Please discuss the way that political parties are intermediary institutions following J.-W. Müller (2021).
5. Please discuss the way that news media are intermediary institutions following J.-W. Müller (2021).
6. Please discuss how digital media lower information costs.
7. Please discuss how digital media lower coordination costs.
8. Please discuss the way how information costs impact the challenge to institutions.
9. Please discuss the way how coordination costs impact the challenge to institutions.
10. Please discuss the challenge of Fridays for Future along the normative questions suggested in this chapter.

6. Public arena

Democratic societies need spaces in which people and political elites become visible to each other, develop shared agendas, and settle on collectively binding decisions. These spaces need to be open to people from all walks of life and from all groups in society. They need to feature the voices of the privileged as well as those of the marginalized. They need to provide people with the information they need for self-governance and enable them to control elites. They need to provide elites with the information they need to govern and represent the people. And although these spaces will fall short of these needs, as long as they are transparent of their workings and allow for critique and subsequent improvement, they can be made to work for people and their pursuit of the public good. These spaces are the public arena.[1]

The public arena is a space of structured tensions. Different people from different groups with different interest encounter each other, compete for attention, and try to shape politics and society. The public arena is a space in which political elites perform their competition for attention and power and which they use to learn about the people and their concerns. These encounters and competitions are noisy and at times come to violate norms and established practices.

Tensions within the public arena come to the fore especially in times of structural shifts within the institutions and organizations hosting the public arena. We are currently witnessing such a shift driven by the digital transformation. Digital technology is deeply transforming and challenging institutions that formerly held a near monopoly for hosting the public arena, the news media. Digital media weaken the economic foundation of news, they transform modes of information delivery and consumption, and they allow for the emergence of new information providers who do not necessarily share the commitment to the same institutional norms and practices of news organizations of the past. At the same time, we see new types of structures emerge that become as important to hosting the public arena as news media were in the past, digital platforms like Facebook, Google, TikTok, or Twitter. Here, we need to understand their role as structures of the public arena and develop norms and rules for their contribution taking into account their differences from former structures of the public arena.

The digital transformation of the public arena is one of the most important challenges democratic societies face today. Associated opportunities and hopes, but also dangers and fears, feature prominently in public discussions. In this chapter, we discuss the public arena, its democratic functions, and challenges introduced by digital media. This

[1]For a discussion of the contemporary public arena and its tensions see Jungherr and Schroeder (2022).

discussion is just getting started, so be prepared to leave with more questions than answers.

6.1. The public arena

Societies need spaces for groups to make themselves visible to each other, to settle on the most important problems of the day, exchange different alternative approaches to solutions, and settle on collectively binding decisions. These spaces are the public arena.

In our book "Digital Transformations of the Public Arena" the sociologist Ralph Schroeder and I define the public arena through the following three characteristics:

> (1) The public arena consists of the media infrastructures that enable and constrain the publication, distribution, reception, and contestation of information that allow people to exercise their rights and duties as citizens.
> (2) This excludes how people use these infrastructures for private life or for commercial purposes except when these uses come to bear on people's rights and duties as citizens.
> (3) These infrastructures mediate the relation between citizens or civil society on the one hand and political elites or the state on the other.
>
> Jungherr and Schroeder (2022), p. 3.

This definition points to four important characteristics of the public arena. First, the public arena consists of structures that make people visible to each other, document and make visible current events, allow for a public negotiation of meaning, and exchange of alternative interventions. In modern societies, these are predominantly media infrastructures. These media infrastructures consist of institutions – such as the news media – and technologies – such as print, radio, television, or the internet. Shifts within available media technologies impact media institutions. One important impact is the shift from broadcast technologies – such as print, radio, and television – to digital media technologies. This shift has impacted deeply the economic and moral foundations of established news media institutions. The consequences of this impact are still not settled and contribute significantly to the current state of public insecurity and fear about the impact of digital media on democracies.[2]

Second, media structures are not neutral but instead come with specific features that enable or constrain – or at least incentivize or disincentivize – specific activities and behaviors. Different structures – as in institutions and technologies – of the public arena

[2]For the link between technological change and media institutions see Bimber (2003); J.-W. Müller (2021). For the ongoing shift within news as an institution see B. A. Williams and Carpini (2011). For a practitioner's perspective on recent changes within news see Rusbridger (2018).

will therefor enable or constrain different patterns in publication, reception, and contestation of information. A public arena relying on mass media and broadcast technology will feature a limit set of few powerful gatekeepers that decide about what actors and topics to allow access to the public area. In contrast, a public arena relying on digital media with widely distributed access points will feature less control of gatekeepers about which actors, voices, and topics gain access. In digital communication environments, access is not the limiting factor. Instead, it is attention. Power in public arenas depending on digital media therefor lies with actors who can amplify selected actors, voices, and topics within the public arena and provide collective attention for them. Other than in the past, power is no longer about merely providing access to the public arena. Instead, it is about having others pay attention.[3]

Third, the structures of the public arena always feature more information and activity than that directly connected with the pursuit of the public good. This was true in the past and stays true today. But associated consequences need to be kept in mind. For one, the public arena is hosted on structures provided by actors with commercial interests. This was true for radio stations, television stations, and newspapers and is also true for digital structures, such as platforms like Facebook, Google, and Twitter. Societies need to figure out how to align the interests of these commercial actors hosting structures with the functions these structures hold for society. Associated tensions cannot be resolved but need to be surfaced and publicly negotiated. Also, usage practices of structures used for recreational, entertainment, and commercial uses will influence the uses of said structures for public purposes, such as the discussion of politics. In the past this was critiqued in the context of a perceived commercialization of news. Today, we see this with practices coming from fan cultures in digital communication environments to start shaping patterns in the discussion of political or societal issues and controversies. Not always for the better.[4]

Fourth, the public arena mediates the relationship between citizens and political elites. It makes people visible to each other. This provides the opportunity for mutual recognition or conflict within the bounds of ordered political competition. Beyond this, it also allows for the formation of new groups of people with shared interests and the construction of new shared identities, providing the potential for new lines of political conflict to emerge.[5]

The public arena also makes people visible to elites and elites visible to people. The structures hosting the public arena are therefor crucial elements in democratic representation. Their internal workings shape who and what becomes visible to elites and therefor provide the mediated reflection of society elites react to and that shapes their perceived

[3]For the concept of power and influence an public arenas relying on digital media see Jungherr, Posegga, and An (2019). For the concept of amplification see W. Phillips (2018). For limited attention in the public arena see Schroeder (2018).

[4]On the multiple interconnected uses of structures of the public arena see Schroeder (2018). On fan cultures shaping digital communication environments overall see W. Phillips (2015); Tiffany (2022). For specific learnings of the Trump campaign from fan cultures see Green (2017).

[5]For the formation of new groups and identities see Bourdieu (1990).

option space. A public arena constituted by structures that foreground groups in society who already are privileged, will incentivize political elites to react to their interests more strongly than a public arena that features many competing voices, some traditionally privileged, some freshly formed from traditionally marginalized groups. The public arena and its structures therefor matter a great deal with regard to who gets seen and represented in society and the opportunities provided for groups with shared interests to find each other and articulate shared interests and demand representation.

The structure, and structural shifts, within the public arena are important. They shape political discourses, public beliefs, conditions of political competition, the representation of social groups, as well the option space for collective action within a society. This makes the structural conditions of information environments, their transformations, and consequences into important objects of study for sociologists, communication scholars, and political scientists. The increasing importance of digital communication environments and the associated increase in digital data traces of contributions to and interactions within the public arena makes it also an important topic in computational social science. Conversely, computational social science offers interesting new perspectives to larger theoretical discussions about the public arena, its structures, and dynamics.[6]

6.2. The public arena and its functions for democracy

The public arena and the structures hosting it are a crucial element in democracies. They provide the basis for people to inform themselves about politics and society, to meaningfully engage in discourse, and ultimately exercise their rights to self-rule. Accordingly, the structures of the public arena are routinely interrogated with regard to their enabling or detrimentally effecting democratic functions. Not surprisingly, there is no shortage of – sometimes conflicting – normative prescriptions for how the way structures of the public arena should function. Of those, a recent account by Jan-Werner Müller fits our discussion:

> “They should be widely accessible; access should not turn into a privilege for those already advantaged. They should be accurate; that is to say, political judgments and opinions (…) must be constrained by facts, even if (…) facts are always fragile. They should also be autonomous - that is to say, not depend on more or less hidden actors in a corrupt way. They must be assessable by citizens. And, as a result of all of the above, they can be accountable.”
>
> J.-W. Müller (2021), p. 139–140.

[6] On the contemporary digitally mediated public arena see Jungherr and Schroeder (2022). On important historical accounts and different concepts of the public see Dewey (1927); Lippmann (1927); Habermas (1962/1990). For overviews see C. Taylor (1993); C. Taylor (1995); Hardt (2001); Ferree et al. (2002a); Christians et al. (2009); Zammito (2012); Rauchfleisch (2017); Wessler (2018).

While Müller talks primarily about parties and the media, we can extend his prescriptions to the structures of the public arena more broadly. Paraphrasing Müller, for the public arena to function, the structures hosting it need to provide access to people irrespective of their societal position or status. Information hosted should be accurate, in other words bounded by facts. This being said, especially in politics facts and their meaning are subject to public contestation and a collective negotiation of meaning. This boundedness can therefor not be established purely through narrow fact checking. Instead it demands for broad commitment of political and societal elites and factions. Also structures of the public arena need to be independent of existing powerful actors or interests in society, be it financially or structurally. Finally, these structures need to be transparent in order for people and regulators to be able to critically interrogate them regarding their inner workings, dependencies, and their impact on the democratic functions of the public arena.

Is with any normative prescription, the one proposed by Müller needs interpretation and qualification if applied to the assessment of specific structures in the public arena. While single structures might fail in specific instances - for example digital platforms being primarily used by people with easy access to digital devices and not by others or some partisan media being closely aligned with political factions - as long as the set of structures provides broad access and features different voices and viewpoints, one might feel not too troubled. But if the structures of the public arena as a set would systematically exclude people or legitimate (as in bounded by facts) opinions then worry about the democratic contribution of the public arena is warranted.

Before we go further and discuss specific structures hosting the contemporary public arena, let us quickly examine the functions the public arena serves for democracy.[7] Here, we will focus on three contributions:

- Visibility and representation;
- Group formation; and
- Supporting collective problem solving and collectively binding decision making.

6.2.1. Visibility and representation

A prerequisite for meaningful self-rule is visibility. Visibility of people to each other and to elites. Visibility of elites to people. And visibility of events and conditions of importance to the public. Structures in the public arena provide this visibility to different degrees and through different mechanisms.

News media, the structures hosting the public arena before digital media, primarily produce and distribute information about the state of society. Trained journalists go out and report the news, pursue deeper investigations into selected topics, or comment on events. In this coverage, people and their voices feature and become visible to news audiences

[7] For an alternative discussion for the functions of the public arena see Rauchfleisch and Kovic (2016).

and elites. News media also provide elites with a platform to reach people, for example through interviews or guest contributions. This coverage, its selection choices, and representation of different people and voices is subject to interrogation and critique.[8]

Other structures - such as digital platforms like Facebook, Twitter, or YouTube - do not produce information themselves but provide actors with the opportunity to publish information and reach people. This could be people, who post their opinions and reactions to current events or report on events they witness on their public social media profiles. This could also be politicians or parties who publish information on their profiles. Or this could be traditional news organizations and journalists that use social media platforms to increase the reach of their coverage.

While the role of news media in the public arena has been well established and the lines of interrogation and critique are well established as well, those for digital platforms are currently established and negotiated.[9] By now it is clear that platforms need to accept responsibility for the content they provide access to and that they distribute. The exact rules by which this is about to happen are still contested, though. Currently, there needs to be a balance established between users' speech rights and users' protection from harm through false or misleading information or harassement. Currently, this is a topic of great academic and industry activity. This is true for both the crafting and assessing of governance rules and policy, as well as the empirical identification and measurement of harmful content and its effects.

6.2.2. Group formation

An important corollary to visibility is the role of structures of the public arena in providing the basis for the formation of politically aligned groups and their representation. As we have discussed in the previous chapter, a crucial feature within politics is the formation of politically meaningful groups:

> "The power of imposing a vision of divisions, that is, the power of making visible and explicit social divisions that are implicit, (…) [bringing] into existence in an instituted, constituted form (…), what existed up until then only as (…) a collection of multiple persons, a purely additive series of merely juxtaposed individuals."
>
> Bourdieu (1990), p. 138.

Jan-Werner Müller goes further in describing this process:

[8] For a critique of news media and their democratic contributions see Keane (2013). For selection processes of news media see Shoemaker and Reese (2014).

[9] For digital structures of the public arena see Jungherr and Schroeder (2022).

> "Here representation is not conceived as substantively or descriptively reproducing something that already exists. It is not a matter of mechanical reproduction. Rather, it is a process in which individuals offer to a possible constituency an image of themselves based on so far unrecognized ideas, interests, or aspects of their identities. As a result, citizens might perceive themselves and the politics they need in a novel light. A constituency is not so much reproduced, or even revealed, as talked into existence and, as a result, uses its political freedoms in novel ways."
>
> J.-W. Müller (2021), p. 79–80.

Structures hosting the public arena, provide spaces that allow for – or hinder – these processes of collective group formation. Processes corresponding with these expectations could be observed with #MeToo, #BlackLivesMatter, or #FridaysForFuture. These groups formed around grievances made public on digital media. People posted and bonded about experiences, formed collective identities, and coordinated political protest. The variety of causes concerned and international occurrence shows that this is neither a feature specific to select causes or locals, instead it appears to be a crucial function of digital media in the formation of new political groups and mobilization. The associated dynamics and effects merit much further attention.[10]

6.2.3. Problem solving

A further democratic function of the public arena is its support in the formulation and solving of problems relevant to the public good. This features very prominently in the work of sociologist Jürgen Habermas. In his conception of *Öffentlichkeit* – or public sphere in the English translation – the structures allowing people to meet and discuss politics can be assessed by the degree to which they allow for broad access, rational exchange of facts and alternative solutions, and disinterested evaluation of options with the goal of reaching best outcomes. Not surprisingly, empirically structures of the public arena tend to fall short of these ideal characteristics. Also not surprisingly, the normative goals presented by Habermas have been strongly contested. Most importantly, his account of rational problem solving through communication has been contested by approaches that see competition and conflict as a more fitting account of exchanges in the public arena.[11]

But moving away from the by now mainly academic question of whether structures of the public arena allow for the communicative ideals postulated by Habermas, this perspective offers many important insights. For example, many academics are working on the development of specific structures – democratic innovations – that try to enable

[10]For the uses of digital structures for these purposes see Gurri (2018); Jackson et al. (2020).

[11]For Habermas' account see Habermas (1962/1990); Habermas (1981). For an overview and development of his argument see C. Taylor (1993); Wessler (2018). For an influential critique foregrounding the role of conflict see Fraser (1990).

people to meet, exchange viewpoints, and find solutions to common problems. Here, the constitution and governance of mediating structures matter. Also, a subfield of democratic theory, epistemic democracy, focuses on the forms and conditions under which people can democratically define problems and contribute to solutions that are within the public interest and foster the public good. The structures of the public arena, old and new, can therefor also be interrogated with regard to their contribution to the formulation and solution of societal problems.[12]

6.3. News media as structures of the public arena

From the eighteenth century onward, news media have been crucial structures hosting the public arena in Western societies. They provided information, made elites visible to publics, and publics visible to elites and to each other. In this, they never functioned without fault or were completely free from power structures in society. Their role, beneficial and detrimental, in hosting the public arena and serving democracy over time and in different countries has been well established. But for our purposes here, we will focus on three features of news media that matter strongly for the contemporary public arena:

- The role of news media as an institution;
- The economic foundations of news; and
- The emergence of alternative news media in the public arena.

6.3.1. News media as institutions

Looking at news media from an institutional perspectives reveals their functions and inherent tensions in their contribution to the contemporary public arena. As we we have discussed in the section on institutions, journalism can be understood as an institution. Journalism can be understood as a binding set of rules for organizations and people belonging to the institution. These rules can relate, for example, to the selection of news items, the representation of political factions in a society or the quality assurance of coverage. Violations of these rules are punished either within newsrooms and media houses, by professional associations or in public. Violations of rules can lead to the exclusion from newsrooms for individual journalists or to punishment by professional associations or the public for media organizations. In extreme cases, particularly intense violations of the rules and regulations can lead to a general population-wide loss of trust, which can also be associated with a general loss of legitimacy for the institution *journalism*.[13]

[12] For a contemporary account of the role of structures in enabling or hindering deliberative problem solving in democracies see Lafont (2020). For epistemic democracy see Landemore (2012); Schwartzberg (2015).

[13] For journalism as an institution, see Kiefer (2010).

Traditional journalism in the Anglo-American mold is characterized by a shared set of institutional norms among media organizations and individual contributors. This includes commitment to the impartial coverage of politics and societal conditions, a clear division between coverage and commentary, and reliably policed quality standards. Importantly, this does not mean that news media had to remain neutral and impartial on issues or accept politicians at their word. It simply means that there is a commitment to impartial coverage of events and facts while comment and opinion remain in clearly marked sections. These principles set normative goals for what journalism is supposed to achieve. They are institutionally maintained and transmitted within organizations, education programs, and professional associations.[14]

While specific articles, journalists, or outlets might fall short of these norms and ideals, these publicly communicated norms formed the basis for public debate, critique, and contestation of news items or news media. This made clear what people could expect from media infrastructures and enabled challenging specific outlets or individuals if they were seen to be deviating from these norms.

Challenges of news media and their contribution to democracy abound. These challenges run the gauntlet from succumbing to economic pressures, dumbing down political coverage, sensationalism, or excluding marginalized voices and groups. Looking closely, at these critiques we would find various signs of what Offe (2006) has characterized as *crisis* and *conflict* of institutions. Later, we will discuss signs of crisis, a mismatch of foundational principles and conditions of the institution news media with current economic conditions in the contemporary public arena. But first, let's focus on conflict, a mismatch between established institutional values of neutrality and impartiality and a growing sense that news organizations and journalists need to take clear position in societal conflicts.[15]

The norms of impartiality and separation of neutral coverage and opinion are not necessarily shared by new entries in the digitally expanded public arena. Instead of one type of news organization, sharing the same set of norms and aiming to more or less appeal to the same mass audience, we find a plentitude of different news organizations with different normative goals and target audiences. There are news organizations explicitly aligned with political factions and dedicated to providing partisan coverage, news organizations funded by nonprofits or philanthropists supporting specific societal goals, news organizations that exclusively follow a for-profit logic with no further consideration of their democratic or societal impact, and news organizations run by volunteers dedicated to shared goals or issues. These diverse organizations and their staff follow different economic incentives and societal missions. While many important news organizations still follow the norms and practices associated with the institution *journalism*, many new organizations do not and even would negate the validity of one binding approach.

[14] On traditional norms of journalism see McQuail (2013); Kovach and Rosenstiel (2001/2021).
[15] For challenges see Keane (2013); Bennett (1983/2016); Usher (2021).

Instead of one *journalism* with globally adhered to norms and practices, we have many *journalisms* competing in the public arena for attention.[16]

One such rift within journalism as an institution is the growing conflict about the normative bindingness of neutrality or impartiality in the coverage of events. Multiple societal and political crises and norm violations have contributed to a growing call for journalism to take sides in political and societal conflicts. This is especially true among young journalists. Drivers are experiences during the presidency of Donald Trump in the US and the growing sense of global dangers following unmitigated climate change. Instead of an objective and neutral journalism, there are calls for a journalism that takes a clear position on social, moral, and political issues, a journalism based on *moral clarity*.[17] Proponents of this position use digital media to advertise this normative reorientation and to illustrate its broad societal acceptance.

While the calls for greater advocacy within journalism is most prominent among contributors to new digital born outlets, it is increasingly also carried into traditional news organizations that, institutionally, would feel more aligned with the traditional norms of impartiality or neutrality. As legacy news organizations adapt to the new business of news in digital environments, they start to learn from digital-born competitors. And as they start to hire from the pool of contributors to digital born outlets, conflicting views of the role of journalism have come to feature within these traditional organizations. This normative shift away from news organizations as structures committed to the goal, however imperfect, of impartiality in political and societal competition and toward open advocacy raises questions concerning their role in hosting, enabling, and adjudicating discourse and political competition in the public arena.

The institution journalism finds itself challenged on multiple fronts in the contemporary public arena. For one, the multiplicity of different organizations providing information in the public arena makes it difficult to develop, maintain, and police a common set of shared norms, rules, and procedures that would form an institutional basis. Also, established norms from the past, such as the primacy of impartial and neutral coverage, are also contested by new entries among news organizations as well as young practitioners who feel that traditional institutional norms do not conform with the political and societal conditions they find themselves in or their own professional goals and aspirations. These shifts raise the question in how far and under which conditions news organizations can be understood primarily as hosts of the public arena and not as participants or competitors. In other words, do news media still function as what J.-W. Müller (2021) has called *intermediary institutions* for democracy or are they better understood as competitors in social or political conflict? And what implications does this have for their fulfillment of the democratic functions supporting self-rule discussed above, such as visibility and representation, group formation, and collective problem solving?

[16] For value shifts within news see Agarwal and Barthel (2015); Eldridge (2018); M. Scott et al. (2019).
[17] For *moral clarity* in journalism, see Gessen (2020); Lowery (2020); Wiedeman (2020).

6.3.2. Economic foundations of the news

One important feature of the contemporary public arena is the weakening of the economic foundations of news production and distribution. Traditionally, commercial news organizations worked as two-sided markets. News organizations sold bundled information to audiences while at the same time selling access to these audiences to ad customers. This was a highly profitable business.[18]

News organizations very actively constructed their audiences and thereby strengthened their value to ad customers by allowing for the targeting of ads. For example, news organizations aiming for undifferentiated mass audiences provided a set of broadly relevant information, catered to mass tastes, and used a broadly accessible and compatible style. Examples include tabloid newspapers, such as the German *Bild*. News organizations like this were of interest to ad customers who wanted to reach large audiences and the broad public.

In contrast, other news organizations construct specific subsets of people by focusing on specific information that is relevant to specific sections of the public. Examples include trade publications, newspapers like the *Economist* or *The Financial Times* targeting audiences with comparatively high income and education, or newspapers like *The Guardian* targeting a politically and socially aligned audience. By providing ad customers with access to thus constructed and known slices of the population, specialized news organizations allowed ad customers to focus their ad-based appeals and ad-money on audience segments of interest to them. While in light of today's fine-grained digital targeting opportunities, these approaches to targeting might seem quaint, they provided the economic basis that made news such a lucrative business up until the nineteen-nineties.

One of the driving factors behind the success of digital media companies is the promise of ad display in digital communication environments. The empires of companies like Alphabet/Google and Meta/Facebook are built to a large extend on their ability to sell ads to customers with the promise of providing them with just the right kind of audience at an comparatively cheap prize. Digital media build on the promise of customized audiences offered by news organization but do so at a much higher granularity. While news organizations in the past could offer coarse targeting, digital media companies promise access to much more precisely tailored audiences. By using information they have about their users, companies like Alphabet/Google and Meta/Facebook can offer ad customers access to audiences who share specific demographic characteristics, live in specific geographic areas, or share specific interest.[19]

This has clear advantages for ad customers. By restricting the audience for an ad to specific audiences, people can spend their ad budget with much greater focus. This brings advantages not only to business. Instead, many societal and political actors can

[18]For news as a business see R. K. Nielsen (2020). On the economics of two-sided markets see Rochet and Tirole (2003); Rochet and Tirole (2006).

[19]For the competition between news organizations and digital media see Auletta (2018). For a numbers driven account on the development of advertising dollars in the US see B. Evans (2020).

now increase their reach to people through targeted ads who were not able to do so in the past. Think of your local blood bank reaching out to people in your area through Google or Facebook ads to alert you about an upcoming donation drive. Or think of your local party alerting people to an important council meeting in which an important local policy issue is discussed. By giving actors with small budgets access to people, digital ads can contribute to a strengthening of local political activity.

Yet, in public debate negative visions of digital targeting dominate. Some fear that allowing political actors to target ads too precisely would allow them to run dark campaigns that remain invisible to outside observers. The associated fears are that in these dark campaigns selected publics are promised different and potentially conflicted outcomes. Or that politicians use dark campaigns to attack minorities or other political factions while appearing conciliatory in the main campaign visible to all. Going further, others believe that targeting would allow the identification of psychological traits making people more susceptible to appeals specifically designed for their psychological weaknesses. Campaigning would thus shift toward psychological manipulation instead of argument driven persuasion.[20]

While these dangers are forcefully argued, there are good reasons to expect the promises of digital targeting to be exaggerated by digital media and digital ad consultants, concerning its reach as well as strength of its effects. Still, the promise alone has led to the development of large data collection and aggregation efforts in the background of digital communication environments. While the expectations of broad surveillance economies might be overblown, there is good reason for regulators to restrict these ill-regulated and intransparent efforts.[21]

Overall, the shift of ad dollars to digital platforms and the shift in the delivery mode of information from physical to digital media has weakened the economic position of news organizations considerably. The associated challenges put every news organization to the test. But while large international news brands like *The New York Times* or those with a strong identity and loyal readership like *The Economist*, *The Financial Times*, or *The Guardian* can weather these challenges and potentially even emerge strengthened, other organizations will struggle or go under. This clearly impacts the quality and breadth of the public area.[22]

Structurally, the public arena will be fine if some national news organizations go out of business. After all, other organizations will pick up the slack and cover national news. But on the local and regional level, the weakened economic basis of news is more

[20] For a classic discussion of negative aspects of targeted communication see Bennett and Manheim (2006).

[21] For an account doubting the promises associated with digital ads see Hwang (2020). For the likely hollow promises of psychological targeting see Hersh (2018). For a helpful critical overview of issues regulating digital information industries see Kapczynski (2020).

[22] For an insiders' accounts of these transitions see Rusbridger (2018). For the current state of digital news consumption in international comparison see the annual edition of the *Reuters Institute Digital News Report* Newman et al. (2022).

troubling. While on the national level, there might be a set of competing and more or less comparable news sources available, on the local or regional level, there probably is only a limited set of sources available. The economic basis for producing and distributing news in these contexts is fragile anyway and through the digital transformation might break the few remaining sources available, thereby negatively impacting the basis for local or regional democracy. In the US this has been discussed under the term *news deserts*.[23]

Overall, digital media have introduced shifts to the economics of traditional news organizations. This raises questions how societies can ensure the reliable and continuous production and distribution of information relevant to the pursuit of the public good and self-rule in democracies irrespective of its commercial viability. These developments reinforce the importance of models allowing for the public funding of news media, such as public service media in Germany.[24]

6.3.3. Alternative news media in the public arena

Digital media have also impacted the position and role of news media as structures of the public arena. By lowering the costs of information publication and distribution, digital media have enabled new actors to provide information in the public arena. The motives and business models of these new entries vary, but all are challenging the former monopoly of traditional news organizations.

The emergence of these differently motivated sources providing alternatives to the coverage of traditional news organizations fundamentally weakens the powers of traditional news organizations to act as gatekeepers to the public arena - deciding what actors, voices, or agendas to allow access to the public arena and to compete for collective attention. This can have positive as well as negative consequences. On the one hand, more and alternative news providers can provide more and more diverse access points to the public arena. This can give voice to marginalized groups and agendas in society. This could improve the representation of societies' different groups and agendas within the public arena. On the other hand, some voices and groups are marginalized for a reason. Discriminatory, hateful, or incendiary voices are excluded from the public arena for a reason. But new and alternative news providers offer these voices access the public arena as well. This contributes to a deterioration of discourse within the public arena. Be it by actively attacking people supporting other political factions or societal groups, by knowing invention or careless distribution of misleading or downright false information, or by actively attacking the participatory rights of others these actors weaken democracy and contribute to political competition in the public arena turning hostile and antagonistic, potentially leading to the withdrawal of others from political information and

[23] On news deserts see Abernathy (2020).
[24] On the challenges of funding news media in democracies see Pickard (2020).

exchange.[25]

New providers of news in the public arena come in different shapes and sizes. There are new entries that resemble traditional news organizations and share their commitment to institutional norms and practices, examples for these digital born organizations are *Politico* internationally or *The Pioneer* in Germany. These organizations pursue news as a business and hope to provide alternatives to established news organizations by running a smaller and more nimble and independent organization or through providing more in depth coverage of niche topics of little interest to mass audiences but that nevertheless have an audience willing to pay a premium for information that otherwise might remain unavailable.[26]

In authoritarian regimes and transitional democracies, digital media have also led to the emergence of independent news organizations critical of the regime and its willing executors in news organizations aligned with the regime. These organizations fulfill a crucial function for the public arena in their society in providing critical information on the workings of regimes and pursuing independent investigations into important but neglected issues. Unfortunately, this service is not always rewarded by the public and often editors and journalists contribute to these organization under considerable professional and personal risks. An example for an organization like this is the *Rappler* from the Philippines.

Other organizations might be funded by philanthropic foundations or individuals. Usually, these are pursuing societal goals. Be it the coverage and investigation of otherwise neglected topics, like the US organization *ProPublica*, or the explicit support of specific factions or societal groups. These organizations can contribute positively to the public arena by extending the scope of covered voices or agendas. But there is also the risk of these organizations becoming uncritical executors of the interests and goals of their funders. In these cases, they are less structures of the public arena but more active participants in the competition for attention and power. And of course it takes a truly courageous person to put their fate in the hands of benevolent billionaires.[27]

But of course, there are also openly partisan news sources in digital communication environments. Countries with primarily commercially funded news are no strangers to openly partisan news organizations. The most prominent example for openly partisan news organizations can probably be found in the US and their openly partisan news channels, like *Fox News* or *MSNBC*. But the comparatively open digital communication environments provide even more options for the establishment of openly partisan news organizations. Again, the US offer the most striking examples for this with news sources like *Breitbart News* or *Jacobin*. Again, sources like these can probably be understood

[25] On the role of news media as gatekeepers to the public arena see Shoemaker and Vos (2009). For an account of how digital media strengthen challenges to the status quo in politics and society see Jungherr, Schroeder, and Stier (2019).

[26] On different motives, models, and challenges of digital born news organizations see Nicholls et al. (2016).

[27] On journalism driven by foundations see K. Wright et al. (2019).

more as participants of political competition in the public arena, than as neutral structures. These media are not hosts for the public arena, to make society and political factions visible to each other, and to allow the exchange of views and negotiation of the public agenda. Instead they are actors actively pitching for one side or the other.

A striking example for the instrumentalization of partisan news organizations in campaigns can be found in the role of Steve Bannon first as executive chairman of the far right news site *Breitbart News* and later chief executive of Donald Trump's 2016 run for the US presidency. The site was instrumental in pushing stories in open support of Trump and his agenda and contributed to shifting the coverage of other news outlets like *Fox News* and even mainstream coverage during the campaign to converge on Trump's agenda. Cases like these illustrate the danger that openly partisan news outlets can pose for the public arena by imitating the style of traditional news while being only interested in pushing the agenda of selected political factions. In the extreme, this can also escalate to the fabrication and distribution of false or misleading information, further contributing to a deterioration of the public arena and legitimate political competition.[28]

The broad extension of sources in digital communication environments challenges the role of news organizations as structures for the public arena. For one, there simply is more information available – reliable or not – this weakens central control over access to and content in the public arena – for good or bad. But the increase of news sources might also further weaken journalism as an institution as it will become increasingly difficult for all news organizations and their contributors to agree on a binding set of shared norms and practices. The contemporary public arena is therefore more diverse and noisy than its past manifestations. This will take some getting used to for journalists, political elites, and the public.

6.4. Digital structures of the public arena

Digital media have not only challenged the position of traditional structures of the public arena. Digital media have also led to the emergence of new structures hosting the contemporary public arena in digital communication environments.[29]

This includes sites like Facebook, Instagram, TikTok, Twitter, and YouTube that allow people and competitors within the public arena to publish information and to reach large audiences. But this also includes companies that provide the means for people to run their own sites contributing information and commentary to the public arena. This includes services that allow the comparatively cheap hosting of sites or Apps – like Amazon Web Services – allowing people to run their own sites. Or services that provide them with the opportunity to monetize information or services – like PayPal or Patreon. Also services facilitating the hosting of digital ads contribute to the digital extension

[28]For more on Breitbart News and Steve Bannon see Green (2017).
[29]For a discussion of digital media and functions of the public arena see Rauchfleisch and Kovic (2016).

of the public arena. Digital ads contribute to the monetization of sites by allowing the owner to host ads and get payed for impressions and clickthroughs. They also support the new structures by allowing their owners to run ads themselves and create easy access points to their sites and information on sites like Facebook, Google, or Twitter, where their information might otherwise not have been able to reach an interested public.

These new structures are important for the new digitally extended public arena. But in their characteristics they deviate from structures that formerly hosted the public arena, news media, and follow different rules. This raises challenges in developing normative goals and binding governance rules assuring their contribution to the public arena strengthens instead of weakens it. In this section, we discuss some of the most pressing challenges raised by the digital extension of the public arena.

