We Haven't Gone Paperless Yet: Why the Printing Press Can Help Us Understand Data and AI

Julian Posada
julian.posada@mail.utoronto.ca
University of Toronto
Canada

Nicholas Weller
nweller@ucr.edu
University of California, Riverside
United States

Wendy H. Wong
wendyh.wong@utoronto.ca
University of Toronto
Canada

ABSTRACT

How should we understand the social and political effects of the datafication of human life? This paper argues that the effects of data should be understood as a constitutive shift in social and political relations. We explore how datafication, or quantification of human and non-human factors into binary code, affects the identity of individuals and groups. This fundamental shift goes beyond economic and ethical concerns, which has been the focus of other efforts to explore the effects of datafication and AI. We highlight that technologies such as datafication and AI (and previously, the printing press) both disrupted extant power arrangements, leading to decentralization, and triggered a recentralization of power by new actors better adapted to leveraging the new technology. We use the analogy of the printing press to provide a framework for understanding constitutive change. The printing press example gives us more clarity on 1) what can happen when the medium of communication drastically alters how information is communicated and stored; 2) the shift in power from state to private actors; and 3) the tension of simultaneously connecting individuals while driving them towards narrower communities through algorithmic analyses of data.

CCS CONCEPTS

• **Social and professional topics** → **History of computing**; **Government technology policy**; *Computing and business*; • **Applied computing** → *Digital libraries and archives*.

KEYWORDS

Printing press, politics of data, corporations, governance

1 INTRODUCTION

We are currently in the era of datafication. Datafication, as we use it here, is the process by which the world is processed, quantified, and stored digitally [24, 60, 61] and converted into binary code. One important development of datafication is the proliferation of AI technologies. Without the rise of "Big Data," data-intensive machine learning techniques would not have been able to make the enormous strides that they have and continue to do. With this increase in data production, analysis, and access, however, we have at the same time struggled to develop strategies to govern data, regulate AI, and articulate the social and political effects of datafication and artificial intelligence. Already, a mounting pile of quandaries have arisen about the scope and ethics around data with very few good answers. In this paper, we show that comparing datafication to the dynamics of the printing press' development in Europe can help to anticipate what might come from datafication and AI.

We argue in this paper that fundamental shifts in information technology – such as our current age of datafication – are characterized by the tension of a sequential decentralization of power enabled by the technology and recentralization of power as actors adapt to the new possibilities. Datafication has given private economic actors the ability to surveil individuals in ways we used to typically understand as belonging to states only [33, 19]. These changes in power are accompanied by changes in the identity of individuals and groups, what we label *constitutive* shifts. As with the printing press in Europe, datafication opened up the floodgates of what we know, and how quickly we can know it. Literacy as a minimal qualification was replaced with access – to devices, to platforms, to the Internet. With this came an emancipatory sense, as in the case of the printing press' break of the Church's stranglehold on words and reproducible knowledge.

Datafication, however, has also ushered in a new centralization of power, as those who developed the means to cull data most efficiently and widely have come to disproportionate amounts of power. Social media firms, such as Facebook and Twitter, are designed for gathering data [26], and they have come to dominate how we think about freedom of expression. This came to the fore with their decisions to allow Donald J. Trump and his allies to continue spreading misinformation throughout his presidency, and their choices to finally cut him off after the January 6, 2021 Capitol riots. Search firms, such as Google, and e-commerce companies, such as Amazon, determine how people see the world and to what they have access. These dynamics have intensified since the COVID-19 pandemic began.

To understand the extent to which datafication and AI might reshape life as we know it, we can look to analogies with shifts

wrought by the printing press. The printing press fundamentally changed life in early modern society and politics in three constitutive ways that are relevant to the current day. First, it moved society from one of oral orientation to one of visual orientation. It made communication individually contemplative, rather than shared. It also committed ideas to paper as a medium [25], which has its own organizational logics regarding how information is conveyed. Second, it shifted power from the Church and state as the keeper of knowledge to private actors. It realigned economic interests and power as well, empowering publishers and printers [79, 95]. Third, it allowed for greater possibilities for human imagination as "communities" with the growth of newspapers and the novel. The archive, as a repository for select documents and publications, created for some social purpose, then became a source of information, truth, and power [95].

The printing press is the most relevant technological change to look for as an analogy because it is changed the production of information, its reception and the way it is used, and its storage. Like the printing press, datafication fundamentally alters our social and political relationships because they store and share data in a much more more efficient format. This allows for the massive gathering of data. AI is the means by which we process and organize those data; AI also trains on that archive of data [47]. First, datafication and AI have moved us to a society of archiving and prediction, of the action taking place on servers and through algorithms indiscernible to the users. It is both more public on its face, and yet private in the way that the data are stored and used. Second, these forces have shifted power from states to private actors that develop and own the algorithms, and it is these actors who invent the categories of data that are to be collected, and implement the algorithms to collect and analyze data from users. While it is true that individuals have more access to data and information, it is also true that data are being collected from individuals in a range of ways that are both voluntary and involuntary [90]. Third, it has both created more opportunities for communities to grow while also providing the opportunity for deeper fractures as algorithms shuffle people into different "filter bubbles." Algorithms have become arbiters of truth and power by determining what data are viewed, by whom.

In Section 2, we review two prominent, alternative ways to think about the effects of datafication and AI, through their economic and ethical implications. In Section 3, we explore the three constitutive ways in which the printing press' effect on social and political structures in Europe hold key lessons for us as we face a widespread shift in how information is conveyed and communications are conducted. In this section, we particularly focus on the simultaneous decentralization forces of the technology that is coupled with the countervailing tendency for actors to centralize access and influence. Section 4 concludes with some proposals on how to integrate insights from the constitutive changes wrought by technological change to policymaking.

2 WAYS OF LOOKING AT DATAFICATION AND AI

Datafication has many different definitional enhancements, depending on the author, but its main point often focuses on what Mejias and Couldry call "[a] contemporary phenomenon which refers to

the quantification of human life though digital information, very often for economic value" [61, 1]. The importance of datafication, however, is not just in the generation of initial data, but the value that can be extracted from that data [61]. Put differently, the value of data lies in both collecting and drawing inferences from that data. Thus, datafication would not have nearly as much influence without the capabilities of AI to leverage the data into predictions.

