Recursive Power: AI Governmentality and Technofutures

Pre-print, final published version forthcoming

McKelvey, Fenwick (Concordia University)[1], Montréal, Canada and Roberge, Jonathan (INRS)[2] *

* Both authors contributed equally

1. Fenwick McKelvey, Concordia University, fenwick.mckelvey@concordia.ca

2. Jonathan Roberge, INRS, jonathan.roberge@inrs.ca

Abstract

We argue that AI is increasingly intractable from the study of governmentality. Yet, AI's governmentality is not singular and, in this chapter, we trace AI's *governmentalities* using two icons of AI governance: Elon Musk and Kevin Kelly. Through a comparison of the supporting discourses behind these two figures, we discuss how two related approaches to AI governmentality mix the symbolic and practical functions of AI. We address through the concepts of doctrines and dispositifs. Together we identify how the recursive loops in AI governmentality created a process for society and, yet, one always in-formation so as to depoliticize the very project of governmentality.

Introduction

Speaking before a crowd at MIT in 2014, the normally blusterous Elon Musk appeared cautious about the future of artificial intelligence. The then-nascent technology was, according to Musk, humanity's "biggest existential risk" and his investments in AI firms sought to control risks in its development. Musk warned that, "with artificial intelligence we are summoning the *demon*" [italics added] (quoted in Gibbs, 2014, np.). As Musk reached for the holy water, technology pundit Kevin Kelly, writing in Wired Magazine, announced that AI was ready to be "unleashed on the world" (2014, np.). Kelly's view of AI was far less demonic but rather something "on the horizon [which] looks more like Amazon Web Services—cheap, reliable, industrial-grade digital smartness running behind everything, and almost invisible except when it blinks off" (2014, np.). Kelly prefigured his view of AI's arrival with the example of IBM's ill-fated Watson – a version of AI referencing the famous detective's assistant. It seemed that according to Kelly, if humanity were Sherlock Holmes, it merely needed its robotic Dr. Watson (Hale, 2011).

We begin with these two men – characters critical to the ongoing coverage of AI – to demarcate our chapter's interest in AI governmentality. Musk and Kelly's differing attitudes towards AI suggest that how AI's power is imagined and deployed is split into distinct camps. These competing camps of technofuturism not only regulate societal imaginations of AI's potential (Hong, 2021), but also co-constitute AI as a novel logic *of* and *for* society. The split between them can be expressed as such:

1. *Musk's demon:* Musk's description of AI as an "existential risk" draws from a loose Internet philosophy of longtermism that argues today's AI is key to the development of superintelligence, an event that determines humanity's future. Superintelligence draws on

Silicon Valley's investment in singularity but adds a certain degree of philosophical legitimacy. As well as a *symbolic* framing device for debates over AI's power, longtermism is a *practical* logic in the terms of venture capital funding as in the case of Musk funding OpenAI. Longtermism defines investment and policy work concerned with the governance of superintelligence and humanity's far-off future. Musk frequently discusses AI along with his plans for the self-driving in Teslas and Mars colonization. His less optimistic peers have are building clocks that run after the end of the world where others are doomsday prepping (Karpf, 2020; Roberts & Hogan, 2019).

2. *Kelly's Watson* exhibits a similar entanglement between the practical and the symbolic though in a more immediate way. IBM scaled back the ambitions of its Jeopardy-winning computer program to configure Watson's system as a less-ambitious model of AI governmentality, one that prioritizes business analytics and "solutions" (Lohr, 2021). This pragmatism moved Watson into the services industry, including banking, insurance, healthcare and the like. Compared to Musk's demon, Watson represents a "de-wilding" of AI as making the technology tame and useful now (Coleman, 2021). Part of the justification for AI being less scary is a constant de-skilling and devolution of human intelligence. Processes are automatable precisely because humans are as unreliable as machines. Watson then draws on natural language processing (NLP) as well as behavioural economics and psychology concerned with heuristics, biases and human fallibility in order to offer a human-machine interaction that can be administrated and systemized.

Is Watson then a prime example of an approach to weak or narrow AI (ANI), in contrast with a strong Artificial General intelligence (AGI)? In what follows, we explore these two approaches

that are neither identical nor incommensurable. We argue that they are not two opposite poles, but rather telling moments and loci within a larger continuum of AI governmentality.

Musk's demon and Kelly's Watson allow us to better understand the ways competition manifests in the funding, regulatory debates, and technological deployment of artificial intelligence. We distinguish between these approaches in a discussion of both their *doctrines*, *dispositifs* and the interactions between — or what we call the *recursivity* of power. We conclude by offering some suggestions for further reflections on what it means for a political technology like AI to ignore its own intrinsic politics, something that more critical and reflexive views can attempt to do. But first, we begin by discussing our approach to AI governmentality, following Foucault, as regime and "conduct of conduct" that simultaneously exists as a strategic *dispositif* and an always-dynamic modus operandi.

