

# Thinking Outside the Black-Box: The Case for “Algorithmic Sovereignty” in Social Media

Social Media + Society  
April-June 2020: 1–12  
© The Author(s) 2020  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2056305120915613  
journals.sagepub.com/home/sms  


Urbano Reviglio<sup>1</sup>  and Claudio Agosti<sup>2</sup>

## Abstract

This article is an interdisciplinary critical analysis of personalization systems and the gatekeeping role of current mainstream social media. The first section presents a literature review of data-driven personalization and its challenges in social media. The second section sheds light on increasing concerns regarding algorithms’ ability to overtly persuade—and covertly manipulate—users for the sake of engagement, introducing the emergence of the exclusive ownership of behavioral modification through hyper-nudging techniques. The third section empirically analyzes users’ expectations and behaviors regarding such data-driven personalization to frame a conceptualization of users’ agency. The fourth section introduces the concept of “algorithmic sovereignty.” Current projects that aim to grant this algorithmic sovereignty highlight some potential applications. Together this novel theoretical framework and empirical applications suggest that, to preserve trust, social media should open their personalization algorithms to a social negotiation as the first step toward a more sustainable social media landscape. To decentralize the immense power of mainstream social media, guarantee a democratic oversight, and mitigate the unintended undesirable consequences of their algorithmic curation, public institutions and civil society could help in developing and researching public algorithms, fostering a collective awareness so as to eventually ensure a fair and accountable “algorithmic sovereignty.”

## Keywords

social media, personalization, algorithm governance, sovereignty, data activism

*As algorithms shape what we read about and how important we consider different issues, they should be recognized as an extension of our will. We should be able to consciously build our own algorithms, change them whenever we want, and not have to delegate this decision to a commercial entity with opaque functioning and objectives.*

—tracking.exposed manifesto<sup>1</sup>

Online personalization is our interface with the infosphere. It is essential to managing information overload so as to preserve individuals’ time and attention. It is, however, an imperfect tool, subject not only to technical challenges but, in particular, also to commercial and political manipulations. Today’s social media landscape is indeed characterized by an oligopolistic market in which corporate non-state actors treat users’ attention as a commodity in what has been described as “surveillance capitalism” (Zuboff, 2015). A small group of platforms—above all Google, YouTube, Twitter and Facebook—act as the ultimate gatekeepers for billions of users worldwide. They provide

a personalized experience, but little individual or group user control over information filtering processes (Harambam et al., 2018). This control asymmetry translates into an industry enriched by collecting, attributing, and inferring data points. The behavioral analysis produced perpetuates an epistemic imbalance in which platforms know people more than people know themselves. As such, social media platforms can exploit people’s vulnerabilities. To sustain this market, short-term pleasure, soft news, homophily, narcissism, entertainment, and advertisement predominate social interactions. In general, the ability to engage users is increasingly of concern in various contexts including possible negative psychological effects

<sup>1</sup>University of Bologna, Italy

<sup>2</sup>University of Amsterdam, The Netherlands

### Corresponding Author:

Urbano Reviglio, University of Bologna, Via Zamboni, 33, 40126 Bologna, Italy.

Email: urbanoreviglio@hotmail.com



(Verduyn et al., 2017), the development of addictive usage patterns (Deibert, 2019), the potential filter bubble effect (Pariser, 2011) and its role in reinforcing echo chambers (Sunstein, 2017) and, in turn, in spreading misinformation (Del Vicario et al., 2016).

In this article, an interdisciplinary critical analysis of the algorithmic black-box problem in social media is provided. As such, we draw from various disciplines, in particular from media and data protection law, sociology and psychology, design and political theory. We conclude highlighting the need for more public discussion about improving data-driven personalization practices in social media through a radical user empowerment and social scrutiny. The article is structured as follows; first, a literature review of data-driven personalization in social media examines its challenges and risks. In particular, algorithms' ability to persuade and eventually manipulate users for the sake of engagement is critically analyzed. Next, an empirical and critical literature review analysis of the expectations and behaviors of users regarding personalization shows how, to fulfill civic rights and duties, the average user cannot be fully responsible to navigate alone all the complexities of online data-driven personalized media consumption. Finally, we introduce the concept of algorithmic sovereignty, preliminarily evaluating its strengths and weaknesses with respect to current potential applications.

## Personalization and Its Challenges

Personalization is ubiquitous online and it shapes every social media engagement as it filters and prioritizes information to make it “fit” each user's profile. As such, it legitimizes social media's gatekeeping role. User interests become more important than content quality or social significance (DeVito, 2017). This rhetoric makes data-driven personalization very appealing in our hyper-individualized societies. Yet, social media platforms are not just communication companies that uniquely shape social networks and their dynamics processes, but also large advertising agencies driven by profit (Fuchs & Marisol, 2015). Personalization can in fact be understood as pursuing the logic of market segmentation until each individual user is reduced to a unique market (Yeung, 2018a).

This process occurs largely beyond the control of users as it is based on implicit personalization—behavioral data collected from subconscious activity<sup>2</sup>—rather than on deliberate and expressed preferences, and this might become a default choice in future personalization (Thurman & Schifferes, 2012).<sup>3</sup> In general, engineers' approach can take two directions: build more advanced AI techniques, or simplify the user interface to convert the recorded human activity in a defined set of possibilities, reducing complexity (Harris, 2019). Modern advancements in technology use both. This inevitably raises several concerns.

To begin with, profiling technologies that allow personalization create a kind of knowledge that is inherently

probabilistic. They cannot produce or detect a sense of self. They can, however, influence the individual's sense of self, especially in everyday social media information consumption. This is concerning for two main reasons.

On the one hand, since long has been warned that the development of identities might be undermined by profiling algorithms that are stuck in the past, that may steer individuals to conform to the status quo of their past actions chosen by past selves. Individuals may start to want what is recommended to them without even recognizing it, in a sort of self-fulfilling prophecy. This may perpetuate existing inequalities and other pathologies, narrowing the “aspirational self,” while threatening the foundational microeconomic principle of preference formation and modification.<sup>4</sup>

On the other hand, profiling is a matter of pattern recognition, which is comparable to categorization, generalization and stereotyping. Personalization thus relies on the comparability or even sameness (by some simplified categories) of the user with others (e.g., collaborative filtering). As such, personalization paradoxically denies individual uniqueness through a “smart” homogenization that negotiates the diversity of humankind (Yeung, 2018a). It is in this context that is particularly discussed the notion of “group privacy” as the right that is held by a group to safeguard information that is not only private to individuals, but which reveals something about that same group (Taylor et al., 2016).