2.1 Economics

The economics of datafication have focused on the financial imperatives and gains for those who create data from observable reality [80, 88]. Terms such as "surveillance capitalism" [96] permeate the public imagination to convey how companies are harnessing the data we generate to create wealth for themselves while also controlling our access to information. Others, such as Fourcade and Healy emphasize the stratification and classification of consumers, through the analysis of the data they generate, and the economy of moral judgment that follows from categorizing "good" and "bad" consumers [33].

The collection and flows of these datasets have been described by Jo and Gebru as influenced by a *laissez-faire* attitude from practitioners and tech companies with little care about the social implications of the data they collect [47]. This culture is related to the influence of the idea that data is a nonrival good (not depleted by consumption) and a by-product of economic activity [49, 89]. Some claim that data can be "owned" privately by firms or consumers and transacted in markets [46]. Put into practice, this conception of data as a market good creates a pervasive environment for the collection and processing of data to train machine learning algorithms that do not account for the rigorous and ethical processes of other data-intensive sectors and fields such as healthcare and social science disciplines [72].

Another influence of economic and business thinking that permeates the datafication process required for AI is the desire to create scalable systems. Scalability is the expansion of a system without the modification of core features [87]. This notion of economic growth through the exploitation of nonrivalrous datafied goods has become fundamental in the startup culture that fuels some recent innovations in artificial intelligence. For Hannah and Park, this scalable thinking permeates the development of predictive systems with the following set of assumptions: that these systems are ethically desirable, depend exclusively on the quantification of human experience, and that a system of any scale requires only a limited number of core functions [38]. In practice, scalable thinking would mean that a system developed in the Silicon Valley could potentially grow to be implemented in other parts of the world without changing its core features, being only necessary to adapt superficial characteristics to the necessities of new markets.

The economic perspective on datafication does not take into account how these data-driven avenues pursued for market advantage fundamentally alter the way that humans interact. The view from here emphasizes why firms choose certain types of data collection and analysis with a simplified (if any) account of the social and political effects of these actions.

2.2 Ethics

There is also the political and ethical perspective on datafication, which can be captured best by the phenomenon of "dataveillance." Dataveillance can be defined as "the name for the disciplinary and control practice of monitoring, aggregating, and sorting data" [76, 124]. Unlike surveillance, which monitors the physical being, "dataveillance watches the shadow that the person casts as they conduct transactions, variously of an economic, social or political nature" [20, 20, 21]. Thus, technology is allowing for greater levels of intrusions upon individual lives, at times justified by security needs.

Since the applications of AI have become too important to be ignored by governments, institutions, and corporations, there has been an increasing interest in how the deployment of this technology would benefit society. In this context, discussions about "ethical AI" have permeated a sector of academia and policy for the past years. Carly Kind, director of the Ada Lovelace Institute, has recently identified three different waves in these discussions, each of them influenced by major academic disciplines [52]. Calling them "waves" does not mean that research from these frameworks has stopped. Instead, new disciplines and perspectives have complemented these different research streams.

The first wave, heavily influenced by philosophy, has focused on general ethical principles that AI systems should follow. These principles came primarily from dominant industry and governmental actors and converged in ideas such as transparency and fairness [48]. However, these principles remained abstract and with no clear consensus on how to implement them, leaving meaningful discussions on policy and regulation behind [14]. Furthermore, these discussions have been criticized as "ethics washing" because of their lack of applicability [39] and funding from "big tech" [1].

A second wave tries to resolve some of the dilemmas from the first technically. Computer science has focused on how algorithms can become "fair" and "unbiased," exploring ways to collect more diverse data while solving algorithmic discrimination in mathematical models. Research on fairness from a computational perspective has focused on developing models that address the difference in "performance" between categories such as "sub-populations" and "individuals" [19, 23, 30, 86]. However, while it is essential for these mathematical models to address the data differently for categories or sub-groups, thinking about "fairness" purely in statistical and mathematical terms does not fully address AI's societal implications. These predictive systems are sociotechnical in nature [52]. An understanding of social contexts and relations that go beyond computational models is necessary for constructing and implementing algorithms [27, 37, 62].

The third wave is much more heavily influenced by the social sciences. It explores the power relations behind the development and deployment of artificial intelligence [8, 50, 63]. A subset of this research on AI from a sociotechnical perspective looks at the data supply chains that allow the creation of machine learning systems from the datafication process for input, training, and feedback data [2, 64]. Beyond the widely discussed issues of privacy in data collection for consumers from the social sciences [5, 58] and computer science [43], these data supply chains also involve a different array of actors, including data workers [75], data brokers [4], and other institutional and governmental actors.

This final wave draws attention to some of the directly-observable trends that datafication and AI have brought on. We start from this perspective, and theorize more broadly about the effects of these technologies. Using the printing press' history as a guide, we "widen the lens" to consider the findings from the third wave for bigger, constitutive effects of datafication and AI.

3 WHY DATAFICATION AND AI ARE ANALOGOUS TO THE PRINTING PRESS AND ARCHIVE

3.1 An Archiving and Predicting Society

Datafication provides the raw material for AI to do what it does best: provide predictions [2]. The data for these predictions, however, are coming from participants in a relationship of unequals between the companies that collect the data and make predictions, and the individuals whose data are being gathered and analyzed. As consumers, we have participated in helping the rise of Big Data and the greater accuracy of machines [67]. Giving away this information has the effect of transferring power to those who archive the data, but it also has ramifications on individual agency. For example, Cathy O'Neil provides an easily-accessible laundry list of ways that discrimination and bias are replicated in the algorithms that determine what jobs we are qualified for and whether we get home loans [69]. Ruha Benjamin has powerfully demonstrated how these dynamics reinforce existing racial inequities in what she calls "the new Jim Code" [7]. However, much of our participation has not occurred in a situation of transparency, awareness, or even the ability to easily disentangle ourselves even if we were aware of the situation.