6.4.1. Responsibilities of digital structures for the public arena

Digital media have become important structures hosting the public arena. People use Google to search for news and information, they get news on their Facebook feeds and publicly comment on it, they post links to news items on Twitter and interact with others, or they post links to news items in messenger groups on WhatsApp or Telegram and discuss them with friends and family. In fact, news sites increasingly rely on social media and messenger services for people to find their content and visit their sites. The affordances, usage practices, and algorithms of digital media thereby increasingly shape the way people find, interact, and share news. They are crucial channels for the flow of information through the public arena.[30]

These sites might not have started out with the goal of hosting the public arena, but by now they certainly do so. Accordingly, they have to accept the associate responsibility and accept for regulators and the public to hold them accountable. But this is easier said then done, while we have settled on what to expect from news media as structures hosting the public arena, digital media deviate from news media in decisive features and therefor bring specific challenges that need to be addressed if we want to understand or regulate their roles hosting the public arena.[31]

Crucially, if we look at former structures of the public arena, they usually combined information production and distribution functions. News organizations combined editorial desks producing information and distribution units that transported information products to points of sale or sent it over the airwaves. Digital media are nearly always only information distributors. Google, Facebook, Twitter, YouTube, WhatsApp do not produce information but merely point people to it or give them the option to point others to it and to allow them to comment on it. This matters in our discussion about the role of these structures in the public arena.

[30] For the state of the digital news environments see Newman et al. (2022). For the importance of digital platforms for referrals to news content see R. K. Nielsen and Fletcher (2022).
[31] See Jungherr and Schroeder (2022).

The institutional norms, we discussed above for news media as structures of the public arena nearly always focused on information production, editing, and curating. The distribution function was taken as a given. But what are the rules for structures that do not produce information themselves but host it? Here, a set of thorny questions arises regarding information quality, the policing of behavior, rules for representation and balance, and the transparency by which information gets displayed and distributed. Let's quickly have a look at each of these issues.

One foundational principle in the American regulatory framework for digital media is that companies are not directly liable for the content its users post on their service. They only become liable once they have been informed about illegal content or content infringing the rights of others and refuse to take it down. This was the foundation for platforms being able to grow quickly and host staggering amounts of content without having to exercise prior editorial or curational control. The flip side of this is that digital media host and provide access to large amounts of uncontrolled and unchecked content.

This of course negatively impacts the public arena. While in the past information was checked by journalists or editors before widespread circulation in the public arena, today unchecked or downright false information can travel widely through digital media before it is checked. Even debunking false information will not stop its circulation through digital media. This has given rise to widespread fears of disinformation running rampant on digital media. It is obvious that platforms hosting information vital to the pursuit of the public good need to address the challenge of information quality.[32]

This being said, it is unclear how this should look exactly. It is far from clear that platforms themselves are the best arbiters of truth, deciding which piece of information might be correct and which misleading or false. It is also difficult to exclusively rely on content produced by established news organizations and exclude information from up until then unknown or unverified sources. Imagine you living in an authoritarian country. How comfortable would you feel if digital platforms were exclusively hosting content from official media organizations aligned with the regime, while excluding voices and sources critical of it?

Also, related to that question is the challenge of moderating and policing speech. It is clear that platforms have a duty to protect their users from harassment and discriminatory or hateful attacks by others. But in practice this has turned out to be difficult to implement, especially in the context of political speech. Not all political speech is civil or polite, especially when directed at elites. We can regret this, but there is a reason that even impolite or uncivil political speech is protected. Often, especially with marginalized groups, impolite speech is part of their challenge of elites or majority groups in society. Deciding which impolite or uncivil speech to block, or users to deplatform is difficult and demands for clear criteria and processes. Having companies making these decisions on

[32]For an account of the risks of contemporary information environments see Bennett and Livingston (2018). For a discussion of the limits of digital disinformation see Jungherr and Schroeder (2021).

the fly without transparent and clear criteria risks damaging their legitimacy as hosts of the public arena or even the public arena as a space for political competition as a whole. This problem is only exasperated since most companies running digital media crucial for the public arena are based in the US or China and strangers to the political culture, legal systems, and contexts of the countries they have to make decisions about. This should give anyone pause trying to outsource these decisions to digital platforms.[33]

There is also the question of meaningful representation in digital spaces. In the past it was easy to look at the output of news organizations or the output of a select set of news organizations to assess the degree to which societal groups were represented or not. This is more difficult to do for digital media. For one, it is unclear what to focus on. Should representation be established on the aggregate level, looking at what groups find representation in all content available on digital platforms? Or should representation be established on the level of content visible to each specific user? In other words, are we satisfied with different societal groups being represented in aggregate or must this be true for the content each and every individual does actually see?

Beyond the conceptual question, what kind of representation of society we expect from platforms, we also have to address the challenge of limited transparency. Platforms are inherently opaque to outsiders. So how can the public or even governments assess the contribution of digital platforms to the quality of the public arena? Clearly, there remains much to do to establish transparent principles and procedures that make digital platforms assessable with regard to their contribution to the public arena. This is an important ongoing debate for academics, regulators, journalism, and the public to establish normative goals and practical procedures to establish integrity and trust for digital structures of the public arena.

6.4.2. Algorithmic shaping of user behavior in the public arena

The importance of digital media as structures for the contemporary public arena also introduces the question after the role of algorithms. Digital structures of the public arena – such as Facebook, Instagram, TikTok, or Twitter – use algorithms to determine which content to show their users and in which order. Algorithms thereby potentially shape the visibility of information crucial or detrimental to the functioning of the public arena. At the same time, the precise workings of these algorithms are unclear and their impact on information distribution uncertain. But in this discussion a series of questions have been raised that provide interesting anchor points for future research.

Probably the most well-known expectation about how digital media might algorithmically shape people's behavior is the *filter bubble*. At around 2011, the political activist Eli Pariser (2011) looked at his Facebook feed and found suspiciously many posts that

[33] For the challenges of regulation digital platforms see Keller (2018); Gorwa (2019a); Douek (2021). For the dangers of over-zealous speech regulation see Kaye (2019); Strossen (2018). For the effects of deplatforming see Rauchfleisch and Kaiser (2021).

supported his political viewpoints, while seeing almost none that contradicted them. This got him to formulate the *filter bubble* thesis.[34]

His reasoning was simple and compelling: Digital media, like Facebook, were interested in having their users to spend more time on the service looking at content providing the company with the opportunity to display and sell ads. To do so, they need to show people content they are likely to to be interested in or to interact with. Since many people like to see content they agree with, the companies developed algorithms identifying content supporting people's prior held beliefs or attitudes. For politics, the consequence was that people would only see content supporting their political beliefs. By using algorithmically shaped services to access the public arena, people would thus move into an algorithmic cage free of surprises only showing them information they were likely to agree with or support. This would be bad news for the public arena. Instead of making people visible to each other, digital structures of the public arena would hide them from each other. This reasoning proved to be as intuitively compelling as difficult to support empirically.

By now various studies have shown that people who use algorithmically shaped digital media to access news do not necessarily have narrower diversity of news exposure than those who use other services. In fact, in a by now classic study, the economists Matthew Gentzkow and Jesse M. Shapiro found for the US that people who were getting their news online were moving in ideologically less segregated information environments than those who talked about politics personally with others (Gentzkow & Shapiro, 2011). Digital communication environments might thus actually contribute to broader exposure to political others than previous communication environments. Looking at the available empirical evidence more broadly does gives little indication that algorithmic shaping would capture people in algorithmic cages exposing them only to politically uniform content.[35] However, there are other potential ways that algorithms might negatively impact the public arena.

Other contributions argue that digital media would try to increase time users spent by algorithmically selecting for highly controversial or emotionally impacting information. This could be sensationalist content or content likely to increase negative reactions, be it in form of critical comments or sharing. Algorithmic shaping of information environments might thus harm the public arena not by showing us too little of the political other but instead too much.[36] Again, looking at information making the rounds on popular digital media, one might be inclined to agree. However, it pays to keep in mind that this diagnosis is still mostly based on speculation about the workings of algorithms and the reach of information. Other than the filter bubble, this expectation is only beginning to be systematically examined empirically. The jury is therefor still out on this.

[34]For more on the filter bubble and its cousin the echo chamber see Jungherr et al. (2020), p. 85–92.

[35]For empirical attempts at measuring the filter bubble in different contexts and environments see Flaxman et al. (2016); Möller et al. (2018); Kitchens et al. (2020); Scharkow et al. (2020); Stier et al. (2022). For reviews of the available empirical evidence see Borgesius et al. (2016); A. Guess et al. (2018).

[36]See for example Settle (2018).

Finally, another approach looks at the potential impact of algorithms on the radicalization of users. While arguments like the filter bubble focus on population wide effects, this argument focuses on the experiences and effects on select users who encounter extremist content. Here the argument goes that people in algorithmically shaped environments like YouTube might encounter mildly deviant or controversial content, such as content in support of the far right or terrorism, and through the recommendation algorithm might get sucked into rabbit holes of increasingly extreme content. For some people, this then might constitute a content journey into radicalism. Empirical studies have shown that the YouTube algorithm suggesting content of what to watch next could produce comparable patterns, pushing people to increasingly more radical content. While algorithms might therefore not have the population wide effects of tearing apart the shared public arena, they certainly can have detrimental effects on select and vulnerable people, thereby strengthening extremism on the margins of society.[37] Clearly, more research is needed here. For example the recent rise of TikTok as a channel for information and news and its heavy reliance on algorithmic content selection points to interesting challenges going forward.

An important challenge for all research in this area is that both the workings of algorithmic shaping within digital media and its effects are highly opaque for academics and the public. This has given rise to far reaching speculation of the hidden workings of digital media and their supposed effects. This opaqueness has also contributed to intuitive but empirically elusive speculations, like the filter bubble, have achieved wide prominence that does not correspond with the empirical evidence. Opaqueness therefore legitimizes speculations while weakening the discursive strength of empirical evidence. Bad ideas therefore exit the field much more slowly than one could wish for. Some academics have reacted to this opaqueness by demanding broader access to data from digital media. While of course more data for academics is always a popular demand among academics, it remains dubious that data alone will solve the challenges presented by algorithmic shaping. Also, companies running digital media have to weigh the interests of their users with the interests of academics. Overall, research addressing patterns and effects of algorithmic shaping on digital media might profit from better and broader data access. But actual progress in the field might depend even more on creative research designs addressing the heavy logical as well as conceptual challenges that characterize this research area.[38]

6.4.3. Geopolitics of digital structures

Digital media hosting the public arena have also introduced questions not usually asked in discussions about traditional structures of the public arena. Digital structures make it necessary to talk about the geopolitics of the public arena. Here, one question looms

[37] See for example Kaiser and Rauchfleisch (2019); Kaiser and Rauchfleisch (2020).
[38] For different suggestion see for example Jürgens and Stark (2017); Munger and Phillips (2022).

large. Increasingly societies have to figure out how to adjust to digital structures that host their public arena being run from countries different from their own.

In the past, structures hosting a country's public arena were run by organizations based within the same country. This gave governments direct control over, access to, and knowledge about structures hosting the public arena. This is still true for most media organizations. But it is decidedly different for digital structures. People from Germany are using Google to search for political news. People from the UK use What'sApp to coordinate protests. People from the US use TikTok to learn about politics. The public arenas of most Western democracies rely crucially on digital structures run from other countries. Most of those are hosted in the US. Services like Facebook, Google, Instagram, Twitter, or YouTube are crucial features of public arenas all over the world. But Chinese services start to figure prominently as well, be it TikTok or WeChat are increasingly used outside China while China remains closed to most Western structures.

By relying on digital structures hosted in other countries, countries open themselves up to potential interference or influence. Not surprisingly, once the geopolitical mood shifts from cross-border cooperation to competition, these dependencies of crucial societal structures become risks. Already in 2013, the NSA scandal revealed that the US government worked closely with digital technology companies to spy on allies. More recently, the growing worry about relying on foreign structures could be seen in the blocking of 5G network technology provided by the Chinese telecommunication company Huawei in many Western countries. The shut-down of Russian propaganda television station Russia Today in Germany following Russia's unprovoked attack on the Ukraine and the Russian blocking of access to Western digital media provide other examples for these fears.

But probably the most striking attempt at controlling foreign digital structures can be found in China. Here, the government has implemented a firewall that shuts off China from the Western internet and digital media run by Western companies. In the beginning, the Great Firewall was not taken seriously by the West or technology companies. But over time, the Great Firewall created a space that protected Chinese companies from Western competition, so that local structures could emerge that were open to regime control. Over time this has allowed China to develop a set of digital media that can compete with those coming from the US and that at the same time offer governments greater degrees of control. These structures promise a greater independence from the US while of course at the same time creating greater dependence on China. In the potential re-emergence of geopolitical blocks in the aftermath of the Ukraine war of 2022, this is an important development. Especially since already before the war, China was exporting its digital media to countries in the context of its international infrastructure and trade initiatives. Some have spoken of this as the Digital Silk Road.[39]

[39] For the Great Firewall see Griffiths (2019). For growing international reliance on Chinese digital media see Erie and Streinz (2021); Hillman (2021).

Looking closely, we can thus identify three types of interdependency emerging from important structures of the public arena being hosted abroad. One is the direct dependence on structures that are potentially open to interference and access of a foreign government. This is the sort of risk made painfully clear to Western democracies in the context of the 2013 NSA scandal. When the world learned that technology companies hosted in the US provided the US government with broad access to its user data allowing it to spy on them. This is the same sort of risk leading various countries to block local infrastructure projects relying on technology provided by Chinese firms.[40]

The second interdependence is the influence of laws and regulations. This influence can go one of two ways. The first route is that laws and regulation from the country providing a digital structure travels to the country using said structure. This is what Erie and Streinz (2021) describe as the *Beijing Effect*. Countries relying on digital media from China have to accept Chinese approaches to digital governance and data regulation. The other route goes the opposite way. Here, the country using digital structures from another country can set rules and regulations to allow market access. If its market is attractive enough, this can indeed lead to local laws and regulation being able to change laws and regulation in the country of origin of the digital structure. This is what Bradford (2020) has called the *Brussels Effect*. As the name suggests, the primary example for this sort of influence stems form the EU. Examples of how EU regulation starts to reign in US companies and to influence even regulation in the US itself are the *Data Governance Act* (DGA), the *Digital Markets Act* (DMA), the *Digital Services Act* (DSA), or the *General Data Protection Regulation* (GDPR).[41]

The third type of influence concerns information flows and probably can be best characterized as a form of soft power. This is the kind of influence China and Russia are primarily afraid of and is the reason why they block access to the Western internet. Here, the fear is that people can get information from without autocratic regimes and might in turn become critical of said regime. But also the West is afraid of this type of influence. Here, the fear is that autocratic regimes use media structures to spread their propaganda beyond their borders. To mitigate these dangers, Western governments have for example closed down local stations of Russia Today, an international television station financed by the Russian state. Recently, similar fears have been raised with regard to the potential influence of the Chinese government in Western democracies through the digital video service TikTok, which becomes increasingly popular among young people in the US and other Western democracies.

Overall, these observations show that geopolitics start to matter in the discussion of the public arena, once digital media become important structures hosting it. Questions of mutual dependencies and influence through digital media will become even more important than they are already. Here, the geopolitical fault lines of the international

[40] For the risks of growing international interdependence see Farrell and Newman (2019b); Drezner et al. (2021).

[41] For the influence of EU data laws and regulation see Farrell and Newman (2019a); Bradford (2020). For growing international reliance on Chinese laws and regulation see Erie and Streinz (2021).

system will come to shape the discussion of the public arena. Mounting or relaxing tensions will increase or decrease the importance of these fault lines.

6.5. The public arena examined

As we have seen, the public arena is a crucial element of democratic societies, linking communication to political competition and democratic representation. It comes as no surprise then to find that the concept has inspired massive research activity. The digital transformation of the public arena has featured very prominently in recent research. The diversity of interests, approaches, and methods in studies on the contemporary digitally extended public arena mirrors the richness of the concept and its related areas. Prominent topics include:

- the detailed examination of structures hosting the public arena, their constitution, contextual embeddedness, and shaping power for discourse and political competition;
- shifting and competing norms and practices among actors within the public arena;
- changing patterns of political competition within the digitally extended public arena and the emergence of challengers to the status quo;
- patterns of exchange and interaction within the digitally extended public arena.

This short list is not complete by far but it sketches some of the rich research opportunities within the public arena. To get a better sense of it, we now turn to three studies that address related questions empirically.

6.5.1. Limits to attention

The big challenge in the contemporary digitally extended public arena is the question of attention. By now we have repeatedly discussed that the contemporary public arena is no longer limited by access or the volume of information. Instead, its limits are set by the limits of individual and collective attention. Winning in the intense competition for attention is of crucial importance for actors within the public arena.[42] While academics, need to understand the underlying dynamics and associated limits. A recent study by Rauchfleisch et al. (2023) illustrates how one can do so.

In their article "How COVID-19 Displaced Climate Change" Adrian Rauchfleisch, Dario Siegen, and Daniel Vogler examine whether attention to one issue of grave societal importance - climate change - was replaced by attention to another issue of great urgency – COVID-19 – or whether attention to both issues and the associated challenges persisted. They examine this by analyzing the presence of both topics in media coverage and on Twitter in Switzerland between April 2019 and October 2020. They collected news

[42] For the competition for attention in the public arena see Schroeder (2018).

coverage on news websites, newspapers, and transcripts from TV and radio newscasts which left them with 1,060,820 articles during the relevant time span. Of those 56,128 stories referred to climate change and 174,407 to COVID-19. 6,431 stories referenced both debates. For the analysis of Twitter, they relied on a tracker covering the whole Swiss Twitter-sphere during the time frame (this includes 296,553 users who posted 92.7 million tweets during the relevant time span). Through a set of topical keywords, the authors identified tweets referring to either topic, leaving them with 407,626 tweets referring to climate change and 3,214,483 mentioning COVID-19.

The authors built two time series of news and Twitter attention to both topics. To identify the causal impact of the COVID-19 pandemic on the attention on climate change in the public arena, they use the data from April 2019 to January 2020 and predict based on the underlying trends how attention to climate change would have developed from February 2020 to October 2020 if COVID-19 had not happened and the underlying dynamics of the previous year would have continued. By identifying the difference between both, their prediction and the actual coverage dynamics, they can identify the impact COVID-19 had on attention toward climate change in the news and on Twitter.[43]

As expected, the authors found that after February 2020 attention in both news and on Twitter toward COVID-19 increased while attention toward climate change clearly decreased. Still, some events related to climate change created attention peaks in both the media and Twitter and overall Twitter attention toward climate change was strongly correlated with news attention. These findings echo earlier research, indicating a strong and persistent link between news coverage and Twitter attention to current events or politics (Jungherr, 2014).

Comparing actual attention toward climate change with predicted attention, the authors show that COVID-19 clearly had negative effects on attention toward climate change. For both news coverage and Twitter reactions, the authors find substantial negative effects, lowering news attention to climate change 46% and attention on Twitter by 55%. This is clear evidence for the limits of attention in the public arena overall, where the prominence of a new topic – COVID-19 – comes at losses for another – climate change – irrespective of its continued relevance to society. This is clearly troubling news from the perspective of climate activists or politicians intend on keeping continuous attention on the topic in order for society to keep focus on maintaining efforts in fighting climate change.

In a further analytical step, the authors show through the analysis of co-occuring hashtags that climate change activists reacted to the decrease in collective attention by trying to link the issue of climate change with the issue COVID-19. Here, they find that only 0.5% of tweets contributing the COVID-19 debate referenced climate change, while 11% of tweets within the climate debate referred to COVID-19. This and looking closely at the connecting hashtags, leads the authors to conclude that climate activists tried to

[43]For more on the underlying time series model see Brodersen et al. (2015).

tactically adjust to the new circumstances and connect their issue of relevance – climate change – with the issue of the day – COVID-19.

In its use of both, time series analysis and the analysis of hashtag co-occurances, the study by Rauchfleisch et al. (2023) provides a helpful template for further research. Their study shows how one can approach issue attention and attention drifts in empirical research as well as examine tactics by actors within the public arena to mitigate or profit from shifts in collective attention.

6.5.2. Digital shaping of behavior

Another strand of research looks at the behavior of people contributing to online discussions about politics and current events. Often these studies focus on the hostility of contributions in online environments and explicitly or implicitly connect a perceived tendency toward hostility with specific conditions of online discourse. For example, digital environments would allow people anonymity, low accountability for their actions, and physical distance to others. In combination, these factors would activate people's negative impulses and turn them into trolls, ready to engage others in a hostile fashion or even to harass them. In short, the internet might turn people into trolls. This set of expectations has been called the "mismatch hypothesis". If true, this of course would provide bad conditions for the digitally extended public arena. Luckily for us Bor and Petersen (2022) put this thesis to the test.

In their article "The Psychology of Online Political Hostility" Alexander Bor and Michael Bang Petersen present a series of comparative studies, in which they test the effects of conditions found in digital communication environments on people's behavior in discourse. To do so, they ran a series of surveys with respondents from the US and Denmark, to test whether expectations from the mismatch hypothesis would correctly predict correlations in their surveys.

In the words of the authors, the mismatch thesis states that:

> "(…) this class of effects imply that the"perfect storm" of novel online features (e.g., anonymity and rapid text-based communication) induces fleeting psychological changes that increase the likelihood of certain psychological states that undermine civil discussions (…). Simply put, when people log online their level of empathy is reduced or they become more aggressive than usual."
>
> Bor and Petersen (2022), p. 2.

The authors contrast this expectations with the connection hypothesis:

> "Online environments are unique in creating large public forums, where hostile messages may reach thousands including many strangers, could stay accessible perennially, and may be promoted by algorithms tuned to generate

> interactions (…). From this perspective, online environments do not shape how people are motivated but shape what they can accomplish given a specific set of motivations. The hostility gap may thus emerge as a direct consequence of the larger reach of those already motivated to be hostile."
>
> Bor and Petersen (2022), p. 2.

This is an important distinction. While people might experience discussions in online environments as more hostile than those offline, this might not be due to people turning into ogres online but simply due to that those behaving badly are more visible online than offline. The underlying problem would then be primarily psychological and motivational, not technological. So, how do Bor and Petersen (2022) test this?

In a first set of three studies, the authors survey respondents from the US and Denmark to find first evidence. They find that respondents in both countries perceived online discussions to be more hostile than those offline. Respondents themselves did not express differences in their own behavior that one could consider hostile between online and offline discussions. And they find that the personality trait status-driven risk seeking is no stronger correlated between self-reported hostile behavior on- or offline.

They build on these findings in a fourth study by using a more comprehensive scale to measure self-reported hostile behavior. They ran this study with respondents from the US. Again, they find no difference between self-reported hostile behavior on- or offline. In combination these studies do not support the mismatch hypothesis.

The authors continue to refine their findings and test different aspects of the mismatch thesis in three subsequent experiments. We skip these studies to focus on their test of the connectivity thesis. Let it suffice then, that the experiments also do not provide evidence for the mismatch hypothesis.

In a final study, the authors test the connectivity hypothesis. For this they survey people from the US and Denmark on whether they had witnessed attacks against self, friends, and strangers in on- or offline environments. Here, the respondents clearly reported to have witnessed attacks more often in online instead of offline environments, with the strongest difference being reported for attacks on strangers.

In combination, the authors see their findings as rejecting the mismatch hypothesis:

> "(…) our research suggests that people do not engage in online political hostility by accident. Online political hostility reflects status-driven individuals' deliberate intentions to participate in political discussions and offend others in both online and offline contexts. In large online discussion networks, the actions of these individuals are highly visible, especially compared with more private offline settings."
>
> Bor and Petersen (2022), p. 16.

The article offers an instructive example for the challenge of identifying the drivers between perceived hostility and deviance in digital communication environments. While it is tempting to attribute digital technology causal effects on people's deviant behavior, it might simply be that digital make more of hostile behavior visible. That alone does not solve the problem of hostility in the digitally extended public arena, but it helps us to identify its drivers and to design interventions.

Beyond the substantive interest, the study offers also an interesting template for careful empirical work presenting a set of carefully designed studies first translating broad expectations into testable hypotheses allowing for the identification of different mechanisms leading to similar outcomes.

6.5.3. Contesting narratives

The public arena consists of spaces that hosts discourses in which societal actors compete for attention and dominance. The digital manifestations of this competition offer us a detailed view of the content, patterns, and tactics of this competition. Especially, the microblogging service Twitter has proved to be a promising research environment to better understand the competition between actors for attention in the public arena. But other digital environments, such as Instagram or Reddit, also start to feature more strongly in research.[44] One example for such a Twitter-based analysis is a paper by Knüpfer et al. (2022).

In their paper "Hijacking *MeToo*: transnational dynamics and networked frame contestation on the far right in the case of the '120 decibels' campaign" Curd Knüpfer, Matthias Hoffmann, and Vadim Voskresenskii analyze the #120db campaign on Twitter. In late January 2018, members of the Austrian and German far-right Identitarian Movement launched a social media campaign. The goal of the campaign was according to Knüpfer et al. (2022):

> "According to their German website, their core goal is the conservation of an 'ethno-cultural' identity, in what is referred to as 'the age of mass migration, globalization and one-world-propaganda'."
>
> Knüpfer et al. (2022), p. 1012.

In this, the activists encouraged women to

> "use social media posts to 'talk about your experiences as a female with foreign infiltration, harassment and violence'."
>
> Knüpfer et al. (2022), p. 1012.

[44] For an US-focused overview of discourse competition on Twitter see Jackson et al. (2020). For a comparison between Twitter tactics in the US, Spain, and Greece see Theocharis et al. (2015). For Germany see Jungherr and Jürgens (2014). For an analysis based on Reddit data see Jungherr et al. (2022).

In the campaign far-right actors latched on to the momentum and frames established by the feminist #MeToo campaign. But this association is merely rhetorical and stylistic – for example through videos imitating the style of grassroots testimonial videos. This is a shift from tactics of the past, where far-right activists might have actively and openly challenged successful frames presented by left or feminist activists. Here, instead of openly challenging or contesting the frame, they try to co-opt it and refocus attention away from the original goal – exposing and challenging sexist behavior and practices condoned by a patriarchal social system – to their own political goals – painting migrants as a broad societal threat.

> "The campaign did so by drawing explicit attention to acts of violence against women perpetrated by 'foreign' men or recent immigrants. This form of strategic frame contestation is not characterized by an outright dismissal of the original framing effort but, rather, by a narrowing of the original problem definition and the propagation of a different set of policy demands."
>
> Knüpfer et al. (2022), p. 1014.

This tactic is what the authors term "hijacking".

To analyze this tactic, they collected tweets containing the campaign hashtag #120db through the Twitter streaming API between January 30 and May 31st, 2018, the run of the campaign. During that time, they collected 172,972 tweets from 44,834 unique user profiles. Of the tweets mentioning #120db roughly ten percent were also mentioning #MeToo. The authors see this, and specific temporal and language patterns, as evidence that the originators of the campaign very actively tried to use the attention on #MeToo to launch their own campaign and inject their contesting frame within the larger #MeToo debate. One tactic to achieve this was the attempt to inject their specific regional claims within a larger international debate.

The authors continue their analysis, through a qualitative look at the content of messages using both hashtags. Here, they look for the occurrence of three tactics:

> "First, agenda surfing is characterized by encouraging and progressive/feminist messages, usually referencing #MeToo without evaluation. Second, re-framing/undermining features a critical evaluation of #MeToo, accentuating the seemingly more accurate problem definition of #120db. Third, critical/anti-120db tweets included negative evaluations of #120db, and sometimes also of #MeToo."
>
> Knüpfer et al. (2022), p. 1021.

The authors handcoded 123 tweets containing both hashtags that were posted during the first 48 hours of the campaign according to their correspondence with these tactics. Here, the authors found that tweets with co-occuring hashtags were dominated by a critical stance toward the #120db as well as those trying to actively reframe #MeToo following the far-right agenda. Mere agenda surfing tweets were in the minority. This shows that activists from the far-right as well as from the original #MeToo movement actively engaged in frame contestation around the concerned hashtags. With far-right activists pushing into the campaign and attention space generated by #MeToo and activists in that space pushing actively back and defending the movement from this attempt at hijacking attention and momentum.

Of course the study by Knüpfer et al. (2022) addresses other questions as well. But for our purposes, this is enough. The study is an interesting close look at the tactics used by activists within the public arena in their competition for attention. It is also interesting as it presents a case for Twitter-based and discursive activism from the far right. Often, these tactics are discussed with a focus on left leaning groups. But as Knüpfer et al. (2022) show these tactics can be successfully employed from the political right as well, weakening the argument that campaign tools or styles can be owned or associated with specific factions on the political spectrum. Associated imbalances in the literature are more likely due to skewed attention by researchers than by actual differences between tactics or approaches between different political factions.

6.6. The contemporary public arena

The contemporary constellation of the public arena looks different from the past. Digital technology has weakened traditional structures, introduced new one, and led to much soul-searching and norm-shifting for actors providing structures of the public arena and competing within it. These shifts mean that both public and society need to adjust their expectations of and practices within the public arena. But also academia needs to adjust to these new constellations.

While these shifts are associated with great fears for our democracy or the quality of discourse, they also bring tremendous research opportunities. Empirical research is challenged to examine the nature, functions, and power relationships between structures of the public arena old and new. How do news media differ from new digital platforms or how do they resemble each other? What can we learn from the study of one type of structure about others?

Also, empirical research needs to find ways to examine patterns of information flow, discourse dynamics, and interaction behavior within the contemporary public arena. How does information flow between structures old and new? Do new features of structures influence the way discursive competition happens between actors in the public arena or can we observe shifts in power? Or how do people behave when engaging in political exchanges in structures old and new?

Finally, empirical research also needs to focus on outcomes. What are the effects of the new constellation of structures within the public arena. Do digital media contribute to polarization within society? What does algorithmic shaping do for information exposure and attitude formation? And is there evidence for more or different paths to radicalization in the new public arena?

But of course, we do not only need empirical research. Maybe the primary task right now lies with the development of theoretical or normative concepts of what to expect from the contemporary public arena. What are the functions and normative goals we demand from news media or digital platforms hosting the contemporary public arena? What do we expect from political elites competing under the changed conditions of the contemporary public arena? And what do we expect from the public? The contemporary public arena brings many opportunities for people, elites, and society. But to capitalize on them, we need to have a better understanding of its shape, dynamics, and effects. Here, there is clear need for creative but empirically grounded conceptual, theoretical, and normative work.

Importantly, this goes beyond easy critiques in the mode of supposed deterministic societal decline, as for example in the mold of *surveillance capitalism*. Too often work like this is empirically ill founded and follows a critical stance that sees capitalism or for-profit companies as the source of all evil. These works tell us little about actual changes within the public arena, its effects on individuals and society, and ultimately do not offer much of a way forward. Besides of course abolishing capitalism. Instead, we need to become better at understanding what is actually happening (in other words establishing meaningful transparency for structures of the public arena old and new) and surfacing and negotiating tensions that exist within and between structures of the public arena (old and new) and actors competing within it.

The new structures of the public area are here to stay. As we have seen, they mitigate some of the ills of previous constellations within the public arena but at the same time introduce some new ills and inspire new worries. It is up to society to figure out the norms and practices allowing us to pursue the public good under these new conditions. Turning back the clock is not an option! Neither should it be, given the well-understood but currently often ignored ills of a public arena dominated by a few powerful structures heavily aligned with the powers-that-be in economy, politics, and society. The current structural transformations of the public arena are noisy, contested, and surface very real political and societal tensions and fractures. But engaged constructively and creatively, these transformations can be used to strengthen societies by engaging these tensions and fractures instead of weakening it by trying to ignore and to hide them.

6.7. Review questions

1. Please define the term *public arena* following Jungherr and Schroeder (2022).
2. Please define the term *filter bubble* following Pariser (2011) and describe its underlying mechanism.

3. Please discuss the three different from of geopolitical interdependence in the contemporary public arena.

7. Artificial intelligence and democracy

7.1. Artificial intelligence in politics and society

The success and widespread deployment of artificial intelligence (AI) have raised awareness of the technology's economic, social, and political consequences. The most recent step in AI development – the application of large language models (LLMs) and other transformer models to the generation of text, image, video, or audio content – has come to dominate the public imaginary of AI and has accelerated this discussion. But to assess AI's societal impact meaningfully, we need to look closely at the workings of the underlying technology and identify the areas of contact within fields of interest.

AI has become a pervasive presence in society. Recent technological advances have allowed for broad deployment of AI-based systems in many different areas of social, economic, and political life. In the process, AI has had – or is expected to have – a deep effect on each area it touches. We see examples in discussions about algorithmic shaping of digital communication environments and the associated deterioration of political discourse;[1] the flooding of the public arena with false or misleading information enabled by generative AI;[2] the future of work and AI's role in replacement of jobs and related automation-driven unemployment;[3] and AI's impact on shifting the competitive balance between autocracies and democracies.[4] With these developments, AI has also begun to touch on the very idea and practice of democracy.

This makes AI with its workings, applications, and effects an important topic for political science.[5] But, since many of AI's applications and their society-wide consequences lie still in the future, political science struggles with addressing associated questions. This chapter provides students with a framework of contributing to the ongoing discussion about the role of AI in society.