Related to the above concerns, personalization of media content can limit the diversity of information, from exposure to discovery. Yet, information filtering processes take place on three intertwined levels; not only technological (e.g., affordances) but also individual (e.g., selective exposure) and social (e.g., sharing practices). The major risk is the creation of “informational bubbles”: at the individual level, filter bubbles (Pariser, 2011) and, at the group level, echo chambers (Sunstein, 2017). The crucial difference is that the former may not depend on user's autonomy and awareness (Bozdag & Timmermans, 2011)—therefore is mainly caused by technological affordances—while the latter pre-existed digital age and thus it is primarily individual and social. These phenomena, however, are poorly defined and, thus, are used more as metaphors.

Yet, personalization—in particular news personalization—could reduce opportunities to self-determine and could negatively affect information finding by reducing the exposure to alternative points of view in the “marketplace of ideas” (Pariser, 2011; Sunstein, 2017) and, more generally, to serendipitous encounters (Reviglio, 2019b). There could be several consequences; from the limitation of personal creativity to a reduction in the ability to build productive social capital. Moreover, personalization could weaken media pluralism, solidarity and make people more politically polarized, narcissistic and vulnerable to (self)propaganda. This, in turn, would make people more susceptible to polarizing messages, help to spread misinformation, and ultimately, erode interpersonal trust (Del Vicario et al., 2016).

Critics, however, argue that these are mere moral panics, and that personalization allows the media to inform people better and more effectively while fostering the cultivation of expert citizens with stronger group identities. In any case, personalized techniques enable potentially manipulative political microtargeting (Kohl et al., 2019). Even so, a prominent concern remains “epistemic inequality.” That is, the wealthier the social networks, the higher the education level, meaning more personalization benefits, and vice versa.

Currently, research on personalization is more often contradictory, ambiguous, and thus unreliable (Bodo et al., 2017; Tucker et al., 2018; Zuiderveen Borgesius et al., 2016). The risks of social media usage (Verduyn et al., 2017) and filter bubbles (Zuiderveen Borgesius et al., 2016), in fact, are very hard to prove and, eventually, to mitigate. Most research is ultimately inconclusive because it is generally survey-based, or is correlational or based on a small or unsatisfactory sample. In light of the fast-changing media landscape, many studies become rapidly outdated. This contributes to the current crisis with regard to the study of algorithms (Bodo et al., 2017). They are indeed “black-boxed” (Pasquale, 2015) which means their functioning is opaque and their interpretability<sup>5</sup> may not even be clear to engineers (Albanie et al., 2017). While insights on the main causes and risks of personalization has been currently understood (Tucker et al., 2018), we still lack evidence with regard to the extent of their consequences. Given the heterogeneity of users, in fact, causes and effects of personalization vary widely. Yet, concerns are growing because of the rise of persuasive technologies and the ability of Big Data to nudge individuals, make them addicted and ultimately deceive them (Yeung, 2018a; Zarsky, 2019).

Another related theoretical issue with personalization and its future development is that they can increasingly gratify users to the extent to which they come to accept manipulation as benevolent. This argument has been discussed in terms of “psychological hedonism” (Gal, 2017): if personalized systems become “pleasure machines,” able to predict our choices and simply grant them to users, are these users willing to give up their autonomous choice? And, if so, under which conditions? Indeed we need to think seriously about the way in which the choices we make affect our values, identities, and the meaning and content of our lives. In other words, whether we as individuals actually strive for hedonism—that is the view that all human action is ultimately motivated by the desire for pleasure and the avoidance of pain—and the extent to which this is beneficial for society at large. Importantly, democracy requires a certain amount of discomfort. This leads to fundamental theoretical questions: where positive individual effects go along with negative societal effects what should be done? Should the state override personal autonomy to safeguard personal autonomy itself? If yes, to what extent and in which cases?

### *Addicted by Design*

Today’s Internet (Alter, 2017) —especially social media (Deibert, 2019)— is already “addictive by design.” This threatens individual autonomy (Gal, 2017; Zarsky, 2019). Human behavior can indeed be manipulated by priming and conditioning, using rewards and punishments. Even algorithms can autonomously explore manipulative strategies that can be detrimental to users (Albanie et al., 2017). For example, they can use A/B testing to experiment with various messages until they find the versions that best exploit individuals’ vulnerabilities (Calo, 2013).<sup>6</sup>

Manipulation and deception become easier thanks to affective computing (or “emotional AI”), captology—the study of computers as persuasive technologies (Fogg et al., 2002)—and the emergence of psychographic techniques focusing on demographic characteristics and “affect detection techniques,” along with diverse types of data such as location-based tracking, real-time data, or keyboard usage. Consider that just dozens of Facebook Likes can reveal useful and highly accurate correlations (Youyu et al., 2015). Implicated data include what the European General Data Protection Regulation (GDPR) classifies as specially protected data: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, use of addictive substances, parental separation, age, gender, and perhaps most importantly, emotions.<sup>7</sup>

Despite the detection of emotions is clearly reductionistic and prone to bias, the resulting persuasive techniques can deeply affect individuals’ self-control, self-esteem, and even self-determination. These can indeed stimulate users in a powerfully subconscious and hormonal way. Facebook’s infamous, large-scale emotional contagion experiment exemplifies this point (Kramer et al., 2014), showing how mainstream social media can affect emotions and exploit vulnerabilities in human psychology.

Compulsion loops are already found in a wide range of social media, and especially online games (Deibert, 2019). Research suggests that such loops may work via “variable-rate reinforcement”<sup>8</sup> in which rewards are delivered unpredictably. This unpredictability affects the brain’s dopamine pathways in ways that magnify rewards. Also, design facets intentionally trigger dopamine rushes or other emotional highs, stimulate popularity contest or implicit social obligations (Kidron et al., 2018)—use a whole host of additional brain hacks, probably many of them not even publicly known. Among those known, moral outrage can be exploited to increase engagement. New York University (NYU) psychology researchers found that each word of moral outrage added to a tweet raises the retweet rate by 17% (Harris, 2019). At times, users may also be captured in a spiral of ever more extreme, conspiratorial content—also known as the “rabbit hole effect.”<sup>9</sup>

Similarly, under the auspices of maintaining Friends’ relationships, Facebook’s “NewsFeed” is primarily moderated

by homophily, or similarity in affinity (DeVito, 2017). Excessive homophily, however, helps to spread misinformation, frequently resulting in homogeneous, polarized clusters reiterating emotionally charged and externally divisive content (Deibert, 2019; Sunstein, 2017). In other words, increasing social and political polarization is actually part of the business model of currently dominant social media platforms. Eventually, most users consume personalized content designed for implicit individual needs such as hedonism, sensationalism, and subtle self-propaganda.

Such Big Data-driven behavioral modification is defined by Yeung (2017) as a technique of “hyper-nudging” which dynamically configures the user’s informational choice context in ways intentionally designed to influence her decisions. For example, “dark patterns” which define instances where designers use their knowledge of human behavior (e.g., psychology) and the desires of end-users to implement deceptive functionality that is not in the user’s best interest (Gray et al., 2018). As such, hyper-nudging—in parallel to personalization—concerns all of the design process, not only algorithmic decision-making. This introduces a new form of power—a new “invisible hand” of behavioral prediction and modification (Zuboff, 2015).

