PREPRINT VERSION
Airoldi, M. (2023). Computational Authority in Platform Society: Dimensions of Power in Machine Learning. In C. Borch, J.P. Pardo-Guerra (eds), *The Oxford Handbook of the Sociology of Machine Learning*. https://doi.org/10.1093/oxfordhb/9780197653609.013.40

Computational Authority in Platform Society: Dimensions of Power in Machine Learning

Massimo Airoldi, Department of Social and Political Sciences, University of Milan

Abstract

A growing body of literature maintains that AI and machine learning (ML) technologies have—or exert—some kind of power: power to act in ways that do not mechanically reflect developers' intentions and values, as well as power over the individuals, social groups, and cultural manifestations subjected to automated predictions and classifications. This chapter aims to dissect the notion of ML power, intended both as power to and power over. First, two complementary scholarly perspectives on ML power to, here labeled "networked" and "cultural" views, are presented and critically discussed. Second, drawing on both classic and contemporary social theories, the chapter outlines four main dimensions of ML power over—as opaque coercion, computational authority, structural conditioning, and data-driven governmentality.

Keywords
power, machine learning, AI, artificial intelligence, nonhuman agency, post-human social theory, computational authority

Introduction

> For a long time social theory has been concerned with defining power relations . . . but it has always found it difficult to see how domination is achieved. . . . In order to understand domination we have to turn away from an exclusive concern with social relations and weave them into a fabric that includes non-human actants, actants that offer the possibility of holding society together as a durable whole. (Latour, 1991, p. 103)

Power—with its forms, mechanisms, and multiple dimensions—represents one of the major concerns, if not the obsession, of modern sociology. Two main views of the concept, as *power to*—that is, the transformative capacity of power—and *power over*—that is, power as domination over others—point to a complex ramification of contrasting theories that continue to animate debates in the social sciences (Haugaard, 2020; Pansardi, 2012).

Social theory sees power essentially as a human "thing" (Latour, 1991). Until recently, only limited scholarly attention was devoted to understanding how power relations are sedimented in and affirmed through technology and material artifacts, in subfields like science and technology studies and actor-network theory (Law, 1991). Something changed at the beginning of the 2000s, when a new breed of nonhuman actants started to be pervasively embedded into the fabric of society. Computer algorithms and basic AI systems were applied to a number of professional fields for some decades already—for example, in financial markets (Pardo-Guerra, 2010). Yet it is only with the development of the Internet and, later, of social media platforms and mobile apps that unprecedented opportunities for profitably automating information processing through the extraction of "big" digital data emerged (Zuboff, 2019; van Dijck et al., 2018). Automated systems such as spam filters, recommender algorithms, search engines, online bots, dynamic pricing, and machine translation models began to populate digital infrastructure and, therefore, to affect our everyday lives as well as culture and society more broadly. As a result, novel theorizations of the shifting relationship between information technologies, power, and society made their appearance in social science journals, giving rise to new and highly multidisciplinary fields of research, such as critical algorithm studies and critical data studies. Lately, rapid advancements in AI research and an increased media coverage of the mythicized successes, blatant failures, and dystopian applications of machine learning (ML) technologies have contributed to bringing the power of ML and AI to the very center of contemporary sociological debates (Burrell & Fourcade, 2021). With large language models creating seemingly original texts, image-generation systems producing award-winning visual art, recommender algorithms directing cultural consumption, and opaque predictive models put to work in almost every field of the social, there is a growing consensus about the fact that sociological theories of power can ignore nonhuman actants no more (Borch, 2022; Schwarz, 2021).

As Latour (1991) early noted, society is invisibly held together by technological artifacts. This is even more evident now that, among many nonhuman actants, there are some endowed with special cultural and social *powers*. Differently from "good, old-fashioned" AI systems and conventional rule-following algorithms, the ML models at the root of current AI applications can inductively learn how to use and manipulate language and culture, thanks to the capturing of human cognitive abilities in the form of training data (Mühlhoff, 2020). In the absence of conscience or humanlike understanding, ML systems can nonetheless connect and process data patterns in ways that successfully simulate cultural competence and communication (Esposito, 2022). The resulting outputs and predictions can have *powerful* consequences on individuals—for instance, in the unpleasant form of rejected job applications, discriminatory social representations, or micro-targeted marketing manipulations. In spite of its documented failures and biases, automated decision-making continues to bear a strong *authority* rooted in the (alleged) objectivity, infallibility, and neutrality of mathematical calculus (Campolo & Crawford, 2020). This other softer, "discursive" power of algorithms and AI is decisively boosted by the technological myths circulating among the "coding elite" (Burrell & Fourcade, 2021) and, increasingly, in public opinion (Beer, 2017).

A growing body of literature in the social sciences maintains that ML technologies have—or exert—power: *power to* act in ways that do not mechanically reflect developers' intentions and values, as well as *power over* the individuals, social groups, and cultural manifestations ordinarily subjected to automated predictions and classifications. The present chapter aims to dissect the notion of ML power, intended both as *power to* and *power over*. While this conceptual distinction is still debated in social theory (Pansardi, 2012), here it will serve solely as a sociological lens to critically map two questions widely resonating in current research: Do ML systems possess some kind of agency? And how do they affect individuals, culture, and society?

I will first outline recent scholarly perspectives on the agency of automated systems, in conversation with human-centric and post-human theoretical perspectives. Then, drawing on classic and contemporary social theories examining the multidimensionality of power as domination, I will highlight four main dimensions of ML *power over*: as opaque coercion (D1), computational authority (D2), structural conditioning (D3), and data-driven governmentality (D4).

Power To: Agency in Machine Learning

Weber (1964) has defined power as the "probability that one actor within a social relationship will be in a position to carry out his own will despite resistance" (p. 152). In this sense, the notion of power refers to an asymmetric social relation (*power over*). Yet another common and more basic interpretation of the term exists: *power* also means the capacity to act and to produce certain outcomes (Pansardi, 2012, p. 75). This second idea of power as *power to* points to the "ability of an individual actor to attain an end or series of ends" (Allen 1999, p. 126). As such, it is often positively connoted as "empowerment" as well as commonly associated with technology (e.g., "a powerful machine").