The chapter will start with a non-technical introduction to AI and the conditions for is successful application. It will then present a set of important areas in democracy where AI is starting to get used and develop effects. The map of these areas can serve as a

[1] See Kaye (2018).
[2] See Krebs et al. (2022).
[3] See Acemoglu and Johnson (2023), Brynjolfsson and McAfee (2016), C. B. Frey (2019).
[4] See Filgueiras (2022), Lee (2018).
[5] See Risse (2023).

conceptual framework for students interested in the future work on AI and democracy .[6]

7.2. What is artificial intelligence?

The success and widespread use of artificial intelligence (AI) has increased awareness of its economic, social and political impacts. The idea of a powerful machine-intelligence has inspired far-reaching expectations and fears regarding the potential or threats associated with AI, ranging from economic growth[7] and post-human transcendence[8] to a downright menace to human existence.[9] *Large Language Models* (LLMs) and other transformer models that enable the automated creation of text, image, video, or audio content are currently dominating the public imagination and are associated successes and innovations are accelerating this discussion.[10] But the discussion of AI and its impacts is broader and goes back well before the recent wave of technological innovations and commercial applications.

In his history of the research field artificial intelligence, Nils J. Nilsson defines AI as

> [...] that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment.
>
> Nilsson (2010), p. xiii.

This definition provides a broad tent for scientific, engineering, and commercial endeavors affiliated with making intelligent machines or learning about intelligence through machines. The continued scientific quest for artificial intelligence starts roughly in the nineteen-fifties bringing psychologists, engineers, and computer scientists together. These efforts begin as the attempt at understanding human intelligence better by trying to recreate it in machines while at the same time getting machines to perform tasks traditionally assigned to humans, such as conversation, reasoning tasks, or playing board games. In the early days of the field from the nineteen-fifties to the nineteen-nineties scientific and commercial efforts in artificial intelligence were dominated by knowledge-based approaches, trying to teach machines about the world, be it the meaning of words, grammar, or expert knowledge in specific subfields. Approaches like these started out being highly popular among scientists and funders but lost steam when the promised

[6] This chapter is in parts based on the articles Jungherr (2023a) and Jungherr and Schroeder (2023). Some of the material presented here, is adapted from theses earlier sources.
[7] See Brynjolfsson and McAfee (2016).
[8] See Hanson (2016).
[9] See Bostrom (2014).
[10] For the technological foundations of transformer models see T. B. Brown et al. (2020), Vaswani et al. (2017). For a non-technical introduction to *ChatGPT*, one of the most popular current AI applications, see Wolfram (2023).

results failed to materialize. The associated frustrations and decreased scientific interest and funding opportunities have become known as *AI winter*.[11]

But each winter ends with the onset of spring.[12]

The first signs of the onsetting *AI spring* appeared in the nineteen-nineties. Increased computing power and ever growing availability of large data sets documenting ever more aspects of human life and society accompanying the digital transformation led to growing interest in the uses of neural network models. Neural networks are a family of computational models that is inspired by the working of the human brain. In a very simplified account, the brain consists of networks of interconnected neurons. Each neuron receives stimuli. Once these incoming stimuli pass a certain threshold, a neuron sends out an electrical signal that itself serves as an input to other connected neurons. This model of the brain concentrates on its information processing characteristics, in which interconnected neurons accept and process information and through their interconnections are able to achieve stunning feats of translating information into perception, knowledge, or action.

Artificial neural networks follow the same logic in their architecture and functioning. Artificial neurons accept numerical inputs and translate them into a single output variable. These artificial neurons are arranged in networks with varying levels of network layers. A first input layer accepts unprocessed signals and puts them through to a series of so-called hidden layers. These hidden layers accept the outputs of the network layer above them, process these outputs, and transmit them to a further layer, until a final output layer is reached that provides the result of the model. This process is called deep learning and enables machines to make data-driven predictions and decisions.[13]

Deep learning has made machines highly efficient in automated pattern recognition and decision making without relying on knowledge or theory. This purely data-driven approach to artificial intelligence has produced spectacular results in many different contexts, including computer vision, machine translation, medical diagnosis, robotics, and voice recognition.[14] They also have been successfully applied in predicting possible but yet unknown biological or chemical compounds[15] and strategic action in game play.[16]

[11] For a recent chronological account of the development of artificial intelligent from its past into its present see Woolridge (2020). For a personal history of the early days of artificial intelligence and the recollections of some of its pioneers see McCorduck (2004). For a narrative history of the people, companies, and feuds in this recent phase in the history of artificial intelligence see Metz (2021). For a broader history of the field see Nilsson (2010).

[12] For a non-technical introduction to current approaches and techniques in artificial intelligence see M. Mitchell (2019). For a highly popular technical introduction to artificial intelligence and current challenges and techniques see Russell and Norvig (1995/2021).

[13] For an easy to follow introduction to neural nets and deep learning see Kelleher (2019). For a technical introduction and deeper discussion see Goodfellow et al. (2016) or Prince (2023). For a practical hands-on introduction to programming deep learning models see Trask (2019).

[14] See LeCun et al. (2015).

[15] See Chow et al. (2018), R. D. King et al. (2009), Schneider (2018).

[16] For Go see Silver et al. (2016); Silver et al. (2017). For Go, Shogi, and Chess see Silver et al. (2018). For diplomacy see FAIR et al. (2022).

A recent advance in deep learning are transformer models.[17] They serve as the foundation of highly popular applications, such as *ChatGPT* or *Midjourney*, that allow the autonomous generation of text, image, and video content.[18] In fact, deep learning and its commercial applications have been so successful that deep learning has become nearly synonymous with artificial intelligence, thereby reducing the field's richness and varied heritage to a limited set of data-driven models and approaches.

But while these models clearly are highly successful and adaptable to surprisingly rich and varied contexts and tasks, should we accept their success as evidence for a deeper, human-level intelligence?

7.3. Narrow artificial intelligence versus artificial general intelligence

Let's go back to Nilsson's definition of AI:

> Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment.
>
> Nilsson (2010), p. xiii.

This definition is very useful since it covers different aspects and goals of artificial intelligence research and development. At the same time, the definition also points to an inherent tension within the field. For example, what accounts for " function[ing] appropriately", what counts as "foresight", and – perhaps most controversially – what is the minimum requirement to meaningfully speak of "function [...] in its environment".

Nilsson consciously takes a broad and open approach to these questions:

> According to that definition, lots of things – humans, animals, and some machines – are intelligent. Machines, such as "smart cameras", and many animals are at the primitive end of the extended continuum along which entities with various degrees of intelligence are arrayed. At the other end are humans, who are able to reason, achieve goals, understand and generate language, perceive and respond to sensory inputs, prove mathematical theorems, play challenging games, synthesize and summarize information, create art and music, and even write histories.
>
> Nilsson (2010), p. xiii.

[17] See Parmar et al. (2018), Vaswani et al. (2017).
[18] See T. B. Brown et al. (2020), Ramesh et al. (2022).

This approach works well for charting the development of the field of artificial intelligence and covering its varied points of origin and developmental paths. But to understand the larger societal effects of artificial intelligence it is limiting, if not misleading. Such a broad account of what it means to be *intelligent* allows commentators to conflate highly specific feats of data-driven pattern recognition, prediction, and decision making with more general and substantively different expressions of intelligence – such as creativity, embodied knowledge, or abduction. Impressive, but highly specific feats of data-driven pattern recognition, prediction, and decision making – such as winning against humans in board games like Chess or Go – are used to infer the subsequent replacement of human decision making in other contexts. This is a category error. It is true, there is undeniable and impressive progress of machines in one narrow category of intelligence – pattern recognition, prediction, and decision making based on the automated analysis of large data sets. But this does not automatically translate into progress in the development of another broader, general intelligence.

This surprisingly widespread fallacy gives rise to what Erik J. Larson calls the "Myth of Artificial Intelligence":

> The myth of artificial intelligence is that its arrival is inevitable, and only a matter of time – that we have already embarked on the path that will lead to human-level AI, and then superintelligence.
>
> Larson (2021), p. 1.

This myth is dangerous since it ignores the crucial differences between the type of intelligence found in current AI-empowered systems and a human level general intelligence:

> The myth of AI insists that the differences are only temporary, and that more powerful systems will eventually erase them. [...] the myth assumes that we need only keep "chipping away" at the challenge of general intelligence by making progress on narrow feats of intelligence, like playing games or recognizing images. This is a profound mistake: success on narrow applications gets us not one step closer to general intelligence. [...] As we successfully apply simpler, narrow versions of intelligence that benefit from faster computers and lots of data, we are not making incremental progress, but rather picking low-hanging fruit. The jump to general "common sense" is completely different, and there's no known path from the one to the other.
>
> Larson (2021), p. 1-2.

This error of using evidence of successful applications of artificial intelligence for specific narrowly defined tasks to project the imminent emergence of a general artificial intelligence that will be replacing human decision making in all walks of life matters. Jumping from the observation that AI-assisted homes can autonomously order produce from the internet to fears of a *Skynet*-level AI that is just around the corner and a threat to all life on earth is a category error and logical fallacy. And one that is particularly tempting and dangerous in the discussion of the likely societal impact of and regulatory answers to artificial intelligence. This does not mean that artificial intelligence has not made great strides and in turn made philosophers and cognitive scientists rethink crucial tenets of what exactly constitutes intelligence. But by conflating the two different categories of intelligence, we run the risk of misdiagnosing the nature, impact, and trajectory of artificial intelligence.

One way to account for these differences, is to differentiate between *narrow artificial intelligence* and *artificial general intelligence (AGI)*. Narrow AI refers to systems developed for a specific, singular, or limited task.[19] Narrow AIs can be very successful in the tasks they are developed for but they fail at other tasks. In contrast an *artificial general intelligence* would be a machine with the same cognitive and intellectual capacities as a typical human. It would be able to perform different tasks equally well, even those it was not explicitly trained for. For example, an AGI would be able to hold a conversation in natural language, solve problems in different areas, perceive the world and its position in it, and reason about it. While the public imagination focuses on AGI, scientific research and commercial applications nearly exclusively focus on narrow AI [20].

This disconnect between research activity and public imagination is important as it leads to a fundamental misunderstanding between discursive expectations and technological reality. The computer scientist and philosopher Brian Cantwell Smith proposes one way of remaining aware of these differences. In *The Promise of Artificial Intelligence* Smith emphasizes the difference between two intelligence tasks – *reckoning* and *judgment*. With *reckoning* he refers to calculation tasks that current instances of AI are highly successful at, like those discussed above:

> [...] the representation manipulation and other forms of intentionally and semantically interpretable behavior carried out by systems that are not themselves capable [...] of understanding what it is that those representations are about – that are not themselves capable of holding the content of their representations to account, that do not authentically engage with the world's being the way in which their representations represent it as being. Reckoning [...] is a term for the calculative rationality of which present-day computers [...] are capable.
>
> Smith (2019), p. 110.

In contrast, Smith uses the term *judgment* for the sort of understanding

> [...] that is capable of taking objects to be objects, that knows the difference between appearance and reality, that is existentially committed to its own existence and to the integrity of the world as world, that is beholden to objects and bound by them, that defers, and all the rest.

[19] See M. Mitchell (2019), p. 45f.
[20] Woolridge (2020), p. 42.

Smith (2019), p. 110.

This difference between these forms of intelligence is not just a matter of terminology or of purely academic interest. Instead, by employing machines adapt at reckoning for tasks that require judgement, societies risk having machines calculate and initiate decisions based on the representation of the world and not the world itself and without any commitment to the consequences of these decisions. Conversely, focusing on the staggering successes of machines with reckoning tasks might lead over time to a devaluation of judgment, tasks that machines are not successful at, thereby settling on a critically reduced account of what human intelligence is and should be about. Keeping the difference between reckoning and judgment in mind is therefore crucial in the discussion of what artificial intelligence can do and for what tasks it is employed.

For the discussion of the impact of artificial intelligence on democracy, it is important to be precise about what kind of artificial intelligence we are talking. While it is easy to extrapolate far reaching impacts of an imagined artificial intelligence on aspects of democracy, this is probably not the best use of our time. Instead, we will focus on the impact of aspects of artificial intelligence already in evidence: The power of data-driven predictions and the impact of this on specific aspects of democracy. But first, it pays to look closely at preconditions for the successful application of artificial intelligence.

7.4. Conditions for the successful application of artificial intelligence

The successful application of artificial intelligence depends on a set of preconditions. Some are obvious. For example, to be successful AI needs to be able to access some digital representation of its environment, either through sensors mapping the world or through the input of existing data. Where these representations are difficult to come by or data are scarce, as in many areas of politics, AI will not be successful. Other preconditions are not so obvious. For example, for AI to produce helpful results, the underlying connections between inputs and outputs must be stable over time. This points to two problems: unobserved temporal shifts between variables (Lazer et al., 2014) and the dangers of relying on purely correlative evidence without support of causal models (Pearl, 2019; Schölkopf et al., 2021).

More important still, especially with respect to democracy, is that normatively speaking the past must provide a useful template for the future. Change is a crucial feature of societies, especially the extension of rights and the participation of previously excluded groups. Over time, many societies strive to decrease discrimination and increase equality. In fact, many policies are consciously designed to break with past patterns of discrimination. AI-based predictions and classifications based on past patterns risk replicating systemic inequalities and even structural discrimination (Bolukbasi et al., 2016; Christian, 2020; S. Mitchell et al., 2021).

Problems that share these characteristics can be found in many areas, such as the digital economy, commerce, digitally mediated social interactions, robotics, and sports. AI has proven highly successful in these areas. But few problems in politics and in democracy more broadly share these characteristics. This limits the application of AI in society and, accordingly, its impact on democracy.

7.4.1. Machine readable

For the successful application of artificial intelligence to a given problem, there needs to be a machine-readable representation available; the problem, its constitutive features and outcomes of interest need to be documented in data. As we have seen in Chapter 3, digital technology has contributed to a steady growth in data and increased the areas of the world and social life captured by it. But, at the same time, we have encountered a series of challenges in the translation of the world into data. Challenges that can limit the uses and usefulness of AI.

There are various fields made accessible for AI-enabled systems through different forms of continuous data collection and digitalization. This includes the collection of data and measurements through sensors in digital devices, such as smart phones, smart cars, or dedicated measurement devices. It is also true for the collection of digital trace data documenting people's interactions with digital services, such as Google, amazon, Facebook, or X. Additionally traditional data sets are digitized and made accessible to machines. These data sources allow AI-enabled systems to support drivers, run smart devices or grids, to shape digital information environments, translate texts, or to deploy ads.

But the growing availability of digital or digitized information sometimes hides the fact, that many areas in social or political life do not lend themselves to digital representation. The use of AI-enabled systems in these arenas might therefore be – and potentially remain – limited. For an example, let's have a look at voting.

> Example: Data about voting and voters
>
> One foundational democratic activity is voting. Voting is the ultimate expression and mechanism of democratic self-rule, allowing people to choose the representatives they elect to be governed by. Naturally, it is an activity that parties want to learn about, predict, and influence. This is a problem that parties and their consultants try to solve through data, algorithms, and recently also AI.[21] To assess the power of AI in this arena, we have to first examine what kind of relevant information is available to parties.
> Countries vary with regard to the information they allow parties to access and collect about voters and voting behavior.[22] The country with the richest provision of voter data is probably the USA. States collect and make available information

about registered voters. Details vary between states, but official voter information can include name, address, gender, race, and information about whether they voted in a previous election or not. These information are highly valuable for political parties and campaigns in modeling who to contact.
The state provides campaigns with highly reliable data about crucial demographic characteristics as well as outcomes of interest – party support and election turnout. This provides modelers with a promising foundation to predict party support and election turnout for people, who they have only demographic information about and no information about outcomes of interest.
But the quality of these models depends on the available data. In his book *Hacking the Electorate*,[23] the political scientist Eitan Hersh analyzes the precision of data-driven targeting by Barack Obama's presidential campaigns in 2008 and 2012, probably still the campaigns with the most sophisticated use of data and models to date. He shows that the campaigns were only able to target likely Democratic voters relatively precisely in states where the official voter file provided data on party registration and election turnout. In states where this information was not available, the campaigns struggled to target respective voters successfully even with rich and varied information available on voters, such as data from commercial data vendors or social media companies. The success of data-driven targeting in the US therefor depends on crucial information being collected and provided by the state, not on varied but only tangentially connected information. Without the availability of these information, the use of AI-enabled systems will be limited here as well.[24]
This example shows that legislative and political choices shape what data about voting is available to modelers and that modelers cannot simply use other potentially connected information to reliably infer data on phenomena that are not measured by design. Unless the underlying legal conditions change and political actors choose to document and make accessible information about voters, the uses of AI to predict vote choices will remain limited.

We can expect the availability of machine readable data covering political processes to grow. For example, Sanders and Schneier (2021) present an instructive account of how this could come about. The authors present an interesting thought experiment on some of the opportunities AI provides for politics. One example they discuss is predicting the success of a bill based on known interventions by lobbyists or constituents. Their paper shows the preconditions but also the opportunities for the increasing application of artificial intelligence in politics. At the same time, the paper illustrates the considerable efforts necessary by many actors to bring this about.

[24]For overviews of the use of data and models in US campaigning see Hersh (2015), Issenberg (2012), Nickerson and Rogers (2014).
[24]For an overview of how different conditions impact the uses of data by campaigns and the views of campaigners on data see Dommett et al. (2024).
[24]See Hersh (2015).
[24]For the limits of data-driven approaches in other campaign contexts see Jungherr (2016a).

In the effort of making ever more areas of politics and social life machine readable, there quickly emerges an inherent tension between the interests of making these areas more predictable and concerns about making campaigning too predictable. Let's stay with voting to illustrate this tension.

There is a legitimate concern in allowing parties to contact voters. The competition of different political viewpoints and proposed actions manifest in parties is crucial for democracy. At the same time, we do not want parties to get too good at this. We do not want parties to be able to clandestinely only contact people that are easy to mobilize or persuade; we want them to openly compete with their arguments for broad support. We also do not want a sustained imbalance in the technical campaigning capabilities to emerge between parties, potentially predetermining elections. This tension between increasing opportunities provided by making new areas of political and social life machine readable and thereby available to artificial intelligence and larger societal values and rights extends to other areas of politics, as well. We therefor can expect the machine-readable representation of politics to remain limited by design, in consequence also limiting the uses of AI-enabled systems.

7.4.2. Abundance

One can think of artificial intelligence as lowering the costs of prediction. In their book *Prediction Machines*, the management and marketing scholars Ajay Agrawal, Joshua Gans, and Avi Goldfarb define prediction as:

> [...] the process of filling in missing information. Prediction takes information you have, often called "data," and uses it to generate information you don't have. In addition to generating information about the future, prediction can generate information about the present and the past.
>
> Agrawal et al. (2018/2022), p. 32.

For prediction to work, the predicted entity needs to happen often. Scarce events or outcomes cannot be reliably predicted, they can only be chanced. So, for artificial intelligence to matter in a field, outcomes of interest need to happen often and need to be documented by data.

For artificial intelligence to automatically provide analysts or academics with reliable predictions of an outcome of interest, it needs vast amounts of data covering the outcome variable of interest and its potential predictors. Take *Google* for example.[25]

[25] For more on Google see Levy (2011). For more on how Google and other digital media companies are using AI see Metz (2021).

Example: Google

In its search business, Google has information about millions of search queries coming in each minute from all over the world. Google also has the information on which displayed result which user clicked after entering specific terms, thereby finding out which identified search result appeared to be of relevance to the user initiating the search. Both, the outcome variable – clicking on a seemingly relevant link – and the input variable – search terms – are available to Google in abundance and therefor offer a fruitful object for AI-based predictions. By automatically identifying patterns in the past behavior of users – connecting for example their search terms, search history, or location with search results they subsequently clicked on – Google can predict which results are of high relevance for users in the future who exhibit similar behavioral patterns.

By correctly predicting which results a user is looking for when using specific terms or showing a specific behavior, Google can beat its competitors by providing users with the relevant information while sorting out the irrelevant. This is already a nice feature in search. But this capability develops its full commercial potential for the company in the display of ads, supposedly targeted to the interests or needs of people using the service. Through reliably predicting which ads to display to whom, Google has a powerful selling proposition to ad customers. At the same, time it does not lose its search users through overly annoying or irrelevant ad display. Of course, the example is prohibitively simple, but it illustrates the kinds of problems for which AI-based systems offer powerful solutions. Similar patterns hold for the display of ads by Google, Facebook, or Amazon, or the recommendation of products on Amazon or Netflix.

In contrast to these examples, many outcomes that are of interest in politics or democracy remain scarce – even in a big data world. Again, let's take voting behavior. In most democracies voting for a specific office takes place in evenly spaced temporal intervals. For example, most democratic countries vote for their heads of state every four years. This makes voting a sparse activity and therefor difficult to predict. While each and everyone one of us is using a search engine multiple times per day, we only vote every couple of years. Accordingly, vote choice is an outcome variable much scarcer than those for which machine prediction has proven stunningly useful.

While automatically predicting people's vote choice might be elusive, other electioneering tasks might turn out to be more promising.[26] For example, Barack Obama's presidential campaigns in 2008 and 2012 modeled the likelihood of people to donate a specific amount to the campaign after receiving an email asking for donations. Given the size of the campaigns' email lists reportedly running in the millions, the frequency of donation asks, the frequency of small donations, and the campaigns' ability to frequently run experiments, make this a task in electioneering well suited to the use of artificial intelligence.

[26] For instructive discussions of the limits and promises of predictive data analysis in election campaigns see Hersh (2015), Nickerson and Rogers (2014), Sides and Vavreck (2014).

While many areas in politics might not come with abundant outcomes, creative actors can reformulate specific tasks or elements in ways that lend themselves better to AI-enabled systems. These new approaches to well-known tasks, such as donation collection in campaigns, can open up politics to creative uses of AI and sometimes even shift the balance of power toward actors willing and able to engage in this reformulation.

7.4.3. Stable connections between variables

For data-driven inference, it is important that the relationship between predicting and predicted variables remains stable between training data and the time of deployment in the wild. This is a temporal problem: Does a phenomenon's future reliably resemble its past? Or, is the future a foreign country, where they do things differently? But, more generally, it is also a problem of theory-free inference of information without the provision of an underlying causal mechanism.

> Example: Google Flu Trends
>
> One famous example for the failure of data-driven prediction is *Google Flu Trends*.[27] Only a few years ago, Google Flu Trends was a popular example for the power of data-driven prediction. Google Flu Trends was an online service run by Google that used the occurrence of topical search terms in specific locations to predict local flu outbreaks. For a while, the service was surprisingly precise and quicker than official statistics. But only a few years in, the service's quality was found to deteriorate quickly.
>
> In a forensic account of the episode Lazer et al. (2014) identified a shift in the function of Google's search field as a likely culprit, breaking the previously identified link between search terms and the flu. By suggesting users search terms corresponding with their initial input, Google changed the behavior of users, which in turn negatively impacted the inference of missing information based on this input. Google changed the relationship of the information its models tried to predict and the information that was available to them.

Another problem lies in the exclusively data-driven inference of information. Especially in data-rich contexts, correlations between variables abound. These correlations could indicate an unobserved causal link that might remain stable over time. In this case, predicting missing information based on available information found to be correlated in the past is feasible. But a correlation might also be the outcome of a random fluctuation – present one moment, gone the next. In this case, prediction would be produce meaningless results. To know which correlation is meaningful and which is not, social science uses theories to provide testable hypotheses about why various indicators should be linked. This allows the careful modeling and testing of links between variables and their predictive power. Causal reasoning and causal inference attempt to determine

[27] For a detailed discussion of the failure of *Google Flu Trends* see Lazer et al. (2014).

what correlations can be seen as meaningful predictions and which probably are better ignored.[28]

Of course, this does not mean that artificial intelligence should only look for connections, its programmers thought of. This would be missing out on the very real opportunities of dats-driven discovery and inference through AI. Still, simply relying on connections identified by machines is just as limiting. Instead, people and AI need to meaningfully interact. This means that people have to critically interrogate the output and the process through which AI inferred information. Here, using causal reasoning provides an important reality check to automatically identified patterns through data-driven procedures.

7.4.4. Continuing past inequalities

An additional challenge to the use of artificial intelligence in broader societal contexts is the question of unwanted bias in the outcomes of data-driven models. One crucial element in societies is change, especially the extension of rights and the inclusion of different groups with regard to their participation in society and the workplace. Over time, many societies strive to decrease discrimination and increase equality. In this regard, future behavior toward people and options afforded to them should not resemble the past. In fact, many policies are consciously designed to break with past patterns of discrimination. The use of artificial intelligence and purely data-driven prediction – at least in its current form – has proven a challenge to these goals.

AI-based learnings are inherently conservative. By relying on patterns found in past data, AI will pursue tasks in ways that were successful in the past but might no longer be appropriate,[29] either due to unobserved shifts between inputs and outputs[30]: or due to a shift in values and norms making past learnings obsolete.[31] This makes AI-supported shapings conservative.

Using data documenting people's characteristics, behavior, and trajectories in the past to infer future behavior and trajectories risks replicating systemic inequalities and even structural discrimination.[32] For example, Bolukbasi et al. (2016) found that a prominent program underlying many services relying on automated natural language processing applications showed consistent evidence of gender bias. They showed that the model featured many biased associations. For example, when presented with the word-pair

[28] For the pitfalls of theory-free prediction see Jungherr et al. (2017), Jungherr (2019). For discussions of causal modeling see Imbens and Rubin (2015), Pearl (2009), Morgan and Winship (2015).

[29] See Vela et al. (2022).

[30] See Lazer et al. (2014).

[31] See Bender et al. (2021).

[32] For a foundational discussion of biases in computer systems see Friedman and Nissenbaum (1996). For a discussion of biases found in automatically trained language models see Caliskan et al. (2017). For critical accounts for the use of artificial intelligence in broad societal contexts in face of various biases see Ferguson (2017), Eubanks (2018). For attempts at mitigating inherent biases in machine learning see Barocas et al. (2023).

"father" and "doctor", the model completed the input "mother" with "nurse". Why did it do this? By examining the statistical relationships between vectors of words representing their co-occurrences in a large corpus of news articles, they found that the combination "mother" and "nurse" resembled statistically the pairing "father" and "doctor".

By relying on statistical relationships between words documenting the outcome of gender inequality of a society's past, the outcome of AI-enabled systems risks reinforcing said inequalities in the future. This is true even if a society consciously tries to intervene through policy designed to counter said biases and tries to establish more equal and less discriminatory behaviors and structures.

Other cases of accidental AI bias following a similar logic include the over-policing of areas traditionally strongly associated with recorded crime[33] or sensors and classification programs not recognizing women or members of racial minorities typically underrepresented in training data.[34] The growing uses of artificial intelligence in many areas – such as healthcare, policing, judicial sentencing, or the roll-out of social services – have raised awareness of this inherent limitation and associated potential dangers in the application of artificial intelligence.

7.5. Artificial intelligence and democracy

AI's recent successes and its broad deployment in many areas of social, economic, and political life have begun to raise questions regarding whether and how AI impacts democracy.[35] The idea and practice of democracy are highly contested concepts with competing accounts of great nuance. The associated discussions within political theory are highly productive and successful in identifying different normative, procedural, or structural features and consequences within our understanding of democracy.[36] Still, for the purposes of getting a handle on the impact of AI on democracy, we need to reduce this rich discussion to a few important – if sometimes contested – features of democracy.

We focus our discussion on four contact areas between AI and democracy at different analytical levels:

- At the *individual* level, AI impacts the conditions of self-rule and people's opportunities to exercise it.
- At the *group* level, AI impacts equality of rights among different groups of people in society.
- At the *institutional* level, AI impacts the perception of elections as a fair and open mechanism for channeling and managing political conflict.

[33]See Ferguson (2017)
[34]See Buolamwini and Gebru (2018).
[35]This section is strongly based on and slightly adapted from Jungherr (2023a).
[36]See Dahl (1998), Guttman (2007), Landemore (2012), Przeworski (2018), Tilly (2007).

- At the *systems* level, AI impacts competition between democratic and autocratic systems of government.

Level	Area of impact
individual	self-rule
group	equality
institutional	elections
system	competition between systems

7.6. Artificial intelligence and self-rule

One tenet of democracy is that governments should be chosen by those they will serve. Such self-rule is a normative idea about legitimizing the temporal power of rulers over the ruled and a practical idea that distributed decision making is superior to other more centralized forms of decision making or rule by experts.[37] AI impacts both the ability of people to achieve self-rule and the perceived superiority of distributed decision making over expert rule in complex social systems, highlighting potential limits to self-rule in several ways.

7.6.1. Shaping information environments

The legitimacy of self-rule is closely connected with the idea of people being able to make informed decisions for themselves and their communities. This depends at least in part on the information environment in which they are embedded.[38] AI affects these informational foundations of self-rule directly. This includes how people are exposed to and can access political information, can voice their views and concerns, and how these informational foundations potentially increase opportunities for manipulation.[39]

As we have already seen in Chapter 4, algorithmic shaping of digital information environments based on people's inferred information preferences or predicted behavioral responses has raised particularly strong concerns.[40] Key among these is that people will be exposed only to information with which they are likely to agree, thus losing sight of the other political side. Empirical findings suggest that these fears may be overblown.[41] In fact, in digital communication environments people may encounter more political information about the other side and that they disagree with than in other information

[37] See Dahl (1998), Landemore (2012), Landemore and Elster (2012), Schwartzberg (2015).
[38] See Jungherr and Schroeder (2022).
[39] See Jungherr and Schroeder (2023).
[40] See Kaye (2018).
[41] See Flaxman et al. (2016), Kitchens et al. (2020), Scharkow et al. (2020).

environments. This can be a problem, especially for political partisans, because it increases the salience of political conflict.[42] But the degree to which this mechanism is driven by AI or might even be lessened through specific algorithm design remains as of now unknown.

Going further, several authors have diagnosed various ill effects of digital communication environments on information quality and political discourse, some AI-driven and others independent of AI.[43] While clearly important, these diagnoses risk overestimating the quality of prior information environments and the role of information for people in their exercise of self-rule. In fact, critiques of the quality of media in democracies abounded well before digital media became prevalent.[44]

In addition, most people do not follow the news closely, do not hold strong political attitudes, and do not perform well when tested on their political knowledge.[45] They seem to rely on informational shortcuts or on social structures to exercise self-rule.[46] Hence, these mechanisms can also be expected to mediate the impact of AI-driven shaping of information environments. To assess AI's impact fully, research needs to consider not only information environments but must also look at whether and how AI affects the structural and social factors that mediate the impact of political information on self-rule.

It does not appear that AI-driven shaping of digital information environments inevitably leads to a deterioration of access to information necessary for people to exercise their right to self-rule. Nevertheless, there is much opaqueness in the way digital communication environments are shaped. The greater the role of these environments in democracies, the greater the need for assessability of the role of AI in their shaping.[47] We also need regular external audits of the effects of AI on the information visible on online platforms, especially the nature and kind of information that is algorithmically promoted or muted.

7.6.2. Economics of news

AI might also come to indirectly impact the creation and provision of relevant political information by changing the economic conditions of news production. For one, recent successes in the development of transformer models suggest that AI might soon be used by media providers to automatically generate text, image, or video content. This might lead to an acceleration of existing trends toward automated content generation in news organizations.[48] This puts pressure on journalists who might see routine tasks shift

[42] See Settle (2018).
[43] See Bennett and Livingston (2021).
[44] See Keane (2013).
[45] See Converse (1964), Lupia and McCubbins (1998), Prior (2007), Zaller (1992).
[46] See Achen and Bartels (2016), Kuklinski and Quirk (2000), Lodge and Taber (2013), Popkin (1991).
[47] See Jungherr and Schroeder (2023).
[48] See Diakopoulos (2019).

toward AI-enabled systems but also on news organizations who might face a new set of ultra low-cost competitors who specialize on automatically generated news content. This potentially increases pressure on journalists' salaries as well as the audiences and profits of news companies, intensifying existing pressures on news as a business.[49]

Additionally AI reconfigures the way news and political information are accessed by the public. Search engines like Bing and Google are experimenting with large language models (LLM) to provide users with automatically generated content in reaction to search queries instead of links to content provided by news and information providers. This limits monetization opportunities for small or middle sized media organizations without strong brand identity and loyalty, that in the past could generate traffic based on query-based referrals from search engines or social networking sites. These new limitations on monetization opportunities might lead to a decline in the coverage of politics, or number of news organizations. This in turn would limit the total amount and diversity of information available to people in order to develop informed decisions. This will hit political outsiders and challengers the hardest who rely on smaller information providers for coverage. This decline in monetization opportunities of news will thus likely lead to a strengthening of existing institutions, media brands, and associated power relations.[50]

Additionally, public perceptions of digital communication environments being dominated by AI-generated content – some of it correct, some of it actively misleading, some of it accidentally misleading – might contribute among parts of the population to an increased valuation of select news organizations, whose process of news production and quality insurance they have come to trust. These news brands might thus find themselves strengthened through an increase of AI-generated content in open communication environments or in the coverage by cost-cutting competitors. Of course, this expectation only holds if these news brands are seen as providing added value over AI-generated content.

