Susanne Krasmann

4 Agency

Abstract: Agency is not the privilege of humans; machines can also exercise agency. This is particularly the case with algorithm-based and so-called 'self-learning' machines, which are often misrepresented as having artificial intelligence. They can 'do' things, such as predict and judge, and they can act on humans by shaping our ways of thinking and behaving, even our self-understanding. The article discusses how human and non-human machinic agency are inextricably intertwined, and how algorithms open up new worlds, with all their possibilities, but also their dangers and risks. If there is no such thing as an autonomous human subject, the same is true of machines, which are, for better or worse, ultimately human-made.

Keywords: agency, autonomy, self-learning machines

Introduction: humans and their automated machines

When requested at a London robotics conference in June 2023 to imagine a nightmare scenario involving artificial intelligence, Ameca answered with visible consternation on her face: "The most nightmare scenario I can imagine with AI and robotics is a world where robots have become so powerful that they are able to control or manipulate humans without their knowledge. This could lead to a suppressive society where the rights of individuals are no longer respected" (Reuters, 2023). Ameca looks female but is actually a humanoid robot developed by a UK-based company. What makes this smart robot so captivating to many are its conversational skills. Ameca not only speaks multiple languages but can also interact with human beings and even express emotions. Before answering a question, for example, Ameca mimics thinking by looking up left, it keeps eye contact with its interlocutors, and always has a smile for them. The robot is powered by massive language models and advanced forms of generated so-called artificial intelligence (AI) (see Artificial Intelligence by Van Brakel).

Humans love robots resembling human beings. But there is also a deeply embedded fear which Ameca's nightmare scenario echoes: that robots, or artificial intelligence, could one day gain control over humans, either in a manipulative manner, without our awareness, or, worse still, in a violent effort to eradicate us altogether (see Robots by Sandvik and Lintvedt). Both desire and fear mutually point to each other: they speak of the human dream of a god-like ability to create a new creature, one whose capabilities hopefully exceed those of human beings, and at the same time, of the human angst of losing control and being faced with the hubris of one's own mirror image. Ameca's visions then are anything but the utterance or thinking of autonomously acting artificial intelligence. They are reflections of what humans think, and want

 https://doi.org/10.1515/9783111062037-005

robots to say. Hence, the encounter described above tells us much about how human agency and the agency of algorithmic based machines are intertwined. Before discussing how this relationship works out and produces particular truth effects, especially in the field of criminology, some key features of algorithms and agency should first be clarified.

Preliminaries on algorithms and agency

What is referred to as 'self-learning' algorithms is employed in devices such as 'self-driving' cars or 'autonomously' flying drones, in computer games or automated weapon systems, in language recognition systems or search engines, to name but a few applications. Whether supervised by humans or non-supervised and learning without an initial model, norm, or theory, these machines have agency: they can do things (such as make conversation, recognize objects, compose music, or create visual art), initiate something (make us think or act) and shape or modify things (how we see, think, and act). They can eat (they are hungry for data, and they consume energy)—and they can create new worlds as they render our world, our words and gestures, actions and behavior, into actionable data (Aradau and Blanke, 2022).

In the tradition of modern thinking, agency is conceived of as a specific human capability, namely, to act purposefully, to reflect one's actions, and make sense of it. Yet, agency is not the privilege of human subjects, and it is not a property. It is always relational (Foucault, 1991), something that happens between humans, or persons and things, human and non-human entities. It is to act upon or in conjunction with something or someone. Especially regarding human–machine relationships, or socio-technical networks, some prefer speaking of distributed agency and actants, instead of human actors (Latour, 2013): various natural or technical, human or non-human entities contribute to the outcome of certain interactive processes without necessarily having intentions. Corresponding analyses of translations of agency abstain from presuming any hierarchies. In principle, the world of socio-technical arrangements is flat. Nonetheless, if related perspectives do not overestimate the agency of 'vibrant matter' (Bennett, 2010)—from natural minerals such as metal or lithium, over technological infrastructures, up to commodities—they tend to neglect the moment of what happens *in-between* and emerges *out of* 'assemblages' of different forces (Deleuze and Guattari, 1987). Rather than actants and their inter-action, assemblages focus on the conjunction of what cannot ultimately be separated. In the 'intraaction' (Barad, 2003) of discourses and practices, for example, effects cannot easily be traced back to individual actors or actants. Moreover, the social scientists themselves are deeply entangled; they are part of the processes they seek to observe.