Recursive Power: on Governmentality, Cybernetics and AI

There is now an established tradition in the social sciences and humanities —in STS in particular— that addresses the inherently political nature of technology. Immediately, we recall Langdon Winner's (1986) famous question "Do artifacts have politics?" as well as Kate Crawford's statement that "AI is politics by other means" (2021, p. 19). Michel Foucault's (1991) writings on the "technologies of power" are central to the tradition behind the works that address a concept now known as "governmentality". Governmentality "refers to a set of institutions, procedures, analyses and reflections, calculations and tactics which permits the exercise of that form of […] power which has the population as its principal target, political economy as its major form of knowledge, and dispositive of security as its essential technical instruments" (Foucault, 2007, p.

108). Of course, many in-depth accounts of the concept exist (Bratich et al., 2003; Rose, 1999; Walters, 2012), but for our discussion of AI, we wish to summarise three key aspects.

First and foremost, Foucault insists that power is a mundane resource. He talks at length about *dispositif* as a strategic rationality and power relation leading to imprisonment and about the architecture of power in the classroom and asylum. Powerful institutions that compared to Mars colonization seem mundane. In all cases, power is defined as productive and relational shaping the conduct of individuals and groups. And while it is difficult to imagine what Foucault himself would have said of today's digital world, it's nonetheless possible to apply his theorization of power to the logic of what has been called 'extensive nudging' (Yeung, 2016).[1] Rouvroy and Berns (2013) have developed such an argument in favor of a renewed understanding of 'algorithmic governmentality' (though we disagree with some aspects of their analysis, as we shall see forthwith). Governmentality encourages attention to practical matters; however, the concept is not only concerned with these practical matters.

A solely strategic reading of Foucault's legacy, however, forecloses the possibility of a world still influenced by pseudo religious "pastoral" logics. Here, we stress a second characteristic of governmentality: one does not have to choose between the practical and the symbolic. Governmentality is both the immanent nature of the here-and-now and the representation of where such a present might be situated within a broader cosmology. As Cooper rightly observes, "pastoralism continues to operate in the algorithmic register" (2020: 29). Foucault himself discusses the durability of mythological views that are both teleological and eschatological —the ideas of innovation and progress among them. We use the concept of the doctrine to refer to this

[1] Or at least the impression of automation (Ananny, 2020; Gorwa et al., 2020; Myers West, 2018; Roberts, 2019). Here we acknowledge a burgeoning field of algorithmic governance (for a review, see Katzenbach & Ulbricht, 2019). Therein, we notice a growing dialogue between AI and algorithms relationship to dispositifs and Foucault's original concept.

second facet of governmentality. Pastoral power represents a battle for souls that does not need a God and/or a Church to exist: what it requires is a shepherd who sees themself as a force for the better good. As we shall see in the next section, individuals such as Musk and movements such as longtermism see their place as shepherds of humanity and as such, they are prime examples of how *dispositifs* are inseparable from doctrines.

The distributed and circulatory aspects of governmentality constitute its third key characteristic. Indeed, we would go as far as to say that governmentality is a cybernetic reality and that Foucault's theories are not incompatible with the ones of people such as mathematician Wiener (1948), his political scientist advocate Karl Deutsch (1966) and others (Fourcade & Gordon, 2020; Rappin, 2018; Roberge, Senneville, et al., 2020). Everything that is governmentality or governance deals with steering in a literal and figurative sense: its etymology comes from the Greek *Kubernêtês* which means ship's pilot whereas *kubernêtiké* signifies the art of both navigating and governing . In cybernetics, the terms at play are 'control' and 'communication' which are not too distant from *dispositif* and doctrine. Musk wields his doctrine of AI's future as a rhetorical device that shapes and controls the technology's current deployment. What is fundamental is the movement, namely the feedback mechanism that loops the elements together. While we talk about the emergence of a recursive form of power which comes to define AI governmentality, we not only mimic the back-propagation central to machine learning, but also the constant back and forth between control and communication, doctrine and dispositive[2]. IBM's Watson, too, is bound by

[2] Our reading of governmentality attempts to synthesize the mythic and regulatory inquiries into artificial intelligence. On one hand, there is a persistent interest in artificial intelligence as a digital sublime that enchants government and industry. The mythic critique of AI, however, parallels a diverse interest in AI as a means of regulation. AI governance is something of a trading zone between scholarship drawing from Deleuze's later work on control societies (Cheney-Lippold, 2017; Deseriis, 2011), a second literature following Scott Lash considering algorithms as post-hegemonic (Beer, 2016; Lash, 2007), a sociological-informed interest in algorithmic governance or algocracy (Aneesh, 2009; Yeung, 2018), and a final intersection between critical race and surveillance studies exemplified by Ruha Benjamin's (2019) formulation of a New Jim Code. What we wish to drawn out, as debated in this literature, is the interpretation of governmentality as if it were a series of successions that forecloses the

such a logic where, for instance, its marketing attracts new clients who feeds the model with new data for it to perform better and thus allowing for more sales. AI governmentality is recursive because it needs to fight degradation and entropy with ever more synergies, and because it needs to adapt in order to maintain and develop. Indeed, to stir is to adapt.