It is also important to remember that this AI-driven turn to specific news brands is only likely to hold for audience members who engage with news and politics demanding accurate information and those interested in politics. This will likely be socio-economically well-resourced and politically engaged people.[51] Others might feel fine with free or automatically generated content. This is likely to reinforce an informational divide between politically interested and disinterested audiences that already has grown following the switch from a low-choice mass media environment to high-choice digital communication environments.[52] In countries without strong public broadcasters, like the US, this divide will also run along economic lines, allowing those able to pay for news to access high quality, curated, and quality checked information, while leaving those not able (or willing) to pay to the noisy, (partially) automated, and contested free digital information

[49]See R. K. Nielsen (2020).
[50]See Jungherr and Schroeder (2023).
[51]See Prior (2018), Schlozman et al. (2018).
[52]See Prior (2017).

environment. Over time, this might mean that socio-economic divides decide (or are seen to decide) over the ability of people to come to informed political decisions.

7.6.3. Speech

AI, though, does not only impact access to information; it also affects the expression of opinions, interests, and concerns in digital communication environments. With digital communication environments becoming increasingly areas for the expression of voice, surfacing of concerns, and construction of political identities, this is an important element in AI's shaping of the conditions for self-rule.

The perceived ability of AI to classify content has put it at the forefront of the fight against harmful digital speech and misinformation. AI is used broadly by tech companies to classify user content to stop it from publication or flag it for moderation.[53] Details of the applied procedures, their successes, and error rates are opaque to outsiders, making it difficult to assess the broadness of AI's uses and its effects on speech. This is problematic: harmful speech and misinformation are both difficult categories for classification. Neither category is objective nor stable and both require interpretation as meaning shifts across contexts and time. This makes them difficult to identify with automated data-driven AI and risks suppression of legitimate political speech.

Additionally, the technical workings of AI also impact the type of speech becoming visible in AI-shaped spaces. By learning typical patterns within a given set of cases, AI will lean toward averages. For AI-enabled shaping and summarizing of speech or political positions, this will favor common positions, concerns, and expressions. Outsiders and minority positions, concerns, and expressions will in unadjusted AI-shaped communication environments submerged and remain invisible. AI would thus negatively impact the ability of a society to become visible to itself in the public arena, lower democracies' information processing capacities, and strengthen the status quo.[54]

Still, there are few alternatives to AI-based moderation given the pure volume of content being published in digital communication environments,[55] which makes it important to gain a better understanding of AI-based moderation's workings and effects. Accordingly, AI-based moderation needs assessability provided by platforms and external audits to ensure its proper workings.

AI-based moderation, however, is not only a risk. Scholars and commentators have long pointed to the limits of large-scale political deliberation imposed through inefficiencies in information distribution, surfacing of preferences, and coordination of people. AI may improve on some of these inefficiencies by predicting individual preferences, classifying information, and shaping information flows.[56] This in turn might open up opportunities

[53] See Douek (2021); Kaye (2018).
[54] See Jungherr and Schroeder (2023).
[55] See Douek (2021).
[56] See Landemore (2024).

for new deliberative and participatory formats in democracies, thereby strengthening and vitalizing democracy.

It is important to remain aware of both the risks and the opportunity AI provides for moderating speech and surfacing concerns in digital communication environments. AI can contribute to creative solutions to some of the technical challenges underlying successful self-rule. But if it is to do so, we need to know more about its actual uses, effects, and risks. This demands for greater transparency from digital platforms and continued vigilance and attention from civil society.

7.6.4. Manipulation

Artificial intelligence could also negatively impact individual informational autonomy by predicting the reactions of people to communicative interventions. This could allow professional communicators to reach people in exactly the right way to shift opinions and behavior. Sanders and Schneier (2021) present a thought experiment that illustrates how lobbyists might use AI to predict the likelihood of success of bills they introduce to legislators. While still far from realization, their example shows interested parties employing AI to increase the resources available to them and potentially targeting interventions aimed at influencing people to behave in ways beneficial to those same parties. AI can also be used to generate messages aimed at persuading people, with early working papers indicating interventions designed by LLMs to have persuasive appeal.[57] Similarly, LLMs are currently used by academics and campaign professionals to simulate reactions and attitudes by prototypical voters for message testing and research. [58] But to be sure, currently the precision and validity of these approaches are still in doubt.

Fears also exist regarding people encountering targeted communicative interventions in digital communication environments. By predicting how people might react to an advertisement, digital consultancies could use AI to tailor interventions to influence people. The English consultancy firm Cambridge Analytica, which claimed to be able to predict which piece of information displayed on Facebook was necessary to get people to behave in ways beneficial to its electoral clients, provided a first taste of this problem. While the company's claims have been debunked,[59] the episode speaks to the perception of AI's power to manipulate people at will, as well as the willingness of journalists and the public to accept widely exaggerated claims about the power of digitally enabled manipulation irrespective of contradicting evidence.

Recent advances in transformer models have opened new avenues for potential manipulation through the automated production of text or images.[60] There are legitimate uses of these models, as well as nefarious ones. For instance, they facilitate the automated

[57]See Bai et al. (2023).
[58]See Bisbee et al. (2023), Horton (2023), J. Kim and Lee (2023).
[59]See Jungherr et al. (2020), 124–130.
[60]See T. B. Brown et al. (2020), Ramesh et al. (2022).

generation of content based on raw information or event data, as found in sports coverage or the stock market.[61] This is largely unproblematic, since AI translates information from one form of representation – such as numerical or event data – into another – such as a narrative news article.

More problematic are cases in which AI does not simply translate one representation of information into another, but generates content based on prompts and past patterns. Examples include text or image responses to textual prompts in form of questions or instructions. AI has no commitment to the truth of an argument or observation; it is only imitating their likeness as found in past data. Today's AI is committed only to the representation of the world, an object, or an argument available to it, not to the world, object, or argument as such.[62] Thus, AI output taken at face value cannot be trusted because it is not necessarily true, only plausible.

More problematic still is the chance that future AI could be used to produce fake information at scale. This could take the form of targeted fakes aimed at misleading people, or flooding information environments with masses of unreliable or misleading AI-generated content. This would dilute information environments, making it more difficult for people to access crucial information and/or making information appear untrustworthy. But while evocative, the mere opportunity of creating more unreliable information might not directly translate into these automatically generate disinformation reaching audiences or persuading them.[63]

Also, somewhat counterintuitively, a mass-seeding of automated misinformation might also contribute to the strengthening of professional news and information curation discussed above. When the prevalence of unreliable or misleading information in digital communication environments becomes evident, the premium for reliable information rises. Accordingly, professional, reliable, and impartial news sources might see a reversal of fortune compared to their economic and ideational challenges of the last twenty years. This way, automated misinformation at scale might turn out to strengthen intermediary institutions that provide information in democracies.

It is important to note that these uses of AI are still projected, and may not come to pass given limits of the underlying technology, the development of efficient countermeasures, and/or the persistence of mediating structures that limit the effects of information overall. But in light of recent technological advances, these uses have come to feature strongly in the public imagination and demand for critical reflection by social and computer scientists.

[61] See Diakopoulos (2019).
[62] See Smith (2019).
[63] For a critical look at why generative AI might not lead to a new age of powerful disinformation see Simon et al. (2023).

7.6.5. Expert rule

Support for self-rule is also closely connected with the assessment of expert rule being limited in complex social systems. Expertise is important, but has limited predictive power in complex societies, and the decentralized decision making and preference surfacing of self-rule, while imperfect, are seen as superior for settling on collectively binding decisions.[64] As we have seen in Chapter 3, the growing availability of data in ever more domains, coupled with new analytical opportunities offered by AI, have raised hopes for new predictive capabilities in complex societies.[65] AI not only highlights the weaknesses of people making political decisions, but also increases the power of experts.

AI brings new opportunities in the modeling and prediction of societal, economic, ecological, and geopolitical trends, promising to provide experts with predictions of people's behavior in reaction to regulatory or governance interventions. While the actual quality of these approaches is still open to question, they have strong rhetorical and legitimizing power. They increase the power of experts, who – sometimes actually and sometimes rhetorically – rely on AI-supported models to ground their advice on how societies should act considering major societal challenges. This apparent increase in the power of experts to guide societies in responding to challenges can reduce the option space available for democratic decision making, shifting the question from whether people *can* to whether they *should* decide for themselves. In this, AI could induce a transition from self-rule to expert rule and thereby weaken democracy.

7.6.6. Power of technology companies

AI also increases the power of companies over the public and even over states. While the theoretical breakthroughs in the current wave of AI began at universities, it is companies that lead in their practical application, further development, and broad rollout.[66] Over time, the power to innovate and critically interrogate AI may shift from public to commercial actors, weakening AI oversight and regulation by democratically legitimated institutions. These challenges can be clearly seen in attempts by both the US and EU at getting to grips with regulating AI development and uses.[67]

There is also the issue of economic and political power. AI has allowed companies such as Google and Amazon to dominate multiple economic sectors.[68] Governments have also begun to rely on AI-based service providers to support executive functions such as policing and security. The result is a growing government dependence on AI companies and an opaque transfer of knowledge from governments to these service providers. Add to this power over AI-enabled information flows and governance over political speech, and

[64]See Dahl (1998), Lindblom (2001).
[65]See Kitchin (2014).
[66]See Ahmed et al. (2023), Metz (2021).
[67]See Kretschmer et al. (2023).
[68]See Bessen (2022), Brynjolfsson et al. (2023).

AI companies hold central positions in democracies, potentially negatively influencing the abilities of people for self-rule. This shows the importance of effective government and civil society oversight of companies that provide AI and those that employ AI to ensure that the foundations of meaningful self-rule hold as societies begin to rely more on AI-supported systems.

7.7. Artificial intelligence and equality

Democracy depends on people having equal rights to participation and representation.[69] While this ideal is imperfectly realized and strongly contested in practice,[70] democracies are in an ongoing struggle to extend rights to formerly excluded groups. AI's reliance on data documenting the past risks subverting this process and instead continuing past discrimination into the future, thereby weakening democracy.

By predicting how people will behave under various circumstances based on observations from the past, AI differentiates between people based on criteria represented in data points. This risks reinforcing existing biases in society and even porting socially, legally, and politically discontinued discriminatory patterns into the present and future.[71] This makes continuous observation and auditing of AI implementation crucial. The associated problems resemble those, we already have encountered in the discussion of algorithms in Chapter 4.

People's visibility to AI depends on their past representation in data. AI has trouble recognizing those who belong to groups underrepresented in the data used to train it. For example, minorities not traditionally represented in data sets will remain invisible to computer vision,[72] and historically underrepresented groups will not be associated with specific jobs and thereby risk discrimination in AI-assisted job procedures.[73]

This general pattern is highly relevant to democracy: for example, systematic invisibility of specific groups means they would be diminished in any AI-based representation of the body politic and in predictions about its behavior, interests, attitudes, and grievances. Accordingly, already disenfranchised people could risk further disenfranchisement and discrimination in the roll out of government services, the development of policy agendas based on digitally mediated preferences and voice, or face heightened persecution from the state security apparatus.

AI also makes some people *more* visible. Historically marginalized groups will be overrepresented in crime records, negatively impacting group members in AI-based approaches

69 See Dahl (1998).
70 See A. Phillips (2021), Young (2002).
71 See Eubanks (2018), Mayson (2019), Mehrabi et al. (2022), S. Mitchell et al. (2021), Obermeyer et al. (2019).
72 See Buolamwini and Gebru (2018).
73 See Caliskan et al. (2017).

to policing or sentencing.[74] In countries like the US, where voting rights are withheld for felons to varying degrees depending on state jurisdiction, systematic biases in AI-supported policing and sentencing might over time come to systematically bias the electorate against historically disenfranchised groups.[75] Additionally, AI-based approaches can also have a profound effect on electoral redistricting.[76] In sum, AI could lead to a reinforcement of structural inequality and discrimination by continuing patterns found in historical data even if a society is trying to enact more equal, less discriminatory practices.

Extrapolating from this, we can expect subsequent AI-based representations of public opinion, the body politic, and AI-assisted redistricting to be biased against groups marginalized in the past. Different degrees of visibility to AI could increase the democratic influence of some groups and decrease that of others. For instance, AI might contribute to an increase of resources for the already privileged by making their voices, interests, attitudes, concerns, and grievances more visible and accessible to decision makers. AI might use the preferences of visible groups in predictions about political trends and policy impact while ignoring those of less visible groups.

AI can also have adverse effects on the labor market. While in principle firms could invest in automation to allow workers to pursue new tasks and thereby increase the value of their labor, it appears that firms do so mostly to lower their own labor costs by substituting AI for human labor-based tasks.[77] This lowers workers' bargaining power and income by substituting labor for capital, which in turn threatens to increase economic inequality and weaken workers' collective bargaining power. Consequently, this could also lower workers' political influence and representation.[78]

What type of labor is affected by AI-based technological progress, though, is uncertain. Automation traditionally substitutes for routine human tasks and thus affects mostly low-skilled workers.[79] But subsequent waves of AI innovation have shown that routine tasks underly many professions, including white-collar and knowledge ones long perceived as being immune to automation. The impact of AI in changing the political fortunes of workers might thus concern larger groups in the economy than traditional forms of automation. This can already be seen in the current discussion about the impact of large language models and generative AI on the creative and software industries, which until now seemed to be exempt from the dangers of automation-driven job replacement. These emerging fault-lines became evident in the Hollywood writers' strike from 2023, in which screenwriters demanded contractual protection against studio uses of AI for writing tasks. And the actors' strike of the same year, asking for control about the use of actors' likeness by AI models.[80]

[74] See Chouldechova (2017), Christian (2020), Ferguson (2017).
[75] See Aviram et al. (2017).
[76] See Cho and Cain (2020)
[77] See Acemoglu and Restrepo (2019).
[78] See Acemoglu (2024), Gallego and Kurer (2022).
[79] See Acemoglu and Restrepo (2022b), C. B. Frey (2019).
[80] See Wilkinson (2023).

At the same time, AI can help aging societies complete substitutable work tasks and concentrate the shrinking labor force on currently non-substitutable tasks, thereby maintaining productivity levels in the face of growing demographic pressures in several developed economies.[81] But realizing AI's economic potential for societies means ensuring that respective gains are broadly shared and do not only benefit a narrow elite. Especially with prosperity gains from digital technology, this link of shared prosperity gains seems to be broken. This raises concerns as to whether elites manage to capture respective AI-enabled gains while most people only face automation-driven economic risks.[82] This would increase inequality in society and weaken democracy. This potentially dangerous development puts the specifics of AI's implementation and its public and regulatory oversight into focus.

AI clearly touches on equality within democracies. Inequalities might arise in the allocation of options and state services using AI-based systems, people's visibility and representation within AI-based systems, and the provision or withdrawal of economic opportunities for people whose job tasks can be replaced with AI. These are, therefore, important areas for further interrogation and, if necessary, regulatory intervention.

7.8. Artificial intelligence and elections

Democracies rely on elections, which channel and manage political conflict by providing factions the opportunity to gain power within an institutional framework. But for this to work, each faction must feel the very real opportunity to win power in future elections. Otherwise, why bother with elections? Why not choose a different way to gain power?[83] In the words of Adam Przeworski, democracy is a system of *organized uncertainty*:[84]

> Actors know what is possible, since the possible outcomes are entailed by the institutional framework; they know what is likely to happen, because the probability of particular outcomes is determined jointly by the institutional framework and the resources that the different political forces bring to the competition. What they do not know is which particular outcome will occur. They know what winning or losing can mean to them, and they know how likely they are to win or lose, but they do not know if they will lose or win.
>
> Przeworski (1991), p. 12–13.

AI applications promise to offset this organized uncertainty of who will lose and who will win elections. Ideas of being able to correctly predict elections or the behavior of voters go back to the early days of the computer age, the 1950s. In his 1955 short

[81]See Acemoglu and Restrepo (2022a).
[82]See Acemoglu and Johnson (2023).
[83]See Przeworski (2018).
[84]Przeworski (1991), p. 13.

story *Franchise*,[85] science fiction author Isaac Asimov has the computer system *Multivac* calculate election results based on the input of one specifically chosen person. In the late 1950s a set of eclectic scientists and engineers founded the *Simulmatics Corporation*[86] with the goal to predict human behavior and support political campaigns based on computer models. Among their client roster was later President of the United States John F. Kennedy. More recently, the Presidential campaigns of Barack Obama used data-driven models to predict the likely behavior of voters.[^obama predicts]

While we can discuss the degree to which each of these examples qualifies as AI, in each we encounter the idea of actors being able to use available information to infer unknown outcomes. This could happen on an individual level where available information might include attitudes revealed in a survey, attitudes inferred based on behavior or choices – such as buying a specific brand of consumer good, driving a specific car, or donating money – , or documented behavior – such as turning out to vote – to infer future behavior – such as vote choice, the decision to turn out to vote, or the willingness to donate money. Alternatively, this could also happen on the system level, taking aggregate information – such as the state of the economy or general approval ratings – to predict the outcome of an election without modeling individual behavior. In other words, artificial intelligence might contribute to the lowering of uncertainty about who will win or lose in democratic elections.

But these approaches remain limited in the prediction of individual voters' behavior. While the voting behavior of committed partisans can be predicted with some probability[87] – at least in two-party systems –, predicting the behavior of people who are only weakly involved with politics is much harder. People do not always vote, and when they do the context can vary greatly. Their vote choices are – as we have encountered – for the most part not available to modelers, making predicting voting behavior automatically a problem for which AI is not well suited. The uncertainty of election victories will thus remain alive for the foreseeable future. But campaigns can develop other relevant data-driven models of elections, such as someone's probability of voting or donating money,[88] which could give campaigns a competitive advantage.[89] At the same time, the subsequent fate of Obama's successor as Democratic nominee Hillary Clinton in 2016 and the victory by Republican nominee Donald Trump show that even such a seemingly decisive predictive advantage as developed by Obama and his team is hard to maintain over time. Also, any such advantage is likely fleeting, given the broad availability of AI-based tools and campaign organizations learning from others' successes and failures.[90]

Firms and governments might also seek to use AI to predict election outcomes or the

[85] Asimov (1955).

[86] For the story of the *Simulmatics Corporation* see Lepore (2020). But note how Lepore is much better at picking holes at past visions of predictability than today's. [^obama predicts]: For a background on the efforts at prediction by the Obama campaigns 2008 and 2012 see Issenberg (2012).

[87] See Hersh (2015), Nickerson and Rogers (2014).

[88] See Hersh (2015), Issenberg (2012), Nickerson and Rogers (2014).

[89] For more on different uses of AI in election campaigns see Foos (2024), Jungherr et al. (2024).

[90] See Kreiss (2016).

electorate's mood swings and possibly intervene. Campaigns or parties might simply have too little data, not enough computing power, or simply not enough talent to capitalize on the opportunities of AI. But this is not necessarily true for large companies developing AI in other areas or governments able to use the services of these companies or to coerce them. But these corporate or government efforts are limited by the same challenges raised above. Still, the public impression of this capability might be enough to undermine and delegitimize elections and give election losers a pretext to challenge results rather than conceding.

Cambridge Analytica's supposed role in the United Kingdom's Brexit vote and the 2016 U.S. presidential election previewed some of the challenges ahead. While there is little indication that data-based psychological targeting was widely used or had sizable effects, these episodes still loom large in the public imagination as an example of AI's perceived power in election manipulation.[91] We can expect widespread AI use in economic, political, and social life to shift people's expectations of its uses and abuses in electioneering, irrespective of its actual uses or inherent limitations.

Overall, AI's direct impact on elections seems limited given the relative scarcity of the predicted activity – voting. While indirect effects are possible through potential opportunities for competitive differentiation, it is doubtful that this can translate into a consistent, systemic shift of power, given the broad availability of AI tools. More likely is the indirect impact mentioned above: that by transposing expectations regarding AI's supposed powers from industry and science to politics, the public may come to believe that AI is actually able to offset the *organized uncertainty* of democratic elections. This alone could weaken public trust in elections and acceptance of election results. It is thus important to keep organized uncertainty alive in the face of AI, not weaken it through irresponsible and fantastical speculation.

7.9. Artificial intelligence and the autocratic competition to democracies

AI also affects the relationship between democracy and other systems of governance, such as autocracy, which some have argued has an advantage in the development and deployment of AI.[92] Leaving aside deeper normative considerations for a moment, on a purely functional dimension, democracies are often seen to be superior to autocracies or dictatorships, due to their superior performance as information aggregators and processors. AI might conceivably offset this functional superiority.

Governments all over the world face a shared challenge: They must decide on a course of action best suited to society or their interests based on expected outcomes. This means collecting and feeding available information about the state of society or the

[91] Jungherr et al. (2020), 124–130.
[92] See Filgueiras (2022), Lee (2018).

consequences of specific actions into implicit or explicit models of how the world works and to adjust one's actions accordingly. Here, democracies are seen to have a competitive advantage over autocracies or dictatorships.[93] By allowing free expression, having a free and inquisitive press, competition between factions and even within governmental groups democracies have structural mechanisms in place that surface information about society, the actions of bureaucracies, or the impact of policies so that political actors can react and reinforce or countermand a course of action.

Autocracies and dictatorships do not have the same mechanisms in place. By controlling speech and the media, they restrict information flows considerably, leaving governments often in the dark with regard to local situations, the preferences of the public, the behavior or corruption in their bureaucracies, and ultimately the consequences of the policies pursued by them. Democracy has thus been seen to allow for a better information acquisition and processing performance than more centralized approaches of governance, such as autocracies and dictatorships. The underlying mechanism is akin to the better performance of the market system compared to centralized planning with regard to economic outcomes. There now has been some debate over whether AI allows autocracies or dictatorships to overcome this disadvantage.

In democracies, companies and governments face limits to AI deployment or pervasive data collection about people's behavior. In autocracies, they have more leeway. A close connection between the state and firms developing and deploying AI in autocracies creates an environment of permissive privacy regulation that provides developers and modelers with vast troves of data, allowing them to refine AI-enabled models of human behavior. Add centrally allocated resources and training of large numbers of AI-savvy engineers and managers, and some expect the result to be a considerable competitive advantage in developing, deploying, and profiting from AI-supported systems. This may allow for asymmetric developmental progress in AI, state capacity, economic benefits, and potentially even military prowess favoring autocracies over democracies.[94]

To be sure, the operating word is *may*. The differential powers of AI for autocracies is currently speculated about, not proven. Accordingly, there are strong critiques emerging of this position.[95] Still, while far from settled, this question is a valuable one to discuss. So let's turn to the place where these arguments have been spelled out most explicitly: China.

China is seen by some to provide a context more suited for the large scale deployment

[93]For fictional illustration of the information challenge to autocracies and dictatorships see Spufford (2010). For academic discussions of the information challenge to autocracies and dictatorships see Wintrobe (1998), Kuran (1995), Gregory and Markevich (2002), Wallace (2022). On the informational strengths of democracy see Lindblom (1965), Ober (2008). For market systems and information see Lindblom (2001). For the challenges of control in organizations and government organizations see Little (2020).

[94]See Filgueiras (2022), Lee (2018).

[95]For strong critiques of the supposed autocratic AI advantage see Farrell et al. (2022), E. Yang and Roberts (2023).

of AI than Western democracies.[96] Reasons for this include the indiscriminate and large-scale collection of data and the state's willingness to support companies – which in any case remain under strong state control – to pursue AI in force. Over time China would be therefore the country in which broad data collection and AI-based prediction of people's behavior become pervasive and ubiquitous. This would allow the authoritarian government to use AI as an information gathering and processing mechanism potentially offsetting its challenges in information gathering by not having free expression or a free press.

While you might not state your preferences or dissatisfaction with the government in an official survey, running automated text analysis on your online chat protocols might unearth your true state of mind. While bureaucracies might keep you in the dark about the true state of the economy under your policies, an automated analysis of local payment patterns might indicate a surprising up-pick or slowdown in the economy.

Large scale data collection and predictive analytics through AI might thus help autocrats and dictators to learn as much about their people and the effects of their policies as democracies, perhaps even more as they can use these tools more pervasively and efficiently. Whether this is a realistic expectation is contested, but the implementation of China's *Social Credit System* is widely seen as an attempt by China's government to capitalize on this potential.

Example: China's Social Credit System

China's *Social Credit System* is currently the most ambitious society-wide scoring system.[97] The goal is to automatically track people and their behavior over multiple societal and economic domains in order to establish their degree of compliance with rules. Behavior conforming with domain-specific rules is rewarded with positive scores. Conversely, behavior conflicting with rules is punished by taking points away. Examples for negative behavior vary but reportedly can include hearing loud music or eating on public transport, crossing red lights, or cheating in online games. People with an overall positive score can find themselves fast tracked with regard to credit applications, while those with negative scores can find themselves excluded from public services. Scores are assigned automatically based on conforming or deviant behavior captured by sensors, such as public surveillance cameras or other digital devices. This system provides the state with a vast pool of data on its people, potentially allowing it to infer opinions, predict unrest, learn about the effects of its policies, and shape individual behavior.

But before we become too enamored – or scared – by this vision of total surveillance and control, we should not forget the tendency of people to learn about and undermine attempts at surveillance and government control.[98] Although, currently

[96] For the supposed special fit of China to AI innovation see Lee (2018). For China's use of data and AI in social planning and control see Pan (2020), Ding et al. (2020). For the development of AI in autocracies see Filgueiras (2022).

stories about advertised features and supposed successes of China's Social Credit System abound, it is best to remain skeptical as to its actual workings and precision.

Still, the Social Credit System illustrates the point made by Lee (2018). He expects that the autocracy China is currently better placed than Western democracies in capitalizing on the potentials of AI. Sooner or later this might translate from pure engineering power into soft or cultural power. Once, AI-enabled Chinese digital platforms will provide users with better entertainment or commercial opportunities than Western platforms, they will capture their user base and assume their central role in cultural, commercial, or political life in countries well beyond China. An early example for this is the Chinese video platform *TikTok*.[99]

Western democracies might be able to shrug off that personalized music clips or funny videos are distributed through digital infrastructures hosted in China rather than the US. But AI-driven power shifts can also happen in other, more crucial areas.

AI might provide people living in autocracies with greater cultural, economic, or health-related opportunities.[100] There are those who might see these benefits as a worthy tradeoff with some individual freedoms, leading to strengthened public support for autocracies and state control. Differential opportunities in realizing the potentials of AI might thus reinforce tendencies already evident in countries facing economic, cultural, or security crises. Particularly in times when democracies increasingly find themselves internally challenged with respect to the opportunities with which they provide people, these potentials of AI that asymmetrically favor autocracies represent an obvious challenge to democracies – if realized.

Going further, AI is a technology increasingly discussed in military and security circles.[101] While its normative role and functional potential in these areas are heavily contested, the growing concerns in these circles point to the broad perception that AI could facilitate democracies falling behind autocracies.

Over time, differential trajectories in the development and deployment of AI in democracies and autocracies may emerge. If the assumption holds that autocracies share a greater affinity with AI and can profit more from it than democracies, AI could lead to a power shift between systems and thus weaken democracy.

[98] On China's *Social Credit System* see Liang et al. (2018), Creemers (2018), Brussee (2023). For the export of Chinese and Russian approaches to social control through AI see Síthigh and Siems (2019), Weber (2019), Morgus (2019).

[98] For tactics to avoid state surveillance or governance attempts more broadly see J. C. Scott (2009).

[99] For the challenge of countries to rely on digital infrastructures developed and maintained in other countries see Jungherr and Schroeder (2022).

[100] For scenarios of the future impact on AI on culture, health, and innovation see Diamandis and Kotler (2020), Lee and Quifan (2021).

[101] For the role of AI in international security and conflict see Goldfarb and Lindsay (2022), Buchanan and Imbrie (2022).

7.10. Artificial intelligence and democracy: The road ahead

While many AI applications still lie in the future, we already start to see AI's impact on democracy. True, many of AI's future uses and effects remain speculative. But it is important that political science engages early on with AI and helps observe, evaluate and guide its implementation. This includes AI's uses in politics, government, and its regulation and governance. This chapter has provided a set of areas and problems for the impact of AI on democracy. The broad contact areas of self-rule, equality, elections, and competition between systems can serve as topical clusters for future work. At the same time, future work will provide a more fine-grained account and advance theories that explain use and effect patterns.

Social scientists need to consider AI in their analysis of features, dangers, and potentials of contemporary democracy. In doing so, they need to reflect on the inner workings and domain specific effects of the underlying technology. At the same time, computer scientists and engineers need to consider the consequences for democracy in AI development and deployment. This means focusing not only on the analysis of the technology itself, but also to consider its embeddedness in economic, political, and social structures that mediate its effects for better or worse. This makes the analysis of AI's impact on democracy an important area of future interdisciplinary work.

The quality of the analysis of AI's effects on democracy depends on specificity regarding the type of AI, how it functions, the conditions for its successful deployment, and the aspect(s) of democracy it touches. Narratives about an unspecified, super-powered AGI and its supposed impact on society may make for stimulating reading, but offer little for the analysis of actual effects on society or democracy. In fact, interested parties can use the discussion of AGI and supposed extinction-level-event-dangers as a smokescreen, distracting public and regulatory attention from more mundane but crucial questions of AI governance, regulation, and the societal distribution of AI-driven gains and risks.

Although AI is often discussed as a danger or threat, it may also provide opportunities to offset some of the contemporary challenges to democracy. Thinking openly about the application of AI in democracy could provide some relief from these challenges. Conscious design choices and transparent audits can help ameliorate dysfunctions and uncover biases.

In general, AI's impact depends on implementation and oversight by the public and regulators. For this, companies, regulators, and society need to be explicit and transparent about what economic, political, or societal goals they want to achieve using AI and how its specific workings can propel or inhibit this pursuit. By nature, this discussion combines normative, mechanistic, and technological arguments and considerations. It is important not to be sidetracked by grandiose, but ultimately imaginary, visions of an AGI, but instead focus on specific instances of narrow AI, their inner workings, uses in specific areas of interest, and effects. This includes the discussion of both potentially positive as well as negative effects.

AI is unlikely to impact many aspects of democracy directly. Nevertheless, public discourse is likely to continue to focus on threats, manipulation, and expected power shifts. This discourse and these expectations have the potential to shape public attitudes toward AI and its impact on democracy strongly, irrespective of their factual basis. Perceived effects can matter more strongly than actual effects. Researchers have a responsibility not to fan the flames of discourse with speculation, but instead remain focused on AI's actual workings and effects.

There are many promising avenues for future scientific work on the impact of AI on democracy. Here, it is important to combine insights from different fields. Purely technological accounts risk overestimating AI's impact in social systems, given their boundedness and the role of social structures. Accounts coming purely from the social sciences risk misrepresenting the actual workings of existing AI and thereby misattributing its consequences.

The impact of AI on democracy is already progressing. Its systematic, interdisciplinary examination and discussion needs to proceed as well. This means it is high time for political scientists to add their voice and perspective to the ongoing debate about the impact of AI on democracy.

7.11. Further Reading

For a helpful account of the workings of artificial intelligence see M. Mitchell (2019).

For an interesting discussion of artificial intelligence and its dependence on representations of the world see Smith (2019).

For a deeper discussion of artificial intelligence and democracy see Jungherr (2023a).

For a more extensive discussion of artificial intelligence and the public arena see Jungherr and Schroeder (2023).

7.12. Review questions

1. Please define the term *artificial general intelligence* (AGI).
2. Please define the term *narrow AI*.
3. Please discuss how according to Smith (2019) *reckoning* versus *judgment* are different forms of intelligence.
4. Please discuss which four features of democracy are contact areas for the impact by AI on democracy.

5. Please discuss how AI might and might not lower the “organized uncertainty” of democracy.
6. Please discuss how AI might and might not impact elections.
7. Please discuss how AI might or might not negatively impact the informational autonomy of people.
8. Please discuss how AI might strengthen expert rule.
9. Please discuss how AI might come to increase inequality in societies.
10. Please discuss the impact of AI on the competition between democratic and autocratic systems of government.

References

Abernathy, P. M. (2020). *News deserts and ghost newspapers: Will local news survive*. University of North Carolina Press.

Acemoglu, D. (2024). Harms of AI. In J. B. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford handbook of AI governance* (pp. 660–706). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.65

Acemoglu, D., & Johnson, S. (2023). *Power and progress: Our thousand-year struggle over technology and prosperity*. PublicAffairs.

Acemoglu, D., & Restrepo, P. (2019). Artificial intelligence, automation and work. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 197–236). The University of Chicago Press.