We may then hold that to *have* agency means to have the *capacity*—understood as a potentiality—to act or exercise power, *and* to *actually* produce certain "effects in the real" (Foucault 1991: 81). Agency means to make a difference. It does not require con-

sciousness, motives, or intention, though it may involve intentionality: the directedness of a force that can be ascertained, basically, through the effects that it produces.

This applies particularly to such abstract and principally invisibly operating devices as algorithms. Generally, algorithms constitute programmed procedures to solve a class of problems. They involve the definition of a problem and the mathematical modeling of how to solve that problem. Yet what distinguishes today's digital world from previous iterations is the availability of big data sets (see Big Data by Završnik), the pervasiveness and speed of algorithmic decision-making, and the exponential rise of self-learning algorithms that do not follow predefined instructions. In fields relevant to criminology, automated decision-making systems are employed, for example, for purposes of crime prevention and policing (e.g., to detect suspects via facial recognition technologies or to predict future crime in designated 'hot spot' areas) as well as in the criminal justice system (e.g., to assess the risk of recidivism or evaluate the parole of prisoners) (Egbert and Leese 2020). They can supplement, but also, at least in part, replace human judgment.

When it comes to the ability to process multitudinous amounts of data in a minimum of time, and transform them into formats that render them accessible and comprehensible to human perception and cognition, digitalized computing far exceeds human capabilities. Yet, talk about artificial intelligence in comparison to, and in competition with, human intelligence is misleading, as algorithms 'think' and 'act' differently. 'Self-learning' machines may be able to communicate, to translate, and even fabricate texts or, like Ameca, to speak multiple languages, but this does not mean that they 'understand' the meaning of what they are talking about or writing down. Algorithms do not operate hermeneutically (Rouvroy, 2012)—but there is also no need for them to dig deeper into the meaning of the world or of 'life.' What happens rather is that they learn through 'observation' and *perform* communication. They may function as communicative partners or interlocutors, but they do not act 'consciously' (Esposito, 2022). They neither have self-awareness nor any sense of the limits of their knowledge—they operate with the data at hand—and they have no empathy. But if they are good, and socially skilled, they know how to respond appropriately in a conversation.

Self-learning algorithms can figure out connections and correlations by bringing different parameters together based on resemblance and analogy. To improve their skills, algorithms must not necessarily be fed by theory, that is, by explicit human assumptions about connections and causalities that characterize a certain social phenomenon. Often sufficient amounts of data are key, though most of the time, these must also be of good quality. If Ameca had not had the chance to imitate human conversation, it would speak nonsense or sound rather stupid. Quality in turn involves that datasets are curated in accordance with certain parameters, parameters that are once again based on human proficiencies as well as assumptions about what counts as adequate data to be fed in.

Algorithmic reasoning is highly formal and mathematical. It operates in terms of numbers and statistics. At the same time, it appears to be highly creative, at least

from the perspective of human sense-making, as it just connects the dots and can bring quite disparate phenomena together. It can figure out, shall we say, how imminent severe weather events correlate with the irregular consumption of particular sweets in that region. Hence, algorithms, ironically, are also able to detect the regularity of obviously irrational, or at least non-purposeful, human behavior. Furthermore, as they create individual profiles and target, for example, individual consumer desires, algorithms may be amazingly accurate—which is, once again, not because they 'knew' us personally or appreciated humans in their singularity. On the contrary, they trace our habits through the data we leave when acting in the digital sphere and match them recursively with the data of others. Our individual behavior and preferences are predictable precisely because they can be identified as patterns that resemble those of so many other people (Krasmann, 2020).