Musk's Demon, Superintelligence and the Rise of Longtermism

The figure of the demon is central to the doctrine and *dispositif* of our first approach to AI governmentality. Elon Musk, in his remarks above, suggested that developing AI was akin to summoning a demon. He then continued, "In all those stories where there's the guy with the pentagram and the holy water, it's like – yeah, he's sure he can control the demon. Doesn't work out" (quoted in Gibbs, 2014, np.). Far from an offhand comment, Musk's remarks invoke the popular trope about the demon as out of control that lays the groundwork for this approach to AI governmentality (Braman, 2002; Canales, 2020; McKelvey, 2018; Roderick, 2007). Indeed, his comments echo the refrain heard in Silicon Valley that summons the demon to articulate artificial intelligence as super intelligence. Musk's investments have then been a self-aware act of reaching for the metaphorical holy water.

Musk may just as likely have drawn inspiration about the threat of a super intelligent demon from Daniel Suarez's techno-thriller novels *Daemon* and its sequel *Freedom*. In them, a deranged millionaire creates a system of programs, collectively called a daemon, to continue developing after his death. The daemon's distributed intelligence disrupts society as it builds autonomous vehicles and launches its own cryptocurrency that runs on a global darknet. First written in 2006,

---

possibility of a world still moved by religious-like "pastoral" logics. Technical innovations do not negate the symbolic or pastoral function of politics; instead, Musk and Kelly have a pastoral function in their articles that seem to shepherd humanity all the while affecting the regulation of and the regulation by AI.

Daemon inspired programmers in Silicon Valley. The foreword to the 2017 edition called the book "the secret handshake" of "technoliterati". Musk himself commented on Twitter that Saurez's *Daemon* was a "great read" just a few months before his holy water comments (Elon Musk [@elonmusk], 2014).

Musk mentioned *Daemon* in a reply to a Twitter thread after endorsing another book, Superintelligence by Nick Bostrom (Bostrom, 2014). Bostrom, a philosopher and Founding Director of the Oxford Future of Humanity Institute, defines the term in an early article as an

> intellect that is *much smarter* than the best human brains in practically every field, including scientific creativity, general wisdom and social skills. This definition *leaves open* how the superintelligence is implemented: it could be a digital computer, an ensemble of networked computers, cultured cortical tissue or what have you. It also *leaves open* whether the superintelligence is conscious and has subjective experiences. (Bostrom, 1998; emphasis added)

Bostrom sees superintelligence as not just an emulation of the human mind, but as an event whereby a system achieves intellectual capacity beyond the human. Superintelligence, both the book and the concept, is central to understanding the doctrinal function of AI as an eschaton event for humanity that warrants priority address by global leaders above other existential threats, even climate change.

Collectively, Bostrom's work contributes to a philosophy of longtermism. Émile P. Torres, probably the most astute observer of the doctrine, summarizes it as the "claim that if humanity can survive the next few centuries and successfully colonize outer space, the number of people who could exist in the future is absolutely enormous" (2022, np.). Torres cites Bostrom and Musk as its

main adherents, but other longtermist projects are cropping up across Silicon Valley such as the Jeff Bezos-funded Clock of the Long Now (Karpf, 2020). Superintelligence is a long-term matter, an existential risk to humankind.

These leaders seem to have embraced their public role as steering humanity to superintelligence safely. A collective superintelligence like the one in Saurez's *Daemon* is only one possible way to this end. Other paths include a eugenics-tinged focus on embryo selection and genetic engineering, human augmentation, and other conventional forms of artificial intelligence. Regardless the outcome, to reach superintelligence is to flirt with singularity and the technological fast-paced autonomous generation it entails, a concept discussed by science fiction writer Vernor Vinge and futurist Raymond Kurzweil (Bostrom, 2005). Yet, superintelligence is also a theme in contemporary thought with variations that resonate with figures in the Dark Enlightenment and other neo-reactionary movements that see advanced intelligences as limits to the present social order (Haider, 2017; Smith & Burrows, 2021).