Definitions of agency in social theory substantially coincide with the above interpretation of *power to*. For instance, according to Giddens (1984), agency is the "capability of the individual to 'make a difference' to a pre-existing state of affairs or course of events[—]that is, to exercise some sort of power" (p. 14). Hence, asking to what extent AI and ML systems have *power to* entails mapping different perspectives on the extent to which they can be interpreted as autonomous "agents" in society.

Nonhuman agency is a long-debated topic in sociology (Cerulo, 2009). In the 1980s, when anthropocentric theories stressed the role of (socially constrained) intentions and consciousness in guiding individual behavior (Boudon & Bourricaud, 1989), actor-network theorists were among the first to propose a truly "symmetric" interpretation of agency, relationally extended to the "missing masses" of material objects, bacteria, ideas, and technological devices (Latour, 1992; Law, 1991). According to this alternative view, the power of "actants" to participate in social life is not rooted in humanlike consciousness or deliberate intentions but rather in networks of sociomaterial "alliances." For instance, in the essay quoted at the beginning of this chapter, Latour (1991) illustrates the agentic role of objects in the exercise of power with the curious example of metal weights that are attached to (old-fashioned) hotel room keys, which actively contribute to increasing the probability that the will of their human allies (i.e., hotel managers) will be carried out against against careless customers who may lose the keys.

The resulting notion of nonhuman agency, intended as the emergent property of ramified arrangements of "sociomaterial" (or "sociotechnical") relations, has spread well beyond the boundaries of Actor Network Theory (ANT) (Cerulo, 2009) to be then widely adopted by the critical literature on platforms, algorithms, and AI (see Neyland, 2019).

According to this first perspective on AI agency, which I will call the *networked view*, ML models are elements within complex more-than-human arrangements made of digital infrastructures, data points, material features, protocols, parameters, cultural values, and arbitrary goals. Algorithms are "heterogeneous and diffuse sociotechnical systems" in themselves, and this implies that, rather than seeing AI technologies as "singular technical objects that enter into many different cultural interactions," sociologists should interpret them as "unstable objects," constantly enacted by sociomaterial practices and, ultimately, "inseparable" from an ecology of sociotechnical relationships (Seaver, 2017, p. 1). As anthropologist Nick Seaver (2017) puts it, "These algorithmic systems are not standalone little boxes, but massive, networked ones with hundreds of hands reaching into them, tweaking and tuning, swapping out parts and experimenting with new arrangements" (p. 5).

This networked perspective on AI agency challenges what Ziewitz (2016) provocatively calls "algorithmic drama"—that is, the widespread idea that algorithms are "powerful and consequential actors in a wide variety of domains," mysterious technological creatures "imbued with agency and impact," which are therefore treated as veritable "subjects" by social science scholars (p. 5). As Ziewitz and others suggest, the frequent fetishization of algorithms as independent, powerful, black-boxed agents is due to a lack of engagement with what an algorithm actually is in the first place. At a closer look, it becomes evident that an algorithm cannot "hold power" by itself: its "apparent" agency resides instead in "a broader set of associations" (Neyland, 2019, pp. 7–8; see also Law, 1991).

Following a *networked view*, we cannot disentangle the agency of an ML system like, say, Spotify's Discover Weekly (Prey, 2018) from a complex set of invisibilized human practices and technical features, just as we cannot conceive the agency of Latour's metal weights in isolation, without hotel, room keys, customers, hotel managers, and their human intentions.

However, several scholars have pointed out that ML systems are substantially different from both material objects and simpler rule-following algorithms and that this difference has important implications when it comes to sociologically defining ML agency and power (Airoldi, 2021; Borch, 2022; Schwarz, 2021; Cardon, 2018). In fact, ML models work by recursively "learning" from patterns in training and feedback data. On the one hand, differently from static artifacts like the Latourian metal weights, ML algorithms *actively* respond to a datafied environment by dynamically adapting their predictions—for instance, platform users' behavioral data at t_0 influence automated content recommendations at t_1. On the other, in contrast with rule-following algorithms, whose actions are mechanically determined by developers' prior choices, current ML applications make autonomous—and, often, inexplicable (Burrell, 2016) or even "alien" (Parisi, 2019)—decisions, following a statistically inductive logic.

Building on these premises, a *cultural view* of AI agency has recently gained traction in sociology (e.g., Esposito, 2022; Borch, 2022; Airoldi, 2021; Fourcade & Johns, 2020). Without neglecting the ramified sociotechnical systems that ML algorithms depend on or replicating "dramatic" understandings of algorithmic power, proponents of this *cultural view* stress the autonomous character and sociocultural roots of automated decision-making. As Borch (2023) notes, "ML systems do not merely perform technologically mediated actions, behaving as human delegates. In some domains, they are equipped with true independent agency" (p. 5). This is the case of financial trading, whereby myriad artificial agents constantly work and interact with each other without any need for human supervision (Borch, 2022; Mackenzie, 2019). Notably, seeing ML systems as autonomous agents does not imply believing in a "general artificial intelligence" (Fjelland, 2020): even without "consciousness, intentionality or meaning in any recognizable human form," ML algorithms retain "some capacity for agency" (Borch, 2022, p. 504). Once embedded in the hybrid ecosystem of digital platforms and devices, these peculiar machines acquire the capability to "make a difference" in the social world through performative predictions and classifications (Airoldi, 2021)—quite like their human counterparts, who also rely on resources, relations, and social structure to exert *power to* (Giddens, 1984; Pansardi, 2012).

I call this second view of AI agency *cultural*, since cultural patterns baked into data are indicated as the very thing that makes AI agency possible in the first place. The regularities in language, visual culture, consumption, and classification that ML systems are ordinarily fed with do not produce solely biased outputs and stereotypical predictions, as critics tend to emphasize (e.g., Noble, 2018); from a cultural perspective, such human-generated patterns are precisely what allow these algorithms to efficiently simulate a humanlike understanding of the social world (Mühlhoff, 2020) and, thus, to behave as "social agents" (Airoldi, 2021). For instance, sociologist Elena Esposito (2017) has noted how "machines parasitically take advantage of the user participation on the web to develop their own ability to communicate competently and informativel" (p. 251). The impressive conversational skills displayed by large language models are certainly a case in point (Weinberg, 2020).