Over time, as smart environments will likely permeate societies, users—especially young people—may be automatically plugged in and guided through life along algorithm-driven pathways nudged in the best interests of whoever owns or pays to use people’s data, at times with dramatic consequences. For example, addictive techniques are already concerning in the case of negative effects on children’s well-being, including increased risk of suicide and depression, conflicts with parents and adverse effects on cerebral and social development (Kidron et al., 2018).

Given this constellation of facts, the resultant, paramount concern is that technology, especially social media, is downgrading our attention spans, a common base of facts, and capacity for complexity and nuanced critical thinking, hindering our ability to construct shared agendas to help to solve our problems and the epochal challenges we all face. For some, this supposed degraded and degrading capacity for collective action arguably represents “the climate change of culture” (Harris, 2019). Still, existing scientific evidences on these consequences remain insufficient, as it seems the current legal framework and political response.

### *Citizens Before Users*

Despite apparently increasing transparency along with the resultant efforts to reform, the current algorithm-driven advertising system that lies at the core of the surveillance economy persists. And it persists in systematically promoting extreme, inaccurate and radical content—regardless of what malicious actors may do to seed it. This is the case of political microtargeting, also known as “dark ads.”<sup>10</sup> Such massive power asymmetry between platforms and potentially

malicious actors using them, on one hand, and individual users on the other, cannot be ignored. Since the Cambridge Analytica scandals implicating manipulative and possibly illegal social media use in Brexit and Trump 2016 campaigning, challenges and potential solutions are being discussed (Kohl et al., 2019).

To access social media platforms, users usually have to give up the data they generate. The profit-generating business of selling the resultant meta-data to advertisers is what ostensibly supports these “free” platforms. But Internet users are not simply or even primarily hedonistic consumers to persuade or even manipulate to maximize data extraction and profit generation. Rather, users are also often producers, citizens, and unpaid workers who generate economic value for these platforms through their “digital labour” (Scholz, 2012). And, still, only a minority of users actually understands how Facebook’s personalization works (Smith, 2018).<sup>11</sup> This is the status quo of the current data-driven society, in which sociality and the “public sphere” are turned into economic value. Users occupy a subordinate position without any reasonable possibility to opt out rather than “agree” or leave their social networks.

One compelling explanation for why this became possible so rapidly is that nothing similar to Information and Communication Technologies development ever happened in the past, so there were few institutional defensive barriers. Information intermediaries have indeed been overestimated and treated as “emissaries of the future,” producing “an overwhelming sense of inevitability” (Zuboff, 2015, p. 85). As such, societies quickly came to depend on these new information and communication tools as necessary resources, and at times even as preconditions for social participation.

Internet regulators have usually approached personalization through the lens of pre-Internet legislation. Personalization algorithms, however, are novel and complex entities, and they require novel solutions and public discussions around their regulation. There are, in fact, three intertwined “paradoxes” that highlight the complexities to fairly govern personalization. First, the “personalization paradox,” that is, a trade-off between privacy and personalization accuracy. Second, and strictly related, the “privacy paradox,” that is, the infamous users’ inconsistent will to protect their own privacy. And third, the “paradox of choice,” that is, the more choices users have, the more easily they rely on simple-to-use personalized tools.<sup>12</sup> In other words, in order for personalization systems to provide a “better service,” users are surveilled and their data captured and exploited. Even if they disagree—as they more often do—they do not proactively react. Even in cases where the users are provided with more agency, they are unlikely to take advantage.

GDPR—the new European Data Regulation in effect since 2018—grants novel rights for data subjects and duties for data controller.<sup>13</sup> Along with the e-Privacy regulation draft, the GDPR actually strengthens “data consumer protection”; users can indeed decide whether to enter into a

contract, be informed, access the data generated, receive information about the logic involved and opt not to be subject to automated decision-making based solely on automated processing. Yet, the right to an explanation (Art. 15) and the right to non-discrimination (Art. 22)—the most relevant in this context—are highly disputable. The major problem actually lies in the apparent belief of lawyers as well as designers that input control alone is sufficient to alleviate any systemic concerns of personalization algorithms (Kohl et al., 2019). Data protection law remains indeed crucial—because personalization requires updated target profiles—but currently no legal framework addresses explicitly the outputs of personalization algorithms.

The approach of user empowerment often relies on informed consent, but this might not be an optimal solution because opt-in approaches lead to uncertainty and context dependence. People in fact cannot be counted on to navigate the complex trade-offs involving privacy self-management (Acquisti et al., 2015). Most people in fact neither read nor understand online privacy policies. These actually have also two inherently contradictory goals; to be understandable to consumers—which requires simplicity and brevity—and say something meaningful about how data are processed—which is complicated and requires a lot of details. Moreover, dark patterns are often employed during terms of conditions and privacy updates (Moen et al., 2018). Although user-centered solutions may better foster the transparency of personalization algorithms, they have significant shortcomings: they shift the responsibility and accountability for the protection of rights and utility to the users, and this usually results in inefficiency such as insufficient privacy protection.

An explanation for the widespread default trust to this system, and the lack of responsiveness toward privacy management can be traced back to behavioral economics as well as psychology. Drawing from the famous work of Kahneman (2011), for example, Thaler and Sunstein (2009) describe two major cognitive systems from which individuals take decisions: an “automatic system” and a “reflective system.” While the latter is the slow, effortful, and controlled way to think, the former is efficient, rapid, largely unconscious, and prone to systematic errors. This bounded rationality affects how people assess the likelihood of future events and how their individual biases and vulnerabilities might be exploited. This is why education on persuasive techniques is probably the most effective deterrent to deception and manipulation (Friestad & Wright, 1994). Similarly, when people have knowledge regarding their own vulnerabilities, the efficacy of persuasive attempts diminish.<sup>14</sup> Developing such a critical and aware approach to persuasion—“persuasion knowledge” (Friestad & Wright, 1994)—is indeed fundamental for citizens of information societies. Yet, it might not be sufficient to resist undesirable consequences of personalization and, ultimately, to fully benefit from its potential.

## Algorithmic Sovereignty in Social Media

Mainstream social media—especially Facebook—explicitly counteract any possibility for its participants to gain sovereignty: it denies all possibilities of participation in the decision-making process of its own algorithms, as well as strictly regulates the opportunities of interoperability for the data it gathers. Therefore, when we speak of algorithmic sovereignty, we intend the moral right of a person to be the exclusive controller of one’s own algorithmic life and, more generally, the right and capacity by citizens as well as democratic institutions to make self-determined choices on personalization algorithms and related design choices. Self-determination is essentially at issue here, and the battlefield is whether or not users will be able to exercise their actual will when it comes to personalization. Until now this premise—essentially that it matters whether people control personalization and its algorithms—has never been seriously commented, stated, or even discussed by social media platforms.