As with the printing press, where the medium upon which ideas were communicated also generated expectations of what knowledge "looked like," datafication has generated expectations about the medium in which communications take place. These communications media are data-intensive. In 2019, 4,497,420 Google searches were performed and 55,140 photos were posted on Instagram every minute [29]. This is based on just over half of the world's human population being online. All of these data are providing new ways in which information is communicated and absorbed by users of the Internet. And these data generate more data about what users want and do not want, how ideas trend, and what kinds of concerns are of the moment. These are all reflections of individual taste, personality, preference – in short, the data give the collectors of that data a good sense of "who" their users are [90]. They can also give away what kinds of relationships people have to one another, and the quality of those relationships, something that has been called "relational big data" [54].

The strength of the prediction is only as good as the data the algorithm trains on. Our society is increasingly geared towards the collection of data, for the purposes of improving prediction through AI. These are quite frequently tied to monetary imperatives, because the entities doing the collecting are mostly companies. Governments have also done their fair share of using data, but often that data is provided by companies with technologies to gather data [59].

The orientation towards sharing in public fora requires that users interact with data-gathering technology. It also has the effect of hiding where all the data go after we have provided them, whether this is our behavior online as tracked by cookies, or the photos we upload and share on social media. One particular example is with facial data, and the use of facial recognition technology. Though not a new technology – it started in the 1960s [77] – it has become commonplace, showing up in apps like Facebook or iPhoto, in smart home products like Google Nest, and in car safety applications like the Subaru DriverFocus Distraction Mitigation System. It is used in airports, on city streets, at Taylor Swift concerts [15], in retail stores, and more [9]. It is also used by police to do their jobs. But police had not had access to the numbers of faces and the volume of facial data until Clearview AI stepped onto the scene. Clearview claims it has amassed 3 billion face images, all scraped from websites such as Facebook, YouTube, and Venmo [45]. Clearview works with over 2400 law enforcement agencies [55], and helps companies with their security too [45]. These data – the archives of faces – are sometimes voluntarily gotten in the first instance, as in the case of Clearview's database, which was taken from photos that users had voluntarily uploaded, albeit for some other purposes. Once in the face databases, it is not clear how one might get out, if ever, creating questions about to whom a face's data belong [92].

An important issue with facial recognition technology is how individuals within a datafied society consent to the collection, archiving, and use of the data they provide. There is not a good way to think about this from a governance perspective, either in terms of how one might consent to the myriad items that app and service creators embed in their terms and conditions; or how devices such as FitBit are collecting vital information that at worst constitute surveillance and social conditioning, but at best are just invasive [34, 21-26]; or from the view of how someone might be able to just forfeit their right to privacy because of the way datafication renders such a right quite difficult to protect [81]. Furthermore, even if individuals did consent, the biases baked into the way our data are collected and the algorithms are written continue to vex the effectiveness and accuracy of technologies such as facial recognition [7, 10, 18, 66]. Recent reports of false arrests of Black men, based on facial recognition technology, for example, highlight how the technique replicates existing discrimination and subjugation of racialized groups [44]. In general, studies found that facial recognition technology was just not very accurate in matching faces in real time to those in the database. In one set of trials in the UK, the accuracy rate was reported to be 19.05% [11].

Not surprisingly, our institutions have not caught up with the sheer volume of datafication, or the technical advances made to improve AI technologies to assess that data. We are being asked to reevaluate how we live our lives, understand our rights, and know where our data are going, when the entire system of datafication of communications, social interactions, and knowledge transfer are happening on a completely new medium that is not the same as the ones with which we have more intimate knowledge, such as paper. The move to datafication and the use of AI to move through those vast stores of data can be better understood if we take the view from the transition from hand-scribing to printing press. Where prior to the printing press, the word was literally sacred because very few could read [25, 49-52], the same applies to data today.

Data have been mysterious and the domain of technical specialists, because they have been the ones creating the tools that are widely used today. The bait-and-switch has been that individuals have all been allowing their lives to be datafied, and very much giving away some part of themselves in this technological transition. Thus, it seems quite reasonable to say that everyone alive today has a very real stake in how human lives are being converted into datafied forms.

3.2 Changes in Power Structures Wrought by Technology

The development of the internet and datafication has been compared to the printing press by a multitude of scholars. New communications media not only establish the forms in which individuals communicate, but also create new ways to interact and the possibilities for new kinds of social ties [85]. Unlike other media innovations, such as the telegraph or the related invention of the telephone, the printing press produces communication on a mass level, intended for a wide audience, and serves medium of communication that fundamentally shifted the way that human being relate to one another in time and space. Datafication, as described above and discussed below, has very much these same qualities. But unlike the printing press, or telephone and telegraph, datafication's link to AI has the additional quality of culling data directly from all who use the technology.

Research on the social and political effects of the printing press are not new. Eisenstein argues that the effects of the printing press were not fully appreciated, because scholars had focused mostly on how the press affected the dissemination of ideas [31]. She demonstrates that the press led to fundamental changes in what people thought and how they thought about the world in such profound ways that it influenced the Reformation, the Renaissance, and the Scientific Revolution.

The printing press has also been connected to other massive changes in social and political organization and the distribution of power. In *Imagined Communities*, Benedict Anderson argues that the newspaper led to the rise of nationalism, as it made it possible for disparate individuals to share a common language, time, calendars, and other factors that built a shared community [3]. Deibert [25] also focuses on the role that the printing press had on power relations to argue that it was an important cause in the development of the modern nation-state system, and he argues that new technologies are likely to disrupt current power relations.

More specific changes in power distribution were wrought by the printing press. In the Russian Revolution context, the printing press created a group of workers who were important to the structure and organization of society but who did not exist prior to the creation of the printing press [79]. In fact, printers were among the first to organize. Their role in the production of reading materials gave them a unique skillset both for the access to printing and also in their high levels of literacy. They were instrumental in undermining censorship laws early in the 20th century.

From Ruud's history, we can see that the printing press created a new class of employees that went from non-existent to politically powerful in a short period of time. In the early 1900s Russian printers took the lead in the development of a free press and therefore

this occupational group which had not even existed 50 years earlier was now leading radical changes in Russian society. Ruud argues that "the modern printing press was itself a powerful – but so far little recognized – 'agent' of political change during the Revolution of 1905 in Russia" [79, 395]. The printing press was not developed to aid in revolutionary political change, but profoundly changed how power was distributed. The case demonstrates how technologies can have unanticipated consequences in terms of usership, skill development, class creation, and group mobilization, all of which can contribute to shifts in power. In the case of Russia, this new class of printers were able to halt the repressive policies of the czarship and foment support for the growing revolution.