The long-term view of superintelligence invites a mixture of ambivalence and imperatives as Bostrom sees a need to act now, but not about immediate, practical concerns. The problem of AI as superintelligence is never mundane, embodied, or even environmental; these are all short-term matters, matters of no consequence in a self-declared rational, if laissez-faire, policy optimization. The fact of the matter is that longtermism calls for more longtermism and that AI call for more AI. This is what we mean by the recursive nature of their power. Through it, longtermism and AI become the causes and the consequences of one another. The outcome is a striking depoliticization of AI made possible through the mixture of elite gatekeeping and a poorly defined problematization of AI. Whether in computer science communities, loosely constituted, elite online debates, or in Bostrom's book *Superintelligence*, the image of "revolution" that is

projected is still a technological one —an engine—rather than the image of a social movement or a civil sphere. This inescapable mythology and iconicity of AI even found its way into the acknowledgement of Bolstrom's book, which reads like a who's who list of AI developers including Yoshua Benigo and Geoffrey Hinton.

Strange as it may seem, the doctrine of longtermism offers a distinct policy agenda for AI governance. Indeed, one could argue that Bostrom's early writing on superintelligence helped legitimate the field's self-fulfilling prophecy, leading to ever-more investment in the technology and a form of autoregulation that resists being constrained by any state government (Wagner, 2018; Mittelstadt, 2019). Musk's demon is part of such a trend and is certainly emblematic of its many inconsistencies. This thread of superintelligence locates the problem in an unpredictable "take-off" period in which AI exceeds thresholds of human and civilization intelligence. The problem, according to Bostrom, is a diabolical one: "The first superintelligence may shape the future of earth originating life, could easily have non-anthropomorphic final goals, and would likely have instrumental reasons to pursue open ended resource acquisition" (Bostrom, 2014, p. 317). (Bostrom's scenario of an AI taking over the world seems lifted from the mind of Saurez. Daemon includes a lot of killer robots). This problematization narrowly frames AI governance as a mostly, if not completely, speculative logic of regulation. The matter of AI governance is not a present concern, but a tautology involving, in Bostrom's words "motivations" and "detonations" that respectively steer superintelligence to anthropomorphic goals.

Politically, addressing an evil demon is not a problem of actual welfare here and now. Longtermism is concerned with future civilizations. Torres notes that longtermism has become part of the Effective Altruism movement—a program of philanthropy active in some technology firms (Matthews, 2015; Torres, 2021). As Torres explains,

> imagine a situation in which you could either lift 1 billion present people out of extreme poverty *or* benefit 0.00000000001 percent of the 10^{23} biological humans who Bostrom calculates could exist if we were to colonize our cosmic neighborhood, the Virgo Supercluster. Which option should you pick? For longtermists, the answer is obvious: you should pick the latter. Why? Well, just crunch the numbers: 0.00000000001 percent of 10^{23} people is 10 billion people, which is *ten times greater* than 1 billion people.(É. P. Torres, 2021)

These calculations drive effective altruism's funding, Out of $416M spent by associated fundings in 2019, $40M (10%) went to "Potential risks from AI" above "Other near-term work (near-term climate change, mental health)" that received $2M (0%) (Todd, 2021). The Effective Altruism Foundation, similarly, has four funds ranked by risk. The "Long-Term Future Fund" has the highest risk profile, over the "Global Health and Development Fund". Longtermism, in short, drives a major part of a growing philanthropic movement.

To be sure, superintelligence is an immaterial approach to intelligence and how it came to define humanity as a historical and social construct. Within such a paradigm, it becomes possible to see societies as computational, as a wired brain composed of neurons that are more trigger and data than flesh and soul. That is what AI represents: namely something other than human, even alien. By continuing with this line of argument, logically, one can see the whole world is a simulation-- another philosophical argument advocated by Bostrom. If AI does not need a body, the philosopher may remove the planet, too. Under this view, the Earth is only a starting place for humanity and one that ultimately does not constitute an existential risk. While acknowledging climate change as a new kind of threat, Bostrom writes,

> Even if humanity were to spend many millennia on such an oscillating trajectory, one might expect that eventually this phase would end, resulting in either the permanent destruction of humankind, or the rise of a stable sustainable global civilization, or the transformation of the human condition into a new 'posthuman' condition. (Bostrom, 2009)

Superintelligence is then defined as a probable fate encounter by present or future civilization, an eschatological time frame longer than a mere matter of climate.

Kelly's Watson and the Looping of *Dispositif* and Doctrine

The maverick co-founder of *Wired* magazine, Kevin Kelly could be considered a prosaic guru of digital technologies. His innate pragmatism comes with a steady dose of enthusiasm that fits in particularly well in Silicon Valley; Kelly's motto reads "over the long term, the future is decided by optimists". In his writing, he has used the now-infamous example of IBM's Watson to introduce his own vision of AI, one that is a decidedly-less ambitious expression of the technology and ensuing regime of governmentality than the longtermists. "Today's Watson," he notes, "[…] no longer exists solely within a wall of cabinets but is spread across a cloud of open-standard servers that run several hundred 'instances' of the AI at once" (Kelly, 2014). Watson, as Kelly explains, exemplifies an AI that infiltrates every aspect of society because it is on-demand, distributed, flexible and adaptative. That is the version of Kelly's Watson that ultimately came to pass, a diminished yet applied and operating AI —or, following our argument, one that is cybernetic from the ground up.