Conceptually, the following analysis is based on a constructionist approach that posits that metaphors organize users’ perceptions and contribute to creating new realities (Krippendorff, 1993). To effectively address new Information and communications technology (ICT)-related challenges, it is indeed needed to enrich our conceptual toolbox (Floridi, 2015). Taking the notion of “algorithmic sovereignty” as a metaphor—first explored by Roio (2018)—we are interested in what could be emphasized or neglected in its use and the extent to which its use may contribute to reiterate or deconstruct mythologies of digital infrastructures, with a specific focus on personalization algorithms in social media.

In the last years, similar related concepts have been introduced and discussed in academia and media (Couture & Toupine, 2017), notably data sovereignty (e.g., De Filippi & McCarthy, 2012), digital sovereignty, and technological sovereignty (DiEM25, 2019). Such increasing interest in the notion of sovereignty has the ultimate goal to assert some form of individual and collective control empowerment over digital technologies. In particular, the notion of technological sovereignty is used to refer to initiatives that create alternatives to commercial and/or military technologies, often with a great emphasis on free and open source software and hardware. As such, it is also framed as an opposition to a hegemonic power, namely the United States and its biggest private tech companies (also called GAFAM: Google, Amazon, Facebook, Apple, Microsoft). In its broad sense, technological sovereignty is often meant as a rupture from “state sovereignty,” while other times—similarly to the notion of “computing sovereignty” by Richard Stallman (Couture & Toupine, 2019)—as the duty of public services to keep total control of their computing tasks in the benefit of citizens.

We can ideally distinguish between two poles of algorithmic sovereignty—weak sovereignty and strong sovereignty—between which different approaches may be situated: “weak algorithmic sovereignty” would refer to the negotiation between states and social media platforms to enact certain technical conditions, while “strong algorithmic sovereignty” would refer to the actual common ownership of algorithms by the people. This latter interpretation widely overlaps with the concept of technological sovereignty. Algorithmic sovereignty can thus be considered as a subgroup of the broader notion of technological sovereignty but, at the same time, it can provide more specific conditions on what the individual and social control over algorithms should be.

In this context, it is easy to think of the free software movement and the open source business models as the paradigms that enable the scrutiny of algorithms by thousands of experts, as well the freedom to modify them and distribute modifications (Roio, 2018).<sup>15</sup> Yet, even when the access to algorithms complies to these major ethical concerns, would that be enough to verify a condition of sovereignty for all living participants? Is transparency, as opposed to secrecy, a sufficient condition to make algorithms functional to the creation of an intelligible society? And how to effectively translate these principles in the context of personalization algorithms in mainstream social media?

### Algorithmic Sovereignty in Practice

Personalization algorithms, more than others, are technologies that inscribe values, thus visions, ideas, and beliefs to satisfy needs and desires. This is a fundamental condition to understand the role of sovereignty and to enable users to relate to it. Exerting more control as hyper-connected citizens on personalization in social media is the collective achievement we call algorithmic sovereignty.

In light of all the consequences of personalization, it is fundamental to recognize all personalization algorithms in which individuals and societies participate as Commons (Ostrom, 2015), or, more precisely, as Knowledge Commons (Vercellone et al., 2015). To do so, it is necessary to give particular attention to features and uses of algorithmic languages, well beyond the legal and economic considerations concerning the contexts in which they are interpreted and executed (Roio, 2018). Similarly, politics for the digital age need to actually consider users as political subjects (Fuchs, 2015). For example, unions could be extended to digital media users. The exploitation of digital labor is in fact connected to the commodification of the commons, including the digital ones.

Nowadays, the only way for participants to gain sovereignty in mainstream social media is to create an entirely new territory.<sup>16</sup> This is hardly suitable due to *network effects* which occur when the value of a platform to any user increases exponentially with the number of already present

users. Normally, this leads to monopolies (Lovink & Rasch, 2013). The history of alternative media is indeed a history of enormous challenges, as the consequence of a political economy that limits the possibilities for civil society because hearing alternative voices is a matter of money and political resources that afford visibility (Fuchs & Marisol, 2015). What has changed in the last decade is the vast societal dependency on information intermediaries’ networks.

The only opportunities for social media users to gain algorithmic sovereignty lies in two possible lines of action (Roio, 2018): either to “fork”<sup>17</sup> the project, with great expense of resources and effort and with an uncertain outcome—this might even be facilitated with the right to portability afforded by the GDPR—or to rely on a strong regulatory framework that opens Facebook’s governance to its participants. Yet, both options are very uncertain and unlikely to succeed. On the one hand, the “critical mass effect” does not make data portability a useful mechanism for migrating to other networks. On the other hand, the focus of the GDPR on data consumer protection does not actually guarantee the protection of collective rights.

To overcome social media monopolies, major media reforms are required. Among many, some scholars advocated proposals for gaining sovereignty such as a “Public Service Internet” that is completely non-profit (Fuchs, 2017) or “Public Service Algorithms,” namely the possibility to provide individuals institutionally coded algorithms (Bennett, 2018). Burri (2015), for instance, proposes a “Public Service Navigator” as a mechanism for “influencing the conditions of access to content, particularly its visibility, discoverability, and usability” (Burri, 2015). A practical example of an exploratory algorithmic tool is represented by the Massachusetts Institute of Technology (MIT) Media lab project *gobo.social*, a social media news aggregator with sliders that users can control to filter information; from political perspectives (left, right, center) to the extent of seriousness, rudeness, gender, and other parameters one can filter information with. This project shows what a user-friendly personalization algorithm might look like.

Yet, the idea that existing legal frameworks such as consumer data protection and antitrust regulation can be applied to the new challenges and that, more generally, media should be left to self-regulation is still very widespread. In the European media policy, it is in fact assumed that in most cases, soft law promotes self-regulation. To sustain at least a fair competition, another solution could be taxing large media (and other) corporations and channeling the resulting income into alternative non-commercial social media (Fuchs & Marisol, 2015). All these measures have their own limits, of course, but they represent constructive steps toward a more sustainable social media landscape. So far, however, there has not been sufficient evidence or political interest to justify any more radical proposals in practice. Certainly, there are still enormous social and technological issues at stake.

Nonetheless, the pursuit for more individual agency and social control of platforms is receiving attention. For example, the *European observatory on Algorithmic Sovereignty*, a

collective place of documentation gathering research, developments, events, and projects related to the topic.<sup>18</sup> Also, the project *Fediverse*—with roughly 5 million users<sup>19</sup>—is an example of a social media where users approach an algorithmic sovereignty. The name is a portmanteau of two words “federation” and “universe,” a common name for federated social networks running on free open software on a myriad of servers across the world. The “Federation” refers to a global social network composed of nodes that talk to each other. Each of them is an installation of software which supports one of the federated social web protocols. Fediverse networks are designed to be run by anybody, free to choose and register on any server and choose the person who will be in charge of its data—the administrator of your server. Selecting a server for its politics of data, however, is an expert choice one cannot expect from the large public.