The Russian example demonstrates how a new technology created a class of private actors that changed society. Although the process is different, datafication has likewise empowered individuals and businesses that have the potential to change society and politics. Corporate actors have become key players in the amassing of data generated by individuals interacting with technology. The biggest players in AI are just nine firms based in China and the US [91]. The voracious data appetites of social media [26], search, and platform providers have been well-documented [96]. Their imperative to keep their user base (and keep growing) is understood [51], as without users, these firms would not have their advantageous positions in data access.

As with the printing press, the advent of datafication as a technology has ushered in actors who have been able to capitalize on the environmental shift that privileges data collection, archiving, and prediction. The current US Department of Justice anti-trust suit against Google speaks to the ways that the company controls search renders it the ubiquitous keeper of information. The outcome of Google's data collection, archiving, and predicting efforts is that they effectively control the public record [66]. However, even as it serves a public purpose in monopoly fashion, Google responds to market incentives: its advertisers [66, 36]. Together with social media giants such as Facebook, Tiktok, and Twitter and app gatekeepers such as Apple and Amazon, they effectively gatekeep what we know, how we know it, and if we know it. This is analogous to how control over the printing press limited what information was received, in what format, and by whom. For example, European colonizers in Latin America regulated printed materials in order to control the spread of information and ideas. Later on, clandestine presses were fundamental in spreading revolutionary ideas that sparked the wars of emancipation in the early 19th century [78].

This stands in contradiction with much early discussion of the internet, which talked about it as a democratizing force in its ability to diffuse power away from the elites and to mass citizens in a way that appears analogous to the printing press from centuries earlier [28]. While clearly information is more accessible than ever before for ever larger segments of society, there is a countervailing trend associated with the collection and analysis of massive amounts of data. "Big Data" has created new economic actors: the (mostly) firms that amass and analyze these data. Just as the press moved power and information away from government and religious institutions and into the hands of private actors such as printers, datafication and AI have given certain kinds of technology companies a leg up (so-called "Big Tech"). Perhaps most importantly, the more we use these technologies, the more powerful they get [71, 14]. The users

use Google search or scroll through their Twitter feeds, the more data these companies collect, the more analysis they provide, and they more they understand about individuals and their users as a collective.

Unlike today's internet landscape, characterized by the prevalence of private enterprises and dominated by a handful of them, sometimes called "Big Tech" or GAFAM (Google, Apple, Facebook, Amazon, and Microsoft), the early infrastructure of the Advanced Research Projects Agency Network (ARPANET), the ancestor of the internet, and subsequent technical developments before its privatization, were financed, developed, and maintained by public entities and resources. The early infrastructure was a public good, financed by the United States government through its military and developed primarily by universities and not-for-profit research centres across the country [82]. Similarly, other developments that became fundamental for today's internet, such as the TCP/IP protocol and the World Wide Web, were also conceived by researchers in international and national publicly-funded institutions such as the European Organization for Nuclear Research (CERN) [35].

The commodification and privatization of the Internet occurred progressively in the 1980s and early 1990s followed the neoliberalization under Reagan in the United States [35, 41]. In the case of the Internet, this shift challenged the idea that single networks supervised by a national regulatory body were required to protect the public interest [82]. The following years after the privatization of the internet saw the emergence of today's internet market and the birth of contemporary Big Tech companies, despite the dotcom bubble of the early 2000s. Today's significant players profited from early setbacks, from the purchase of patents from extinct or failing businesses and to their acquisition and integration into major corporations [32].

Several economic, political, and social conditions coalesced that brought about Big Tech. These conditions were produced by federal regulators not addressing the increasing size and power of these companies, their steady accumulation of capital, and their tax avoidance practices [82]. Along with these characteristics, as mentioned above, major technology companies also based their growth on their reliance on non-rivalrous goods [49], the reduction of production costs, notably through the exploitation of labour [16], and a "winner-take-all" form of competition characterized by the acquisition of competitors or their suppression through coercion and other means [32, 82]. These dynamics make oligopolies more likely. Globally, the strategy of Big Tech has also been characterized by its dependence on scalability, where their growth is not dependent on major changes to their operations and business models [38].

Another critical component of the concentration of wealth and power in contemporary datafication is Big Tech's reliance on the socioeconomic and infrastructural model of the platform to coordinate markets, social relationships, and information [17]. In this context, platforms are "(re-)programmable digital infrastructures that facilitate and shape personalised interactions among end-users and complementors" [74, 3]. Thus, as platforms, Big Tech controls the informational and material exchanges between the many actors in its markets, for instance, by regulating information flows through algorithms with end-users [13, 53] and the interplay between platforms and back-end users [65]. Big Tech sits at the nexus

of datafication, having created the systems that others must use, and the monitoring and maintaining of those systems.

This concentration of power and resources also influenced recent developments in AI research and have stirred the direction of the field. For instance, current discussions on the ethical implications of AI development were propelled by Big Tech funding, especially after the Cambridge Analytica scandal gave evidence of the large-scale social implications of platform power [42]. In terms of technical and scientific developments in the field, Big Tech also increased the number of partnerships with universities and publicly-funded research institutions, the same that were fundamental in the development of the publicly funded infrastructures of ARPANET during the dawn of the internet. However, these relationships between academia and industry transacted by Big Tech are now characterized by power imbalances in favour of these private corporations. For instance, Ochigame argues that, in the "Partnership on AI" initiative, a not-for-profit coalition between major tech companies and universities, the latter has less decision-making power than Big Tech members, contributing to the idea that some ethical and research initiatives are funded primarily for the benefit of the private sector [68].

Regarding the funding of academic research by Big Tech, Abdalla and Abdalla found that faculty members in the areas of computer science, AI ethics, and AI fairness from four leading R1 universities (MIT, Stanford, Berkeley, and the University of Toronto), who disclosed funding sources, received grants from major technology companies [1]. Counting the percentage of individuals who received Big Tech funding during any moment of their careers, including their doctoral studies, the number increased to 84% for computer sciences faculty, 88% for those working on AI, and 97% for those working on AI ethics [1]. Funding from private entities in academic research raises questions about what type of advances are pursued and funded and what outcomes are privileged over others.