Watson's mundane status today is at odds with its beginnings. Watson was ahead of the curve in 2011. Debuting to America as a contestant on Jeopardy!, the supercomputer surpassed all

expectations and went on to become the game's new champion. It was a media-pop-culture stunt, the kind of which Deep blue and later AlphaGo built upon (see Binder, 2021). Watson's success depended not only on data —troves and troves of data—but on a new way to organize and make sense of it: namely, Natural Language Processing, or NLP. NLP communicates in a 'smart' way, for instance being able to 'understand' an answer in Jeopardy and (re)translating it into a query.

NLP is one of the two major branches of machine learning and neural nets that launched today's AI Spring. Just a year after Watson bested Jeopardy!, Geoffrey Hinton's team revolutionized the other branch with its win at the ImageNet competition (Cardon et al., 2018), setting off another round of investment in AI —a round that IBM seemed well-poised to capitalize on with Watson. Not so; in reality, IBM had little to sell. Adapting a symbolic and theoretical machine to applied issues proved to be too difficult and the rewards too limited. Unlike its competitors operating in a more speculative mode, IBM pivoted to a "revised A.I. strategy —a pared-down, less world-changing ambition" (Lohr, 2021).

Watson could be less ambitious because it relied on the feedback loop between a doctrine less concerned with saving humanity and a pragmatic skepticism of human intelligence. Kelly's Watson was not superintelligent, but neither were the humans it sought to replace. 'Smartness' is diminished in a definition of artificial intelligence that draws on decades of variations of Cold War rationality that sought to blur the distinctions between human and machine intelligence, but to lower the overall threshold. The human then becomes programmable, optimizable, and interchangeable with fraught, biased, and unreliable forms of AI (Mirowski, 2002). As Erikson et al. nicely summarize, "emphasizing the divergence between actual human reasoning and standards of formal rationality […] implicitly reinforced the normative authority of the latter" (2013: 24). AI does not need to be more intelligent, just smarter, better structured, and more efficient than humans.

Kelly's Watson offers a version of AI governmentality with its own circular-logic and self-referential pastoral doctrine. Watson works to formally operationalize and scale up solutions framed as optimizable (McKelvey & Neves, 2021). Kelly imagines such work as taking the form of algorithms and models that can be indefinitely replicated if tweaked just enough. IBM's effort to colonize the data- and money-rich environment of healthcare is a case in point. Watson was supposed to be able to do it all: macro-calculations in the form of genomics and image diagnostics as well as adapted-precision medicine and human care via personal assistant and chatbot for patients. In such an environment, the patient's ability to emotionally connect to someone/something or to have their needs and mood be "understood" is priceless, even if made possible by multiple forms of deception, nudging and monitoring interventions. Lately, 'Watson Work' launched, combining parts of Watson Health with the optimization of the workplace in order to "help business navigate […] with the ongoing COVID-19 crisis as effectively as possible." As the press release reads, "applying AI […] is especially useful in this context, where there are so many different sources of information, and every aspect of the situation is in flux" (Quoted in Mashable News, 2020). In fact, the very definition governmentality, too, is in flux here: what counts as power deploys recursively on all the different fronts of dispositive and doctrine, control and communication, knowledge, and action as well as the capacities to influence the conduct of both masses and individuals.

Kelly's Watson operates in a very immediate temporality, both in terms of daily, mundane adoption, but also within the broader yet still short-term contexts of its commercial roll out. The scope of time is one without much deep consideration of AI's existential risks, the kind of which is central to longtermism. There is a fundamental tension and ambiguity here between the two perspectives, a conflict that nevertheless might signal a convoluted form of dialogue. Even for

Kelly, time appears to be moving in a spiral, one in which the flexible inclusion of short-term elements permits them finding their way into a loose sense of the long-term. Time, in other words, is as dynamic and unified as power itself. It is all about finding a sweet spot, a moment in which the management of such temporal flows seem to hold —the possibilities of start-ups to bank on their innovations and immediate utility for customers, or, more likely in this landscape, on their acquisition by bigger and more established companies. Today's AI technologies are being deployed in cycles that have the characteristics of being both flexible and opaque. Watson and others evolve and morph precisely because they are black boxes (Bucher, 2018; Pasquale, 2015; Roberge, Morin, et al., 2020). And this never ceases to represent a challenge for outside oversight and political regulation. As Cuéllar (2017) notes, what we are witnessing nowadays is an enhanced and faster process of "cyberdelegation" where more traditional means and meanings of legitimacy and accountability are being refurbished and pushed away. AI's deployment, the management of its short-term/long term tension and everything that deals with its inherent autopoietic nature announces an "escape for regulation" (Wagner, 2018).