Another significant approach is taken by the project *tracking.exposed*. By scraping and collecting the data individually recommended to the user on supported platforms (Facebook and YouTube above all), it collects and stores evidence which is usable in new ways. A result of such data reuse is the comparison of personalized experiences. By assuming everyone has an individualized perception of the public discourse, the project extrapolates the topics and let users compare, which might allow for improved critical judgment regarding the quality of information received. Also, these data are shared for research purposes, with non-governmental organizations (NGOs) and academic teams. By creating new profiles, with as many of the possible variables (age, friends, pages followed, likes) under control, it is possible to test how algorithms lead to divisive perceptions. As a side effect, it has collected a library of public information. Inspired by the political concept of the European Data Commons (DiEM25, 2019), the project keeps data in the collective interest, as far as it does not expose any individual without his or her consent. Yet, full anonymity cannot be offered at the current state of research. Therefore, the team has to make a privacy impact assessment for each form of data reuse.<sup>20</sup>

All together, these proposals and projects question—and to some extent could even disrupt—the current power asymmetry between users and the public toward mainstream social media platforms and their personalization algorithms. Yet, they could also renew trust between platforms and users. To ensure algorithmic sovereignty, however, a comprehensive approach to personalization algorithms is needed: one that complements the principles advocated by the free software movement to accommodate the needs, skills, and motivations of the average user, not only experts, activists, or enthusiasts of civil society.

### Algorithmic Sovereignty in Theory

With “algorithmic sovereignty” in social media, we generally intend to regulate the use of information filtering and personalization design choices according to democratic principles,

setting their scope for private purposes and harnessing their power for the public good. In other words, to open black-boxed personalization algorithms of (mainstream) social media to users and public institutions.

While there have been discussions on the ethics of recommender systems and personalization (Bozdag & Timmermans, 2011; Helberger, 2019; Milano et al., 2020), and on how to fulfill main algorithmic principles such as fairness, transparency, accountability, accuracy, and privacy (algo:aware, 2018), in this hypothetical radical scenario, algorithms could even be exchanged, remixed, tested, plugged, and even sold or rented. On top of this opening, it is expected that not all users would have the knowledge to build their own algorithms. Moreover, one could assume that some influential organizations (such as news media, political parties, ideological groups) would also spread their own algorithms to their followers. Audience fragmentation and engagement optimization may still dominate the media landscape. Therefore, we identify and discuss prominent intertwined preconditions, principles, and goals—certainly not exhaustive—to grant in social media what we intend as algorithmic sovereignty.

First of all, and more obviously, there is a general need for “algorithmic literacy.” We can define it as the basic knowledge on how filtering mechanisms and design choices function and what their impact on one’s own life is and may be. This can be cultivated with regular updates on personalization research, visualization tools, and dashboards. Some paradigmatic experiments have already been done: (1) in general, through browser extensions (Reviglio, 2019a), (2) with interactive sliders (e.g., *gobo.social*), or (3) showing users their filter bubbles and help to “burst” them (Nagulendra & Vassileva, 2016). Clearly, to gain more control, algorithmic literacy need to go hand in hand with digital and media literacy.

Another significant aim to achieve is “content neutrality,” meaning platforms should become agnostic toward content. These should in fact return machine-readable, unfiltered, chronologically ordered data. Then, a reader device would receive these data, and an algorithm would filter and prioritize. This could be a fundamental architectural requirement for a strong algorithmic sovereignty. By making sure that the filtering happens on the client side, it is ensured that the platform becomes effectively neutral, and nobody, except the individual, will end up knowing what was watched and for how long. Such neutrality would also help to prevent deception by design and, eventually, to cultivate a related aspirational goal: “algorithmic neutrality,” which is the idea that content should spread freely without biases.

Content and algorithmic neutrality might be also achieved with a “right to profile transparency,” through an infrastructure that supports what the philosopher Mireille Hildebrandt (2015) defines as “counter-profiling,” meaning “to conduct data mining operations on the behaviours of those that are in the business of profiling whether ‘those’ are humans,

computing systems or hybrid configurations” (p. 223). This is a general strategy to figure out how one is actually targeted. In addition, platforms and algorithm developers could routinely disclose to their users and the public any experiment that the users were subjected to.<sup>21</sup> An independent board to review and approve experiments in advance could also be established. This would be particularly in line with two intertwined newly discussed human rights: the “right to not be measured, analysed or coached”<sup>22</sup> (van Est et al., 2017) and a “right to cognitive sovereignty” that ought to protect individuals’ right to mental self-determination (Yeung, 2018b).

More generally, algorithm auditing needs to be supported as well as data sharing. In particular, it is fundamental to guarantee a strong separation between filtering mechanism and data. Filtering mechanisms should be, as much as possible, stateless and idempotent, that is, capable to return the same result if executed twice. This is a formal way to describe a more easily auditable machine. On the other hand, personalization algorithms that use neural networks keep updating their internal state, making it impossible to perform certain analyses either *ex post* or *ex ante*. If the same result can be accomplished using a simpler computing system, it would also be simpler to explain, standardize, teach, and assess outcomes. If it is possible to keep a simpler computing system as standard, the technical debt caused by the lack of explainability of algorithms might be avoided. Moreover, metadata which describes the content produced is fundamental to control. Functionalities of research, analysis, and prioritization might perform better with metadata usable by end-users. Even if it is a laborious process, mainly for skilled individuals, metadata represent an opportunity to differently index and retrieve content so as to eventually increase content quality over quantity.

There are, then, user-centered preconditions to negotiate. Being essential to personalized information consumption, design choices, and affordances should be *to some extent* adjustable, especially for what concerns attention management. For example, disabling the auto-loading of posts (similar to the auto-play function for videos which is usually set by default) and decide to scroll *n* posts anytime one logs in. Another interesting example is the possibility to hide metrics in order not to be influenced by comparative quantifications (such as the number of likes and shares).<sup>23</sup> The quantification and maximization of social interactions, in fact, contribute to create a “culture of performance” (Castro, 2016) that seems to be negatively correlated with well-being (Verduyn et al., 2017). In essence, all those basic design choices that ultimately affect information behavior and personalize one’s information diet ought to be socially negotiated, implemented, and ultimately, adjustable.

The development of “natural metrics” usable to evaluate the informative experience produced by the algorithm is also another critical issue. For example, the percentage ratio of photos/videos/text, the time spent on the platform, the number of posts encountered, or which sources have been consumed. This information should be collected on the

device and given back to the user to primarily increase self-awareness.