The two perspectives on AI agency described above are complementary, since they can help address research problems from different yet equally useful angles. Research adopting a *cultural view* invites us to think of ML systems as "data-hungry" agents participating in forms of feedback-based social learning (Fourcade & Johns, 2020)—what in my work I call "machine socialization" (Airoldi, 2021). Inspired by classical theoretical perspectives like symbolic interactionism (Mackenzie, 2019; Yolgörmez, 2021), Luhmann's theory of communication (Esposito, 2022, 2017), or Bourdieu's relational sociology (Airoldi, 2021), scholars sharing a cultural view of AI agency are interested in exploring possible directions for a truly post-human sociology. As Borch (2023) counterintuitively suggests, a networked approach might be illsuited for this ambitious purpose: Focusing on how human agency is entangled with nonhuman actants, "ANT fails to account for the ways in which distinctively inter-algorithmic activities play out" (p. 10) (e.g., among high-frequency trading systems; see Mackenzie, 2019).

However, adopting a cultural view also entails limitations. Emphasizing the contingently autonomous, culturally driven, and intrinsically opaque functioning of ML can degenerate into a soft anthropomorphism and risks complicating the important question of who is ultimately accountable for the exercise of ML power (Campolo & Crawford, 2020). If the human "principals"—for example, developers, managers, tech entrepreneurs—can claim no control over their artificial "agents," then it becomes harder to establish clear "chains of power" and responsibility (Reed, 2020), especially in the case of controversial or discriminatory outcomes of automated decision-making (Burrell & Fourcade, 2021). Conversely, studies presenting a *networked view* of AI agency generally concentrate on the situated development and organizational deployment of these technologies, often by means of immersive ethnographic studies (e.g., Seaver, 2022). This different analytical angle can fruitfully illuminate the human hands silently directing the fragile functioning of ML systems (Seaver, 2017; Neyland, 2019) as well as the invisibilized processes of labor exploitation—for example, data annotation, AI impersonation, and AI verification (Tubaro et al., 2020)—which are indispensable for sustaining the myth of a truly "artificial" and "intelligent" machine (Crawford, 2021; Broussard, 2018).

Power Over: ML and Domination

In the social sciences, the power of ML systems has been mainly theorized as *power over* human subjects and society in general (Burrell & Fourcade, 2021; Schwarz, 2021). Sociological takes on this "social power of algorithms" (Beer, 2017) have proliferated in the past decades, in close connection with a broader debate on the societal implications of digital platforms and datafication (van Dijck et al., 2018). Research in this area has extensively covered topics such as the effects of bots and online algorithms on political behavior and public opinion (Gandini et al., 2022), the consequences of recommender systems on cultural consumption and classification (Prey, 2018), or the role of automated systems in the reproduction of social inequalities and systemic racism (Eubanks, 2018). Scholars have characterized the modalities of this novel "algorithmic dominion" (Burrell & Fourcade, 2021, p. 225) through a vast repertoire of concepts and labels, such as "algocracy" (Aneesh, 2009), "hypernudging" (Yeung, 2017), or "soft biopower" (Cheney-Lippold, 2011). Some of these notions, such as "filter bubble" (Pariser, 2011) or "surveillance capitalism" (Zuboff, 2019), have spread well beyond academia to become popular buzzwords, indicating the rise of novel forms of domination that are enabled by increasingly autonomous calculative technologies.

In the 2000s, the main characters of this then emerging body of literature were computer algorithms, generally speaking David Beer (2009) was among the first sociologists to notice how, with the diffusion of social media platforms and the multiplication of online-based human-machine interactions, these systems acquire "the capacity to shape social and cultural formations and impact directly on individual lives" (p. 994). Algorithmic *power* was largely portrayed as an abstract, black-boxed force *over* Internet users (Amoore & Piotukh, 2016; Pasquale, 2015; Mackenzie, 2006). More recently, technologically deterministic and "dramatic" (Ziewitz, 2016) accounts of algorithmic power have gradually given way to closer investigations of the peculiar types of power relations established by AI and ML (Schwarz, 2021; Fourcade & Johns, 2020; Campolo & Crawford, 2020) as well as to a growing scholarly interest in the imaginaries and forms of resistance displayed by the dominated humans in the loop (Bucher, 2017; Velkova & Kaun, 2021).

Below, I critically review a selection of the vast body of literature on ML *power over*. Drawing inspiration from human-centric theorizations of the multiple dimensions of power (Haugaard, 2020; Lukes, 1974), I illustrate how existing research on the social power of AI and ML can be summarized as pointing to four main dimensions of ML power: over individual lives (D1, D2) and over culture and society (D3, D4).

D1: ML Power as Opaque Coercion

Classic social theory argues that power is latent within social relations and manifests itself either as coercive force or as institutionalized authority (Haugaard, 2020; Weber, 1978). While authority depends upon belief, force has "a physical existence irrespective of meaning" (Haugaard, 2021, p. 155). The physical action of violence is the clearest example of a coercive form of power, commonly used as a threat in order to ensure compliance—as in a "your money or your life" situation (a, p. 733).