More concretely, the trend seems to be that large models trained with huge datasets that depend on the existing infrastructures of major technology corporations are privileged over other types of research. This was evidenced by the winner of best paper award at the most recent Conference on Neural Information Processing Systems (NeurIPS), one of the major conferences in artificial intelligence, which presented OpenAI's Generative Pre-trained Transformer 3 (GPT-3) language model [40]. Large models like GPT-3, however, present environmental, economic, and social concerns. First, because of their requirement of enormous processing power and electricity that, in the case of language processing models, can emit up to 626,000 pounds of carbon dioxide, the equivalent of around five times the average emissions of a car in the United States [84]. And, second, because these "state of the art" models are also trained with large amounts of data that privilege quantity over quality of content, paying little attention to social context, and being difficult to audit [6]. Bender and colleagues also point out that these large models can manipulate the application of their outputs, potentially benefiting the companies behind them in social and economic settings [6].

As in the case of the printing press, the initial euphoria of a technology capable of breaking up powers in the status quo shifted the distribution of power to a new equilibrium. The printing press created new centers of power (around states, printers) and gutted

the old centers of power (religious institutions). Datafication and AI initially gave way to relatively frictionless means to communicate and link up with others (the Internet) that was created by the state. As the shift towards datafication has intensified, however, the corporations have seized power by controlling the currency of power: data. Coupled with largely corporate-fronted advances in AI, the corporations have taken on new roles in defining and directing the lives of billions of people.

3.3 Communities of Individuals?

The printing press and the communication revolution it created laid the groundwork for what Anderson calls "imagined communities" in which members of a large geographic area develop a socially-constructed concept of themselves as members of a group, in that context a nation-state [3]. The internet and datafication has ushered in a time in which geographic proximity is no longer essential to the ability to create community. This creates opportunities for communities to develop while also providing the opportunity for these communities to coalesce around potentially narrower interests and information creating concern among some about the development of "filter bubbles."[1]

At the same time, the communities that are created (by conscious choice and by interaction with unknown algorithms) can now involve ever narrower groups of people with similar beliefs, interests and behaviors, because geography is no longer a condition for interaction. In one sense, this allows people to find and identify with the people with whom they share the most similarity and can be very valuable in that regard. For example, this gave activists an edge in online protests, using methods such as doxing and DDoS attacks on targets [93]. On the other hand, the ubiquity of information combined with datafiction and predictive algorithms that generate internet search results, recommended new purchases, and help introduce our online personas to others can lead to the creation of "filter bubbles," which are a "personal ecosystem of information that's been catered by these algorithms" [70].

The printing press facilitated the opening of communication and knowledge, breaking the stranglehold of the Church on "the word" and letting new social classes rise with the newfound capability to acquire and spread information. The Internet, in part, has done this as well [25]. Media companies, which previously held the key to accessing mass communications such as news and entertainment, were subsumed by the competition created by the entry of Internet-based firms [94]. Yet, because of the way that search, advertising, and newsfeed technologies have pivoted to data-intensive processes of personalized advertising and exposure to information, this has resulted in the somewhat discordant result of simultaneously allowing for new sources of information to emerge and driving people into groups that share a common data stream with each other but not one that is broadly shared with others. Thus, there are both more data sources and fewer options at the same time, driven by

---

[1]There is some disagreement about whether filter bubbles are an appropriate term or more akin to a term to incite concern about a new technology [12], but for our purpose we use the term because it connotes the idea that there is a filtering process (both by individuals and algorithms) that can create an information and interaction bubble for people.

datafication and AI. We both know more about the individuals participating in digital technologies and are better able to push them towards their respective interests.

In studying the online media and information environment, Sood and Lelkes discuss how the development of narrowcasting and the replacement of a few, well-known media sources with a multitude of less-known media sources has given consumers greater choice but also allowed them to choose to consume their information from more congenial information sources, which can create a filter bubble of information [83]. In the context of the current information environment with its thousands of possible websites and sources of information, datafication and predictive algorithms play an indispensable (if unappreciated by many) role in influencing what people see, read, or otherwise consume. The information sources to which people are exposed are necessarily heavily influenced by the personalization algorithms that determines what choices are available to us.

The choice of more congenial media sources could be driven by either a preference for congenial news or the belief that congenial information sources are more trustworthy [83]. While there is some evidence that in the aggregate people are reasonable arbiters of quality news sources [73], this does not imply that all individuals are good arbiters or that when presented with an algorithmically chosen list of news stories/sources that we choose wisely. In fact, Luca et al. demonstrate that some individuals actually prefer low-quality "click bait" stories because they believe them to be more trustworthy [56]. Perceived trust in an information source is a necessary condition for persuasion [57] and therefore these sources can still influence behavior and beliefs.

The combination of preferences for low-quality (but trusted) sources and predictive algorithms may be particularly negative as it pushes people towards inferior information sources. The algorithms are designed to encourage clicks rather than information quality and therefore, if an individual chooses information from a less reliable source and the algorithm then feeds the person more information from similar sources, then they will be increasingly likely to see stories from unreliable, but perceived to be trustworthy, sources. Although people have agency over what stories to click on and what to read, the options presented to them are determined by algorithms that have no concern with the accuracy of the information or its effects.