Ethics is a case in point. Not surprisingly, ethical and responsible AI became the two central occupations of Kelly's Watson given its applied governmentality. IBM claims, for example, that it sells, "world-changing AI, built responsibly". "Mitigating bias" is one of six IBM positions on AI, a position that begins by acknowledging that "There's no question that human biases could influence algorithms and result in discriminatory outcomes" (Hobson & Dortch, 2021, np.) Not unlike the discussions of Cold War rationality above, the circular effect here is that as AI is measured for biases, these measures find their way to testing humans who come up lacking too. AI is measured against a diminishing attitude toward human intelligence.

What comes lacking, too —and maybe more importantly— is a proper sense of what counts as politics. The current profusion of optimistic, voluntary and often naïve discourses and ethical declarations have become something of a trope that is a pale version of politics of goodwill, a so-called regulation without coercion in general, and State regulation in particular (Jobin et al., 2019; Mittelstadt et al., 2016; Stark et al., 2021). To reframe this in Foucault's language: people and organizations the likes of Kelly and IBM have consecrated the absence of a State *dispositif* into a doctrine. The stakes become so low that State regulation is unnecessary so that another corporate political program can then swoop in, one that blurs the distinction between the public and the private all the while it establishes a new sense of legitimacy and accountability. This latest form of AI governmentality is in many ways a continuation of the present neoliberal or slightly post-neoliberal moment that uses AI as a part of an overall logic of societal optimization and techno-solutionism.

Conclusion

Our main argument in this piece is that far from contradictory, Musk's demon and Kelly's Watson form the conditions of possibility for one another. They highlight and indeed reinforce each other (though part of a larger metastability of AI governmentality we are tracing). There is no simple antagonism between of weak-narrow versus strong-general AI — the ANI vs. AGI trope. What we find is much more complex. While dealing with superintelligence and the prospect of humanity itself, the demon narrative is not only representational and mythical, but performative *here* and *now*. While offering the kind of "cheap, reliable, industrial-grade digital smartness" Watson does for business, it too deploys views and understandings of how the world works that makes it

political. In the end, Musk's demon and Kelly's Watson are prime examples of how doctrines and dispositifs co-constitute today's AI governmentality.

A key, yet still underestimated element of AI capability to steer society involves the refinement of feedback loops and the advent of an enhanced form of recursive power. To stay the course for technologies such as artificial intelligence is to rapidly move on and constantly adapt to the different fields. Lags and dysfunctions are to be overcome by being recycled; "bizarre" outputs refurbished as inputs and so on and so forth. This is what gives rise to its often free-floating, self-referential and self-perpetuating logic. AI governmentality can indeed be defined as being caught in its own spiral. Here we want to argue alongside people such as Louise Amoore when she notes that "the advent of deep learning is *generative* of new norms and thresholds of what 'good', normal', and stable [political] orders look like" [emphasis added] (Amoore, 2022, p. 2). By doing so, we also want to push against what is every so often the static, iron-cage-like thesis of the literature on algorithmic governmentality. If control and *dispositif* are fundamental, as this literature points out, so too are the issues dealing with communication, doctrines. What is more, when taken together, these allow us to better understand how and why today's AI governmentality repeatedly goes unchallenged. As seen above, it is rare that States go against such technological deployment. In investing in research, trying to implement AI in their management and in promoting ethical endeavors, States have less shaping power over AI than they are shaped by it. Marion Fourcade and Jeff Gordon are right when they observe that what we are witnessing is a "deeper transformation in statecraft itself" (2020 :80). The same can be said about intellectual discourses: philosophers, pundits and scholars such as Bolstrom have done little to challenge and question the type of power that AI is gaining. Reflexivity and criticism still appear to be sparse

resources even if, for both today and tomorrow, they are central in the capability to connect with a more deliberative civil society and ultimately defend society itself.

References

Amoore, L. (2022). Machine learning political orders. *Review of International Studies*, 1–17. https://doi.org/10.1017/S0260210522000031

Ananny, M. (2020). Making Up Political People: How Social Media Create the Ideals, Definitions, and Probabilities of Political Speech. *Georgetown Law Technology Review*, *4*(2), 352–366.

Aneesh, A. (2009). Global Labor: Algocratic Modes of Organization. *Sociological Theory*, *27*(4), 347–370. https://doi.org/10.1111/j.1467-9558.2009.01352.x

Beer, D. (2016). The social power of algorithms. *Information, Communication & Society*, 1–13. https://doi.org/10.1080/1369118X.2016.1216147

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

Binder, W. (2021). AlphaGo's Deep Play: Technological Breakthrough as Social Drama. In J. Roberge & M. Castelle (Eds.), *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies* (pp. 167–195). Springer International Publishing. https://doi.org/10.1007/978-3-030-56286-1_6

Bostrom, N. (1998). How Long Before Superintelligence? *International Journal of Futures Studies*, *2*.