Leaving aside the noteworthy case of AI-driven killer robots, how an ML model could coerce human behavior through force might not be entirely evident at first sight. The "narrow" AI systems we ordinarily encounter in our digital peregrinations seem quite innocuous after all (e.g., music recommender systems, helpful machine translation tools, playful image and text generation models). Still, other and probably less benign computational models are put to work by corporations and public administrations in order to flag potential criminals, recognize human faces in public squares, rank consumer credit, monitor work performance, or scrutinize job applications (Burrell & Fourcade, 2021; Bucher, 2018). Such systems have tremendous power over the individual lives of citizens, workers, and consumers. Their non-neutral and certainly not infallible predictions can instantaneously shape one's life-chances—for instance, by deciding who will be eligible or not for a loan or health insurance (Fourcade & Healy, 2013). And even in the case of the apparently innocuous algorithms distributing content and ads online, "softer" forms of coercion are in place, aimed at directing user behavior in highly personalized ways. In the context of platform-based user-machine interactions, ML systems gently "force" us to do (or not do) things all the time—to watch funny videos, to remove social media posts, to check notifications, or to take the third street on the left. Instead of simply anticipating our needs, these systems end up manufacturing them, "nudging" our behavior in (profitable) directions (Yeung, 2017). For example, automated recommendations in Netflix are estimated to generate about 80% of hours of streamed content (Gomez-Uribe & Hunt, 2015), and comparable figures are likely to characterize other digital services. It is no surprise then that, without making explicit reference to terms like "force" or "violence," scholars have nonetheless frequently depicted the social power of algorithms and AI systems as a form of coercion over datafied human subjects, whose agency is seen by definition as "contested in and through algorithms" (Mackenzie, 2006, p. 44). For instance, papers by Bucher (2012) and Cotter (2019) narrate, respectively, Facebook users and Instagram influencers as the victims of a constant "threat of invisibility"—as in a digital and algorithmic version of the aforementioned "your money or your life" situation. Still, a recent sociological literature notes how ML-based coercion mechanisms partially differ from both the coercive exercise of social power portrayed in anthropocentric theories and the mere augmentation of human force enabled by less autonomous technological artifacts (Schwarz, 2021; Cardon, 2018).

As I have argued above, ML methods are designed to pursue a mathematically formulated goal through an inductive (supervised or unsupervised) learning process rooted in data examples. If the goal is decided by the developers and, thus, reflects their values and intentions, the final formulation of the ML model and its contingent implementation do not follow any predetermined script; conversely, they statistically emerge in a dynamic relation with data (Airoldi, 2021). This "operational autonomy" (Borch, 2022) bears important implications when reflecting on whether and how ML systems can coerce human action.

First, it implies a specific kind of opacity, rooted in the complex calculations and high-dimensional data flows characterizing ML (Borch & Hee Min, 2022; Burrell, 2016). ML methods base their decision-making on the computational analysis of thousands, if not millions, of data "features"—for example, the properties of each single pixel that allows a neural network to recognize visual elements in previously unseen images. At this scale, manually inspecting the datasets is an almost impossible task, as it is the post hoc interpretation of the final predictions, regardless of the level of human expertise. As Burrell (2016) has noticed, "Machine optimizations based on training data do not naturally accord with human semantic explanations" (p. 10). Borch and Hee Min (2022) argue the same, also highlighting the challenges that this poses to computer science research on "explainable AI."

In addition to the complexity-driven inscrutability of ML methods, Burrell (2016) identifies two other forms of opacity characterizing algorithmic systems more generally: intentional corporate or state secrecy—also discussed by Pasquale (2015)—and the opacity stemming from users' technical illiteracy—explored by recent research on digital inequalities and "algorithmic awareness" (Gran et al., 2021). As a result, interactions between automated systems and datafied human subjects are generally marked by important informational asymmetries (Airoldi, 2021, pp. 89–90), which obscure the rationale and, sometimes, the very exercise of ML power. Such opaque interactional configurations regularly happen, for instance, in the context of "algorithmic media" like digital platforms (Bucher, 2018) or at the expense of workers, who are subjected to various kinds of automated coercion, threat, retaliation, and control (Kellogg et al., 2020). On this line, Beer (2009) has noted how the opacity of algorithmic domination makes it a sort of "technological unconscious," an unknowable force that invisibly "produces" everyday life. "Being constantly subjected to algorithmic decision-making makes it appear as the natural order of things . . . and not as an exercise of power," the sociologist Ori Schwarz (2021, p. 138) writes similarly. In his *Sociological Theory for Digital Society*, Schwarz (2021, pp. 145–146) argues that the opaque rules at the root of ML power, while equally abstract, differ from the regulatory rules of Weberian bureaucracies insofar as they are secret, unintelligible, and, paradoxically, not "calculable."

This leads to my second and closely related point: ML systems operate within digital infrastructure according to what theorist Scott Lash (2007) has called "generative rules"—that is, "virtuals that generate a whole variety of actuals" (p. 71). Through these generative rules, more than merely "mediating" social life, algorithms instantaneously "constitute" it (Beer, 2009)—for instance, by composing a TikTok feed suited to my datafied taste, which will subtly encapsulate my whole platformized experience. ML systems do not simply "afford" specific types of human conduct or passively "script" their possible choices; their conscienceless statistical predictions performatively shape action. As Schwarz (2021) explains, "Generative rules are artificial human-made rules that operate in a way similar to the rules of nature" (p. 131). In contrast with theorizations of power as a symbolic authority depending on the recognition of the governed (Weber, 1978), the sociological literature building on Lash's work describes ML power essentially as "post-hegemonic"—that is, a "naked force without legitimation" (Schwarz, 2021, p. 133), closer to an earthquake than to a court ruling and, as such, similar to what Zuboff (2019) calls "uncontracts," contracts that are enforced and formulated unilaterally and automatically, regardless of platform users' will or awareness.

D2: ML Power as Computational Authority

Like physical violence, the opaque and coercive dimension of ML domination outlined above exists independently from any human legitimation, understanding, or belief (Haugaard, 2020). Whenever I purchase a plane ticket online, I am inevitably subjected to the hidden generative rules of dynamic pricing systems, forcing me to buy at a tailored and probably unfair price. Still, as a user, I retain some agency in this human-machine interaction. I try to hypothesize which digital actions may result in an increase of the final price—repeated online searches coming from my IP address could be used as indicator of interest on my part!—and act differently in response. If I suspect that the dynamic pricing system is unfair, I may decide to change websites or even airlines. Trust (or faith) in the automated technology is needed also on the part of the other humans involved in the sociotechnical assemblage surrounding the ML agent: computer scientists maintaining the system operational, investors seeking profit, managers pursuing monthly sales targets or curating customer relations.