Furthermore, there are information self-determination and media pluralism concerns to cope with. In particular, the capacity to reach a balanced information diet from both an individual and a political perspective (Eskens et al., 2017) and, at the same time, make sure that a range of viable “perspective widening” tools are provided (Delacroix & Veale, 2020).

On the one hand, a “right to receive information”—as guaranteed by Article 10 of the European Court of Human Rights—recontextualized in the digital context could help to establish what news consumers legitimately may expect from the news media with respect to the diversity or relevance of personalized recommendations (Eskens et al., 2017). In practice, this right could result in an effective explicit personalization—especially related to political news.

On the other hand, a “right to information explorability” to increase serendipitous encounters and to reduce potential filter bubbles and echo chambers (Reviglio, 2019b). This could be achieved increasing information findability and discoverability, for instance, with the possibility to choose the accuracy and diversity of information filtering. In practice, this might imply more interactive control of the algorithmic outputs and, as such, discoverability throughout sliders, topic categories, filters to navigate the information by source and keyword, or even *algorithmic recommender personae* as specific avatars to filter information (Harambam et al., 2018). As Hildebrandt (2019) advocates, societies must demand companies to explore and enable alternative ways of datafying and modeling the same person. Multiple and dynamic profiling could be paramount to this scope.

### Final Considerations

Certainly, the above outlined general suggestions need to be discussed further as they may have potentially unintended and undesirable consequences, impossible to systematically assess or even speculate in a short article such as this. Yet, we briefly identify some major weaknesses and critical issues.

If institutions empower users who cannot exercise their algorithmic sovereignty, they may rely on default choices, and this may only leave aside the weaker users and enable other actors as well as mainstream social media. This scenario is not different from a free market where the players are not equal, and it might even legitimize current mainstream platform at the expense of potential emerging competitors. Now, if we imagine the “liberalization” of algorithms would happen tomorrow, it is not hard to believe that a technocratic group of skilled individuals would try to claim their algorithm to be the best. By leveraging existing conditions of influences, they might just repropose the same form of algorithm oppression but with a variety of small actors in the market.

Also, this liberal market approach might lead to the misleading message that an algorithm can be better than another

one. The perfect algorithm, in absolute terms, does not exist. Every one of us has different priorities, interests, and time availability. **The correct algorithm is the one that best fulfills the needs of the individual, and this might not be true anymore if one begins to change.** The fitting algorithm thus cannot be permanent. Instead, the above-mentioned natural metrics, dashboards, and tools are fundamental, and should be imagined as the regulation imposed in the food industry (to declare the allergens, ingredients, kilocalories etc.). It can be assumed that negotiating these parameters and design choices should be an effort made and/or supported by the current democratic institutions and expert network bodies, such as the World Wide Web Consortium (w3c), Institute of Electrical and Electronics Engineers (IEEE), or Internet Engineering Task Force (IETF). In any case, it is desirable for a multistakeholder governance to meaningfully involve all the relevant actors.

Because algorithm sovereignty is a political challenge, the solution cannot be only technical. Algorithmic sovereignty indeed implicitly calls for more responsibility over citizens which ought to be able to decide on their own instead of delegating to someone else. We believe that institutions like schools, academia, and public service media ought to be proactively responsible for the most significant preconditions: digital, media, and algorithmic literacy.

## Conclusion

Personalization algorithms not only provide fruitful opportunities but also dystopian realities. Social network effects have indeed resulted in a few large transnational companies controlling the vast majority of social media use. Given this control, it is very difficult to establish alternatives that question the very principles on which the current business model and its perverse consequences exist. Nowadays personalization algorithms in social media tend to represent individuals and society as a body without contradictions and complexities, entities that can be reduced to a calculation to assure profitability. The outlined risks are fundamentally threatening individual progress and societal cohesion. There is a clear and increasing need to reimagine social media. To regain trust between users and platforms, personalization algorithms need to be seen not as a form of legitimate hedonistic subjugation, but as an opportunity for new forms of individual liberation and social awareness. For these reasons, a basic—and admittedly underdeveloped—notion of algorithmic sovereignty has been introduced and discussed: we pose that the individual and collective autonomy to govern personalization algorithms is a necessary step to imagine new forms of algorithmic arbitrage in social media.

Achieving such algorithmic sovereignty will likely be hard; there is a fundamental need to devise strategies that interrupt and in some cases outlaw surveillance capitalism's foundational mechanisms. Opening the black-box of personalization algorithms is only one piece of the puzzle. Through

further discussion, we need to realize how to reach a fair, effective, and accountable algorithmic sovereignty in social media. In this article, we outlined general and preliminary preconditions and principles. Despite vast technical, social, and political issues, the major problematic ultimately hinges on balancing conflicting democratic values, and translate them into code, design, and social practices. We believe that such endeavor is within reach and that public institutions and civil society can eventually sustain its realization.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Claudio Agosti worked in this project as part of ALEX, a project funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No 825974-ALEX).

## ORCID iD

Urbano Reviglio  <https://orcid.org/0000-0001-5948-1476>

## Notes

1. See the project tracking.exposed which hosts different interrelated projects with the aim to foster greater transparency in the context of personalization algorithms. For example, facebook.tracking.exposed, youtube.tracking.exposed, and eu19.tracking.exposed.
2. Implicit personalization determines user preferences from data collected (Thurman & Schifferes, 2012). It can actually increase political selective exposure—people's ability to see information that conforms to their pre-existing ideas and priorities—as it makes information avoidance less psychologically costly (Dylko et al., 2018).
3. Clearly, there is a fair amount of unpredictability in communication technology development, preventing precise predictions regarding what future implementations of personalization will look like. Yet, it is expected the rising of Ambient Intelligence related to the Internet of Things. This construct offers a vision in which automatic smart online and offline environments and devices interact with each other, taking an unprecedented number of decisions for us and about us to cater to our inferred preferences, representing a new paradigm in the construction of knowledge (Hildebrandt & Koops, 2010).
4. Individuals have different orders of preferences: “first-order preference” is expressed in how we behave in the moment that a stimulus or temptation affects our consciousness. In contrast, “second-order preference” is the choice we make for ourselves upon further reflection, generally separated from the immediate temptation. Think of the snake offering the forbidden fruit in the Garden of Eden (first-order preference: eat fruit), and the initial resistance (second-order preference: do not eat it). Online behavior seems to enable much more fluid expression of first-order preferences. In this way, personalization