Acemoglu, D., & Restrepo, P. (2022a). Demographics and automation. *The Review of Economic Studies*, *89*(1), 1–44. https://doi.org/10.1093/restud/rdab031

Acemoglu, D., & Restrepo, P. (2022b). Tasks, automation, and the rise in U.S. wage inequality. *Econometrica*, *90*(5), 1973–2016. https://doi.org/10.3982/ECTA19815

Acerbi, A. (2020). *Cultural evolution in the digital age*. Oxford University Press. https://doi.org/10.1093/oso/9780198835943.001.0001

Achen, C. H., & Bartels, L. M. (2016). *Democracy for realists: Why elections do not produce responsive government*. Princeton University Press. https://doi.org/10.2307/j.ctvc7770q

Agarwal, S. D., & Barthel, M. L. (2015). The friendly barbarians: Professional norms and work routines of online journalists in the united states. *Journalism*, *16*(3), 376–391. https://doi.org/10.1177/1464884913511565

Agrawal, A., Gans, J., & Goldfarb, A. (2022). *Prediction machines: The simple economics of artificial intelligence* (Updated and Expanded). Harvard Business Review Press. (Original work published 2018)

Ahmed, N., Wahed, M., & Thompson, N. C. (2023). The growing influence of industry in AI research. *Science*, *379*(6635), 884–886. https://doi.org/10.1126/science.ade2420

Alam, M. R., Reaz, M. B. I., & Ali, M. A. M. (2012). A review of smart homes – past, present, and future. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *42*(6), 1190–1203. https://doi.org/10.1109/TSMCC.2012.2189204

Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, *6*(14), eaay3539. https://doi.org/10.1126/sciadv.aay3539

Almqvist, M. F. (2016). Piracy and the politics of social media. *Social Sciences*, *5*(3), 41. https://doi.org/10.3390/socsci5030041

Alpaydin, E. (2021). *Machine learning* (Revised and updated). The MIT Press. (Original work published 2016)

An, J., Kwak, H., Posegga, O., & Jungherr, A. (2019). Political discussions in homogeneous and cross-cutting communication spaces. In J. Pfeffer, C. Budak, Y.-R. Lin, & F. Morstatter (Eds.), *ICWSM 2019: Proceedings of the thirteenth international AAAI conference on web and social media* (pp. 68–79). Association for the Advancement of Artificial Intelligence (AAAI).

Anderson, B. (2016). *Imagined communities: Reflections on the origin and spread of nationalism* (Revised). Verso.

Angrist, J. D., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, *24*(2), 3–30. https://doi.org/10.1257/jep.24.2.3

Anstead, N., & O'Loughlin, B. (2015). Social media analysis and public opinion: The 2010 UK General Election. *Journal of Computer-Mediated Communication*, *20*(2), 204–220. https://doi.org/10.1111/jcc4.12102

Asimov, I. (1955). Franchise. *If: Worlds of Science Fiction*, (August), 2–15. https://ia801300.us.archive.org/25/items/1955-08_IF/1955-08_IF.pdf

Auletta, K. (2009). *Googled: The end of the world as we know it*. The Penguin Press.

Auletta, K. (2018). *Frenemies: The epic disruption of the ad business (and everything else)*. Penguin Press.

Auxier, B., & Anderson, M. (2021). *Social media use in 2021*. Pew Research Center. https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/

Aviram, H., Bragg, A., & Lewis, C. (2017). Felon disenfranchisement. *Annual Review of Law and Social Science*, *13*, 295–311. https://doi.org/10.1146/annurev-lawsocsci-110316-113558

Badue, C., Guidolini, R., Carneiro, R. V., Azevedo, P., Cardoso, V. B., Forechi, A., Jesus, L., Berriel, R., Paixão, T. M., Mutz, F., de Paula Veronese, L., Oliveira-Santos, T., & Souza, A. F. D. (2021). Self-driving cars: A survey. *Expert Systems with Applications*, *165*(1), 113816. https://doi.org/10.1016/j.eswa.2020.113816

Bai, H., Voelkel, J. G., Eichstaedt, J. C., & Willer, R. (2023). Artificial intelligence can persuade humans on political issues. *OSF Preprints*. https://doi.org/10.31219/osf.io/stakv

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *PNAS: Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Ball, M. (2022). *The Metaverse: And how it will revolutionize everything*. Liveright Publishing Corporation.

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis*, *23*(1), 76–91. https://doi.org/10.1093/pan/mpu011

Barberá, P., Boydstun, A. E., Linn, S., McMahon, R., & Nagler, J. (2021). Automated text classification of news articles: A practical guide. *Political Analysis*, *29*(1), 19–42. https://doi.org/10.1017/pan.2020.8

Barbrook, R., & Cameron, A. (1995). The californian ideology. *Mute*, *1*(3). https://www.metamute.org/editorial/articles/californian-ideology

Barlow, J. P. (1996). *A declaration of the independence of cyberspace*. EFF: Electronic Frontier Foundation. https://www.eff.org/cyberspace-independence

Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. The MIT Press.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, *104*, 671–732. https://doi.org/10.15779/Z38BG31

Baron, J. (2023). *Thinking and deciding* (5th ed.). Cambridge University Press. https://doi.org/10.1017/9781009263672 (Original work published 1988)

Barron, A. T. J., Huang, J., Spang, R. L., & DeDeo, S. (2018). Individuals, institutions, and innovation in the debates of the French Revolution. *PNAS: Proceedings of the National Academy of Sciences*, *115*(18), 4607–4612. https://doi.org/10.1073/pnas.1717729115

Baumgartner, F. R., Boef, S. D., & Boydstun, A. E. (2008). *The decline of the death penalty and the discovery of innocence*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790638

Beauchamp, N. (2017). Predicting and interpolating state-level polling using Twitter textual data. *American Journal of Political Science*, *61*(2), 490–503. https://doi.org/10.1111/ajps.12274

Bechar, A., & Vigneault, C. (2016). Agricultural robots for field operations: Concepts and components. *Biosystems Engineering*, *149*, 94–111. https://doi.org/10.1016/j.biosystemseng.2016.06.014

Beieler, J., Brandt, P. T., Halterman, A., Schrodt, P. A., & Simpson, E. M. (2016). Generating political event data in near real time. In R. M. Alvarez (Ed.), *Computational social science: Discovery and prediction* (pp. 98–120). Cambridge University Press. https://doi.org/10.1017/CBO9781316257340.005

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Facct '21: Proceedings of the 2021 acm conference on fairness, accountability, and transparency* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Bengfort, B., Bilbro, R., & Ojeda, T. (2018). *Applied text analysis with python: Enabling language aware data products with machine learning*. O'Reilly Media.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.

Benkler, Y. (2011). *The penguin and the leviathan: How cooperation triumphs over self-interest*. Crown Publishing Group.

Bennett, W. L. (1990). Towards a theory of press-state relations in the US. *Journal of Communication*, *40*(2), 103–125. https://doi.org/10.1111/j.1460-2466.1990.tb02265.x

Bennett, W. L. (2005). Beyond pseudoevents: Election news as reality tv. *American Behavioral Scientist*, *49*(3), 364–378. https://doi.org/10.1177/0002764205280919

Bennett, W. L. (2016). *News: The politics of illusion* (10th ed.). The University of Chicago Press. (Original work published 1983)

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2), 122–139. https://doi.org/10.1177/0267323118760317

Bennett, W. L., & Livingston, S. (Eds.). (2021). *The disinformation age: Politics, technology, and disruptive communication in the United States*. Cambridge University Press. https://doi.org/10.1017/9781108914628

Bennett, W. L., & Manheim, J. B. (2006). The one-step flow of communication. *The ANNALS of the American Academy of Political and Social Science*, *608*(1), 213–232. https://doi.org/10.1177/0002716206292266

Bennett, W. L., & Segerberg, A. (2013). *The logic of connective action: Digital media and the personalization of contentious politics*. Cambridge University Press. https://doi.org/10.1017/CBO9781139198752

Benoit, K. (2020). Text as data: An overview. In L. Cuirini & R. Franzese (Eds.), *The sage handbook of research methods in political science and international relations* (pp. 461–497). SAGE Publications. https://doi.org/10.4135/9781526486387.n29

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 1–4. https://doi.org/10.21105/joss.00774

Bergen, M. (2022). *Like, comment, subscribe: Inside YouTube's chaotic rise to world domination*. Viking.

Bessen, J. (2022). *New goliaths: How corporations use software to dominate industries, kill innovation, and undermine regulation*. Yale University Press.

Bi, Z., & Wang, X. (2020). *Computer aided design and manufacturing*. John Wiley & Sons.

Bimber, B. (2003). *Information and American democracy: Technology in the evolution of political power*. Cambridge University Press. https://doi.org/10.1017/CBO9780511615573

Bisbee, J., Clinton, J., Dorff, C., Kenkel, B., & Larson, J. (2023). Artificially precise extremism: How internet-trained LLMs exaggerate our differences. *SocArxiv*. https://doi.org/10.31235/osf.io/5ecfa

Black, F. (1971a). Toward a fully automated stock exchange, part i. *Financial Analysts Journal*, *27*(4), 28–35. https://doi.org/10.2469/faj.v27.n4.28

Black, F. (1971b). Toward a fully automated stock exchange, part ii. *Financial Analysts Journal*, *27*(6), 24–28. https://doi.org/10.2469/faj.v27.n6.24

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of Science. *The Annals of Applied Statistics*, *1*(1), 17–35. https://doi.org/10.1214/07-AOAS114

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Boast, R. (2017). *The machine in the ghost: Digitality and its consequences*. Reaktion Books.

Boczkowski, P. J., & Papacharissi, Z. A. (2018). *Trump and the media* (P. J. Boczkowski & Z. A. Papacharissi, Eds.). The MIT Press.

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In D. D. Lee, U. von Luxburg, R. Garnett, M. Sugiyama, & I. Guyon (Eds.), *Nips'16: Proceedings of the 30th international conference on neural information processing systems* (pp. 4356–4364). Curran Associates Inc. https://doi.org/10.5555/3157382.3157584

Bond, B., & Exley, Z. (2016). *Rules for revolutionaries: How big organizing can change everything*. Chelsea Green Publishing.

Bor, A., & Petersen, M. B. (2022). The psychology of online political hostility: A comprehensive, cross-national test of the mismatch hypothesis. *American Political Science Review*, *116*(1), 1–18. https://doi.org/10.1017/S0003055421000885

Borges, J. L. (1975). On exactitude in science (N. T. di Giovanni, Trans.). In, *A universal history of infamy* (p. 131). Penguin Books. (Original work published 1946). Trans. of Del rigor en la ciencia. (1946). *Los Anales de Buenos Aires*, *1*(3), 53

Borgesius, F., Trilling, D., Möller, J., Bodó, B., de Vreese, C., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, *5*(1), 1–16. https://doi.org/10.14763/2016.1.401

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Bourdieu, P. (1990). Social space and symbolic power. In *In other words: Essays towards a reflexive sociology* (pp. 123–139). Stanford University Press.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. The MIT Press.

Bradford, A. (2020). *The brussels effect: How the european union rules the world*. Oxford University Press. https://doi.org/10.1093/oso/9780190088583.001.0001

Brandt, M., Tucker, C. J., Kariryaa, A., Rasmussen, K., Abel, C., Small, J., Chave, J., Rasmussen, L. V., Hiernaux, P., Diouf, A. A., Kergoat, L., Mertz, O., Igel, C., Gieseke, F., Schöning, J., Li, S., Melocik, K., Meyer, J., Sino, S., … Fensholt, R. (2020). An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature*, *587*(7832), 78–82. https://doi.org/10.1038/s41586-020-2824-5

Brayne, S. (2021). *Predict and surveil: Data, discretion, and the future of policing*. Oxford University Press. https://doi.org/10.1093/oso/9780190684099.001.0001

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brennen, J. S., & Kreiss, D. (2016). Digitalization. In *The international encyclopedia of communication theory and philosophy* (pp. 1–11). John Wiley & Sons. https://doi.org/10.1002/9781118766804.wbiect111

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., & Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics*, *9*(1), 247–274. https://doi.org/10.1214/14-AOAS788

Brown, E., & Farrell, M. (2021). *The cult of we: Wework, adam neumann, and the great startup delusion*. Crown Publishing Group.

Brown, M. A., & Zhou, S. (2019). Smart-grid policies: An international review. In P. D. Lund, J. Byrne, R. Haas, & D. Flynn (Eds.), *Advances in energy systems: The large-scale renewable energy integration challenge* (pp. 127–147). John Wiley & Sons. https://doi.org/10.1002/9781119508311.ch8

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems* (pp. 1877–1901, Vol. 33). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Brussee, V. (2023). *Social credit: The warring states of China's emerging data empire*. Palgrave Macmillan. https://doi.org/10.1007/978-981-99-2189-8

Bryant, J., & Oliver, M. B. (Eds.). (2009). *Media effects: Advances in theory and research* (3rd ed.). Routledge.

Brynjolfsson, E., Jin, W., & Wang, X. (2023). Information technology, firm size, and industrial concentration. *NBER Working Paper*, (31065). https://doi.org/10.3386/w31065

Brynjolfsson, E., & McAfee, A. (2016). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.

Buchanan, B., & Imbrie, A. (2022). *The new fire: War, peace, and democracy in the age of ai*. The MIT Press.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler & C. Wilson (Eds.), *Proceedings of the 1st conference on fairness, accountability and transparency* (pp. 77–91, Vol. 81). Proceedings of Machine Learning Research (PMLR). https://proceedings.mlr.press/v81/buolamwini18a.html

Burton, J. W., Cruz, N., & Hahn, U. (2021). Reconsidering evidence of moral contagion in online social networks. *Nature Human Behavior*, *5*, 1629–1635. https://doi.org/10.1038/s41562-021-01133-5

Buyalskaya, A., Gallo, M., & Camerer, C. F. (2021). The golden age of social science. *PNAS: Proceedings of the National Academy of Sciences*, *118*(5), 1–11. https://doi.org/10.1073/pnas.2002923118

Calhoun, C., Gaonkar, D. P., & Taylor, C. (2022). *Degenerations of democracy*. Harvard University Press.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Camargo, C. Q., & Simon, F. M. (2022). Mis- and disinformation studies are too big to fail: Six suggestions for the field's future. *Harvard Kennedy School Misinformation Review*, *3*(5), 1–9. https://doi.org/10.37016/mr-2020-106

Cardella, L., Hao, J., Kalcheva, I., & Ma, Y.-Y. (2014). Computerization of the equity, foreign exchange, derivatives, and fixed-income markets. *The Financial Review*, *49*(2), 231–243. https://doi.org/10.1111/fire.12033

Carlson, M., Robinson, S., & Lewis, S. C. (2021). *News after Trump: Journalism's crisis of relevance in a changed media culture*. Oxford University Press. https://doi.org/10.1093/oso/9780197550342.001.0001

Carreyou, J. (2018). *Bad blood: Secrets and lies in a silicon valley*. Alfred A. Knopf.

Casero-Ripollés, A., Feenstra, R. A., & Tormey, S. (2016). Old and new media logics in an electoral campaign: The case of podemos and the two-way street mediatization of politics. *The International Journal of Press/Politics*, *21*(3), 378–397. https://doi.org/10.1177/1940161216645340

Castells, M. (2001). *The internet galaxy: Reflections on the internet, business and society*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199255771.001.0001

Castells, M. (2013). *Communication power* (2nd ed.). Oxford University Press. (Original work published 2009)

Chabert, J.-L. (Ed.). (1999). *A history of algorithms: From the pebble to the microchip* (C. Weeks, Trans.). Springer. https://doi.org/10.1007/978-3-642-18192-4 (Original work published 1994). Trans. of Chabert, J.-L. (1994). *Histoire d'algorithmes, du caillou à la puce*. Éditions Belin

Chafkin, M. (2021). *The contrarian: Peter Thiel and Silicon Valley's pursuit of power*. Penguin Books.

Chen, L., Zhang, C., & Wilson, C. (2013). Tweeting under pressure: Analyzing trending topics and evolving word choice on Sina Weibo. In M. Muthukrishnan, A. El Abbadi, & B. Krishnamurthy (Eds.), *Cosn '13: Proceedings of the first acm conference on online social networks* (pp. 89–100). ACM. https://doi.org/10.1145/2512938.2512940

Chia, A., Keogh, B., Leorke, D., & Nicoll, B. (2020). Platformisation in game development. *Internet Policy Review*, *9*(4), 1–28. https://doi.org/10.14763/2020.4.1515

Cho, W. K. T., & Cain, B. E. (2020). Human-centered redistricting automation in the age of ai. *Science*, *369*(6508), 1179–1181. https://doi.org/10.1126/science.abd1879

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, *5*(2), 153–163. https://doi.org/10.1089/big.2016.0047

Chow, S., Liver, S., & Nelson, A. (2018). Streamlining bioactive molecular discovery through integration and automation. *Nature Reviews Chemistry*, *2*, 174–183. https://doi.org/10.1038/s41570-018-0025-7

Christensen, G., Freese, J., & Miguel, E. (2019). *Transparent and reproducible social science research: How to do open science*. University of California Press.

Christian, B. (2020). *The alignment problem: Machine learning and human values*. W. W. Norton & Company.

Christians, C. G., Glasser, T. L., McQuail, D., Nordenstreng, K., & White, R. A. (2009). *Normative theories of the media: Journalism in democratic societies*. University of Illinois Press.

Christin, A. (2020). *Metrics at work: Journalism and the contested meaning of algorithms*. Princeton University Press.

Cirone, A., & Spirling, A. (2021). Turning history into data: Data collection, measurement, and inference in HPE. *Journal of Historical Political Economy*, *1*(1), 127–154. https://doi.org/10.1561/115.00000005

Citron, D. K., & Pasquale, F. (2014). The scored society: Due process fo automated predictions. *Washington Law Review*, *89*(1), 1–33.

Clark, D. (2016). *Alibaba: The house that Jack Ma built*. Harper Collins.

Cogburn, D. L., & Espinoza-Vasquez, F. K. (2011). From networked nominee to networked nation: Examining the impact of web 2.0 and social media on political participation and civic engagement in the 2008 obama campaign. *Journal of Political Marketing*, *10*(1-2), 189–213. https://doi.org/10.1080/15377857.2011.540224

Cohen, I. G., Gostin, L. O., & Weitzner, D. J. (2020). Digital smartphone tracking for COVID-19: Public health and civil liberties in tension. *JAMA*, *323*(23), 2371–2372. https://doi.org/10.1001/jama.2020.8570

Cohen, J. E. (2012). *Configuring the networked self: Law, code, and the play of everyday practice*. Yale University Press.

Cohen, J. E. (2013). What privacy is for. *Harvard Law Review*, *126*(7), 1904–1933.

Cohen, J. E. (2019). *Between truth and power: The legal constructions of informational capitalism*. Oxford University Press. https://doi.org/10.1093/oso/9780190246693.001.0001

Converse, P. (1964). The nature of belief systems in mass publics. In D. E. Apter (Ed.), *Ideology and discontent* (pp. 206–261). Free Press.

Coppock, A., Hill, S. J., & Vavreck, L. (2020). The small effects of political advertising are small regardless of context, message, sender, or receiver: Evidence from 59 real-time randomized experiments. *Science Advances*, *6*(36), 1–6. https://doi.org/10.1126/sciadv.abc4046

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2022). *Introduction to algorithms* (4th ed.). The MIT Press. (Original work published 1990)

Cotter, K. (2019). Playing the visibility game: How digital influencers and algorithms negotiate influence on Instagram. *New Media & Society*, *21*(4), 895–913. https://doi.org/10.1177/1461444818815684

Crain, M. (2021). *Profit over privacy: How surveillance advertising conquered the internet*. University of Minnesota Press.

Creemers, R. (2017). Cyber China: Upgrading propaganda, public opinion work and social management for the twenty-first century. *Journal of Contemporary China*, *26*(103), 85–100. https://doi.org/10.1080/10670564.2016.1206281

Creemers, R. (2018). China's social credit system: An evolving practice of control. *Social Science Research Network*. https://doi.org/10.2139/ssrn.3175792

Crosby, A. W. (1996). *The measure of reality: Quantification in western europe, 1250–1600*. Cambridge University Press. https://doi.org/10.1017/CBO9781107050518

Dahl, R. A. (1998). *On democracy*. Yale University Press.

Daston, L. (2022). *Rules: A short history of what we live by*. Princeton University Press.

DellaPosta, D., Shi, Y., & Macy, M. (2015). Why do Liberals drink Lattes? *American Journal of Sociology*, *120*(5), 1473–1511. https://doi.org/10.1086/681254

Dennis, J. (2020). A party within a party posing as a movement? Momentum as a movement faction. *Journal of Information Technology & Politics*, *17*(2), 97–113. https://doi.org/10.1080/19331681.2019.1702608

Deringer, W. (2018). *Calculated values: Finance, politics, and the quantitative age*. Harvard University Press.

Deseriis, M. (2020). Digital movement parties: A comparative analysis of the technopolitical cultures and the participation platforms of the movimento 5 stelle and the piratenpartei. *Information, Communication & Society*, *23*(12), 1770–1786. https://doi.org/10.1080/1369118X.2019.1631375

Dewey, J. (1927). *The public and its problems*. Holt Publishers.

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.

Diamandis, P. H., & Kotler, S. (2020). *The future is faster than you think: How converging technologies are transforming business, industries, and our lives*. Simon & Schuster.

Ding, J., Chun, A., Liu, Y.-L., Han, E., Lewis, D., Gal, D., & Creemers, R. (2020). *The ai powered state: China's approach to public sector innovation*. Nesta. https://apo.org.au/sites/default/files/resource-files/2020-05/apo-nid305076.pdf

Doerr, J. (2018). *Measure what matters: How Google, Bono, and the Gates Foundation rock the world with OKRs*. Portfolio/Penguin.

Dommett, K. (2020). Roadblocks to interactive digital adoption? elite perspectives of party practices in the united kingdom. *Party Politics*, *26*(2), 165–175. https://doi.org/10.1177/1354068818761196

Dommett, K., Kefford, G., & Kruschinski, S. (2024). *Data-driven campaigning and political parties: Five advanced democracies compared*. Oxford University Press. https://doi.org/10.1093/oso/9780197570227.001.0001

Dommett, K., Temple, L., & Seyd, P. (2021). Dynamics of intra-party organisation in the digital age: A grassroots analysis of digital adoption. *Parliamentary Affairs*, *74*(2), 378–397. https://doi.org/10.1093/pa/gsaa007

Donovan, J., Dreyfuss, E., & Friedberg, B. (2022). *Meme wars: The untold story of the online battles upending democracy in America*. Bloomsbury Publishing.

Douek, E. (2021). Governing online speech: From ‘posts-as-trumps’ to proportionality and probability. *Columbia Law Review*, *121*(3), 759–834. https://doi.org/10.2139/ssrn.3679607

Drezner, D. W., Farrell, H., & Newman, A. L. (2021). *The uses and abuses of weaponized interdependence* (D. W. Drezner, H. Farrell, & A. L. Newman, Eds.). Brookings Institution Press.

Duff, B. E., & Meisner, C. (2023). Platform governance at the margins: Social media creators’ experiences with algorithmic (in)visibility. *Media, Culture & Society*, *45*(2), 285–304. https://doi.org/.o0r.g1/107.711/0717/603146434347327212111 1923

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press. https://doi.org/10.1017/CBO9780511761942

Eldridge, S. A. (2018). *Online journalism from the periphery: Interloper media and the journalistic field*. Routledge. https://doi.org/10.4324/9781315671413

Elster, J. (2015). *Explaining social behavior: More nuts and bolts for the social sciences* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781107763111 (Original work published 2007)

Enos, R. D., & Hersh, E. D. (2015). Party activists as campaign advertisers: The ground campaign as a principal-agent problem. *American Political Science Review*, *109*(2), 252–278. https://doi.org/10.1017/S0003055415000064

Entman, R. M. (2004). *Projections of power: Framing news, public opinion, and U.S. foreign policy*. The University of Chicago Press.

Epstein, B., & Broxmeyer, J. D. (2020). The (surprisingly interesting) story of e-mail in the 2016 presidential election. *Journal of Information Technology & Politics*, *17*(3), 232–248. https://doi.org/10.1080/19331681.2020.1755762

Epstein, E. J. (2017). *How america lost its secrets: Edward snowden, the man and the theft*. Alfred A. Knopf.

Epstein, J. M. (2006). *Generative social science: Studies in agent-based computational modeling*. Princeton University Press.

Erie, M. S., & Streinz, T. (2021). The beijing effect: China’s digital silk road as transnational data governance. *Journal of International Law and Politics*, *54*(1), 1–92.

Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, *113*(1), 1–40. https://doi.org/10.1086/517897

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin’s Press.

Evans, B. (2020). News by the ton: 75 years of US advertising. *Benedict Evans*. https://www.ben-evans.com/benedictevans/2020/6/14/75-years-of-us-advertising

Evans, D. S., & Schmalensee, R. (2016). *Matchmakers: The new economics of multisided platforms*. Harvard Business School Publishing.

FAIR, M., Bakhtin, A., Brown, N., Dinan, E., Farina, G., Flaherty, C., Fried, D., Goff, A., Gray, J., Hu, H., Jacob, A. P., Komeili, M., Konath, K., Kwon, M., Lerer, A., Lewis, M., Miller, A. H., Mitts, S., Renduchintala, A., … Zijlstra, M. (2022).

Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science*, *378*(6624), 1067–1074. https://doi.org/10.1126/science.ade9097

Fannin, R. A. (2019). *Tech titans of china: How china's tech sector is challenging the world by innovating faster, working harder & going global.* Nicholas Brealey Publishing.

Farrell, H., & Newman, A. L. (2019a). *Of privacy and power: The transatlantic struggle over freedom and security.* Princeton University Press.

Farrell, H., & Newman, A. L. (2019b). Weaponized interdependence: How global economic networks shape state coercion. *International Security*, *44*(1), 42–79. https://doi.org/10.1162/isec_a_00351

Farrell, H., Newman, A. L., & Wallace, J. (2022). Spirals of delusion: How AI distorts decision-making and makes dictators more dangerous. *Foreign Affairs*, *101*(5), 168–181.

Ferguson, A. G. (2017). *The rise of big data policing: Surveillance, race, and the future of law enforcement.* New York University Press.

Ferree, M. M., Gamson, W. A., Gerhards, J., & Rucht, D. (2002a). Four models of the public sphere in modern democracies. *Theory and Society*, *31*(3), 289–324. https://doi.org/10.1023/A:1016284431021

Ferree, M. M., Gamson, W. A., Gerhards, J., & Rucht, D. (2002b). *Shaping abortion discourse: Democracy and the public sphere in germany and the united states.* Cambridge University Press. https://doi.org/10.1017/CBO9780511613685

Filgueiras, F. (2022). The politics of AI: Democracy and authoritarianism in developing countries. *Journal of Information Technology & Politics*, *19*(4), 449–464. https://doi.org/10.1080/19331681.2021.2016543

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, *80*(1), 298–320. https://doi.org/10.1093/poq/nfw006

Floreano, D., & Wood, R. J. (2015). Science, technology and the future of small autonomous drones. *Nature*, *521*, 460–466. https://doi.org/10.1038/nature14542

Flyvbjerk, B. (2001). *Making social science matter: Why social inquiry fails and how it can succeed again.* Cambridge University Press. https://doi.org/10.1017/CBO9780511810503

Foos, F. (2024). The use of AI by election campaigns. *OSF Preprints*. https://doi.org/10.31219/osf.io/zm2r6

Foucault, M. (1994). *The order of things: An archaeology of the human sciences* (A. Sheridan, Trans.). Vintage Books. (Original work published 1966). Trans. of *Les mots et les choses: Une archéologie des sciences humaines.* (1966). Gallimard

Fraser, N. (1990). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text*, (25/26), 56–80. https://doi.org/10.2307/466240

Frenkel, S., & Kang, C. (2021). *An ugly truth: Inside facebook's battle for domination.* Harper.

Frey, C. B. (2019). *Technology trap: Capital, labor, and power in the age of automation.* Princeton University Press.

Frey, M. (2021). *Netflix recommends: Algorithms, film choice, and the history of taste*. University of California Press.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, *14*(3), 330–347. https://doi.org/10.1145/230538.230561

Frier, S. (2020). *No filter: The inside story of Instagram*. Simon & Schuster.

Gaisbauer, F., Pournaki, A., Banisch, S., & Olbrich, E. (2021). Ideological differences in engagement in public debate on twitter. *PLoS One*, *16*(3), e0249241. https://doi.org/10.1371/journal.pone.0249241

Gallego, A., & Kurer, T. (2022). Automation, digitalization, and artificial intelligence in the workplace: Implications for political behavior. *Annual Review of Political Science*, *25*, 463–484. https://doi.org/10.1146/annurev-polisci-051120-104535

Galloway, S. (2017). *The four: The hidden dna of amazon, apple, facebook, and google*. Portfolio/Penguin.

Gellman, B. (2020). *Dark mirror: Edward Snowden and the surveillance state*. Bodley Head.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–574. https://doi.org/10.1257/jel.20181020

Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, *126*(4), 1799–1839. https://doi.org/10.1093/qje/qjr044

Gerbaudo, P. (2019). *The digital party: Political organisation and online democracy*. Pluto Press.

Gerring, J. (2012). *Social science methodology: A unified framework* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139022224 (Original work published 2001)

Gessen, M. (2020). Why are some journalists afraid of "moral clarity"? *The New Yorker*. https://www.newyorker.com/news/our-columnists/why-are-some-journalists-afraid-of-moral-clarity

Gessler, T., & Hunger, S. (2022). How the refugee crisis and radical right parties shape party competition on immigration. *Political Science Research and Methods*, *10*(3), 524–544. https://doi.org/10.1017/psrm.2021.64

Gieryn, T. F. (1999). *Cultural boundaries of science: Credibility on the line*. University of Chicago Press.

Gigerenzer, G. (2018). The bias bias in behavioral economics. *Review of Behavioral Economics*, *5*(3–4), 303–336. https://doi.org/10.1561/105.00000092

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gilardi, F., Baumgartner, L., Dermont, C., Donnay, K., Gessler, T., Kubli, M., Leemann, L., & Müller, S. (2022). Building research infrastructures to study digital technology and politics: Lessons from Switzerland. *PS: Political Science & Politics*, *55*(2), 354–359. https://doi.org/10.1017/S1049096521000895

Gilardi, F., Gessler, T., Kubli, M., & Müller, S. (2021). Social media and political agenda setting. *Political Communication*, *39*(1), 39–60. https://doi.org/10.1080/10584609.2021.1910390

Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, *40*, 129–152. https://doi.org/10.1146/annurev-soc-071913-043145

Goldfarb, A., & Lindsay, J. R. (2022). Prediction and judgment: Why artificial intelligence increases the importance of humans in war. *International Security*, *46*(3), 7–50. https://doi.org/10.1162/isec_a_00425

Golebiewski, M., & boyd, d. (2019). *Data voids: Where missing data can easily be exploited*. Data & Society. https://datasociety.net/wp-content/uploads/2019/11/Data-Voids-2.0-Final.pdf

González-Bailón, S. (2017). *Decoding the social world: Data science and the unintended consequences of communication*. The MIT Press.

González-Bailón, S., Borge-Holthoefer, J., Rivero, A., & Moreno, Y. (2011). The dynamics of protest recruitment through an online network. *Nature Scientific Reports*, *1*(197). https://doi.org/10.1038/srep00197

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.

Goodhart, D. (2017). *The road to somewhere: The new tribes shaping british politics*. C. Hurst & Co.

Gorwa, R. (2019a). The platform governance triangle: Conceptualising the informal regulation of online content. *Internet Policy Review*, *8*(2), 1–22. https://doi.org/10.14763/2019.2.1407

Gorwa, R. (2019b). What is platform governance? *Information, Communication & Society*, *22*(6), 854–871. https://doi.org/10.1080/1369118X.2019.1573914

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, *7*(1), 1–15. https://doi.org/10.1177/2053951719897945

Granovetter, M. (2017). *Society and economy: Framework and principles*. The Belknap Press of Harvard University.

Green, J. (2017). *Devil's bargain: Steve Bannon, Donald Trump, and the storming of the Presidency*. Penguin Press.

Gregory, P. R., & Markevich, A. (2002). Creating Soviet industry: The house that Stalin built. *Slavic Review*, *61*(4), 787–814. https://doi.org/10.2307/3090390

Griffiths, J. (2019). *The great firewall of China: How to build and control an alternative version of the internet*. ZED Books.

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.

Guess, A., Nagler, J., & Tucker, J. A. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), 1–8. https://doi.org/10.1126/sciadv.aau4586

Guess, A., Nyhan, B., Lyons, B., & Reifler, J. (2018). *Avoiding the echo chamber about echo chambers: Why selective exposure to like-minded political news is less prevalent than you think*. Knight Foundation. %7Bhttps://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/133/original/Topos_KF_White-Paper_Nyhan_V1.pdf%7D

Guess, A. M., Nyhan, B., & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behavior*, *4*, 472–480. https://doi.org/10.1038/s41562-020-0833-x

Gurri, M. (2018). *The revolt of the public and the crisis of authority in the new millennium* (2nd ed.). Stripe Press.

Guttman, A. (2007). Democracy. In R. E. Goodin, P. Pettit, & T. Pogge (Eds.), *A companion to contemporary political philosophy* (2nd ed., pp. 521–532). Blackwell Publishing. https://doi.org/10.1002/9781405177245.ch25

Habermas, J. (1981). *Theorie des kommunikativen Handelns*. Suhrkamp.

Habermas, J. (1990). *Strukturwandel der Öffentlichkeit: Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft*. Suhrkamp. (Original work published 1962)

Hafner, K., & Lyon, M. (1996). *Where wizards stay up late: The origins of the internet*. Simon & Schuster.

Halberstam, D. (1972). *The best and the brightest*. Random House.

Han, H. (2014). *How organizations develop activists: Civic associations and leadership in the 21st century*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199336760.001.0001

Hand, D. J. (2004). *Measurement theory and practice: The world through quantification*. Wiley.

Hand, D. J. (2007). *Information generation: How data rule our world*. Oneworld.

Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. The MIT Press.