Of course the information streams are not fully individualized, rather datafication allows the creation of communities of individuals who share the same media sources (and which may be quite low quality) and who may be geographically distant from one another. Pushing people into communities and information environments that are persuasive but full of misinformation can have significant real-world implications. For example, Hillary Clinton, as we know, lost the 2016 Presidential election to Donald Trump. Yet, that was not the end of her importance in that year, at least among some circles. An online conspiracy theory flourished on platforms such as Reddit, 4Chan, and Facebook, based on the leaked Clinton emails. The theory put Clinton at the center of a child trafficking ring that was centered in the basement of Comet Ping Pong, a pizzeria on Connecticut Avenue in Washington, D.C. Reddit users concluded that "cheese pizza" was code phrase for "child pornography." These theories became so widespread, and the allegations so animated

that on December 4, 2016, North Carolinian Edgar Maddison Welch walked in with a military-style rifle and handgun and fired several shots. No one was hurt. In January 2019, Californian Ryan Jaselskis set Comet Ping Pong on fire. Although other platforms had shut down Pizzagate by this time, a group on Facebook, "PizzagateUncompromised," remained alive and well, even after the theory had been debunked by multiple sources, including The New York Times and Snopes.com. The Pizzagate example demonstrates the power of datafication and the creation of communities that turned online interaction into "real-world" social action. While the Pizzagate episode demonstrates a dangerous combination of community and information, it serves as a vivid illustration of how datafication will affect social and political worlds.

4 GOING FORWARD

The datafication of the world portends social, economic, and political changes that are foreshadowed by the upheaval caused by the printing press. These changes will have both normatively good and bad implications, and our focus in this paper is to argue that we think carefully about the changes that datafication will cause; not because we can stop it (or might even want to) but rather because we need to build a new legal and regulatory infrastructure [36]. Already, rules are shifting as they adapt to the information age [22], but our analysis shows that the slow shift as laws catch up to realities may be too leisurely for anticipating the deep social and political changes that are possible.

An important takeaway from our argument is that in thinking about how to address the challenges of datafication from a policy perspective we need to recognize that the changes wrought by it will be no less profound. Therefore, when designing policy responses to manage datafication we need to move away from thinking about datafication as just a faster printing press and instead we need to incorporate technical and social insights. It is one thing to think about regulating a technology, such as facial recognition, but another thing to truly assess what it means to have data about a person's face sitting, even regulated, on servers. What status quo powers are disrupted, or decentralized, and what actors are poised to seize power in the vacuum and create a world to their liking? Already, Big Tech's major players have taken advantage of ideological frameworks, a relatively sanguine attitude towards datafication, and a hands-off attitude towards AI to shape the world as they see fit. The many of us are captive in those platforms without much protection, but that is not the only way.

The way we discuss potential regulatory frameworks about these issues should harken back to the analog world of the printing press. What rules became possible in a world of print that was impossible in a world of spoken words? What rules have become impossible in a world of digital data that were possible in the world of print? A number of basic human rights – freedom of expression, right to privacy/consent – have already come to the fore as problems of portability into the digital era. We simply cannot assume that our rules from twenty years past will apply now or twenty years hence. Instead, learning from the effects of the printing press, we should look to who holds the power in a digital world, why, and what constitutive changes to expect from the technology.

REFERENCES

[1] Mohamed Abdalla and Moustafa Abdalla. 2020. The Grey Hoodie Project: Big Tobacco, Big Tech, and the threat on academic integrity. *arXiv* (2020). arXiv:2009.13676 http://arxiv.org/abs/2009.13676

[2] Jay Agrawal, Joshua Gans, and Avi Goldfarb. 2018. *Prediction Machines: The Simple Economics of Artificial Intelligence.* Harvard Business Review Press, Cambridge, MA.

[3] Benedict Anderson. 1983. *Imagined Communities: Reflections on the Origin and Spread of Nationalism.* Verso, London. 160 pages.

[4] Gary Anthes. 2015. Data brokers are watching you. *Commun. ACM* 58, 1 (2015), 28–30.

[5] Payal Arora. 2019. Decolonizing Privacy Studies. *Television & New Media* 20, 4 (may 2019), 366–378.

[6] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Conference on Fairness, Accountability, and Transparency (FAccT '21)* (mar 2021).

[7] Ruha Benjamin. 2019. *Race after technology: abolitionist tools for the New Jim Code.* Polity Press, Cambridge, UK.

[8] Abeba Birhane. 2020. Algorithmic Colonization of Africa. *Scripted* 17, 2 (2020), 389–409.

[9] Lane Brown. 2019. There Will Be No Turning Back on Facial Recognition. *New York Magazine* (nov 2019). https://nymag.com/intelligencer/2019/11/the-future-of-facial-recognition-in-america.html

[10] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (2018).

[11] Matt Burgess. 2019. The Met Police's facial recognition tests are fatally flawed. *WIRED* (2019). https://www.wired.co.uk/article/met-police-london-facial-recognition-test

[12] Ryan Burns. 2019. New Frontiers of Philanthro-capitalism: Digital Technologies and Humanitarianism. *Antipode* 51, 4 (2019), 1101–1122.

[13] Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* 3, 1 (jan 2016).

[14] Ryan Calo. 2017. Artificial Intelligence Policy: A Roadmap. *SSRN Electronic Journal* (2017), 1–28.

[15] Gabrielle Canon. 2019. How Taylor Swift showed us the scary future of facial recognition. https://www.theguardian.com/technology/2019/feb/15/how-taylor-swift-showed-us-the-scary-future-of-facial-recognition

[16] Antonio A. Casilli. 2019. *En attendant les robots.* Éditions du Seuil, Paris. 401 pages.

[17] Antonio A. Casilli and Julian Posada. 2019. The Platformisation of Labor and Society. In *Society and the Internet* (2 ed.), Mark Graham and William H. Dutton (Eds.). Oxford University Press, Oxford.

[18] Niraj Chokshi. 2019. Facial Recognition's Many Controversies, From Stadium Surveillance to Racist Software. *The New York Times* (2019). https://www.nytimes.com/2019/05/15/business/facial-recognition-software-controversy.html

[19] Alexandra Chouldechova and Aaron Roth. 2018. The frontiers of fairness in machine learning. *arXiv* (2018). arXiv:1810.08810 https://arxiv.org/abs/1810.08810

[20] Roger Clarke. 1988. Information Technology and Dataveillance. *Communication ACM* 31, 5 (1988), 498–512.

[21] Roger Clarke and Graham Greenleaf. 2017. Dataveillance Regulation: A Research Framework. *Journal of Law, Information and Science* 25, 1 (2017), 104–122.

[22] Julie E. Cohen. 2019. *Between truth and power: the legal constructions of informational capitalism.* Oxford University Press, New York, NY.