- algorithms can shrink the individual “aspirational self.” Furthermore, there are also “third-order preferences,” meaning people make different decisions collectively than as individuals. These can also be affected by personalization algorithm.
5. Interpretability is the degree to which a human can understand the cause of a decision or consistently predict the model’s result (Miller, 2019).
  6. “In an A/B test, the experimenter sets up two experiences: ‘A’, the control, is usually the current system and considered the ‘champion’, and ‘B’, the treatment, is a modification that attempts to improve something—the ‘challenger’. Users are randomly assigned to the experiences, and key metrics are computed and compared” (<https://hbr.org/2017/09/the-surprising-power-of-online-experiments>).
  7. These techniques often rely on a highly contested scientific paradigm that argues all humans, everywhere, experience the same basic emotions, and express those emotions in the same way. Those emotions include happiness, anger, sadness, disgust, surprise, and fear. This paradigm of universal emotions is insufficiently evidence-based and poorly regarded in the relevant scientific communities. However, it could become a self-fulfilling prophecy, as a vast range of human expressions around the world are “coarsened” into a narrow set of machine-readable bins. As a result, the many and diverse ways of emotional expression, which vary from culture to culture, may be simplified and thus impoverished.
  8. A variable ratio reinforcement schedule occurs when, after  $n$  number of actions, a certain reward is achieved. Slot machines are a real-world example of a variable ratio.
  9. This article made the definition a common reference: <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>. And even if executives from YouTube denied <https://nymag.com/intelligencer/2019/03/youtube-exec-denies-the-existence-of-rabbit-hole-effect.html>, it really seems that YouTube drives such conspirational content (see [algotransparency.org](http://algotransparency.org)).
  10. During the Brexit referendum, the cross-party “Vote Leave” campaign commissioned 1,433 customized adverts promoting a more or less explicit pro-Brexit message (Kohl et al., 2019).
  11. Only about 36% of Facebook users intentionally tried to influence their newsfeed, while just 14% believe they have a lot of control over it (Smith, 2018).
  12. A paradigmatic example is YouTube’s recommendations which already drive more than 70% of the time spent in the video sharing platform and 90% of the “related content” is indeed personalized.  
See <https://www.cnet.com/news/youtube-ces-2018-neal-mohan> and <https://youtube.tracking.exposed/data>.
  13. See <https://gdpr-info.eu/>
  14. As such, a psychological analysis of people’s behavior into the platform may be provided to users to help them understand themselves. An example was the project *dataselfie* which analyzed users’ interaction with Facebook to offer them a glimpse into their inferred personality traits.
  15. More than 30 years ago, the free software movement put forward an ethical and legal framework that establishes four fundamental freedoms: (1) The freedom to run the program as you wish, for any purpose, (2) the freedom to study how the program works and to change it as you wish, (3) the freedom to redistribute copies so you can help your neighbor, and (4) the freedom to distribute copies of your modified versions to others.
  16. A directory to all alternative platforms can be found on <https://switching.software>.
  17. In software engineering, a project fork occurs when developers take a copy of source code from one software package and start independent development on it, creating a distinct and separate piece of software.
  18. See <https://algorithms.org/>
  19. See <https://fediverse.party/en/post/fediverse-in-2019>
  20. See <https://facebook.tracking.exposed/data-reuse>
  21. In United States, a bill about this issue (called “Deceptive Experiments To Online User Reduction”) has been recently proposed and discussed.
  22. The right is intended as “the right to decide whether or not they want to participate in experiments carried out by other actors (which usually goes hand in hand with surveillance) or other activities that involve registering or otherwise observing people’s lives and influencing their behaviour with technological means.” (van Est et al., 2017, p. 43).
  23. Efforts in this direction have been recently taken, for example, by Instagram. See <https://www.bbc.co.uk/newsround/48,134,723>

## References

- Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221), 509–514.
- Albanie, S., Shakespeare, H., & Gunter, T. (2017). Unknowable manipulators: Social network curator algorithms. *arXiv preprint arXiv:1701.04895*.
- algo:aware. (2018). *Raising awareness on algorithms*. Procured by the European Commission’s Directorate-General for Communications Networks, Content and Technology.
- Alter, A. (2017). *Irresistible: The rise of addictive technology and the business of keeping us hooked*. Penguin.
- Bennett, J. (2018). Public service algorithms. In D. Freedman & V. Goblet (Eds.), *A future for public service TV* (pp. 112–120). London: University of Goldsmith Press.
- Bodo, B., Helberger, N., Irion, K., Zuiderveen Borgesius, F., Moller, J., van de Velde, B., & de Vreese, C. (2017). Tackling the algorithmic control crisis -The technical, legal, and ethical challenges of research into algorithmic agents. *Yale Journal of Law and Technology*, 19, Article 3.
- Bozdag, E., & Timmermans, E. (2011, September 6). *Values in the filter bubble ethics of personalization algorithms in cloud computing* [Conference session]. 1st International Workshop on Values in Design—Building Bridges between RE, HCI and Ethics, Lisbon, Portugal.
- Burri, M. (2015). Contemplating a ‘public service navigator’: In search of new (and better) functioning public service media. *International Journal of Communication*, 9, 1341–1359.
- Calo, R. (2013). Digital market manipulation. *George Washington Law Review*, 82, 995–1051.
- Castro, J. C. L. (2016). Social networks as dispositives of neoliberal governmentality. *Journal of Media Critiques*, 2(7), 85–102.
- Couture, S., & Toupin, S. (2019). What does the notion of “sovereignty” mean when referring to the digital?. *new media & society*, 21(10), 2305–2322.