Bostrom, N. (2005). *A History of Transhumanist Thought*. 25.

Bostrom, N. (2009). The future of humanity. In *New waves in philosophy of technology* (pp. 186–215). Springer.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Braman, S. (2002). Posthuman Law: Information Policy and the Machinic World. *First Monday*, *7*(12). http://firstmonday.org/ojs/index.php/fm/article/view/1011

Bratich, J. Z., Packer, J., & McCarthy, C. (Eds.). (2003). *Foucault, cultural studies, and governmentality*. State University of New York Press.

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

Canales, J. (2020). *Bedeviled*. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691175324/bedeviled

Cardon, D., Cointet, J.-P., & Mazières, A. (2018). La revanche des neurones. *Reseaux*, *n° 211*(5), 173–220.

Cheney-Lippold, J. (2017). *We are data: Algorithms and the making of our digital selves*. New York University Press.

Coleman, B. (2021). Technology of the Surround. *Catalyst: Feminism, Theory, Technoscience*, *7*(2), Article 2. https://doi.org/10.28968/cftt.v7i2.35973

Cooper, R. (2020). Pastoral Power and Algorithmic Governmentality. *Theory, Culture & Society*, *37*(1), 29–52. https://doi.org/10.1177/0263276419860576

Crawford, K. (2021). *Atlas of AI*. Yale University Press. https://yalebooks.yale.edu/book/9780300209570/atlas-ai

Cuéllar, M.-F. (2017). Cyberdelegation and the Administrative State. In N. R. Parrillo (Ed.), *Administrative Law from the Inside Out: Essays on Themes in the Work of Jerry L. Mashaw* (pp. 134–160). Cambridge University Press. https://doi.org/10.1017/9781316671641.006

Deseriis, M. (2011). The General, the Watchman, and the Engineer of Control. *Journal of Communication Inquiry*, *35*(4), 387–394. https://doi.org/10.1177/0196859911415677

Deutsch, K. W. (1966). *The Nerves of Government*. Free Press.

Elon Musk [@elonmusk]. (2014, August 3). *@drwave @itsDanielSuarez Yeah, Daemon is a great read* [Tweet]. Twitter. https://twitter.com/elonmusk/status/495771005634482176

Erickson, P., Klein, J. L., Daston, L., Lemov, R., Sturm, T., & Gordin, M. D. (2013). *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality*. University of Chicago Press.

Foucault, M. (1991). *Governmentality* (G. Burchell, C. Gordon, & P. M. Miller, Eds.; pp. 87–104). The Univesity of Chicago Press.

Foucault, M. (2007). *Security, Territory, Population: Lectures at the College de France, 1977-78* (M. Senellart, Ed.; G. Burchell, Trans.). Palgrave Macmillan.

Fourcade, M., & Gordon, J. (2020). Learning Like a State: Statecraft in the Digital Age. *Journal of Law and Political Economy*, *1*(1). https://escholarship.org/uc/item/3k16c24g

Gibbs, S. (2014, October 27). Elon Musk: Artificial intelligence is our biggest existential threat. *The Guardian*. https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, *7*(1), 205395171989794. https://doi.org/10.1177/2053951719897945

Haider, S. (2017, March 28). *The Darkness at the End of the Tunnel: Artificial Intelligence and Neoreaction*. Viewpoint Magazine. https://viewpointmag.com/2017/03/28/the-darkness-at-the-end-of-the-tunnel-artificial-intelligence-and-neoreaction/

Hale, M. (2011, February 8). Actors and Their Roles for $300, HAL? HAL! *The New York Times*. https://www.nytimes.com/2011/02/09/arts/television/09nova.html

Hobson, S., & Dortch, A. (2021, May 26). *Mitigating Bias in Artificial Intelligence*. IBM Policy Lab. https://www.ibm.com/policy/mitigating-ai-bias/

Hong, S. (2021). Technofutures in Stasis: Smart Machines, Ubiquitous Computing, and the Future That Keeps Coming Back. *International Journal of Communication*, *15*(0), 21.

Hookway, B. (1999). *Pandemonium: The Rise of Predatory Locales in the Postwar World*. Princeton Architectural Press.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Karpf, D. (2020, January 29). The 10,000-Year Clock Is a Waste of Time. *Wired*. https://www.wired.com/story/the-10000-year-clock-is-a-waste-of-time/

Katzenbach, C., & Ulbricht, L. (2019). Algorithmic governance. *Internet Policy Review*, *8*(4). https://policyreview.info/concepts/algorithmic-governance

Kelly, K. (2014, October 27). The Three Breakthroughs That Have Finally Unleashed AI on the World. *Wired*. https://www.wired.com/2014/10/future-of-artificial-intelligence/

Lash, S. (2007). Power after Hegemony: Cultural Studies in Mutation? *Theory, Culture & Society*, *24*(3), 55–78.