This example suggests that it is possible to derive a second ideal-typical dimension of ML power, as an authority that is conferred and acknowledged by the interacting human subjects (Weber, 1978). This computational authority is linked—and yet irreducible—to human trust toward artificial agents (Sundin et al., 2017) and has been conceptualized by a multidisciplinary literature. Lustig and Nardi (2015) define "algorithmic authority" as "the legitimate power of algorithms to direct human action" (p. 743) and examine it in the case of the blockchain cryptocurrency system of Bitcoin. Other works deploy this notion in relation to human-machine collaborations in healthcare to explore how the computational authority of AI-based diagnostic tools ends up affecting the practices and epistemic authority of human clinicians (e.g., Racine et al., 2019). In the critical and sociological literature, the notion of computational authority has been evoked in relation to the epistemic power of search engines, which dictate what we commonly deem as "true" or "relevant" (Rogers, 2013, pp. 97–100), as well as in broader theorizations of the rise of an (authoritarian) "algorithmic culture" (Striphas, 2015) characterized by the "offloading of cultural work" to ubiquitous data processing.

All contributions emphasizing the relevance of computational authority point to the fact that "conclusions described as having been generated by an algorithm wear a powerful legitimacy" (Gillespie, 2016, p. 23). Since the times of Leibniz and Babbage, the promises and fears of calculability and automation have fueled mythical cultural discourses linked to modern processes of rationalization and have culminated with the buzz around “big data.” In this centennial story, AI is the undisputed protagonist, the narrative keystone bridging Internet countercultures and positivism, science fiction dystopias, and business prophecies (Natale & Ballatore, 2020). Driven by mathematical calculus rather than by “irrational” emotions or “weak” flesh, thinking machines were and still are narrated as infallible and objective (Crawford, 2021). This widespread ideological discourse—intentionally blind to the sociocultural roots of code and data—justifies the ubiquitous implementations of ML techniques by magnifying their powers and, thus, makes the exercise of computational authority possible in the first place (Broussard, 2018; Beer, 2017).

Burrell and Fourcade (2021) note that “the notion that a mechanistic, impersonal process is superior to one rooted in the discretion of individuals” is “not an invention of the computer age” (p. 222). In fact, they maintain that the discursive legitimization of computational authority presents important similarities with the rise of a rational-legal authority, theorized by Weber (1978) and embodied by the expert functionary following “calculable rules” and deciding “without regard for persons” (p. 975). Still, according to Campolo and Crawford (2020), computational authority is not entirely reducible to this Weberian ideal-type: Its social legitimacy derives instead from a paradoxical combination of modern disenchantment and an “enchanted epistemology” rooted in magic thinking. The ambivalence of this “enchanted determinism” is particularly evident in the case of deep learning: “When the disenchanted predictions and classifications of deep learning work as hoped, we see a profusion of optimistic discourse that characterizes these systems as magical, appealing to mysterious forces and superhuman power” (Campolo & Crawford, 2020, p. 5).

However, ML models do not always work as hoped. Social media recommendations can be “faulty” and “out of sync” with users’ interests and beliefs (Bucher, 2017, p. 35). AI systems make naïve mistakes all the time, sometimes producing harmful and discriminatory results (Broussard, 2018; Noble, 2018; Eubanks, 2018). Therefore, individuals may start questioning their computational authority and become themselves disenchanted toward autonomous systems and their mythicized power (Airoldi, 2021, p. 94). Several scholars have argued that experiences of glitched, offensive, or simply “irritating” behaviors by ML systems can open up spaces for grassroots forms of resistance against algorithmic domination (Ruckenstein, 2023; Velkova & Kaun, 2021). Authority is “inextricably linked to the performance of it” (Haugaard, 2021, p. 155), and so is computational authority. As such, it is fragile, subjected to trust withdrawn or disbelief, and constantly threatened by resistance.

I argue that, while apparently incompatible, D1 and D2 often coexist in human-machine interactions. We can see them as the ideal-typical poles of a dynamic continuum describing the extent to which human subjects are conscious of the contingent exercise of ML power (Airoldi, 2021). In the case of strong informational asymmetries, such as when there is no direct interaction between the classified human and the classifying machine, D1 becomes prevalent. ML power is then an opaque coercion, a “technological unconscious” (Beer, 2009), a “naked force without legitimation” (Schwarz, 2021, p. 133): As my credit score is computed, an automated system instantaneously neglects the loan application (Fourcade & Healy, 2013). On the contrary, when human interactants are perfectly aware of the machine behind the screen, the power relation assumes the characteristics of D2: The surgeon consciously decides to perform the operation in accordance with the AI-driven suggestions (Racine et al., 2019). A computational authority is then recognized—or, as it happens, resisted.

D3: ML Power as Structural Conditioning

In *Power: A Radical View*, Steven Lukes (1974) articulates three dimensions of *power over*. The first consists in the ability of openly influencing the actions of the dominated; the second is about non-decision-making—that is, the behind-the-eye capacity of "setting the agenda" and constraining agency. It can be argued that these first two dimensions bear some similarities with my characterization of, respectively, D2 and D1—though a systematic comparison is beyond the scope of the present chapter.

Luke's third and final dimension of power is of particular relevance here, since it concerns culture and social structure. In his view, three-dimensional power consists in the ideological manipulation of commonsense and practical experience, which makes dominated subjects unwarily participate in the social reproduction of their own subordinate condition. This modality of power is close to Gramsci's "hegemony" and Bourdieu's "symbolic violence" (Bourdieu & Wacquant, 1992), since it works in and through culture by means of hierarchical relations inscribed in tacit social knowledge—which are naturalized as such (Haugaard, 2020).

Critical research on AI, algorithms, and platforms often indicates ML systems as key technical and discursive elements of the hegemonic domination characterizing surveillance capitalism (Markham, 2021; Zuboff, 2019), which opposes a ruling "coding elite" to a fragmented "cybertariat" (Burrell & Fourcade, 2021). Moreover, papers and books denouncing the risks and social implications of bias in ML overall agree on the fact that social inequalities have become increasingly "automated" and, as such, further invisible (Eubanks, 2018; Benjamin, 2019; Broussard, 2018).