- De Filippi, P., & McCarthy, S. (2012). Cloud computing: Centralization and data sovereignty. *European Journal of Law and Technology*, 3, 1–18.
- Deibert, R. J. (2019). The road to digital unfreedom: Three painful truths about social media. *Journal of Democracy*, 30(1), 25–39.
- Delacroix, S., & Veale, M. (2020). Smart technologies and our sense of self: Going beyond epistemic counter-profiling. In M. Hildebrandt & K. O'Hara (Eds.), *Life and the law in the era of data-driven agency* (pp. 80–99). Edward Elgar Publishing.
- Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). Echo chambers: Emotional contagion and group polarization on Facebook. *Scientific Reports*, 6, Article 37825.
- DeVito, M. A. (2017). From editors to algorithms: A values-based approach to understanding story selection in the Facebook news feed. *Digital Journalism*, 5(6), 753–773.
- DiEM25. (2019). *Progressive agenda for Europe* (Technological Sovereignty: Democratizing Technology and Innovation Green Paper No. 3).
- Dylko, I., Dolgov, I., Hoffman, W., Eckhart, N., Molina, M., & Aaziz, O. (2018). Impact of customizability technology on political polarization. *Journal of Information Technology & Politics*, 15(1), 19–33.
- Esken, S., Helberger, N., & Moeller, J. (2017). Challenged by news personalisation: Five perspectives on the right to receive information. *Journal of Media Law*, 9(2), 259–284.
- Floridi, L. (2015). *The onlife manifesto*. Springer.
- Fogg, B. J., Lee, E., & Marshall, J. (2002). Interactive technology and persuasion. In J. P. Dillard & M. Pfau (Eds.), *The handbook of persuasion: Developments in theory and practice* (pp. 765–788). SAGE.
- Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of consumer research*, 21(1), 1–31.
- Fuchs, C. (2017). Towards the public service internet as alternative to the commercial internet. *ORF Texte*, 20, 43–50.
- Fuchs, C., & Marisol, S. (2015). The political economy of capitalist and alternative social media. In C. Atton (Ed.), *The Routledge companion to alternative and community media* (pp. 165–175). Routledge.
- Gal, M. S. (2017). Algorithmic challenges to autonomous choice. *Michigan Telecommunications and Technology Law Review*, 25, 59–104.
- Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018, April). *The dark (patterns) side of UX design* [Conference session]. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montréal, QC, Canada.
- Harambam, J., Helberger, N., & van Hoboken, J. (2018). Democratizing algorithmic news recommenders: How to materialize voice in a technologically saturated media ecosystem. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180088.
- Harris, T. (2019). U.S. Senate. Committee on Commerce, Science & Transportation. Hearing before the Subcommittee on Communications, Technology, Innovation, and the Internet on: Optimizing for Engagement: Understanding the Use of Persuasive Technology on Internet Platforms. [https://www.commerce.senate.gov/public/\\_cache/files/96e3a739-dc8d-45f1-87d7-ec70a368371d/35DE175BA77F95FB2EFC5E94018ABF1A.06-25-19harris-testimony.pdf](https://www.commerce.senate.gov/public/_cache/files/96e3a739-dc8d-45f1-87d7-ec70a368371d/35DE175BA77F95FB2EFC5E94018ABF1A.06-25-19harris-testimony.pdf)
- Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, 7, 993–1012.
- Hildebrandt, M. (2015). *Smart technologies and the end(s) of law: Novel entanglements of law and technology*. Edward Elgar Publishing.
- Hildebrandt, M. (2019). Privacy as protection of the incomputable self: From agnostic to agonistic machine learning. *Theoretical Inquiries in Law*, 20(1), 83–121.
- Hildebrandt, M., & Koops, B. J. (2010). The challenges of ambient law and legal protection in the profiling era. *The Modern Law Review*, 73(3), 428–460.
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
- Kidron, B., Evans, A., Afia, J., Adler, J. R., Bowden-Jones, H., Hackett, L., & Scot, Y. (2018). *Disrupted childhood: The cost of persuasive design*. 5Rights.
- Kohl, U., Davey, J., & Eisler, J. (2019). *Data-driven personalisation and the law—A primer: Collective interests engaged by personalisation in markets, politics and law*. [https://eprints.soton.ac.uk/428082/1/Data\\_Driven\\_Personalisation\\_and\\_the\\_Law\\_A\\_Primer.pdf](https://eprints.soton.ac.uk/428082/1/Data_Driven_Personalisation_and_the_Law_A_Primer.pdf)
- Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788–8790.
- Krippendorff, K. (1993). Major metaphors of communication and some constructivist reflections on their use. *Cybernetics & human knowing*, 2(1), 3.
- Lovink, G., & Rasch, M. (2013). *Unlike us reader: Social media monopolies and their alternatives* (INC Reader No. 8). [https://monoskop.org/images/7/7b/Lovink\\_Geert\\_Rasch\\_Miriam\\_eds\\_Unlike\\_Us\\_Reader\\_Social\\_Media\\_Monopolies\\_and\\_Their\\_Alternatives.pdf](https://monoskop.org/images/7/7b/Lovink_Geert_Rasch_Miriam_eds_Unlike_Us_Reader_Social_Media_Monopolies_and_Their_Alternatives.pdf)
- Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *AI & Society*.
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38.
- Moen, G. M., Ravna, A. K., & Myrstad, F. (2018). *Deceived by design*. Forbrukerrådet.
- Nagulendra, S., & Vassileva, J. (2016). Providing awareness, explanation and control of personalized filtering in a social networking site. *Information Systems Frontiers*, 18(1), 145–158.
- Ostrom, E. (2015). *Governing the commons*. Cambridge University Press.
- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Reviglio, U. (2019a). Improving user experience by browser extensions: A new role of public service media? In S. El Yacoubi, F. Bagnoli, & G. Pacini (Eds.), *Internet Science. INSCI 2019. Lecture Notes in Computer Science*, vol 11938 (pp. 257–271). Springer.
- Reviglio, U. (2019b). Serendipity as an emerging design principle of the infosphere: Challenges and opportunities. *Ethics and Information Technology*, 21(2), 151–166.
- Roio, D. (2018). *Algorithmic sovereignty* [Doctoral dissertation]. University of Plymouth.
- Scholz, T. (Ed.). (2012). *Digital labor: The internet as playground and factory*. Routledge.

- Smith, A. (2018). *Many Facebook users don't understand how the site's news feed works*. Pew Research Center, Journalism and Media.
- Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press.
- Taylor, L., Floridi, L., & Van der Sloot, B. (Eds.). (2016). *Group privacy: New challenges of data technologies* (Vol. 126). Springer.
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.
- Thurman, N., & Schifferes, S. (2012). The future of personalization at news websites: Lessons from a longitudinal study. *Journalism Studies*, 13(5–6), 775–790.
- Tucker, J., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., & Nyhan, B. (2018). *Social media, political polarization, and political disinformation: A review of the scientific literature*. William and Flora Hewlett Foundation. <https://ssrn.com/abstract=3144139>
- van Est, Q. C., Gerritsen, J., & Kool, L. (2017). *Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality*. The Rathenau Institute.
- Vercellone, C., Bria, F., Fumagalli, A., Gentilucci, E., Giuliani, A., Griziotti, G., & Vattimo, P. (2015). *Managing the commons in the knowledge economy*. European Union's Seventh Framework Programme.
- Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., & Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? A critical review. *Social Issues and Policy Review*, 11(1), 274–302.
- Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society*, 20(1), 118–136.
- Yeung, K. (2018a). Five fears about mass predictive personalisation in an age of surveillance capitalism. *International Data Privacy Law*, 8, 258–269.
- Yeung, K. (2018b). *A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework*. <https://ssrn.com/abstract=3286027>
- Youyu, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4), 1036–1040.
- Zarsky, T. (2019). Privacy and manipulation in the digital age. *Theoretical Inquires in Law*, 20(1), 157.
- Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75–89.
- Zuiderveen Borgesius, F. J., Trilling, D., Moeller, J., Bodó, B., De Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, 5(1), 102–114.

### Author Biographies

**Urbano Reviglio** is a PhD candidate of the International Joint Doctorate in Law, Science, and Technology, University of Bologna, Italy. His background is interdisciplinary, in particular, the ethics and governance of social media. His research includes the study of filter bubbles, echo chambers, and the value of digital serendipity.

**Claudio Agosti** is a fellow researcher at the University of Amsterdam. He has a hacker background and security professional. He researches, implements, and promotes technology in the public interest. In the last decade, he addressed issues such as communication protection for whistleblowers, analysis of the web-tracking phenomenon, and algorithm analysis.