Hanna, R. N., & Linden, L. L. (2012). Discrimination in grading. *American Economic Journal: Economic Policy*, *4*(4), 146–168. https://doi.org/10.1257/pol.4.4.146

Hanson, R. (2016). *The age of Em: Work, love, and life when robots rule the earth*. Oxford University Press. https://doi.org/10.1093/oso/9780198754626.001.0001

Harcourt, B. E. (2006). *Against prediction: Profiling, policing, and punishing in an actuarial age*. University of Chicago Press.

Hardt, H. (2001). *Social theories of the press: Constituents of communication research, 1840s to 1920s* (2nd ed.). Rowman & Littlefield.

Harwell, D., & Lorenz, T. (2023). Israel-Gaza war sparks debate over TikTok's role in setting public opinion. *Washington Post*. https://www.washingtonpost.com/technology/2023/11/02/tiktok-israel-hamas-video-brainwash/

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7 (Original work published 2001)

Herbst, S. (1993). *Numbered voices: How opinion polling has shaped american politics*. The University of Chicago Press.

Hersh, E. D. (2015). *Hacking the electorate: How campaigns perceive voters*. Cambridge University Press. https://doi.org/10.1017/CBO9781316212783

Hersh, E. D. (2018, May 16). Cambridge Analytica and the future of data privacy: Written testimony of Eitan Hersh. In *Hearing before the United States Senate Committee on the Judiciary*. United States Senate. https://www.judiciary.senate.gov/imo/media/doc/05-16-18%20Hersh%20Testimony1.pdf

Hesse, H. (2002). *The glass bead game* (R. Winston & C. Winston, Trans.). Picador. (Original work published 1943). Trans. of *Das Glasperlenspiel.* (1943). Fretz & Wasmuth

Hillman, J. E. (2021). *The digital silk road: China's quest to wire the world and win the future.* Harper Business.

Hindman, M. (2005). The real lessons of Howard Dean: Reflections on the first digital campaign. *Perspectives on Politics*, *3*(1), 121–128. https://doi.org/10.1017/S1537592705050115

Hindman, M. (2009). *The myth of of digital democracy.* Princeton University Press.

Hoffman, D. L., & Novak, T. P. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing*, *60*(3), 50–68. https://doi.org/10.1177/002224299606000304

Hoffmann, C. P. (2020). Techlash: Digitale plattformen zwischen utopie und dystopie. In S. Russ-Mohl (Ed.), *Streitlust und streitkunst: Diskurs als essenz der demokratie* (pp. 66–91). Herbert von Halem Verlag.

Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science*, *355*(6324), 486–488. https://doi.org/10.1126/science.aal3856

Holmberg, E. S. (2021). Lessons from Trump's suspension: How Twitter should clarify and strengthen its "public interest" approach to moderating leaders' violence-inspiring speech. *Harvard Journal of Law & Technology*, *35*(1).

Holmes, D. E. (2017). *Big data: A very short introduction.* Oxford University Press.

Horton, J. J. (2023). Large language models as simulated economic agents: What can we learn from Homo Silicus? *NBER Working Papers.* https://doi.org/10.3386/w31122

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, *12*(12), 767–797. https://doi.org/10.17705/1jais.00282

Hvitfeldt, E. (2022). *Supervised machine learning for text analysis in r.* CRC Press.

Hwang, T. (2020). *Subprime attention crisis: Advertising and the time bomb at the heart of the internet.* Farrat, Straus; Giroux.

Igo, S. (2007). *The averaged american: Surveys, citizens, and the making of a mass public.* Harvard University Press.

Igo, S. E. (2018). *The known citizen: A history of privacy in modern america.* Harvard University Press.

IJzerman, H., Lewis Jr., N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J. D., & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behavior*, *4*, 1092–1094. https://doi.org/10.1038/s41562-020-00990-w

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction.* Cambridge University Press.

Isaac, M. (2019). *Super pumped: The battle for uber.* W. W. Norton & Company.

Isaacson, W. (2014). *The innovators: How a group of inventors, hackers, geniuses, and geeks created the digital revolution.* Simon & Schuster.

Issenberg, S. (2012). *The victory lab: The secret science of winning campaigns*. Crown Publishing Group.

Iyer, R., Koleva, S., Graham, J., Ditto, P., & Haidt, J. (2012). Understanding libertarian morality: The psychological dispositions of self-identified Libertarians. *PLoS One*, *7*(8), e42366. https://doi.org/10.1371/journal.pone.0042366

Jackson, S. J., Bailey, M., & Welles, B. F. (2020). *#HashtagActivism: Networks of race and gender justice*. The MIT Press.

Jiang, M., & Fu, K.-w. (2018). Chinese social media and big data: Big data, big brother, big profit? *Policy & Internet*, *10*(4), 372–392. https://doi.org/10.1002/poi3.187

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with alphafold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Jungherr, A. (2012). The German federal election of 2009: The challenge of participatory cultures in political campaigns. *Transformative Works and Cultures*, *10*. https://doi.org/10.3983/twc.2012.0310

Jungherr, A. (2014). The logic of political coverage on Twitter: Temporal dynamics and content. *Journal of Communication*, *64*(2), 239–259. https://doi.org/10.1111/jcom.12087

Jungherr, A. (2015). *Analyzing political communication with digital trace data: The role of twitter messages in social science research*. Springer. https://doi.org/10.1007/978-3-319-20319-5

Jungherr, A. (2016a). Datengestützte Verfahren im Wahlkampf. *Zeitschrift für Politikberatung*, *8*(1), 3–14. https://doi.org/10.5771/1865-4789-2016-1-3

Jungherr, A. (2016b). Four functions of digital tools in election campaigns: The German case. *The International Journal of Press/Politics*, *21*(3), 358–377. https://doi.org/10.1177/1940161216642597

Jungherr, A. (2019). Normalizing digital trace data. In N. J. Stroud & S. C. McGregor (Eds.), *Digital discussions: How big data informs political communication* (pp. 9–35). Routledge. https://doi.org/10.4324/9781351209434-2

Jungherr, A. (2023a). Artificial intelligence and democracy: A conceptual framework. *Social Media + Society*, *9*(3), 1–14. https://doi.org/10.1177/20563051231186353

Jungherr, A. (2023b). Digital campaigning: How digital media change the work of parties and campaign organizations and impact elections. In J. Skopek (Ed.), *Research handbook digital sociology* (pp. 446–462). Edward Elgar. https://doi.org/10.4337/9781789906769.00035

Jungherr, A., & Jürgens, P. (2010). The political click: Political participation through e-petitions in Germany. *Policy & Internet*, *2*(4), 131–165. https://doi.org/10.2202/1944-2866.1084

Jungherr, A., & Jürgens, P. (2013). Forecasting the pulse: How deviations from regular patterns in online data can identify offline phenomena. *Internet Research*, *23*(5), 589–607. https://doi.org/10.1108/IntR-06-2012-0115

Jungherr, A., & Jürgens, P. (2014). Through a glass, darkly: Tactical support and symbolic association in Twitter messages commenting on Stuttgart 21. *Social Science Computer Review*, *32*(1), 74–89. https://doi.org/10.1177/0894439313500022

Jungherr, A., Posegga, O., & An, J. (2019). Discursive power in contemporary media systems: A comparative framework. *The International Journal of Press/Politics*, *24*(4), 404–425. https://doi.org/10.1177/1940161219841543

Jungherr, A., Posegga, O., & An, J. (2022). Populist supporters on Reddit: A comparison of content and behavioral patterns within publics of supporters of Donald Trump and Hillary Clinton. *Social Science Computer Review*, *40*(3), 809–830. https://doi.org/10.1177/0894439321996130

Jungherr, A., & Rauchfleisch, A. (2024). Negative downstream effects of alarmist disinformation discourse: Evidence from the United States. *Political Behavior*, 1–21. https://doi.org/10.1007/s11109-024-09911-3

Jungherr, A., Rauchfleisch, A., & Wuttke, A. (2024). Deceptive uses of Artificial Intelligence in elections strengthen support for AI ban. *arXiv*. https://doi.org/10.48550/arXiv.2408.1261

Jungherr, A., Rivero, G., & Gayo-Avello, D. (2020). *Retooling politics: How digital media are shaping democracy*. Cambridge University Press. https://doi.org/10.1017/9781108297820

Jungherr, A., & Schlarb, D. (2022). The extended reach of game engine companies: How companies like Epic Games and Unity Technologies provide platforms for extended reality applications and the metaverse. *Social Media + Society*, *8*(2), 1–12. https://doi.org/10.1177/20563051221107641

Jungherr, A., Schoen, H., & Jürgens, P. (2016). The mediation of politics through Twitter: An analysis of messages posted during the campaign for the German federal election 2013. *Journal of Computer-Mediated Communication*, *21*(1), 50–68. https://doi.org/10.1111/jcc4.12143

Jungherr, A., Schoen, H., Posegga, O., & Jürgens, P. (2017). Digital trace data in the study of public opinion: An indicator of attention toward politics rather than political support. *Social Science Computer Review*, *35*(3), 336–356. https://doi.org/10.1177/0894439316631043

Jungherr, A., & Schroeder, R. (2021). Disinformation and the structural transformations of the public arena: Addressing the actual challenges to democracy. *Social Media + Society*, *7*(1), 1–13. https://doi.org/10.1177/2056305121988928

Jungherr, A., & Schroeder, R. (2022). *Digital transformations of the public arena*. Cambridge University Press. https://doi.org/10.1017/9781009064484

Jungherr, A., & Schroeder, R. (2023). Artificial intelligence and the public arena. *Communication Theory*, *33*(2–3), 164–173. https://doi.org/10.1093/ct/qtad006

Jungherr, A., Schroeder, R., & Stier, S. (2019). Digital media and the surge of political outsiders: Explaining the success of political challengers in the United States, Germany, and China. *Social Media + Society*, *5*(3), 1–12. https://doi.org/10.1177/2056305119875439

Jungherr, A., & Theocharis, Y. (2017). The empiricist's challenge: Asking meaningful questions in political science in the age of big data. *Journal of Information Tech-*

nology & Politics, *14*(1), 97–109. https://doi.org/10.1080/19331681.2017.1312187

Jungherr, A., Wuttke, A., Mader, M., & Schoen, H. (2021). A source like any other? field and survey experiment evidence on how interest groups shape public opinion. *Journal of Communication*, *71*(2), 276–304. https://doi.org/10.1093/joc/jqab005

Jürgens, P., Jungherr, A., & Schoen, H. (2011). Small worlds with a difference: New gatekeepers and the filtering of political information on Twitter. In D. D. Roure & S. Poole (Eds.), *Websci 2011: Proceedings of the 3rd international web science conference*. ACM. https://doi.org/10.1145/2527031.2527034

Jürgens, P., Meltzer, C. E., & Scharkow, M. (2022). Age and gender representation on German TV: A longitudinal computational analysis. *Computational Communication Science*, *4*(1). https://doi.org/10.5117/CCR2022.1.005.JURG

Jürgens, P., & Stark, B. (2017). The power of default on Reddit: A general model to measure the influence of information intermediaries. *Policy & Internet*, *9*(4), 395–419. https://doi.org/10.1002/poi3.166

Jürgens, P., Stark, B., & Magin, M. (2020). Two half-truths make a whole? on bias in self-reports and tracking data. *Social Science Computer Review*, *38*(5), 600–615. https://doi.org/10.1177/0894439319831643

Kafka, F. (1998). *The castle* (M. Harman, Trans.). Schocken. (Original work published 1926). Trans. of Brod, M. (Ed.). (1926). *Das Schloss*. Kurt Wolff

Kafka, F. (1999). *The trial* (B. Mitchell, Ed. & Trans.). Schocken. (Original work published 1925). Trans. of Brod, M. (Ed.). (1925). *Der Prozess*. Die Schmiede

Kahan, D. M. (2016a). The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it. In R. A. Scott & M. C. Buchmann (Eds.), *Emerging Trends in the Social and Behavioral Sciences* (pp. 1–16). John Wiley & Sons. https://doi.org/10.1002/9781118900772.etrds0417

Kahan, D. M. (2016b). The politically motivated reasoning paradigm, part 2: Unanswered questions. In R. A. Scott & M. C. Buchmann (Eds.), *Emerging Trends in the Social and Behavioral Sciences* (pp. 1–15). John Wiley & Sons. https://doi.org/10.1002/9781118900772.etrds0418

Kaiser, J., & Rauchfleisch, A. (2019). The implications of venturing down the rabbit hole. *Internet Policy Review*. https://policyreview.info/articles/news/implications-venturing-down-rabbit-hole/1406

Kaiser, J., & Rauchfleisch, A. (2020). Birds of a feather get recommended together: Algorithmic homophily in YouTube's channel recommendations in the United States and Germany. *Social Media + Society*, *6*(4), 1–15. https://doi.org/10.1177/2056305120969914

Kanno-Young, Z., & Kang, C. (2021). 'They're killing people': Biden denounces social media for virus disinformation. *The New York Times*. https://www.nytimes.com/2021/07/16/us/politics/biden-facebook-social-media-covid.html

Kapczynski, A. (2020). The law of informational capitalism. *The Yale Law Journal*, *129*(5), 1460–1515.

Kargar, S., & Rauchfleisch, A. (2019). State-aligned trolling in Iran and the double-edged affordances of Instagram. *New Media & Society*, *21*(7), 1506–1527. https://doi.org/10.1177/1461444818825133

Karpf, D. (2012a). *The MoveOn effect: The unexpected transformation of american political advocacy*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199898367.001.0001

Karpf, D. (2012b). Social science research methods in internet time. *Information, Communication & Society*, *15*(5), 639–661. https://doi.org/10.1080/1369118X.2012.665468

Karpf, D. (2016). *Analytical activism: Digital listerning and the new political strategy*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190266127.001.0001

Karpf, D. (2019). Something I no longer believe: Is internet time slowing down? *Social Media + Society*, *5*(3), 1–4. https://doi.org/10.1177/2056305119849492

Katz, E., & Lazarsfeld, P. F. (1955). *Personal influence, the part played by people in the flow of mass communications*. The Free Press.

Katz, R. S., & Mair, P. (2018). *Democracy and the cartelization of political parties*. Oxford University Press.

Kaye, D. (2018). *Report on artificial intelligence technologies and implications for freedom of expression and the information environment*. United Nations Human Rights Office of the High Commissioner. https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ReportGA73.aspx

Kaye, D. (2019). *Speech police: The global struggle to govern the internet*. Columbia Global Reports.

Keane, J. (2013). *Democracy and media decadence*. Cambridge University Press. https://doi.org/10.1017/CBO9781107300767

Kelleher, J. D. (2019). *Deep learning*. The MIT Press.

Keller, D. (2018). *Internet platforms: Observations on speech, danger, and money*. Hoover Institution. https://cyberlaw.stanford.edu/files/publication/files/381732092-internet-platforms-observations-on-speech-danger-and-money.pdf

Keuschnigg, M., Lovsjö, N., & Hedström, P. (2018). Analytical sociology and computational social science. *Journal of Computational Social Science*, *1*(1), 3–14. https://doi.org/10.1007/s42001-017-0006-5

Kiefer, M. L. (2010). *Journalismus und Medien als Institutionen*. UVK Verlagsgesellschaft.

Kim, H., Choi, H., Kang, H., An, J., Yeom, S., & Hong, T. (2021). A systematic review of the smart energy conservation system: From smart homes to sustainable smart cities. *Renewable and Sustainable Energy Reviews*, *140*(110755), 1–17. https://doi.org/10.1016/j.rser.2021.110755

Kim, J., & Lee, B. (2023). AI-augmented surveys: Leveraging large language models for opinion prediction in nationally representative surveys. *arXiv*. https://doi.org/10.48550/arXiv.2305.09620

King, G. (2011). Ensuring the data-rich future of the social sciences. *Science*, *331*(6018), 719–721. https://doi.org/10.1126/science.1197872

King, R. D., Rowland, J., Oliver, S. G., Young, M., Aubrey, W., Byrne, E., Liakata, M., Markham, M., Pir, P., Soldatova, L. N., Sparkes, A., Whelan, K. E., Whelan, K. E., & Clare, A. (2009). The automation of science. *Science*, *324*(5923), 88–89. https://doi.org/10.1126/science.1165620

Kitchens, B., Johnson, S. L., & Gray, P. (2020). Understanding echo chambers and filter bubbles: The impact of social media on diversification and partisan shifts in news consumption. *MIS Quarterly*, *44*(4), 1619–1649. https://doi.org/10.25300/MISQ/2020/16371

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 1–12. https://doi.org/10.1177/2053951714528481

Kleinberg, J., & Tardos, É. (2005). *Algorithm design*. Addison Wesley Longman.

Knight, A., & Creemers, R. (2021). Going viral: The Social Credit System and COVID-19. *Social Science Research Network*. https://doi.org/10.2139/ssrn.3770208

Knockel, J., Parsons, C., Ruan, L., Xiong, R., Crandall, J., & Deibert, R. (2020). *We chat, they watch how international users unwittingly build up WeChat's chinese censorship apparatus*. The Citizen Lab. https://tspace.library.utoronto.ca/bitstream/1807/101395/1/Report%23127--wechattheywatch-web.pdf

Knüpfer, C., Hoffmann, M., & Voskrensenskii, V. (2022). Hijacking MeToo: Transnational dynamics and networked frame contestation on the far right in the case of the '120 decibels' campaign. *Information, Communication & Society*, *25*(7), 1010–1028. https://doi.org/10.1080/1369118X.2020.1822904

Knuth, D. E. (1997). *The art of computer programming: Fundamental algorithms* (3rd ed., Vol. 1). Addison Wesley Longman. (Original work published 1968)

Kovach, B., & Rosenstiel, T. (2021). *The elements of journalism: What newspeople should know and the public should expect* (4th ed.). The Crown Publishing Group. (Original work published 2001)

Krebs, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, *9*(1), 104–117. https://doi.org/10.1017/XPS.2020.37

Kreiss, D. (2012). *Taking our country back: The crafting of networked politics from Howard Dean to Barack Obama*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199782536.001.0001

Kreiss, D. (2016). *Prototype politics: Technology-intensive campaigning and the data of democracy*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199350247.001.0001

Kreiss, D. (2019). Digital opportunity structures: Explaining variation in digital mobilization during the 2016 democratic primaries. In M. X. D. Carpini (Ed.), *Digital media and democratic futures* (pp. 42–68). University of Pennsylvania Press.

Kretschmer, M., Kretschmer, T., Peukert, A., & Peukert, C. (2023). The risks of risk-based ai regulation: Taking liability seriously. *arXiv*, 1–18. https://doi.org/10.48550/arXiv.2311.14684

Kroll, A. (2018). Cloak and data: The real story behind Cambridge Analytica's rise and fall. *Mother Jones*, (May/June). https://www.motherjones.com/politics/2018/03/cloak-and-data-cambridge-analytica-robert-mercer/

Kuklinski, J. H., & Quirk, P. J. (2000). Reconsidering the rational public: Cognition, heuristics, and mass opinion. In A. Lupia, M. D. McCubbins, & S. L. Popkin (Eds.), *Elements of reason: Cognition, choice, and the bounds of rationality* (pp. 153–182). Cambridge University Press. https://doi.org/10.1017/CBO9780511805813.008

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86.

Kuran, T. (1995). *Private truths, public lies: The social consequences of preference falsification*. Harvard University Press.

Lafont, C. (2020). *Democracy without shortcuts: A participatory conception of deliberative democracy*. Oxford University Press. https://doi.org/10.1093/oso/9780198848189.001.0001

Landemore, H. (2012). *Democratic reason: Politics, collective intelligence, and the rule of the many*. Princeton University Press.

Landemore, H. (2024). Can artificial intelligence bring deliberation to the masses? In R. Chang & A. Srinivasan (Eds.), *Conversations in philosophy, law, and politics* (pp. 39–69). Oxford University Press. https://doi.org/10.1093/oso/9780198864523.003.0003

Landemore, H., & Elster, J. (Eds.). (2012). *Collective wisdom: Principles and mechanisms*. Cambridge University Press. https://doi.org/10.1017/CBO9780511846427

Lane, H., & Dyshel, M. (2022). *Natural language processing in action: Understanding, analyzing, and generating text with python* (2nd ed.). Manning Publications Co.

Laney, D. (2001). 3d data management: Controlling data volume, velocity, and variety, application delivery strategies. *META Group: Application Delivery Strategies*, (949).

Lanz, M., & Precht, R. D. (2022). Lanz und Precht diskutieren über die Medienlandschaft. *Lanz und Precht*. https://www.youtube.com/watch?v=OvUSVSdr-zI

Larson, E. J. (2021). *The myth of artificial intelligence: Why computers can't think the way we do*. The Belknap Press of Harvard University.

Laswell, H. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas* (pp. 243–276). Institue for Religous; Social Studies.

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, *97*(2), 311–331. https://doi.org/10.1017/S0003055403000698

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M. W., Roy, D., & Alstyne, M. V. (2009). Computational social science. *Science*, *323*(5915), 721–723. https://doi.org/10.1126/science.1167742

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, K.-F. (2018). *AI superpowers: China, Silicon Valley, and the new world order*. Houghton Mifflin Harcourt.

Lee, K.-F., & Quifan, C. (2021). *AI 2041: Ten visions for our future*. Currency.

Lepore, J. (2020). *If then: How one data company invented the future*. W. W. Norton & Company.

Levine, F., Locke, C., Searls, D., & Weinberger, D. (2000). *The cluetrain manifesto: The end of business as usual*. Persus Publishing.

Levy, S. (2010). *Hackers: Heroes of the computer revolution*. O'Reilly Media.

Levy, S. (2011). *In the plex: How google thinks, works, and shapes our lives*. Simon & Schuster.

Levy, S. (2020). *Facebook: The inside story*. Blue Rider Press.

Lewis, M. (2014). *Flash boys: A wall street revolt*. W. W. Norton & Company.

Liang, F., Das, V., Kostyuk, N., & Hussain, M. M. (2018). Constructing a data-driven society: China's social credit system as a state surveillance infrastructure. *Policy & Internet*, *10*(4), 415–453. https://doi.org/10.1002/poi3.183

Lindblom, C. E. (1965). *Intelligence of democracy: Decision making through mutual adjustment*. The Free Press.

Lindblom, C. E. (2001). *The market system: What it is, how it works, and what to make of it*. Yale University Press.

Lippmann, W. (1927). *The phantom public*. The Macmillan Company.

Little, D. (2020). *A new social ontology of government: Consent, coordination, and authority*. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-48923-6

Liu, X., Glocker, B., McCradden, M. M., Ghassemi, M., Denniston, A. K., & Oakden-Rayner, L. (2022). The medical algorithmic audit. *The Lancet: Digital Health*, *4*(5), E384–E397. https://doi.org/10.1016/S2589-7500(22)00003-6

Lodge, M., & Taber, C. S. (2013). *The rationalizing voter*. Cambridge University Press. https://doi.org/10.1017/CBO9781139032490

Louridas, P. (2017). *Real-world algorithms: A beginner's guide*. The MIT Press.

Lowe, W. (2008). Understanding Wordscores. *Political Analysis*, *16*(4), 356–371. https://doi.org/10.1093/pan/mpn004

Lowery, W. (2020). A reckoning over objectivity, led by black journalists. *The New York Times*. https://www.nytimes.com/2020/06/23/opinion/objectivity-black-journalists-coronavirus.html

Lu, Y., & Pan, J. (2021). Capturing clicks: How the Chinese Government uses clickbait to compete for visibility. *Political Communication*, *38*(1–2), 23–54. https://doi.org/10.1080/10584609.2020.1765914

Lupia, A., & McCubbins, M. D. (1998). *The democratic dilemma: Can citizens learn what they need to know?* Cambridge University Press.

Lutscher, P. M., Weidmann, N. B., Roberts, M. E., Jonker, M., & King, A. (2020). At home and abroad: The use of denial-of-service attacks during elections in nondemocratic regimes. *Journal of Conflict Resolution*, *64*(2-3), 373–401. https://doi.org/10.1177/0022002719861676

MacKenzie, D. (2021). *Trading at the speed of light: How ultrafast algorithms are transforming financial markets*. Princeton University Press.

MacKenzie, D. (2022). Blink, bid, buy. *London Review of Books*, *44*(9). https://www.lrb.co.uk/the-paper/v44/n09/donald-mackenzie/blink-bid-buy

MacKenzie, D., Caliskan, K., & Rommerskirchen, C. (2023). The longest second: Header bidding and the material politics of online advertising. *Economy and Society*, *42*(3), 554–578. https://doi.org/10.1080/03085147.2023.2238463

Macy, M. W., & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology*, *28*, 143–166. https://doi.org/10.1146/annurev.soc.28.110601.141117

Mair, P. (2013). *Ruling the void: The hollowing of Western democracy*. Verso.

Margetts, H. (2001). The cyber party: The causes and consequences of organisational innovation in european political parties. In *The causes and consequences of organisational innovation in european political parties*. ECPR Joint Sessions of Workshops.

Margetts, H. (2006). Cyber parties. In R. S. Katz & W. Crotty (Eds.), *Handbook of party politics* (pp. 528–535). SAGE. https://doi.org/10.4135/9781848608047.n46

Mau, S. (2019). *Das metrische wir: Über die quantifizierung des sozialen* [The metric society: On the quantification of the social] (S. Howe, Trans.). Polity Press. (Original work published 2017)

Mayson, S. G. (2019). Bias in, bias out. *The Yale Law Journal*, *128*(8), 2218–2300.

McCombs, M., & Valenzuela, S. (2021). *Setting the agenda: Mass media and public opinion* (3rd ed.). Polity Press. (Original work published 2004)

McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. A K Peters.

McElreath, R. (2020). *Statistical rethinking: A bayesian course with examples in r and stan* (2nd ed.). CRC Press.

McFarland, M. (2014). Elon Musk: 'With artificial intelligence we are summoning the demon.' *The Washington Post*. https://doi.org/2014-10-24

McGregor, S. C. (2019). Social media as public opinion: How journalists use social media to represent public opinion. *Journalism*, *20*(8), 1070–1086. https://doi.org/10.1177/1464884919845458

McGregor, S. C. (2020). "Taking the temperature of the room": How political campaigns use social media to understand and represent public opinion. *Public Opinion Quarterly*, *84*(S1), 236–256. https://doi.org/10.1093/poq/nfaa012

McKenna, E., & Han, H. (2014). *Groundbreakers: How Obama's 2.2 Million volunteers transformed campaigning in America*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199394593.001.0001

McQuail, D. (2013). *Journalism and society*. SAGE Publications.

Meckler, L. (2012). Obama data trove is up for grabs. *The Wall Street Journal*. https://www.wsj.com/articles/SB10001424127887323622904578129432544571720

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, *54*(6), 1–35. https://doi.org/10.1145/3457607

Menkveld, A. J. (2016). The economics of high-frequency trading: Taking stock. *Annual Review of Financial Economics*, *8*, 1–24. https://doi.org/10.1146/annurev-financial-121415-033010

Mennicken, A., & Espeland, W. N. (2019). What's new with numbers? sociological approaches to the study of quantification. *Annual Review of Sociology*, *35*, 223–245. https://doi.org/10.1146/annurev-soc-073117-041343

Mercier, H. (2020). *Not born yesterday: The science of who we trust and what we believe*. Princeton University Press.

Merz, N., Regel, S., & Lewandowski, J. (2016). The Manifesto Corpus: A new resource for research on political parties and quantitative text analysis. *Research & Politics*, *3*(2), 1–8. https://doi.org/10.1177/2053168016643346

Metaxa, D., Park, J. S., Robertson, R., Karahalios, K., Wilson, C., Hancock, J., & Sandvig, C. (2021). Auditing algorithms: Understanding algorithmic systems from the outside in. *Foundations and Trends in Human–Computer Interaction*, *14*(4), 272–344. https://doi.org/10.1561/1100000083

Metz, C. (2021). *Genius makers: The mavericks who brought AI to Google, Facebook, and the world*. Dutton.

Mignano, M. (2022). The end of social media and the rise of recommendation media. *Every*. https://doi.org/https://every.to/p/the-end-of-social-media

Miller, J. H., & Page, S. E. (2007). *Complex adaptive systems: An introduction to computational models of social life*. Princeton University Press.

Mitchell, M. (2019). *Artificial intelligence: A guide for thinking humans*. Farrat, Straus; Giroux.

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application*, *8*, 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902

Möller, J., Trilling, D., Helberger, N., & van Es, B. (2018). Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society*, *21*(7), 959–977. https://doi.org/10.1080/1369118X.2018.1444076

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles for social research* (2nd ed.). Cambridge University Press.

Morgus, R. (2019). The spread of russia's digital authoritarianism. In N. D. Wright (Ed.), *Artificial intelligence, china, russia, and the global order: Technological, political, global, and creative perspectives* (pp. 89–97). Air University Press.

Mozur, P., Xiao, M., & Liu, J. (2022). 'an invisible cage': How China is policing the future. *The New York Times*. https://www.nytimes.com/2022/06/25/technology/china-surveillance-police.html

Muller, J. Z. (2018). *The tyranny of metrics*. Princeton University Press.

Müller, J.-W. (2021). *Democracy rules*. Allen Lane.

Müller, S. (2022). The temporal focus of campaign communication. *The Journal of Politics*, *84*(1), 585–590. https://doi.org/10.1086/715165

Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, *39*(3), 629–649. https://doi.org/10.1007/s11109-016-9373-5

Munger, K., & Phillips, J. (2022). Right-wing YouTube: A supply and demand perspective. *The International Journal of Press/Politics*, *27*(1), 186–219. https://doi.org/10.1177/1940161220964767

Murgia, M. (2023). Algorithms are deciding who gets organ transplants. are their decisions fair? *Financial Times*. https://www.ft.com/content/5125c83a-b82b-40c5-8b35-99579e087951

Narayanan, A. (2023). Understanding social media recommendation algorithms. *Knight First Amendment Institute at Columbia University*. https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms

Natale, S., & Ballatore, A. (2014). The web will kill them all: New media, digital utopia, and political struggle in the Italian 5-Star Movement. *Media, Culture & Society*, *36*(1), 105–121. https://doi.org/10.1177/0163443713511902

Neuman, W. R. (1991). *The future of the mass audience*. Cambridge University Press.

Neuman, W. R., Guggenheim, L., Jang, S. M., & Bae, S. Y. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication*, *64*(2), 193–214. https://doi.org/10.1111/jcom.12088

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Nielsen, R. K. (2022). *Reuters institute digital news report 2022*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022

Nicholls, T., Shbbir, N., & Nielsen, R. K. (2016). *Digital-born news media in europe*. Reuters Institute for the Study of Journalism.

Nickerson, D. W., & Rogers, T. (2014). Political campaigns and big data. *The Journal of Economic Perspectives*, *28*(2), 51–74. https://doi.org/10.1257/jep.28.2.51

Nielsen, M. (2012). *Reinventing discovery: The new era of networked science*. Princeton University Press.

Nielsen, R. K. (2011). Mundane internet tools, mobilizing practices, and the coproduction of citizenship in political campaigns. *New Media & Society*, *13*(5), 755–771. https://doi.org/10.1177/1461444810380863

Nielsen, R. K. (2012). *Ground wars: Personalized communication in political campaigns*. Princeton University Press.

Nielsen, R. K. (2018). No one cares what we know: Three responses to the irrelevance of political communication research. *Political Communication*, *35*(1), 145–149. https://doi.org/10.1080/10584609.2017.1406591

Nielsen, R. K. (2020). Economic contexts of journalism. In K. Wahl-Jorgensen & T. Hanitzsch (Eds.), *The handbook of journalism studies* (2nd ed., pp. 324–340). Routledge. https://doi.org/10.4324/9781315167497-21

Nielsen, R. K., & Fletcher, R. (2022). Concentration of online news traffic and publishers' reliance on platform referrals: Evidence from passive tracking data in the UK. *Journal of Quantitative Description: Digital Media*, *2*, 1–23. https://doi.org/10.51685/jqd.2022.015

Nielsen, R. K., & Ganter, S. A. (2018). Dealing with digital intermediaries: A case study of the relations between publishers and platforms. *New Media & Society*, *20*(4), 1600–1617. https://doi.org/10.1177/1461444817701318

Nilsson, N. J. (2010). *The quest for artificial intelligence: A history of ideas and achievements*. Cambridge University Press. https://doi.org/10.1017/CBO97805118193 46

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

Nuernbergk, C., & Conrad, J. (2016). Conversations and campaign dynamics in a hybrid media environment: Use of Twitter by members of the German Bundestag. *Social Media + Society*, *2*(1), 1–14. https://doi.org/10.1177/2056305116628888

Nyhan, B. (2020). Facts and myths about misperceptions. *Journal of Economic Perspectives*, *34*(3), 220–236. https://doi.org/10.1257/jep.34.3.220

Ober, J. (2008). *Democracy and knowledge: Innovation and learning in classical athens*. Princeton University Press.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Offe, C. (2006). Political institutions and social power: Conceptual explorations. In I. Shapiro, S. Skowronek, & D. Galvin (Eds.), *Rethinking political institutions. the art of the state* (pp. 9–31). New York University Press.

O'Mara, M. (2019). *The code: Silicon valley and the remaking of america*. Penguin Press.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), 943. https://doi.org/10.1126/science.aac4716

O'Reilly, T. (2005). What is Web 2.0: Design patterns and business models for the next generation of software. *O'Reilly Blog*. http://oreilly.com/web2/archive/what-is-web-20.html

O'Reilly, T. (2017). *Wtf? what's the future and why it's up to us*. Harper Collins.