Lohr, S. (2021, July 16). What Ever Happened to IBM's Watson? *The New York Times*. https://www.nytimes.com/2021/07/16/technology/what-happened-ibm-watson.html

Mashable News. (2020, June 19). *IBM'S Newly Launched 'Watson Works' Uses AI To Help Firms Manage New Work Challenges*. Mashable India.

https://in.mashable.com/tech/14925/ibms-newly-launched-watson-works-uses-ai-to-help-firms-manage-new-work-challenges

Matthews, D. (2015, August 10). *I spent a weekend at Google talking with nerds about charity. I came away ... worried.* Vox. https://www.vox.com/2015/8/10/9124145/effective-altruism-global-ai

McKelvey, F. (2018). *Internet daemons: Digital communications possessed*. University of Minnesota Press.

McKelvey, F., & Neves, J. (2021). Introduction: Optimization and its discontents. *Review of Communication*, *21*(2), 95–112. https://doi.org/10.1080/15358593.2021.1936143

Mirowski, P. (2002). *Machine dreams: Economics becomes a cyborg science*. Cambridge University Press.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2). https://doi.org/10.1177/2053951716679679

Myers West, S. (2018). Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society*, *20*(11), 4366–4383. https://doi.org/10.1177/1461444818773059

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Rappin, B. (2018). Algorithme, management, crise: Le triptyque cybernétique du gouvernement de l'exception permanente. *Quaderni. Communication, technologies, pouvoir*, *96*, 103–114. https://doi.org/10.4000/quaderni.1182

Roberge, J., Morin, K., Senneville, M., & Sudmann, A. (2020). Deep Learning's Governmentality: The Other Black Box. In *The Democratization of Artificial Intelligence* (pp. 123–142). transcript Verlag. https://doi.org/10.1515/9783839447192-008

Roberge, J., Senneville, M., & Morin, K. (2020). How to translate artificial intelligence? Myths and justifications in public discourse. *Big Data & Society*, *7*(1), 205395172091996. https://doi.org/10.1177/2053951720919968

Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press.

Roberts, S. T., & Hogan, M. (2019). Left Behind: Futurist Fetishists, Prepping and the Abandonment of Earth. *B2o: An Online Journal*, *4*(2). https://escholarship.org/uc/item/8sr8n99w

Roderick, I. (2007). (Out of) Control Demons: Software Agents, Complexity Theory and the Revolution in Military Affairs. *Theory & Event*, *10*(2).

Rose, N. S. (1999). *Powers of Freedom: Reframing Political Thought*. Cambridge University Press.

Smith, H., & Burrows, R. (2021). Software, Sovereignty and the Post-Neoliberal Politics of Exit. *Theory, Culture & Society*, *38*(6), 143–166. https://doi.org/10.1177/0263276421999439

Stark, L., Greene, D., & Hoffmann, A. L. (2021). Critical Perspectives on Governance Mechanisms for AI/ML Systems. In J. Roberge & M. Castelle (Eds.), *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies* (pp. 257–280). Springer International Publishing. https://doi.org/10.1007/978-3-030-56286-1_9

Todd, B. (2021, August 9). *How are resources in effective altruism allocated across issues?* 80,000 Hours. https://80000hours.org/2021/08/effective-altruism-allocation-resources-cause-areas/

Torres, É. P. (2021, July 28). The Dangerous Ideas of "Longtermism" and "Existential Risk." *Current Affairs*. https://www.currentaffairs.org/2021/07/the-dangerous-ideas-of-longtermism-and-existential-risk

Torres, P. (2022, April 30). *Elon Musk, Twitter and the future: His long-term vision is even weirder than you think*. Salon. https://www.salon.com/2022/04/30/elon-musk-twitter-and-the-future-his-long-term-vision-is-even-weirder-than-you-think/

Wagner, B. (2018). Ethics As An Escape From Regulation.: From "Ethics-washing" To Ethics-shopping? In E. Bayamlioğlu, I. Baraliuc, L. Janssens, & M. Hildebrandt (Eds.), *Being Profiled* (pp. 84–89). Amsterdam University Press; JSTOR. https://doi.org/10.2307/j.ctvhrd092.18

Walters, W. (2012). *Governmentality: Critical encounters*. Routledge.

Wiener, N. (1948). *Cybernetics or, Control and Communication in the Animal and the Machine*. J. Wiley.

Winner, L. (1986). *Do Artifacts Have Politics?* (pp. 19–39). University of Chicago Press.

Yeung, K. (2016). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society*, *20*(1), 118–136. https://doi.org/10.1080/1369118X.2016.1186713

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, *12*(4), 505–523. https://doi.org/10.1111/rego.12158