[23] Sam Corbett-Davies and Sharad Goel. 2018. The measure and mismeasure of fairness: A critical review of fair machine learning. *arXiv* (2018). arXiv:1808.00023 https://arxiv.org/abs/1808.00023

[24] Nick Couldry and Jun Yu. 2018. Deconstructing datafication's brave new world. *New Media & Society* 20, 12 (Dec. 2018), 4473–4491.

[25] Ronald Deibert. 1997. *Parchment, printing, and hypermedia: communication in world order transformation.* Columbia University Press, New York.

[26] Ronald J. Deibert. 2020. *Reset: Reclaiming the Internet for Civil Society.* House of Anansi Press, Toronto, ON.

[27] Emily Denton, Alex Hanna, Razvan Amironesei, Andrew Smart, Hilary Nicole, and Morgan Klaus Scheuerman. 2020. Bringing the People Back In: Contesting Benchmark Machine Learning Datasets. *arXiv* (2020). arXiv:2007.07399 http://arxiv.org/abs/2007.07399

[28] James Dewar. 1998. *The Information Age and the Printing Press: Looking Backward to See Ahead.* RAND Corporation. https://www.rand.org/pubs/papers/P8014.html

[29] Domo. 2020. Data Never Sleeps 7.0. *Domo* (2020). https://www.domo.com/learn/data-never-sleeps-5

[30] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. *ITCS 2012 - Innovations in Theoretical Computer Science Conference* (2012), 214–226. arXiv:1104.3913

[31] Elizabeth L. Eisenstein. 1979. *The printing press as an agent of change: communications and cultural transformations in early-modern Europe. Volumes I and II.* Cambridge University Press, Cambridge.

[32] Rana Foroohar. 2019. *Don't Be Evil.* Currency, New York, NY.

[33] Marion Fourcade and Kieran Healy. 2017. Seeing like a market. *Socio-Economic Review* (2017).

[34] Brett M. Frischmann and Evan Selinger. 2018. *Re-engineering humanity.* Cambridge University Press, Cambridge.

[35] James M Gillies and Robert Cailliau. 2000. *How the Web was born: The story of the World Wide Web.* Oxford University Press, Oxford.

[36] Gillian K. Hadfield. 2016. *Rules for a Flat World: Why Humans Invented Law and How to Reinvent It for a Complex Global Economy.* Oxford University Press, Oxford.

[37] Dylan Hadfield-Menell, McKane Andrus, and Gillian K. Hadfield. 2019. Legible normativity for AI alignment: The value of silly rules. *AIES 2019 - Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (2019), 115–121. arXiv:1811.01267

[38] Alex Hanna and Tina M. Park. 2020. Against Scale: Provocations and Resistances to Scale Thinking. *arXiv* (2020). arXiv:2010.08850 http://arxiv.org/abs/2010.08850

[39] Karen Hao. 2020. In 2020, let's stop AI ethics-washing and actually do something. *MIT Technology Review* (2020). https://www.technologyreview.com/2019/12/27/57/ai-ethics-washing-time-to-act/

[40] Karen Hao. 2021. Five ways to make AI a greater force for good in 2021. *MIT Technology Review* (2021). https://www.technologyreview.com/2021/01/08/1015907/ai-force-for-good-in-2021/

[41] David Harvey. 2007. *A brief history of neoliberalism.* Oxford University Press, USA.

[42] Anne Helmond, David B. Nieborg, and Fernando N. van der Vlist. 2019. Facebook's evolution: development of a platform-as-infrastructure. *Internet Histories* 3, 2 (2019), 123–146.

[43] Mireille Hildebrandt. 2019. Privacy as Protection of the Incomputable Self : From Agnostic to Agonistic Machine Learning. *Theoretical Inquiries in Law* 20, 83 (2019), 83–121.

[44] Kashmir Hill. 2020. Flawed Facial Recognition Leads To Arrest and Jail for New Jersey Man. *The New York Times* (2020). https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html

[45] Kashmir Hill. 2020. The Secretive Company That Might End Privacy as We Know It. *The New York Times* (2020). https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html

[46] Shota Ichihashi. 2019. Non-competing Data Intermediaries. *Bank of Canada Working Paper* (2019), 1–41.

[47] Eun Seo Jo and Timnit Gebru. 2020. Lessons from archives: Strategies for collecting sociocultural data in machine learning. *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (2020), 306–316. arXiv:1912.10389

[48] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1, 9 (2019), 389–399.

[49] Charles I. Jones and Christopher Tonetti. 2020. Nonrivalry and the Economics of Data. *American Economic Review* 110, 9 (2020), 2819–2858.

[50] Pratyusha Kalluri. 2020. Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature* 583, 7815 (jul 2020), 169–169.

[51] Tero Karppi. 2018. *Disconnect: Facebook's affective bonds.* University of Minnesota Press, Minneapolis.

[52] Carly Kind. 2020. The term 'ethical AI' is finally starting to mean something. *Venture Beat* (aug 2020). https://venturebeat.com/2020/08/23/the-term-ethical-ai-is-finally-starting-to-mean-something/

[53] Rob Kitchin. 2017. Thinking critically about and researching algorithms. *Information, Communication & Society* 20, 1 (Jan. 2017), 14–29.

[54] Karen Levy. 2013. Relational Big Data. *SSRN Scholarly Paper* (Sept. 2013). https://papers.ssrn.com/abstract=2832921

[55] Elizabeth Lopatto. 2020. Clearview AI CEO says 'over 2,400 police agencies' are using its facial recognition software. *The Verge* (2020). https://www.theverge.com/2020/8/26/21402978/clearview-ai-ceo-interview-2400-police-agencies-facial-recognition

[56] Mario Luca, Kevin Munger, Jonathan Nagler, and Joshua A Tucker. 2021. You Won't Believe Our Results! But They Might: Heterogeneity in Beliefs About the Accuracy of Online Media. *Journal of Experimental Political Science* (jan 2021), 1–11.

[57] Arthur Lupia and Mathew D. McCubbins. 1998. *The Democratic Dilemma*. Cambridge University Press, Cambridge.

[58] Mary Madden. 2017. *Privacy, Security, and Digital Inequality*. Number September 27. Data & Society, New York, NY. 125 pages.