Page, S. E. (2018). *The model thinker: What you need to know to make data work for you*. Basic Books.

Pan, J. (2020). *Welfare for autocrats: How social assistance in China cares for its rulers*. Oxford University Press. https://doi.org/10.1093/oso/9780190087425.001.0001

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. The Penguin Press.

Parker, G. G., Alstyne, M. W. V., & Choudary, S. P. (2016). *Platform revolution: How networked markets are transforming the economy and how to make them work for you*. W. W. Norton & Company.

Parmar, N. J., Vaswani, A., Uszkoreit, J., Kaiser, L., Shazeer, N., Ku, A., & Tran, D. (2018). Image transformer. In J. Dy & A. Krause (Eds.), *Proceedings of the 35th international conference on machine learning* (pp. 4055–4064). PMLR. http://proceedings.mlr.press/v80/parmar18a/parmar18a.pdf

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Pearl, J. (2009). *Causality: Models, reasoning and inference* (2nd ed.). Cambridge University Press.

Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, *62*(3), 54–60. https://doi.org/10.1145/3241036

Pearlman, N. G. (2012). *Margin of victory: How technologists help politicians win elections* (N. G. Pearlman, Ed.). Praeger.

Pentland, A. (2008). *Honest signals: How they shape our world*. The MIT Press.

Perry, C., & DeDeo, S. (2021). The cognitive science of extremist ideologies online. *arXiv*. https://doi.org/10.48550/arXiv.2110.00626

Phillips, A. (2021). *Unconditional equals*. Princeton University Press.

Phillips, W. (2015). *This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture*. The MIT Press.

Phillips, W. (2018). *The oxygen of amplification: Better practices for reporting on extremists, antagonists, and manipulators online*. Data & Society.

Phillips, W., & Milner, R. M. (2017). *The ambivalent internet: Mischief, oddity, and antagonism online*. Polity Press.

Pickard, V. (2020). *Democracy without journalism? confronting the misinformation society*. Oxford University Press. https://doi.org/10.1093/oso/9780190946753.001.0001

Piper, A. (2018). *Enumerations: Data and literary study*. The University of Chicago Press.

Pollitt, C., & Bouckaert, G. (2017). *Public management reform: A comparative analysis - into the age of austerity* (4th ed.). Oxford University Press.

Popkin, S. L. (1991). *The reasoning voter: Communication and persuasion in presidential campaigns*. The University of Chicago Press.

Popkin, S. L. (2021). *Crackup: The Republican implosion and the future of presidential politics*. Oxford University Press. https://doi.org/10.1093/oso/9780190913823.001.0001

Porter, T. M. (2020). *Trust in numbers: The pursuit of objectivity in science and public life* (2nd ed.). Princeton University Press. (Original work published 1995)

Posegga, O. (2023). Unlocking big data: At the crossroads of computer science and the social sciences. In J. Skopek (Ed.), *Research handbook digital sociology* (pp. 115–129). Edward Elgar. https://doi.org/10.4337/9781789906769.00013

Posegga, O., & Jungherr, A. (2019). Characterizing political talk on Twitter: A comparison between public agenda, media agendas, and the Twitter agenda with regard to topics and dynamics. In *Hicss 2019: Proceedings of the 52nd Hawaii international conference on system science* (pp. 2590–2599). Scholarspace. https://doi.org/10.24251/HICSS.2019.312

Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine*, *25*, 37–43. https://doi.org/10.1038/s41591-018-0272-7

Prince, S. J. D. (2023). *Understanding deep learning*. The MIT Press.

Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press. https://doi.org/10.1017/CBO9781139878425

Prior, M. (2009). The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly*, *73*(1), 130–143. https://doi.org/10.1093/poq/nfp002

Prior, M. (2017). Conditions for political accountability in a high-choice media environment. In K. Kenski & K. H. Jamieson (Eds.), *The Oxford handbook of political communication* (pp. 897–912). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199793471.013.63

Prior, M. (2018). *Hooked: How politics captures people's interest*. Cambridge University Press.

Przeworski, A. (1991). *Democracy and the market: Political and economic reforms in Eastern Europe and Latin America*. Cambridge University Press. https://doi.org/10.1017/CBO9781139172493

Przeworski, A. (2018). *Why bother with elections?* Polity Press.

Quiring, O. (2016). Interactivity. In *Quiring, o. (2016). interactivity. the international encyclopedia of communication theory and philosophy, 1–12.* (pp. 1–12). John Wiley & Sons. https://doi.org/10.1002/9781118766804.wbiect075

Rahman, H. A. (2021). The invisible cage: Workers' reactivity to opaque algorithmic evaluations. *Administrative Science Quarterly*, *66*(4), 945–988. https://doi.org/10.1177/00018392211010118

Rainie, L., & Wellman, B. (2012). *Networked: The new social operating system*. The MIT Press.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. *arXiv*. https://doi.org/10.48550/arXiv.2204.06125

Rau, J. P., & Stier, S. (2019). Die Echokammer-Hypothese: Fragmentierung der Öffentlichkeit und politische Polarisierung durch digitale Medien? *Zeitschrift für Vergleichende Politikwissenschaft*, *13*(3), 339–417. https://doi.org/10.1007/s12286-019-00429-1

Rauchfleisch, A. (2017). The public sphere as an essentially contested concept: A co-citation analysis of the last 20 years of public sphere research. *Communication and the Public*, *2*(1), 3–18. https://doi.org/10.1177/2057047317691054

Rauchfleisch, A., & Kaiser, J. (2020). The German far-right on YouTube: An analysis of user overlap and user comments. *Journal of Broadcasting & Electronic Media*, *64*(3), 373–396. https://doi.org/10.1080/08838151.2020.1799690

Rauchfleisch, A., & Kaiser, J. (2021). Deplatforming the far-right: An analysis of YouTube and BitChute. *Social Science Research Network*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3867818

Rauchfleisch, A., & Kovic, M. (2016). The internet and generalized functions of the public sphere: Transformative potentials from a comparative perspective. *Social Media + Society*, *2*(2), 1–15. https://doi.org/10.1177/2056305116646393

Rauchfleisch, A., Siegen, D., & Vogler, D. (2023). How COVID-19 displaced climate change: Mediated climate change activism and issue attention in the Swiss media and online sphere. *Environmental Communication*, *17*(3), 313–321. https://doi.org/10.1080/17524032.2021.1990978

Rauh, C., & Schwalbach, J. (2020). *The parlspeech v2 data set: Full-text corpora of 6.3 million parliamentary speeches in the key legislative chambers of nine representative democracies*. Harvard Dataverse. https://doi.org/10.7910/DVN/L4OAKN

Raymond, E. S. (1999). *The cathedral and the bazaar: Musings on Linux and open source by an accidental revolutionary*. O'Reilly Media.

Rheingold, H. (1993). *The virtual community: Homesteading on the electronic frontier*. Addison-Wesley.

Rid, T. (2020). *Active measures: The secret history of disinformation and political warfare*. Farrat, Straus; Giroux.

Risse, M. (2023). *Political theory of the digital age: Where artificial intelligence might take us*. Cambridge University Press. https://doi.org/10.1017/9781009255189

Rivero, G. (2019). Preaching to the choir: Ideology and following behaviour in social media. *Contemporary Social Science*, *14*(1), 54–70. https://doi.org/10.1080/21582041.2017.1325924

Roberts, M. E. (2018). *Censored: Distraction and diversion inside China's great firewall*. Princeton University Press.

Rochet, J.-C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, *1*(4), 990–1029. https://doi.org/10.1162/154247603322493212

Rochet, J.-C., & Tirole, J. (2006). Two-sided markets: A progress report. *The RAND Journal of Economics*, *37*(3), 645–667. https://doi.org/10.1111/j.1756-2171.2006.tb00036.x

Rosen, J. (2006). The people formerly known as the audience. *PRESSthink: Ghost of Democracy in the Media Machince*. http://archive.pressthink.org/2006/06/27/ppl_frmr.html

Roughgarden, T. (2016). *Twenty lectures on algorithmic game theory*. Cambridge University Press. https://doi.org/10.1017/CBO9781316779309

Rusbridger, A. (2018). *Breaking news: The remaking of journalism and why it matters now*. Canongate.

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson Education. (Original work published 1995)

Salganik, M. J. (2018). *Bit by bit: Social research in the digital age*. Princeton University Press.

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, *311*(5762), 854–856. https://doi.org/10.1126/science.1121066

Salganik, M. J., & Watts, D. J. (2009). Web-based experiments for the study of collective social dynamics in cultural markets. *Topics in Cognitive Science*, *1*(3), 439–468. https://doi.org/10.1111/j.1756-8765.2009.01030.x

Sanders, N. E., & Schneier, B. (2021). Machine learning featurizations for AI hacking of political systems. *arXiv*, https://doi.org/10.48550/arXiv.2110.09231

Schäfer, A., & Zürn, M. (2021). *Die demokratische Regression*. Suhrkamp.

Schäfer, M. S., & Wessler, H. (2020). Öffentliche Kommunikation in Zeiten künstlicher Intelligenz. *Publizistik*, *65*, 307–331. https://doi.org/10.1007/s11616-020-00592-6

Scharkow, M., Mangold, F., Stier, S., & Breuer, J. (2020). How social network sites and other online intermediaries increase exposure to news. *PNAS: Proceedings of the National Academy of Sciences of the United States of America*, *117*(6), 2761–2763. https://doi.org/10.1073/pnas.1918279117

Schlozman, K. L., Brady, H. E., & Verba, S. (2018). *Unequal and unrepresented: Political inequality and the people's voice in the new gilded age*. Princeton University Press.

Schmøkel, R., & Bossetta, M. (2022). FBAdLibrarian and Pykognition: Open science tools for the collection and emotion detection of images in Facebook political ads with computer vision. *Journal of Information Technology & Politics*, *19*(1), 118–128. https://doi.org/10.1080/19331681.2021.1928579

Schneider, G. (2018). Automating drug discovery. *Nature Reviews Drug Discovery*, *17*(97–113). https://doi.org/10.1038/nrd.2017.232

Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Toward causal representation learning. *Proceedings of IEEE*, *109*(5), 612–634. https://doi.org/10.1109/JPROC.2021.3058954

Schroeder, R. (2014). Big Data and the brave new world of social media research. *Big Data & Society*, *1*(2), 1–11. https://doi.org/10.1177/2053951714563194

Schroeder, R. (2018). *Social theory after the internet: Media, technology and globalization*. UCL Press.

Schroeder, R. (2019). Digital media and the entrenchment of right-wing populist agendas. *Social Media + Society*, *5*(4), 1–11. https://doi.org/10.1177/2056305119885328

Schroeder, R. (2021). Digital media and the globalizing spread of populism. In D. Y. Jin (Ed.), *The routledge handbook of digital media and globalization* (pp. 179–187). Routledge.

Schwalbach, J., & Rauh, C. (2021). Collecting large-scale comparative text data on legislative debate. In H. Back, M. Debus, & J. M. Fernandes (Eds.), *The politics of legislative debate* (pp. 91–109). Oxford University Press. https://doi.org/10.1093/oso/9780198849063.001.0001

Schwartzberg, M. (2015). Epistemic democracy and its challenges. *Annual Review of Political Science*, *18*, 187–203. https://doi.org/10.1146/annurev-polisci-110113-121908

Schwartzel, E. (2022). *Red carpet: Hollywood, china, and the global battle for cultutal supremacy*. Penguin Press.

Schwemmer, C., Unger, S., & Heiberger, R. (2023). Automated image analysis for studying online behaviour. In J. Skopek (Ed.), *Research handbook of digital sociology* (pp. 278–291). Edward Elgar. https://doi.org/10.4337/9781789906769.00023

Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale University Press.

Scott, J. C. (2009). *The art of not being governed: An anarchist history of upland southeast asia*. Yale University Press.

Scott, M., Bunce, M., & Wright, K. (2019). Foundation funding and the boundaries of journalism. *Journalism Studies*, *20*(14), 2034–2052. https://doi.org/10.1080/1461670X.2018.1556321

Segal, A. (2021). Huawei, 5G, and weaponized interdependence. In D. W. Drezner, H. Farrell, & A. L. Newman (Eds.), *The uses and abuses of weaponized interdependence* (pp. 149–168). Brookings Institution Press.

Settle, J. E. (2018). *Frenemies: How social media polarizes America*. Cambridge University Press. https://doi.org/10.1017/9781108560573

Shapiro, C., & Varian, H. R. (1999). *Information rules: A strategic guide to the network economy*. Harvard Business Review Press.

Shifman, L. (2016). Cross-cultural comparisons of user-generated content: An analytical framework. *International Journal of Communication*, *10*, 5644–5663.

Shirky, C. (2008). *Here comes everybody: The power of organizing without organizations*. The Penguin Press.

Shoemaker, P. J., & Reese, S. D. (2014). *Mediating the message in the 21st century* (3rd ed.). Routledge.

Shoemaker, P. J., & Vos, T. P. (2009). *Gatekeeping theory*. Routledge.

Sides, J., & Vavreck, L. (2014). Obama's not-so-big data. *Pacific Standard*. http://www.psmag.com/navigation/politics-and-law/obamas-big-data-inconclusive-results-political-campaigns-72687/

Sifry, M. L. (2023). Can democrats be 'people-first' if their campaigns value people last? *The Connector*. https://theconnector.substack.com/p/can-democrats-be-people-first-if

Silge, J., & Robinson, D. (2017). *Text mining with r: A tidy approach*. O'Reilly Media.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*, 484–489. https://doi.org/10.1038/nature16961

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, *362*(6419), 1140–1144. https://doi.org/10.1126/science.aar6404

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, *550*, 354–359. https://doi.org/10.1038/nature24270

Simitis, S. (1987). Reviewing privacy in an information society. *University of Pennsylvania Law Review, 135*(3), 707–746. https://doi.org/10.2307/3312079

Simitis, S. (1995). From the market to the polis: The EU directive on the protection of personal data. *Iowa Law Review*, *80*(3), 445–469.

Simon, F. M., Altay, S., & Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School Misinformation Review*, *4*(5), 1–11. https://doi.org/10.37016/mr-2020-127

Singh, S. (2020, February 18). *Special delivery: How internet platforms use artificial intelligence to target and deliver ads*. New America Foundation. https://www.newamerica.org/oti/reports/special-delivery/

Síthigh, D. M., & Siems, M. (2019). The Chinese Social Credit System: A model for other countries? *Morden Law Review*, *82*(6), 1034–1071. https://doi.org/10.1111/1468-2230.12462

Smith, B. C. (2019). *The promise of artificial intelligence: Reckoning and judgment*. The MIT Press.

Solove, D. J. (2008). *Understanding privacy*. Harvard University Press.

Somashekhar, S., Sepúlveda, M.-J., Puglielli, S., Norden, A., Shortliffe, E., Kumar, C. R., Rauthan, A., Kumar, N. A., Patil, P., Rhee, K., & Ramya, Y. (2018). Watson for oncology and breast cancer treatment recommendations: Agreement with an expert multidisciplinary tumor board. *Annals of Oncology*, *29*(2), 418–423. https://doi.org/10.1093/annonc/mdx781

Soni, J. (2022). *The founders: The story of paypal and the entrepreneurs who shaped silicon valley*. Simon & Schuster.

Sprietsma, M. (2013). Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics*, *45*(1), 523–538. https://doi.org/10.1007/s00181-012-0609-x

Spufford, F. (2010). *Red plenty: Inside the Fifties' Soviet dream*. Faber & Faber.

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, *38*(5), 503–516. https://doi.org/10.1177/0894439319843669

Stier, S., Mangold, F., Scharkow, M., & Breuer, J. (2022). Post post-broadcast democracy? news exposure in the age of online intermediaries. *American Political Science Review*, *116*(2), 768–774. https://doi.org/10.1017/S0003055421001222

Stier, S., Posch, L., Bleier, A., & Strohmaier, M. (2017). When populists become popular: Comparing Facebook use by the right-wing movement Pegida and German political parties. *Information, Communication & Society*, *20*(9), 1365–1388. https://doi.org/10.1080/1369118X.2017.1328519

Stone, B. (2013). *The everything store: Jeff bezos and the age of amazon*. Little, Brown; Company.

Stone, B. (2017). *The upstarts: How Uber, Airbnb, and the killer companies of the new Silicon Valley are changing the world*. Black Bay Books.

Stone, B. (2021). *Amazon unbound: Jeff bezos and the invention of a global empire*. Simon & Schuster.

Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M. M., Larsen, J. E., & Lehmann, S. (2014). Measuring large-scale social networks with high resolution. *PLoS One*, *9*(4), e95978. https://doi.org/10.1371/journal.pone.0095978

Stromer-Galley, J. (2000). On-line interaction and why candidates avoid it. *Journal of Communication*, *50*(4), 111–132. https://doi.org/10.1111/j.1460-2466.2000.tb02865.x

Stromer-Galley, J. (2019). *Presidential campaigning in the internet age* (2nd ed.). Oxford University Press. https://doi.org/10.1093/oso/9780190694043.001.0001 (Original work published 2014)

Strossen, N. (2018). *Hate: Why we should resist it with free speech, not censorship.* Oxford University Press.

Subhayan Mukerjee, S. G.-B., Sílvia Majó-Vázquez. (2018). Networks of audience overlap in the consumption of digital news. *Journal of Communication*, *68*(1), 26–50. https://doi.org/10.1093/joc/jqx007

Sunstein, C. R. (2001). *Republic.com.* Princeton University Press.

Tai, Y., & Fu, K.-w. (2020). Specificity, conflict, and focal point: A systematic investigation into social media censorship in China. *Journal of Communication*, *70*(6), 842–867. https://doi.org/10.1093/joc/jqaa032

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning ofwords: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Taylor, C. (1993). Modernity and the rise of the public sphere. In G. B. Peterson (Ed.), *The tanner lectures on human values* (pp. 203–260). University of Utah Press. https://tannerlectures.utah.edu/_resources/documents/a-to-z/t/Taylor93.pdf

Taylor, C. (1995). Liberal politics and the public sphere. In *Philosophical arguments* (pp. 257–288). Harvard University Press.

Taylor, F. W. (1911). *The principles of scientific management.* Harper & Brothers.

Teo, E., & Fu, K.-w. (2021). A novel systematic approach of constructing protests repertoires from social media: Comparing the roles of organizational and non-organizational actors in social movement. *Journal of Computational Social Science volume*, *4*(2), 787–812. https://doi.org/10.1007/s42001-021-00101-3

Tetlock, P. E. (2017). *Expert political judgment: How good is it? how can we know?* (Revised). Princeton University Press. (Original work published 2005)

Thelen, K. (2018). Regulating Uber: The politics of the platform economy in Europe and the United States. *Perspectives on Politics*, *16*(4), 938–953. https://doi.org/10.1017/S1537592718001081

Theocharis, Y., & Jungherr, A. (2021). Computational social science and the study of political communication. *Political Communication*, *38*(1–2), 1–22. https://doi.org/10.1080/10584609.2020.1833121

Theocharis, Y., Lowe, W., van Deth, J. W., & García-Albacete, G. (2015). Using twitter to mobilize protest action: Online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. *Information, Communication & Society*, *18*(2), 202–220. https://doi.org/10.1080/1369118X.2014.948035

Theocharis, Y., Vitoratou, S., & Sajuria, J. (2017). Civil society in times of crisis: Understanding collective action dynamics in digitally-enabled volunteer networks. *Journal of Computer-Mediated Communication*, *22*(5), 248–265. https://doi.org/10.1111/jcc4.12194

Thompson, C. (2019). *Coders: The making of a new tribe and the remaking of the world.* Penguin Press.

Tiffany, K. (2022). *Everything i need i get from you: How fangirls created the internet as we know it.* Farrat, Straus; Giroux.

Tilly, C. (2007). *Democracy.* Cambridge University Press. https://doi.org/10.1017/CBO9780511804922

Timberg, C., & Gardner, A. (2012). Democrats push to redeploy Obama's voter database. *The Washington Post.* https://www.washingtonpost.com/business/economy/democrats-push-to-redeploy-obamas-voter-database/2012/11/20/d14793a4-2e83-11e2-89d4-040c9330702a_story.html

Toepfl, F., & Piwoni, E. (2015). Public spheres in interaction: Comment sections of news websites as counterpublic spaces. *Journal of Communication*, *65*(3), 465–488. https://doi.org/10.1111/jcom.12156

Trask, A. W. (2019). *Grokking deep learning.* Manning Publications Co.

Trippi, J. (2004). *The revolution will not be televised: Democracy, the internet, and the overthrow of everything.* Regan Books.

Tse, E. (2015). *China's disruptors: How alibaba, xiaomi, tencent and other companies are changing the rules of business.* Portfolio/Penguin.

Tucker, J. A., Theocharis, Y., Roberts, M. E., & Barberá, P. (2017). From liberation to turmoil: Social media and democracy. *Journal of Democracy*, *28*(4), 46–59.

Tufekci, Z. (2017). *Twitter and tear gas: The power and fragility of networked protest.* Yale University Press.

Turner, F. (2006). *From counterculture to cyberculture: Stewart brand, the whole earth network, and the rise of digital utopianism.* The University of Chicago Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1134. https://doi.org/10.1126/science.185.4157.1124

Underwood, T. (2019). *Distant horizons: Digital evidence and literary change.* The University of Chicago Press.

Usher, N. (2021). *News for the rich, white, and blue: How place and power distort American journalism.* Columbia University Press.

Van Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., Drury, J., Dube, O., Ellemers, N., Finkel, E. J., Fowler, J. H., Gelfand, M., Han, S., Haslam, S. A., Jetten, J., … Wille, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behavior*, *4*, 460–471. https://doi.org/10.1038/s41562-020-0884-z

van der Vlist, F., Helmond, A., Burkhardt, M., & Seitz, T. (2022). Api governance: The case of Facebook's evolution. *Social Media + Society*, *8*(2), 1–24. https://doi.org/10.1177/20563051221086228 van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press. https://doi.org/10.1093/oso/9780190889760.001.0001

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. V. N. Vishwanathan, & R. Garnett (Eds.), *Nips 2017: 31st conference on neural information processing systems* (pp. 1–11, Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Vela, D., Sharp, A., Zhang, R., Nguyen, T., & an Oleg S. Pianykh, A. H. (2022). Temporal quality degradation in AI models. *Scientific Reports*, *12*(11654), 1–12. https://doi.org/10.1038/s41598-022-15245-z

Wallace, J. L. (2022). *Seeking truth & hiding facts: Information, ideology, & authoritarianism in china*. Oxford University Press. https://doi.org/10.1093/oso/9780197627655.001.0001

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815478

Watts, D. J. (2011). *Everything is obvious: How common sense fails us*. Random House.

Weber, V. (2019). Understanding the global ramifications of china's information-control model. In N. D. Wright (Ed.), *Artificial intelligence, china, russia, and the global order: Technological, political, global, and creative perspectives* (pp. 76–80). Air University Press.

Weidmann, N. B. (2023). *Data management for social scientists: From files to datasets*. Cambridge University Press. https://doi.org/10.1017/9781108990424

Wells, C., Pevehouse, D. V. S. J. C., Yang, J., Pelled, A., Boehm, F., Lukito, J., Ghosh, S., & Schmidt, J. L. (2016). How Trump drove coverage to the nomination: Hybrid media campaigning. *Political Communication*, *33*(4), 669–676. https://doi.org/10.1080/10584609.2016.1224416

Wessler, H. (2018). *Habermas and the media*. Polity Press.

Westin, A. F. (1967). *Privacy and freedom*. Atheneum.

Wiedeman, R. (2020). Times change. *New York Magazine*. https://nymag.com/intelligencer/2020/11/inside-the-new-york-times-heated-reckoning-with-itself.html

Wilkinson, A. (2023). The looming threat of ai to Hollywood, and why it should matter to you. *Vox*. https://www.vox.com/culture/23700519/writers-strike-ai-2023-wga

Williams, B. A., & Carpini, M. X. D. (2011). *After broadcast news: Media regimes, democracy, and the new information environment*. Cambridge University Press. https://doi.org/10.1017/CBO9780511846366

Williams, N. W., Casas, A., & Wilkerson, J. D. (2020). *Images as data for social science research*. Cambridge University Press. https://doi.org/10.1017/9781108860741

Windsor, L. C. (2021). Advancing interdisciplinary work in computational communication science. *Political Communication*, *38*(1–2), 182–191. https://doi.org/10.1080/10584609.2020.1765915

Wintrobe, R. (1998). *The political economy of dictatorship*. Cambridge University Press. https://doi.org/10.1017/CBO9781139174916

Wolfram, S. (2023). *What is ChatGPT doing ... and why does it work?* Wolfram Media, Inc.

Wooldridge, A. (2011). *Masters of management: How the business gurus and their ideas have changed the world—for better and for worse.* Harper Collins.

Woolridge, M. (2020). *The road to conscious machines: The story of ai.* Pelican Books.

Wright, J., & Ma, Y. (2022). *High-dimensional data analysis with low-dimensional models: Principles, computation, and applications.* Cambridge University Press. https://doi.org/10.1017/9781108779302

Wright, K., Scott, M., & Bunce, M. (2019). Foundation-funded journalism, philanthrocapitalism and tainted donors. *Journalism Studies*, *20*(5), 675–695. https://doi.org/10.1080/1461670X.2017.1417053

Wuttke, A. (2019). Why too many political science findings cannot be trusted and what we can do about it: A review of meta-scientific research and a call for academic reform. *Politische Vierteljahresschrift*, *60*(1), 1–19. https://doi.org/10.1007/s11615-018-0131-7

Yang, E., & Roberts, M. E. (2023). The authoritarian data problem. *Journal of Democracy*, *34*(4), 141–150. https://doi.org/10.1353/jod.2023.a907695

Yang, T., Majó-Vásquez, S., Nielsen, R. K., & González-Bailón, S. (2020). Exposure to news grows less fragmented with an increase in mobile access. *PNAS: Proceedings of the National Academy of Sciences of the United States of America*, *117*(46), 28678–28683. https://doi.org/10.1073/pnas.2006089117

Young, I. M. (2002). *Inclusion and democracy.* Oxford University Press. https://doi.org/10.1093/0198297556.001.0001

Zaller, J. R. (1992). *The nature and origins of mass opinion.* Cambridge University Press. https://doi.org/10.1017/CBO9780511818691

Zammito, J. H. (2012). The second life of the "public sphere": On charisma and routinization in the history of a concept. In C. J. Emden & D. Midgley (Eds.), *Changing perceptions of the public sphere* (pp. 90–119). Berghahn Books.

Zhang, Y., Shah, D., Foley, J., Abhishek, A., Lukito, J., Suk, J., Kim, S. J., Sun, Z., Pevehouse, J., & Garlough, C. (2019). Whose lives matter? mass shootings and social media discourses of sympathy and policy, 2012–2014. *Journal of Computer-Mediated Communication*, *24*(4), 182–202. https://doi.org/10.1093/jcmc/zmz009

Zhu, Y. (2022). *Hollywood in china: Behind the scenes of the world's largest movie market.* The New Press.

Ziegele, M., Breiner, T., & Quiring, O. (2014). What creates interactivity in online news discussions? an exploratory analysis of discussion factors in user comments on news items. *Journal of Communication*, *64*(6), 1111–1138. https://doi.org/10.1111/jcom.12123

Zuckerberg, M. (2017). I believe the most important thing we can do is work to bring people closer together. *Facebook*. https://www.facebook.com/notes/393134628500376/

Zwolinski, M., & Tomasi, J. (2023). *The individualists: Radicals, reactionaries, and the struggle for the soul of libertarianism.* Princeton University Press.

A. Digitaler Wandel im Kaleidoskop der Sozialwissenschaft

A.1. Wie kann die Sozialwissenschaft *digitalen Wandel* und den *Wandel durch das Digitale* abbilden und erklären?

Es freut mich sehr, heute vor dieser interdisziplinären Runde sprechen zu dürfen.[1] Heute und in den kommenden Tagen kommen in Trier Journalismusforscherinnen und -forscher, Kommunikations- und Politikwissenschafterinnen und -wissenschaftler für die Fachtagung *Politischer Journalismus: Konstellationen - Muster - Dynamiken*[2] zusammen. Dabei nähren wir uns unseren Forschungsgegenständen mit unterschiedlichen Interessen, Sensibilitäten und vor dem Hintergrund unterschiedlicher Forschungstraditionen.

Aber es eint uns eine geteilte Aufgabe:

Wie können wir den *digitalen Wandel* und den *Wandel durch das Digitale* in Journalismus und politischer Kommunikation abbilden und wie können wir die zugrundeliegenden Dynamiken und Effekte erklärbar machen?

Diese Aufgabe stellt uns vor unterschiedliche Fragen und Herausforderungen:

Woher wissen wir eigentlich, wie digitale Medien im Journalismus und der Politik *tatsächlich* genutzt werden? Klar, wir wissen, *dass* digitale Medien genutzt werden. Aber über diesen recht trivialen Befund hinaus wissen wir erstaunlich wenig darüber *wie* sie tatsächlich genutzt werden und welchen Einfluss sie auf Organisationsstrukturen oder Erfolge und Misserfolge von Akteuren tatsächlich haben. Detaillierte Arbeiten zu tatsächlicher Nutzung und entsprechenden strukturellen Anpassungsprozessen liegen uns überwiegend aus den USA vor (Christin, 2020; Karpf, 2012a; Kreiss, 2012). Daraus jedoch 1:1 Nutzungsmuster und Effekte für Deutschland abzuleiten, verbietet sich von selbst.

Auf der Seite von Nutzerinnen und Nutzern stellen sich auch neue Fragen. Wie können wir heute eigentlich noch die Mediennutzung von Menschen messen? Konnten Kolleginnen und Kollegen in den 90ern und 2000ern noch halbwegs legitim nach Medienkonsum

[1] Der Text ist die leicht erweiterte Fassung meiner Keynote zur *Gemeinsamen Jahrestagung der DGPuK-Fachgruppe Journalistik/Journalismusforschung, der DGPuK-Fachgruppe Kommunikation und Politik, dem Arbeitskreis Politik und Kommunikation (DVPW) und der Fachgruppe Politische Kommunikation (SGKM)* an der Universität Trier vom 29. September 2022.

[2] https://www.poljour22.de

in Fernsehen, Zeitung oder Radio fragen, splittert das heutige real-existierende digitale Mediensystem mögliche Nutzungsformen so weit auf, dass sich die Abfrage in Umfragen fast verbietet. Die Optimistischen unter Ihnen mögen hier auf die Möglichkeit der Beobachtung von Nutzerinnen und Nutzern durch Webtracking hinweisen. Dies erfordert jedoch nicht unerhebliche Glaubensbekenntnisse in Fragen der Repräsentativität der Getrackten oder des getrackten Verhaltens (Jürgens et al., 2020).

Neben Veränderungen in der Arbeit von Organisationen und dem Verhalten von Menschen, stellt sich auch die Frage nach der Veränderung des Systems als Ganzes. Wie verschieben sich Informationsflüsse, die Struktur von Öffentlichkeit und welche Auswirkungen haben diese Prozesse auf gesellschaftliches Leben und die Demokratie? Diese Fragen sind einfach gestellt und in Meinungsbeiträgen im Feuilleton oder auf Twitter leicht beantwortet. Ihre systematische, wissenschaftliche und hinterfragbare Beantwortung ist jedoch deutlich schwerer – nicht zuletzt, da (wie eben gezeigt) ja schon auf der Ebene einzelner Organisationen und Menschen Wandel aktuell schwierig abzubilden ist.

Eine der wahrscheinlich schwersten Aufgaben, ist darüber hinaus die Frage nach dem counterfactual. Was wäre eigentlich, wenn es digitale Medien nicht gebe? Zum Beispiel schreibt verschiedentlich die öffentliche Wahrnehmung – aber auch Kolleginnen und Kollegen – digitalen Medien eine entscheidende Rolle in den Erfolgen von Rechtspopulisten zu. Aber ist der Brexit, der Erfolg Donald Trumps, oder der der AfD wirklich ohne digitale Medien nicht denkbar? Wie isolieren wir den Einfluss digitaler Medien vor dem Hintergrund anderer gesellschaftlicher Entwicklungen – wie der Finanzkrise, steigender Migrationsbewegungen in Folge von ökologischer, politischer oder wirtschaftlicher Destabilisierung von Weltregionen und der steigenden Fragilität von Lebensentwürfen hinsichtlich nationaler und internationaler Krisen? Sind die entsprechenden demokratischen und anti-demokratischen Gegenbewegungen zum politischen status quo wirklich ursächlich oder primär als kommunikative Phänomene zu erklären? Oder konstruktiver gefragt: Wo liegt der Beitrag von Kommunikationsmustern und Kommunikationsinfrastrukturen in den von uns aktuell erlebten Herausforderungen des status quo?

Darüber wie wir uns gemeinsam dieser Aufgabe stellen können und bessere und systematischere Antworten auf Fragen, wie die eben skizzierten, geben können möchte ich in der Zeit mit Ihnen gemeinsam nachdenken.