The opaque functioning and mythical aura of AI technologies tend to obscure the power differences segmenting society, turning them into "objective" mathematical facts sharply separated from it and, thus, contributing to their "reification"—that is, the process "whereby the social constructedness of structures is denied" (Haugaard, 2021, p. 165). For instance, Noble (2018) has painstakingly shown how search engines swiftly reinforce race and gender discriminations through their computational outputs, which are nonetheless perceived by Internet users as neutral and authoritative. Studies have also shed light on how the stereotypes and biases present in police reports are technologically amplified when such data are used to train predictive policing models (Burrell & Fourcade, 2021, p. 233). These and other contributions point to a fundamental question raised by D3: How do ML systems shape the symbolic and social structure of society?

The structural conditioning that ML exerts on and within society, which in my work I call "techno-social reproduction" (Airoldi, 2021), has been described by various scholars as recursive. Beer (2022) notes that "where algorithms are present then actions are taken based upon, informed by or shaped by the presence of data from previous actions," this process leading to "multiple feedback loops, each endlessly feeding into the next" (p. 1). The recursive reproduction of pre-existing societal patterns by automated systems is exemplified by so-called filter bubbles (Pariser, 2011) on digital platforms: The more one watches videos of kittens online, the more this type of content will be recommended by platform-based ML models, and the more it will be likely to be viewed by the user in turn. Sociologists have noted how similar loops do not amplify individual patterns only, but rather they "play a role in sustaining intersecting hierarchies of race, class, gender, and other modes of domination and axes of inequality" (Fourcade & Johns, 2020, p. 812).

Haugaard (2020) has recently revised Lukes's (1974) original theorization, arguing that three-dimensional power operates based on what Giddens (1984) calls "practical consciousness" and Bourdieu (1977) "habitus"—that is, a pre-reflexive, embodied, and generative knowledge derived from socialization. Encapsulating unequal social conditions in the form of cultural dispositions, habitus tacitly orients individual behavior, cognition, and life trajectories in directions that are likely to reproduce, rather than challenge, social and symbolic inequalities (Bourdieu & Wacquant, 1992). If we acknowledge that the socially structured cultural patterns embodied as class or gender habitus are now massively turned into digital traces fed to ML models whose predictions powerfully affect human habits and practices in turn, new sociological questions at the intersection of social and ML emerge (Airoldi, 2021; Fourcade & Johns, 2020). What about music consumption and its social stratification, in the age of streaming platforms and recommender systems? What are the structural outcomes of human-machine feedback loops in dating, policing, or language use? Are ML systems destined to inexorably reinforce a human-organized social order? Or could they transform it in unexpected, "alien" directions (Parisi, 2019)? These are only some out of many possible questions for a post-human sociological agenda.

D4: ML Power as Data-Driven Governmentality

Haugaard (2020) adds a fourth dimension to Lukes's three-dimensional account of *power over*. This final and more abstract dimension concerns the social construction of social subjects—that is, the process of "subjection" theorized by Foucault (1982) and Butler (1997). Here, power subjugates individuals by "making" them—as patients, prisoners, workers, students, foreigners, abnormal persons, consumers, and, now increasingly, users and data sources. As Foucault (1982) notes, this internalized, knowledge-based form of power "applies itself to immediate everyday life which categorizes the individual, marks him by his own individuality, attaches him to his own identity, imposes a law of truth on him" (p. 781).

Modern subjects are "made" of and through data. Statistics, demographic assessments, and forecasting techniques, along with the panoptic architectures of prisons, hospitals, and schools, were all indicated by Foucault as powerful mechanisms for governing populations (Cheney-Lippold, 2011; Foucault, 1982). As Arvidsson (2004) early noticed, since the advent of the Internet and rapid diffusion of online tracking devices, we have all become "subject to a virtually ever-present 'panoptic sort'" (p. 457). This "surveillant gaze" soon expanded, boosted by the spread of mobile devices and rampant platformization of social life (van Dijck et al., 2018). According to Zuboff (2019), the transformation of digital consumers' "data exhaust" into precious fuel for machine intelligence inaugurated the prediction-based power regime of "surveillance capitalism." From a different yet equally critical angle, Couldry and Mejias (2019) theorize the rise of "data colonialism" as a pervasive process of economic extraction that colonizes everyday life by transforming the subject into the object of constant data tracking and nudging.

Subjection is connected to the Foucauldian notions of discipline, biopolitics, and, especially, "governmentality"—that is, a totalizing power regime distinct from sovereignty that governs social relations and life, broadly speaking (Erlenbusch, 2013). As Introna (2016) explains, "The concept of governmentality focuses our attention on how practices, knowledge, and power become interconnected to enact particular governed subjects" (p. 28). Several contributions in the field of critical algorithm studies use the idea of governmentality to illuminate the epistemological and performative effects of prediction and automation over society and social subjects (Neyland, 2019). For instance, Introna (2016) makes a Foucauldian analysis of the use of plagiarism detection system Turnitin in academic writing, showing how it is "deployed as a technology of government to constitute domains of knowledge and how such knowledge regimes become internalized by subjects" (p. 20). With the normalization of the use of this software and internalization of its computational logics, students and academics are algorithmically "produced" to enact the subjectivity of the "self-governed original writer" (p. 35).