[59] Bill Marczak, John Scott-Railton, Adam Senft, Irene Poetranto, and Sarah McKune. 2015. Pay No Attention to the Server Behind the Proxy. *Citizen Lab* 15 (oct 2015).

[60] Viktor Mayer-Schönberger and Kenneth Cukier. 2014. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Eamon Dolan Mariner Books, Boston, MA. 272 pages.

[61] Ulises A. Mejias and Nick Couldry. 2019. Datafication. *Internet Policy Review* 8, 4 (2019), 1–10.

[62] Milagros Miceli, Martin Schuessler, and Tianling Yang. 2020. Between Subjectivity and Imposition. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–25. arXiv:2007.14886

[63] Shakir Mohamed, Marie-Therese Png, and William Isaac. 2020. Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology* 33, 4 (dec 2020), 659–684. arXiv:2007.04068

[64] Gemma Newlands. 2021. Far from Effortless: Invisible Labour in AI-as-a-Service. *Working Paper* (2021).

[65] David B. Nieborg and Thomas Poell. 2018. The platformization of cultural production: Theorizing the contingent cultural commodity. *New Media & Society* 20, 11 (nov 2018), 4275–4292.

[66] Safiya Umoja Noble. 2018. *Algorithms of Oppression. How Search Engines Reinforce Racism*. NYU Press, New York, NY.

[67] Illah Reza Nourbakhsh and Jennifer Keating. 2020. *AI & humanity*. The MIT Press, Cambridge, Massachusetts.

[68] Rodrigo Ochigame. 2019. The invention of "Ethical AI": How Big Tech Manipulates Academia to Avoid Regulation. *The Intercept* (2019). https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/

[69] Cathy O'Neil. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books, Portland, OR.

[70] Lynn Parramore. 2010. The Filter Bubble. *The Atlantic* (2010). https://www.theatlantic.com/daily-dish/archive/2010/10/the-filter-bubble/181427/

[71] Frank Pasquale. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, Cambridge, MA. 320 pages.

[72] Amandalynne Paullada, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, and Alex Hanna. 2020. Data and its (dis)contents: A survey of dataset development and use in machine learning research. *NeurIPS 2020 Workshop: ML Retrospectives, Surveys & Meta-analyses (ML-RSA)* (2020). arXiv:2012.05345

[73] Gordon Pennycook and David G. Rand. 2019. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences* 116, 7 (feb 2019), 2521–2526.

[74] Thomas Poell, David B. Nieborg, and José van Dijck. 2019. Platformisation. *Internet Policy Review* 8, 4 (2019).

[75] Julian Posada. 2020. The Future of Work Is Here: Toward a Comprehensive Approach to Artificial Intelligence and Labour. *Ethics in Context* (2020).

[76] Rita Raley. 2013. Dataveillance and counterveillance. In *Raw Data Is an Oxymoron*, Lisa Gitelman (Ed.). MIT Press, Cambridge, MA, 121–46.

[77] Shaun Raviv. 2020. The Secret History of Facial Recognition Shaun. *Wired* (2020). https://www.wired.com/story/secret-history-facial-recognition/

[78] Eugenia Roldán Vera. 2013. The History of the Book in Latin America (including Incas and Aztecs). In *The Book. A Global History*, Michael F. Suarez and H.R. Woudhuysen (Eds.). Oxford University Press, Oxford, 656–670.

[79] Charles A Ruud. 1980. The Printing Press as an Agent of Change. *Imago Mundi* 32, 1 (1980), 95–97.

[80] Jathan Sadowski. 2019. When data is capital: Datafication, accumulation, and extraction. *Big Data & Society* 6, 1 (Jan. 2019).

[81] G. Alex Sinha. 2018. A Real-Property Model of Privacy. *DePaul Law Review* 68, 3 (2018), 567–614.

[82] Nikos Smyrnaios. 2017. *Les GAFAM contre internet : une économie politique du numérique*. InaGlobal, Paris.

[83] Gaurav Sood and Yphtach Lelkes. 2018. Don't Expose Yourself: Discretionary Exposure to Political Information. In *Oxford Research Encyclopedia of Politics*. Oxford University Press, 1–49.

[84] Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2020. Energy and policy considerations for deep learning in NLP. *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference* (2020), 3645–3650. arXiv:1906.02243

[85] John B. Thompson. 1995. *The media and modernity: a social theory of the media*. Stanford University Press, Stanford, CA.

[86] Antonio Torralba and Alexei A. Efros. 2011. Unbiased look at dataset bias. In *CVPR 2011*. IEEE, 1521–1528.

[87] A. L. Tsing. 2012. ON NONSCALABILITY: The Living World Is Not Amenable to Precision-Nested Scales. *Common Knowledge* 18, 3 (oct 2012), 505–524.

[88] Jose van Dijck. 2014. Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society* 12, 2 (May 2014), 197–208.

[89] Hal Varian. 2019. Artificial Intelligence, Economics, and Industrial Organization. In *The Economics of Artificial Intelligence: An Agenda*, Ajay Agrawal, Joshua Gans, and Avi Goldfarb (Eds.). University of Chicago Press, Chicago, IL, 399–419.

[90] Carissa Véliz. 2020. *Privacy is Power: Why and how you should take back control of your data*. Bantam Press, New York.

[91] Amy Webb. 2019. *The Big Nine: how the tech titans and their thinking machines could warp humanity*. PublicAffairs, New York.

[92] Wendy H. Wong. 2021. As U.S. Capitol investigators use facial recognition, it begs the question: Who owns our faces? *The Conversation* (2021). http://theconversation.com/as-u-s-capitol-investigators-use-facial-recognition-it-begs-the-question-who-owns-our-faces-153253

[93] Wendy H. Wong and Peter A. Brown. 2013. E-Bandits in Global Activism: WikiLeaks, Anonymous, and the Politics of No One. *Perspectives on Politics* 11, 4 (2013), 1015–1033.

[94] Tim Wu. 2010. *The master switch: the rise and fall of information empires*. Alfred A. Knopf, New York.

[95] Elizabeth Yale. 2015. The History of Archives: The State of the Discipline. *Book History* 18, 1 (2015), 332–359.

[96] Shoshana Zuboff. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs, New York, NY.