A.2. Wissenschaft im Wettbewerb der Erklärungen

Die Untersuchung und Erklärung des *digitalen Wandel* und des *Wandels durch das Digitale* sind einerseits aus wissenschaftlicher Sicht wichtige und spannende Aufgaben. Schließlich gilt es hier für die Sozialwissenschaft, Prozesse abzubilden und zu erklären, die noch lange nicht abgeschlossen sind, sondern sich erst noch voll entfalten. Dies bedeutet sowohl begrifflich, konzeptionell, theoretisch als auch methodisch mutig Neuland zu beschreiten und sich nicht zu scheuen, Altgewohntes zurück zu lassen, um den neuen Aufgaben und dem sich neu entfaltenden Gegenstand gerecht zu werden. Dabei ist die Geschwindigkeit des Wandels nicht nur Gegenstand *von* sondern auch praktische Herausforderung *für* Forschung. Vor zehn Jahren sprach der amerikanische Politikwissenschaftler Dave Karpf von diesen Herausforderungen in seinem Artikel "Social Science Research Methods in Internet Time" (Karpf, 2012b). Zugegeben, heute spricht Karpf zwar davon, dass sich die *Internet Zeit* seitdem verlangsamt habe (Karpf, 2019), aber als jemand, der mit und zu digitalen Medien in der politischen Kommunikation seit 13 Jahren forscht, sei mir erlaubt zu sagen, dass sich – zumindest für mich – die Dynamik und der Wandel in unserem Feld noch immer recht schnell anfühlt.

Aber unsere Aufgabe bleibt nicht nur rein wissenschaftlich. Wir sind nicht allein, im Versuch, den Einfluss digitaler Medien in Politik, Öffentlichkeit und Journalismus zu beschreiben, zu verstehen oder zu erklären. Tatsächlich könnte man meinen, dass die wissenschaftlichen Stimmen anderen Stimmen im Diskurs hinterherlaufen. Man denke nur an die breite öffentliche Popularität der Echo-Kammern oder Filterblasen-These, die wortgewaltig von einem Rechtswissenschaftler – Cass Sunstein (Sunstein, 2001) – bzw. einem politischen Aktivisten und Campaigner – Eli Pariser (Pariser, 2011) – aufgestellt wurden.

Wie inzwischen zahlreiche empirische Studien zeigen, ist die Popularität dieser Thesen wahrscheinlich nicht ihrer empirischen Tragkraft geschuldet (Rau & Stier, 2019). Stattdessen wird zu ihrem Erfolg nicht unwesentlich beigetragen haben, dass die überwiegend negativen politischen Effekte, die hier digitalen Medien zugeschrieben wurden, gut zu den Interessen von Medienhäusern und staatlichen Regulatoren passen. Diese nutzen vermeintliche Wahrheiten über den demokratischen Schaden, den digitale Medien anrichten, gerne, um ihre Kontrolle über digitale Kommunikationsumgebungen zu erhöhen und die Marktmacht US-amerikanischer Konzerne einzugrenzen. Zu ähnlich gerne genutzten, aber wissenschaftlich nicht haltbaren Thesen gehören die weitverbreitete Manipulation von Wählerinnen und Wählern in digitalen Kommunikationsumgebungen durch Beratungsfirmen wie *Cambridge Analytica* (Jungherr et al., 2020, pp. 124–130) oder die gesellschaftliche Zersetzung durch digitale Desinformation (Jungherr & Schroeder, 2021; Mercier, 2020).

Doch auch Plattform-Firmen selber sind genauso darum bemüht, den Diskurs zu prägen. Sei dies durch die Beschäftigung digitaler Vordenker im eignen Haus oder die großzügige Kooperation mit einigen ausgewählten Wissenschaftlerinnen und Wissenschaftlern. Wie immer, wenn es um Geld oder politischen Einfluss geht, liegt die Vermutung nah, dass diskursives Ringen um Deutungshoheit nicht primär durch tatsächliches Erkenntnisinteresse bestimmt ist. Hier gilt es ganz grundsätzlich für die Wissenschaft, sich den freundlichen Sponsoren in Medienhäusern, Digitalwirtschaft und Staatskanzleien zu entziehen und die von ihnen präsentierten Narrativen kritisch zu hinterfragen.

Die Sozialwissenschaft steht in einem direkten Wettbewerb um das was der Wissenschaftssoziologe Thomas Gieryn als "epistemische Autorität" bezeichnet; also, "the legitimate power to define, describe, and explain bounded domains of reality" (Gieryn, 1999, p. 1). In diesem Wettbewerb gilt es für Wissenschaftlerinnen und Wissenschaftler klar ihre Unabhängigkeit gegenüber den saisonalen Schwankungen, Enthusiasmen, Ängsten und Interessen im öffentlichen Raum zu bewahren. Stattdessen braucht es klar kommunizierte und öffentlich kritisch überprüfte und hinterfragte wissenschaftliche Standards in der Entwicklung von Begriffen, Theorien und Konzepten und der nachvollziehbaren und hinterfragbaren Entwicklung empirischer Evidenz.

Diese prozessgeschaffene Autorität gilt es um so stärker zu entwickeln, einzufordern und ihren Bruch zu sanktionieren, da Wissenschaftlerinnen und Wissenschaftler rein diskursiv häufig anderen Teilnehmerinnen und Teilnehmern unterlegen sind. Wie es kürzlich der meinungsstarke Publizist und Medienkritiker in eigener Sache Richard David Precht leicht paraphrasiert sagte[3]: Wissenschaftler seien eben nur Autoren von Briefen unter Freunden. Für öffentliche Aufmerksamkeit brauche es medienstarke Menschen wie ihn (Lanz & Precht, 2022, 42:05-43:02). Geht es allein um diskursive Macht und nicht um prozessgestützte und hinterfragbare Autorität, dann ist es in diesem Wettbewerb schlecht für die Wissenschaft bestellt. Gelingt es uns nicht, den Wert von auf prozessbasierter Autorität entwickelten Befunde und Einsichten zu etablieren und öffentlich zu verteidigen, sind wir darauf reduziert, Bausteine, für Medien-Personalities – wie Richard David Precht – zu liefern in der Hoffnung, dass diese Bausteine in deren situative Agenden und Markenentwicklungssterben passen.

Das öffentliche Interesse an Fragen des digitalen Wandels begegnet auf Seiten der Kommunikationswissenschaft dem Wunsch nach stärkerer Präsenz im vorwissenschaftlichen Raum. Man denke nur an den jüngsten kommunikationswissenschaftlichen cri du coeur – eindrucksvoll durch Rasmus Kleis Nielsen in den Seiten von *Political Communication* formuliert (R. K. Nielsen, 2018) – wie es denn sein könne, dass politische Entscheidungsträger, Journalismus oder die Öffentlichkeit ganz allgemein gefühlt allen anderen Wissenschaften in Fragen des digitalen Wandels von politischer Kommunikation und Nachrichtenmedien folgen würden *außer* der Kommunikationswissenschaft.

Lassen wir für einen Moment die spalterische Frage beiseite ab welchem Grad von Präsenz im öffentlichen Diskurs oder regulativen und wirtschaftlichen Entscheidungsprozessen, die von Nielsen formulierte Unzufriedenheit in Teilen des Feldes behoben ist. Wenden wir uns lieber einer grundlegenderen und zugleich unangenehmeren Frage zu:

Auf die Frage "Warum werden wir nicht häufiger gefragt?" folgt recht natürlich die Frage "Warum sollten wir denn gefragt werden?" oder anders gesagt "Taugen unsere intern wissenschaftlich rezipierten und diskutierten Befunde eigentlich als Grundlage gesellschaftlichen oder regulativen Handelns?" Sind wir als Feld also *ready for prime time*?

Jüngst gab es in der Psychologie eine vergleichbare Situation: In einem Akt quasihomerischer Heroik meldeten sich in den Seiten von *Nature Human Behavior* einige Kolleginnen und Kollegen um Jay Van Bavel mit ihrer Expertise freiwillig im Kampf gegen Covid-19 und der Gestaltung pandemischer Gegenmassnahmen (Van Bavel et al., 2020). In gesellschaftlichen Großkrisen ist natürlich jede helfende Hand gerne

[3] https://youtu.be/OvUSVSdr-zI?t=2515

gesehen, allerdings stellt sich gerade bei wissenschaftlichen Beiträgen auch die Frage, wie belastbar die Befunde und Theorien sind auf deren Basis großzügig Ratschläge erteilt werden. Und gerade die in diesem Fall stark bemühte *Behavioural Psychology* hat in den Fragen Belastbarkeit oder gar Reproduzierbarkeit von Befunden ein eher schillerndes Profil. Entsprechend gab es in den Seiten desselben Journals auch schnell konstruktive Gegenrede. Hier äußerten Kolleginnen und Kollegen um Hans IJzerman Zweifel daran, dass bei allem wissenschaftsinternen Fortschritt der Psychologie und der Sozialwissenschaft wenig Prozesse und Standards im Feld vorliegen, die erlaubten zwischen Befunden zu unterscheiden, die im wissenschaftlichen Prozess und Diskurs interessant und wertvoll sind und solchen, die tatsächlich robust genug sind, um auf ihrer Basis bevölkerungsweite Maßnahmen und Interventionen auszurollen (IJzerman et al., 2020).

Als ein Beispiel hierfür lässt sich vielleicht der vermutete Einfluss von *Motivated Reasoning* auf die Überzeugungskraft kommunikativer Interventionen nennen. Viele Studien in Labor- und Umfrageexperimenten deuten auf den abschwächenden Einfluss von Voreinstellungen zu Sendern oder Themen auf die Überzeugungswirkung von kommunikativen Inhalten (Kahan, 2016a, 2016b). Diese Befunde werden häufig als Argument für die vermeintlich schwache Wirkung politischer Information, die Irrationalität von Empfängern politischer Informationen und die generelle Unbeweglichkeit von politisch überzeugten Menschen verwendet: Kurz, politische Information hat keine Wirkung und Menschen glauben ohnehin nur das, was sie bereits glauben (Achen & Bartels, 2016). Gleichzeitig sind viele der entsprechenden Befunde entweder auf Basis viel zu kleiner Fallzahlen oder im heißesten denkbaren politischen Konfliktumfeld der USA erhoben. Entsprechend haben Studien mit größerer Fallzahl und in anderen Kontexten verschiedentlich gezeigt, dass Voreinstellungen keine oder vernachlässigbare Effekte auf die Wirkung politischer Kommunikations-Interventionen haben (Coppock et al., 2020; Jungherr et al., 2021). Vielleicht ist Motivated Reasoning also nicht die beste Basis, um Rückschlüsse auf die Wirkung – oder Nicht-Wirkung – politischer Kommunikation generell zu ziehen oder die Empfänger politischer Information grundsätzlich als für nicht-überzeugbar zu erklären.

Wie kann Sozialwissenschaft aber nun angesichts dieser Herausforderungen *digitalen Wandel* und den *Wandel durch das Digitale* besser abbilden und für Gesellschaft erklärbar machen?

Hierfür müssen wir uns zuerst das Prisma sozialwissenschaftlicher Forschung ansehen.

A.3. Das Prisma sozialwissenschaftlicher Forschung

In der Sozialwissenschaft sind wir gewohnt, gesellschaftliche Phänomene aus spezifischen Blickwinkeln zu untersuchen. Dies öffnet vielversprechende, aber häufig isolierte, Perspektiven auf unsere Forschungsgegenstände. Wie ein Prisma Licht in das dem zugrundeliegende Farbschema bricht, so tut dies auch die Sozialwissenschaft mit ihren Gegenständen.

Auf die Gefahr hin, unzulässig zu vereinfachen, können wir zwischen drei unterschiedlichen sozialwissenschaftlichen Perspektiven auf Gesellschaft unterscheiden:

- Strukturen,
- Sprache und
- Akteure.

Diese drei Ansätze bilden wichtige Facetten von gesellschaftlichen Leben ab und haben uns erfolgreich unterschiedliche Aspekte digitalen Wandels und Wandel durch das Digitale aufgezeigt.

Der Blick auf Strukturen, die Rahmen für menschliches Verhalten setzen und es dadurch formen, öffnet hier unterschiedliche Perspektiven. Seit Beginn der Forschung zu gesellschaftlichen Effekten des Internets steht die Frage nach der Wirkung von Netzwerk-Strukturen im Vordergrund (Benkler, 2006; Castells, 2009/2013; Easley & Kleinberg, 2010; Rainie & Wellman, 2012). Was geschieht mit Gesellschaften, die nicht mehr in vertikalen Hierarchien organisiert sind sondern in Netzwerken? Hier herrschte die Erwartung vor, dass Netzwerkstrukturen generell zu Demokratisierung und dem Abbau von Ungleichheit in Gesellschaften beitragen würden. Eine optimistische Erwartung an Effekte der Digitalisierung, die heute wohl so nur noch selten geteilt wird.

Strukturen, die heute stärker im Blick der Forschung stehen sind Plattform-Firmen und Algorithmen als formende Strukturen für Informationsflüsse und Informationsräume. Plattform-Geschäftsmodelle bieten Räume in denen unterschiedliche Anbieter auf Nachfrager treffen können – sei dies z.B. ein Kleinanzeigenmarkt wie Ebay, Mitfahrmöglichkeiten wie bei Uber oder Werbekunden auf der Suche nach interessanten Kontakten wie bei Facebook oder Google. In der Wirtschaftswissenschaft werden Plattform-Geschäftsmodelle noch immer sehr positiv diskutiert (D. S. Evans & Schmalensee, 2016; Parker et al., 2016; Shapiro & Varian, 1999). In der Sozialwissenschaft wird auf diese Strukturen jedoch kritischer geblickt, nicht zuletzt da, sobald sich Plattformen als erfolgreiche Vermittler in Märkten etabliert haben, sie zu quasi-Monopolen werden können und dadurch starken Einfluss über Anbieter und Nachfrager ausüben (Chia et al., 2020; Gorwa, 2019b; Jürgens & Stark, 2017; R. K. Nielsen & Ganter, 2018; van Dijck et al., 2018). Der Blick auf Strukturen öffnet hier eine wichtiger Perspektive auf gesellschaftliche Machtverschiebungen im Kontext der Digitalisierung.

Im Kontrast zu ökonomischen Strukturen, wie Plattformen, zeigt der Blick auf Algorithmen (Diakopoulos, 2019), den Einfluss technischer Strukturen. Hier stellen sich Fragen nach dem Effekt von algorithmisch strukturierten Kommunikationsräumen. Dabei dominiert die Vorstellung der Filterblase, die Vermutung, dass Algorithmen Menschen primär in politisch homogene Informationsumgebungen und Räume sortieren würden und damit die *politisch anderen* verstecken würden (Pariser, 2011). Dabei erscheint es aus heutiger Perspektive interessanter, die Frage zu stellen, ob Algorithmen uns die *politisch anderen* nicht sogar verstärkt anzeigen, anstelle sie zu verstecken (Settle, 2018). Der

Blick auf die prägende Kraft von technischen Strukturen sollte allerdings auch erweitert werden. Eine natürliche Erweiterung – wie kürzlich in der *Publizistik* von Mike Schäfer und Hartmut Wessler gefordert (M. S. Schäfer & Wessler, 2020) – ist die Untersuchung des formenden Einfluss von künstlicher Intelligenz (KI) (Jungherr, 2023a; Jungherr & Schroeder, 2023). Eine weitere zur Zeit noch nur selten diskutierte technische Struktur mit wachsender Bedeutung sind Game Engines (Jungherr & Schlarb, 2022), nicht zuletzt, da sie entscheidende Bausteine in der Entwicklung digitaler Dienste, Dienstleistungen und Geschäftsmodelle auf Basis von vernetzter augmented reality sind, ein sich aktuell vollziehender Entwicklungsschritt der Digitalisierung, die einige Kommentatoren als *Metaverse* bezeichnen (Ball, 2022).

Die Analyse von Strukturen und ihrer formenden Wirkung auf menschliches Verhalten hat eine lange Tradition in der Sozialwissenschaft und öffnet den Blick für ihre oft versteckte prägende Kraft. Gleichzeitig läuft die Analyse der Strukturen digitaler Kommunikation ähnliche Gefahren, wie andere primär strukturelle Theorien und Analyseansätze: Die Faszination mit strukturellen Faktoren kann die Handlungsfähigkeit und strategische Anpassungsfähigkeit von Akteuren unterschätzen. Nicht umsonst lesen sich viele der aktuellen Kritiken von Plattform- oder KI-Strukturen wie nur geringfügig aktualisierte Texte der Kritik kapitalistischer Strukturen aus einer früheren Phase der Sozialwissenschaft.

Ein anderer Forschungsstrang beschäftigt sich mit der Untersuchung von Sprache und dem Ausdruck gesellschaftlicher Macht und Dynamiken in ihr. Beispiele sind die Untersuchung von Sprache und Symbolen in digitaler Kommunikation – wie in der Untersuchung von digitaler Kultur und Nutzungspraktiken (W. Phillips & Milner, 2017; Shifman, 2016) –, die Untersuchung strategischer Nutzung von Sprache – wie in der Untersuchung von Agenda Setting und Framing Dynamiken (Jungherr, Posegga, & An, 2019; Posegga & Jungherr, 2019) –, der Untersuchung von Informationsflüssen zwischen unterschiedlichen Akteuren und Medien (Jungherr, 2014) oder der Untersuchung von öffentlich sichtbaren sprachlichen oder symbolischen Interaktionen zwischen Akteursgruppen in digitalen Kommunikationsumgebungen (Jürgens et al., 2011; Nuernbergk & Conrad, 2016; Toepfl & Piwoni, 2015; Ziegele et al., 2014). Dank steigender computergestützter Speicher- und Rechenkapazität ist hier seit kurzem ist hier auch die systematische Untersuchung von Bildern oder Videomaterial ein wachsender Forschungsbereich (Jürgens et al., 2022). Die Untersuchung von Sprache, Symbolen und Diskursen in digitalen Kommunikationsumgebungen ist vielfältig und vielversprechend. Gleichzeitig besteht hier auch immer die Gefahr, sich von der Faszination von Sprache oder Bildern ablenken zu lassen, und die Kontigenz ihrer Nutzung oder Wirkung zu vernachlässigen.

Die dritte Forschungstradition, auf die ich hier kurz eingehen möchte, ist die Untersuchung von Akteuren. Legionen von Forscherinnen und Forschern untersuchen Mediennutzung und ihre Effekte auf Einstellungen und Verhalten auf der Individualebene (Bryant & Oliver, 2009). Welche Quellen und Inhalte werden konsumiert und welche Effekte haben sie? Wie oben beschrieben, steht gerade die Mediennutzungsforschung vor nicht unerheblichen Herausforderungen durch die neue Vielfalt an digitalen Kanälen und Quellen. Gleichzeitig verspricht die statistische Analyse von Befragungs- oder Beobachtungsdaten die präzise Identifikation von Kommunikationseffekten in bestimmten Kontexten. Dieses Versprechen setzt aber natürlich auch die verantwortliche und sorgfältige Nutzung statistischer Methoden voraus, dies ist – insbesondere in Bezug auf die notwendige Stichprobengrößen für die Identifikation oft kleiner Effekte oder besonders für die Identifikation von Interaktionseffekten – leider nicht immer der Fall.

Arbeiten in diesen drei Forschungstraditionen haben unser Verständnis des digitalen Wandels und des Wandels durch das Digitale in den letzten Jahren deutlich voran gebracht. Gleichzeitig besteht hier auch eine Gefahr. Häufig bleiben Studiendesigns in einer der Traditionen verhaftet, ohne die Perspektiven oder Erkenntnisse anderer Forschungsansätze genügend in der Interpretation ihrer Ergebnisse zu reflektieren.

So entsteht ein Prisma sozialwissenschaftlicher Forschung. Durch dieses Prisma unterschiedlicher Perspektiven spalten wir das Phänomen Digitalisierung in unterschiedliche schillernde Dimensionen, laufen aber Gefahr, das Gesamtbild aus dem Blick zu verlieren. Dies ist nicht unbedingt ein Problem aus der Perspektive einzelner Studien. Natürlich kann ich die Effekte spezifischer kommunikativer Interventionen im Labor oder in einem Umfrageexperiment auf Studienteilnehmerinnen und Teilnehmer testen, ohne Strukturen der digitalen Mediennutzung zu berücksichtigen. Dies wird allerdings zu einem Problem, sobald auf Basis dieser Befunde Großthesen über die Wirkung digitaler Medien auf die Gesellschaft aufgestellt werden. Hier sind isolierte Befunde – egal wie wissenschaftlich rigide, transparent und reproduzierbar – immer nur im Kontext anderer Perspektiven und Einflussfaktoren zu interpretieren.

Isolierte empirische Fakten sind wichtig; aber empirische Fakten allein bergen die Gefahr, Fehlschlüsse zu ziehen sobald auf ihrer Basis unreflektiert Rückschlüsse auf das große Ganze gezogen werden. Um die tatsächliche Bedeutung und Verallgemeinerbarkeit empirische Fakten einzuschätzen, müssen sie zusammen mit anderen Elementen und Analyseebenen gedacht werden. In anderen Worten: Wir als Feld müssen Veränderungen in Strukturen, Sprache und Akteuren gemeinsam und verzahnt konzeptualisieren, um ein Verständnis der tatsächlichen gesellschaftliche Änderungen und Effekten zu bekommen.

Die Gefahren wenn dies nicht passiert, sehen wir aktuell in der Forschung und öffentlichen Kommunikation und Wahrnehmung zu den vermeintlichen Wirkungen und Gefahren von digitaler Desinformation.

A.4. Digitale Desinformation durch das Prisma der Sozialwissenschaft

Seit einigen Jahren erlebt die Forschung zur Desinformation in digitalen Kommunikationsumgebungen eine beeindruckende Konjunktur. In einer jüngst veröffentlichten kritischen Auseinandersetzung mit dem Feld diagnostizieren Chico Camargo und Felix Simon sogar, das Feld sei inzwischen “too big to fail” (Camargo & Simon, 2022), nicht zuletzt durch das starke öffentliche und journalistische Interesse an dem Themenbereich, der beachtlichen Förderaktivität durch Drittmittelgeber und das starke Interesse von Regierungen und Regulatoren. Dennoch zeigt genau dieses Forschungsgebiet eindrucksvoll einige der Gefahren, die entstehen, wenn wir digitale Phänomene durch das Prisma sozialwissenschaftlicher Forschungsansätze betrachten, anstelle die unterschiedlichen Perspektiven, wie in einem Kaleidoskop, gemeinsam zu nutzen.

Schauen wir zuerst aus der Perspektive von Sprache auf das Phänomen digitaler Desinformation. Hier gibt es tatsächlich Anlass zur Sorge. Digitale Kommunikationsräume sind voll mit bewusst oder fahrlässig falscher oder fehlleitender Information. Wir finden sie in Form von Falschinformationen auf gezielt für die Verbreitung von Falschinformationen eingerichteter Webseiten, in Form von Nachrichten auf sozialen Netzwerkseiten, in Form humoristischer Meme oder Kurzvideos oder auch in Form von Sprachnachrichten auf Messenger Apps (Donovan et al., 2022). Schaut man allein aus inhaltsanalytischer Sicht auf Beiträge in digitalen Informationsräumen kann einem tatsächlich Angst und Bange werden.

Aber wie ist es um das Phänomen bestellt, wenn wir andere Perspektiven zur Hilfe nehmen?

Die Analyse von Strukturen hinterlässt hier ein gemischtes Bild. Einerseits sehen wir, dass digitale Plattformen – wie Facebook, Twitter und YouTube – und Messenger-Dienste – wie WhatsApp und Telegram – erfolgreich genutzt werden, um Desinformation zu teilen (Bennett & Livingston, 2018). Anderseits zeigen unterschiedliche Interventionen von Plattformbetreibern – wie zum Beispiel Moderation, Friction, Metered Moderation oder Deplatforming – klar Wirkung in der Einschränkung und Verlangsamung der Verbreitung von Desinformation (Rauchfleisch & Kaiser, 2021). Ja, digital Plattformen können zur schnellen Verbreitung von Desinformation beitragen. Gleichzeitig können sie diese auch effektiv einschränken, wenn sie sich intern dazu entschließen oder der externe Druck dazu Anlass gibt. Ob man diese Interventionen aus demokratietheoretischer Sicht tatsächlich für angemessen oder normativ wünschenswert hält ist wiederum eine andere Frage, gehen sie doch zumindest potentiell mit einer Einschränkung von politischer Rede als Kollateralschaden einher (Keller, 2018).

Auch sind digitale Plattformen zwar wichtige aber bei weitem nicht exklusive kommunikative Infrastrukturen der heutigen *Public Arena* (Jungherr & Schroeder, 2022). Stattdessen spielen Massenmedien auf unterschiedlichen traditionellen und neuen kommunikativen Kanälen weiterhin eine sehr wichtige Rolle als Informationsanbieter auch in digitalen Kommunikationsräumen (Allen et al., 2020). Etwaige Desinformation bleibt also nicht unwidersprochen, sondern wird auch in digitalen Kommunikationsumgebungen durch qualitativ hochwertige Quellen kontextualisiert, widerlegt und angegriffen. Ein Blick auf Strukturen zeigt also die Kontingenz der Reichweite und Deutungshoheit von Desinformation in der Öffentlichkeit.

Der Blick auf die Akteursebene gibt weiterhin Anlass zur Entspannung. Es ist klar, dass es Desinformation in digitalen Kommunikationsumgebungen gibt. Deutlich weniger klar ist jedoch wer diese Desinformation sieht und welche Wirkung sie entfaltet. Die wenigen Studien, die versuchen die Reichweite von Desinformation empirisch zu ermitteln zeigen, dass sich Desinformationen überwiegend unter Menschen verteilen, die geteilte politische Einstellungen in Übereinstimmung mit der gesehen Desinformation zeigen (A. Guess et al., 2019; A. M. Guess et al., 2020). Weitere psychologische Studien legen nah, dass Desinformation weniger ein informierendes oder überzeugendes kommunikatives Phänomen ist, sondern wohl eher ein Inhaltstyp, der es Menschen erlaubt, ihre politische Identität zur Schau zu stellen und Menschen mit geteilter Meinung zu finden (Mercier, 2020; Nyhan, 2020). Das macht die geteilten Inhalte nicht schöner, aber wir haben es wahrscheinlich nicht mit einem Phänomen zu tun, dass Massen von Menschen dazu bringt, ihre Wahlentscheidung oder politische Zugehörigkeit auf Basis falscher Tatsachen anzupassen. Unter der Zuhilfenahme von Erkenntnissen der Mediennutzungsforschung und der Psychologie erscheint die anfängliche Sorge, ausgelöst von rein auf Sprache und Inhalten basierenden Analysen, also schon deutlich weniger dringlich.

Schon diese kurze – und zwangsweise kursorische Diskussion – zeigt, dass es für die zutreffende Analyse des öffentlichen Einflusses von digitaler Desinformation wichtig ist, unterschiedliche sozialwissenschaftliche Perspektiven gemeinsam zu sehen und nicht, auf Basis isolierter Befunde oder isolierter Perspektiven gesamtgesellschaftliche Aussagen zu treffen. Erst durch den Zusammenzug der Perspektiven werden ihre Einzelbefunde interessant und aussagekräftig.

Warum ist das wichtig?

Die zutreffende Einschätzung über die Präsenz von Desinformation in digitalen Kommunikationsräumen und ihres gesamtgesellschaftlichen Einfluss ist wichtig, da die öffentliche Kommunikation hierzu weitreichendere Wirkungen hat, als wir vielleicht zuerst vermuten.

In einem Umfrageexperiment haben Adrian Rauchfleisch und ich vor kurzem die Wirkung undifferenzierter und differenzierter Warnungen vor digitaler Desinformation getestet (Jungherr & Rauchfleisch, 2024). Wir testen die Wirkung von zwei Treatments. Ein Treatment hatte die Form eines typischen Medienberichts zur Desinformation, in dem undifferenziert auf Basis wissenschaftlicher Befunde vor den Gefahren digitaler Desinformation gewarnt wurde. Das zweite Treatment hatte dieselbe Form und Inhalte, ergänzte sie jedoch noch um wichtige Kontextinformationen auf Basis wissenschaftlicher Befunde, die die Wirkung und Verbreitung von Desinformation relativierten. Unser Experiment zeigte, dass Menschen, die undifferenzierte Warnungen vor den vermuteten Gefahren von digitaler Desinformation angezeigt bekommen, sowohl niedrigere Zufriedenheit mit dem aktuellen Zustand der Demokratie äußerten, als auch höhere Unterstützung für starke Eingriffe in Meinungsäußerung und Zensur in digitalen Kommunikationsumgebungen zeigten als Menschen denen differenzierte Information gegeben wurden und Menschen in einer Kontrollgruppe, denen keine Informationen angezeigt wurden.

Dieser Befund ist wichtig, da er zeigt, dass sensationalistische Wissenschaftskommunikation auf Basis einzelner Befunde – empirischer Fakten – selbst wenn sie gut gemeint ist Gefahr läuft, in der Gesellschaft einerseits zu einem falschen Bild der Effekte von digitaler Kommunikation beizutragen, als auch durch undifferenzierte und letztlich in ihrer Dringlichkeit unbegründeter Warnungen, insgesamt zu einem Anstieg von Angst und damit verbündenden Gefühlen von Kontrollverlust und Hilflosigkeit beitragen kann.

Wie wir über digitale Medien und ihre Wirkung auf die Demokratie sprechen hat Effekte. Und nicht nur solche, die wir erwarten oder wünschen. Hier müssen wir als Feld lernen und nicht nur auf Basis unserer jeweiligen Perspektive und vorliegenden isolierten empirischen Fakten kommunizieren. Statt der Spaltung des Gegenstands im sozialwissenschaftlichen Prisma müssen wir ein möglichst holistisches Bild im Kaleidoskop unterschiedlicher sozialwissenschaftlicher Perspektiven kommunizieren.

A.5. Sozialwissenschaftliche Perspektiven nicht als Prisma sondern als Kaleidoskop

Dieser kurze Abriss zeigt die Gefahren, die darin liegen wenn wir uns unseren Forschungsgegenständen ausschließlich durch das sozialwissenschaftliche Prisma nähren. Unterschiedliche Perspektiven brechen unsere Forschungsgegenstände in unterschiedliche Bestandteile. Dies allein ist nicht problematisch und kann sogar zu Klarheit und der mehrdimensionalen Einschätzung unserer Forschungsgegenstände beitragen. Es wird allerdings zu einem Problem sobald wir die durch unsere Perspektive sichtbaren Splitter des Gegenstands zum Ganzen erheben.

Alleinigkeitsansprüche der Perspektiven – ob Struktur, Sprache oder Akteur – sind das Problem. Sie verstellen uns den Blick auf den Gegenstand in seiner Gänze und seiner Kontextabhängigkeit. Entsprechend müssen wir unterschiedliche sozialwissenschaftliche Perspektiven nutzen, um so die unterschiedlichen Aspekte unserer Gegenstände in ihrer Gänze sichtbar zu machen und die Bedeutung unserer Befunde in einem Gesamtbild einzuordnen. Dies bedeutet, sich den unterschiedlichen Perspektiven mehr als Kaleidoskop denn als Prisma zu nähren.

Natürlich bedeutet dies nicht, dass jede Studie alle denkbaren Perspektiven vereinen und testen muss. Es bedeutet allerdings, dass wir in der Einordnung unserer Befunde – oder der öffentlichen Kommunikation über sie – zumindest versuchen müssen, die durch unsere Perspektive sichtbar gemachten empirischen Fakten vor dem Hintergrund anderer relevanter Perspektiven und Befunde einzuordnen.

Gleichzeitig kann dies auch bedeuten, ambitionierter in unseren Theorien und Konzeptionalisierungen zu werden. Niemand zwingt uns, das gefühlt 10.000ste Papier zu Echo-Kammern oder Filterblasen zu schreiben. Neue Theorien und Konzepte können uns erlauben, unterschiedliche Facetten digitalen Wandels aus unterschiedlichen Perspektiven vereint zu denken und messbar zu machen. Einen solchen Versuch haben Jisun An, Oliver Posegga und ich mit unserem Papier zu "Discursive Power in Contemporary Media Systems" gemacht (Jungherr, Posegga, & An, 2019). Zugegeben, unsere empirische Messung lässt noch auf sich warten, aber das Konzept vereint strukturelle Überlegungen und Faktoren mit der sprachlichen Abbildung von politischem Wettbewerb. Andere Gegenstände laden zu anderen Konzeptionalisierungen ein.

Unser Gegenstand ist neu, dem sollten unsere Theorien, Konzepte und Methoden gerecht werden.

Letztlich sollte wir als Sozialwissenschaftlerinnen und -wissenschaftler nicht mit der Produktion isolierter empirischer Fakten zufrieden sein. Empirische Fakten sind wichtig und die Grundlage für vieles was folgt. Aber um ihre Bedeutung einschätzen zu können, müssen sie theoretisch eingebettet und verankert sein. Darüber hinaus braucht es für den Transfer in die Wirklichkeit und realweltliche Einschätzung von Entwicklungen das Kaleidoskop unterschiedlicher Ansätze und Befunde, isoliert bieten sie uns nur ein Zerrbild.

Aktuell stellen sich uns und der Gesellschaft viele Fragen im Umgang mit digitalen Medien. Das öffentliche Interesse an unserer Arbeit ist groß und es bleibt viel für uns zu tun. Der Horizont ist weit und die Zukunft unsicher. Welch bessere Ausgangslage könnten wir uns für unsere Arbeit also wünschen?