Papers investigating modes of algorithmic governance tend to portray algorithms as elements within networked sociomaterial entanglements, wholly redefining what we mean as "subject," "knowledge," "life," "culture," and more (Amoore & Piotukh, 2016; Ziewitz, 2016; Striphas, 2015). Only a few works explore how the operational autonomy that distinguishes ML from simpler calculative devices may specifically contribute to the enactment of governmentality. For instance, Cheney-Lippold (2011) argues that contingent and recursive data-driven predictions performatively transform the subjects' identity categories in the Foucaldian direction of "soft" biopower and biopolitics. ML models produce new and continuous forms of categorization, for instance by estimating one's degree of "maleness" through the dynamic analysis of behavioral data flows. These "algorithmic identities," invisibly associated to our "data doubles" (Couldry & Mejias, 2019), exert a "modular" control over subjects, creating not individuals but endlessly subdivide-able "dividuals" (Cheney-Lippold, 2011). Prey (2018) presents a similar interpretation of the platform-based governmentality experienced by music listeners, who are enacted as multiplicities and "individuated" through automated classifications and recommendations. Also, Mackenzie (2013) elaborates on how ML affects the subjectivities of programmers and software developers, in the context of a "regime of anticipation" obsessed with prediction. Overall, the contingent and recursive "thinking" of ML computational behavior (Burrell, 2016) adds a layer of complexity to societal mechanisms of subjection and governmentality, which deserve further and more systematic sociological investigation.

Conclusion

This chapter has aimed to discuss the critical literature on the power of AI and ML systems in light of sociological theories of power as *power to*—that is, the agentic dimension of power—and *power over*—that is, power as domination over individuals and society. First, I have argued that recent social science research understands ML agency in two main ways: as the situated enactment of ramified socio-technical systems (*networked view*) or as the result of the increasingly autonomous and culturally driven operations of learning machines (*cultural view*). Second, I have attempted to extend human-centric theorizations of the multidimensionality of domination to the inanimate realm of ML, organizing the existing literature on algorithms and AI along four ideal-typical dimensions of ML *power over*: opaque coercion (D1), computational authority (D2), structural conditioning (D3), and data-driven governmentality (D4).

Without pretending to be exhaustive, this short overview of the rapidly growing research on ML power in society aims to overcome counterproductive and yet long-lasting distinctions between “the social” and “the technical” in social theory. In fact, as Law (1991) argued more than three decades ago, “A simple distinction between the material (or the technological) on the one hand, and the social on the other, does not catch the subtlety of the way in which power (or agency) effects are generated” (p. 176).

ML is here to stay, and algorithmic governance will likely “play an ever-increasing role in the exercise of power, a means through which to automate the disciplining and controlling of societies and to increase the efficiency of capital accumulation” (Kitchin, 2017, p. 15). I am not alone in believing that, given the strength of its theoretical tradition, sociology is particularly well equipped to decode, measure, and interpret the many challenges that this shift toward computational authority is already presenting as well as the novel forms of resistance that will inevitably emerge.

References

Airoldi, M. (2021). *Machine habitus: Toward a sociology of algorithms*. Polity.

Allen, A. (1999). *The power of feminist theories: Domination, resistance, solidarity*. Routledge.

Amoore, L., & Piotukh, V. (2016). *Algorithmic life: Calculative devices in the age of big data*. Routledge.

Aneesh, A. (2009). Global labor: Algocratic modes of organization. *Sociological Theory*, *27*(4), 347–370.

Arvidsson, A. (2004). On the "pre-history of the panoptic sort": Mobility in market research. *Surveillance & Society*, *1*(4), 456–474.

Beer, D. (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society*, *11*(6), 985–1002.

Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, *20*(1), 1–13.

Beer, D. (2022). The problem of researching a recursive society: Algorithms, data coils and the looping of the social. *Big Data & Society*, *9*(2). https://doi.org/10.1177/20539517221104997

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim Code*. Polity.

Bierstedt, R. (1950). An Analysis of Social Power. *American Sociological Review*, *15*(6), 730.

Borch, C. (2022). Machine learning and social theory: Collective machine behavior in algorithmic trading. *European Journal of Social Theory*, *25*(4), 503–520.

Borch, C. (2023). Machine learning and postcolonial critique: Homologous challenges to sociological notions of human agency. *Sociology*, Article 00380385221146877.

Borch, C., & Hee Min, B. (2022). Toward a sociology of machine learning explainability: Human–machine interaction in deep neural network-based automated trading. *Big Data & Society*, *9*(2). https://doi.org/10.1177/20539517221111361

Boudon, R., & Bourricaud, F. (1989). *A critical dictionary of sociology*. University of Chicago Press.

Bourdieu, P. (1977). *Outline of a theory of practice*. Cambridge University Press.

Bourdieu, P., & Wacquant, L. (1992). *An invitation to reflexive sociology*. Polity.

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. MIT Press.

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, *14*(7), 1164–1180.

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of facebook algorithms. *Information, Communication & Society*, *20*(1), 30–44.

Bucher, T. (2018). *If . . . then: Algorithmic power and politics*. Oxford University Press.

Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1). https://doi.org/10.1177/2053951715622512

Burrell, J., & Fourcade, M. (2021). The society of algorithms. *Annual Review of Sociology*, *47*(1), 213–237.

Butler, J. (1997). *The psychic life of power: Theories in subjection*. Stanford University Press.

Campolo, A., & Crawford, K. (2020). Enchanted determinism: Power without responsibility in artificial intelligence. *Engaging Science, Technology, and Society*, *6*, 1–19.
Cardon, D. (2018). The power of algorithms. *Pouvoirs*, *164*(1), 63–73.
Cerulo, K. A. (2009). Nonhumans in social interaction. *Annual Review of Sociology*, *35*(1), 531–552.
Cheney-Lippold, J. (2011). A new algorithmic identity: Soft biopolitics and the modulation of control. *Theory, Culture & Society*, *28*(6), 164–181.
Cotter, K. (2019). Playing the visibility game: How digital influencers and algorithms negotiate influence on Instagram. *New Media & Society*, *21*(4), 895–913.
Couldry, N., & Mejias, U. A. (2019). Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media*, *20*(4), 336–349.
Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
Erlenbusch, V. (2013). The place of sovereignty: Mapping power with Agamben, Butler, and Foucault. *Critical Horizons*, *14*(1), 44–69.
Esposito, E. (2017). Artificial communication? The production of contingency by algorithms. *Zeitschrift für Soziologie*, *46*(4), 249–265.
Esposito, E. (2022). *Artificial communication.* MIT Press.
Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
Fjelland, R. (2020). Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*, *7*(1), 1–9.
Foucault, M. (1982). The subject and power. *Critical inquiry*, *8*(4), 777-795.
Fourcade, M., & Healy, K. (2013). Classification situations: Life-chances in the neoliberal era. *Accounting, Organizations and Society*, *38*(8), 559–572.
Fourcade, M., & Johns, F. (2020). Loops, ladders and links: The recursivity of social and machine learning. *Theory and Society*, *49*(5), 803–832.
Gandini, A., Gerosa, A., Gobbo, B., Keeling, S., Leonini, L., Mosca, L., . . . Splendore, S. (2022). *The algorithmic public opinion: A literature review*. https://doi.org/10.31235/osf.io/m6zn8
Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. University of California Press.
Gillespie, T. (2016). Algorithm. In B. Peters (Ed.), *Digital keywords* (pp. 18–30). Princeton University Press.
Gomez-Uribe, C. A., & Hunt, N. (2015). The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems*, *6*(4), 1–13.
Gran, A. B., Booth, P., & Bucher, T. (2021). To be or not to be algorithm aware: A question of a new digital divide? *Information, Communication & Society*, *24*(12), 1779–1796.
Haugaard, M. (2020). *The four dimensions of power*. Manchester University Press.
Haugaard, M. (2021). The four dimensions of power: Conflict and democracy. *Journal of Political Power*, *14*(1), 153–175.
Introna, L. D. (2016). Algorithms, governance, and governmentality: On governing academic writing. *Science, Technology & Human Values*, *41*(1), 17–49.

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals, 14*(1), 366–410.

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), 14–29.

Lash, S. (2007). Power after hegemony: Cultural studies in mutation? *Theory, Culture & Society*, *24*(3), 55–78.

Latour, B. (1991). Technology is society made durable. In J. Law (Ed.), *A sociology of monsters: Essays on power, technology, and domination* (pp. 103–131). Routledge.

Latour, B. (1992). Where are the missing masses? In W. Bijker & J. Law (Eds.), *Shaping technology/building society: Studies in sociotechnical change* (pp. 225–258). MIT Press.

Law, J. (Ed.). (1991). *A sociology of monsters: Essays on power, technology, and domination*. Routledge.

Lukes, S. (1974). *Power: A radical view*. Macmillan.

Lustig, C., & Nardi, B. (2015, January). Algorithmic authority: The case of bitcoin. *Proceedings of the 2015* 48th Hawaii International Conference on System Sciences (pp. 743–752).

Mackenzie, A. (2006). *Cutting code: Software and sociality*. Peter Lang.

Mackenzie, A. (2013). Programming subjects in the regime of anticipation: Software studies and subjectivity. *Subjectivity*, *6*, 391–405.

MacKenzie, D. (2019). How algorithms interact: Goffman's "interaction order" in automated trading. *Theory, Culture & Society*, *36*(2), 39–59.

Markham, A. (2021). The limits of the imaginary: Challenges to intervening in future speculations of memory, data, and algorithms. *New Media & Society*, *23*(2), 382–405.

Mühlhoff, R. (2020). Human-aided artificial intelligence: Or, how to run large computations in human brains? Toward a media sociology of machine learning. *New Media & Society*, *22*(10), 1868–1884.

Natale, S., & Ballatore, A. (2020). Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence*, *26*(1), 3–18.

Neyland, D. (2019). *The everyday life of an algorithm*. Springer International.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

Pansardi, P. (2012). Power to and power over: two distinct concepts of power? *Journal of Political Power*, *5*(1), 73–89.

Pardo-Guerra, J. P. (2010). Creating flows of interpersonal bits: The automation of the london stock exchange, 1955–90. *Economy and Society*, *39*(1), 84–109.

Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.

Parisi, L. (2019). The alien subject of AI. *Subjectivity*, *12*(1), 27–48.

Pasquale, F. (2015). *The black box society*. Harvard University Press.

Prey, R. (2018). Nothing personal: Algorithmic individuation on music streaming platforms. *Media, Culture & Society*, *40*(7), 1086–1100.

Racine, E., Boehlen, W., & Sample, M. (2019). Healthcare uses of artificial intelligence: Challenges and opportunities for growth. *Healthcare Management Forum*, *32*(5), 272–275.

Reed, I. A. (2020). *Power in modernity*. University of Chicago Press.

Rogers, R. (2013). *Digital methods*. MIT Press.

Ruckenstein, M. (2023). *The feel of algorithms*. University of California Press.

Schwarz, O. (2021). *Sociological theory for digital society*. Polity.

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, *4*(2).

Seaver, N. (2022). *Computing taste: Algorithms and the makers of music recommendation*. University of Chicago Press.

Striphas, T. (2015). Algorithmic culture. *European Journal of Cultural Studies*, *18*(4–5), 395–412.

Sundin, O., Haider, J., Andersson, C., Carlsson, H., & Kjellberg, S. (2017). The search-ification of everyday life and the mundane-ification of search. *Journal of Documentation*, *73*(2), 224–243.

Tubaro, P., Casilli, A. A., & Coville, M. (2020). The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society*, *7*(1).

Van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

Velkova, J., & Kaun, A. (2021). Algorithmic resistance: Media practices and the politics of repair. *Information, Communication & Society*, *24*(4), 523–540.

Weber, M. (1964). *The theory of social and economic organization*. Free Press.

Weber, M. (1978). *Economy and society*. University of California Press.

Weinberg, J. (2020, July 30). Philosophers on GPT-3 (updated with replies by GPT-3). Daily Nous. https://dailynous.com/2020/07/30/philosophers-gpt-3

Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society*, *20*(1), 118–136.

Yolgörmez, C. (2021). Machinic encounters: A relational approach to the sociology of AI. In J. Roberge & M. Castelle (Eds.), *The cultural life of machine learning: An incursion into critical AI studies* (pp. 143–166). Springer International.

Ziewitz, M. (2016). Governing algorithms: Myth, mess, and methods. *Science, Technology & Human Values*, *41*(1), 3–16.

Zuboff, S. (2019). *The age of surveillance capitalism*. Profile Books.