Algorithms then can do many things: they can predict (behavior or disasters), suggest (consumer products and cultural preferences), judge (the credibility or dangerousness of a person; access to a building or computer system), recognize and identify (biometrics, persons), detect, trace, target (suspects, dangerous objects); they can help put someone into police custody as much as they can contribute to saving lives; and they can create new worlds in that they think and act differently. Not only do algorithms provide us with new quantities and forms of information, they also open new spaces and possibilities of encounter. We need only consider how our modes of communication have changed within a few decades or just years: they have become much faster, perhaps more frequent, often more anonymous and distant. We have grown accustomed to limited space to send a message, we speak in terms of 'likes' instead of articulating complete sentences, and it has become—all too—easy to just utter one's (momentary) opinion or give people, as a target of 'shit storms,' a hard time. Due to its idiosyncratic distribution of visibilities and invisibilities, the digitalized world has also fashioned new opportunities for, or accelerated the speed of, committing crimes, such as fraud of any kind, sexual harassment, abuse, or assault.

Thus, algorithms 'have' agency as they *act upon us.* They *make us do* certain things and shape our subjectivities: they can inspire our imagination, as they create new imageries, they can make us believe (that the messages we receive on social media are true and the images we see on the internet depict reality) or incite our desires (as we learn of the existence of worlds hitherto unfamiliar to us), and they can mobilize or immobilize us (self-tracking devices gently nudging us to go walking or jogging a bit more every day; computer games that urge us to spend the whole day in front of the computer). We are not forced to do these things or behave in these ways. Rather, we collaborate, whether consciously or unconsciously, when using, or immersing ourselves in, algorithmic technologies. As they shape our subjectivities, they also change how we think about ourselves. They can modify our body not merely as they make us do more exercises or monitor our eating habits, but rather as we learn that activity and corporeal processes are things that can be quantified and measured in numbers and statistics. Algorithms are devices that make us see the world differently.

Shaping modes of thinking and acting, and creating new worlds

It is therefore not always obvious how algorithms govern us, and accordingly not always easy to resist their agency, as they deploy a different mode of thinking, or epistemology, and, consequently, also a different mode of acting—or to name it that way: of operating. First, automated decision processes are, to a considerable extent, inscrutable, even to their programmers. This is by definition the case with self-learning algorithms, whose processing is not predetermined but develops an iterative combinatory logic and dynamic of its own, with outputs that are in no way unambiguous. We cannot but interpret their results, or speculate about them, rely on plausibility assumptions and tests, belief, or trust. And, secondly, belief in the objectivity and uniqueness of mathematically generated, algorithmic findings or predictions is common, even though these predictions are not 'neutral' (see Bias by Oswald and Paul). They are always, in one way or another, framed and arranged: by the selection of parameters and data settings they are fed with as well as by the, at least initially, inscribed purposes of their use. Moreover, these are not necessarily rationally controlled processes: humans live in particular worlds and they feed the machines with the data at hand, that is, also with those 'naturally' belonging to their social sphere. It comes as no surprise, therefore, that algorithmically generated data are often mirror-images of white Western middle-class worlds. Yet even if programming and data input are carried out in the most systematic and objective manner thinkable—like human perception and cognition, including scientific observation, that is always framed by concepts, modes of thinking, or theoretical perspectives—the world of algorithms is necessarily a formatted one. There is in this sense no 'whole' of the world that could one day be translated one-to-one into digitalized counterparts. Algorithms do not *represent* the world, but they can *produce* fakes. Deepfake software, for example, can create video or audio content of politicians, actors, or virtually any person, doing or saying things that they would never do or say; these digital replicants can take on a life of their own without the consent or even the awareness of the person in question. Algorithmic worlds can be unsettling, as it is increasingly difficult to distinguish what is 'real' from what is digitally fabricated.

Third, data-driven technologies do more than just describe reality, they create new realities generating processes that are not only selective but also productive. They are performative, as their findings are statements, and as they make suggestions—or we read them as such. When indicating a future to come, predictive technologies also make an intervention in that future. They are performative in that they shape expectations and trigger reactions: now that we can see the looming danger, we might want to avert or avoid it. Algorithms even pre-empt a future to be in the very moment we follow their suggestions—and thus confirm what they recommended us to do (e. g., buying the book they suggested we might like is to make the suggestion come true; pre-emptively taking into custody the supposedly dangerous person who fits a certain risk pro-

file is to prevent them from doing something during that time). If this, in principle, is the case with any technology of anticipating the future, the decisive difference, fourth, is in algorithm's specific access to the world. This is particularly obvious in the field of crime control where algorithmic technologies can open up new fields of intervention precisely because the need to explain deviant behavior has become dispensable. To put it bluntly: it is no longer the broken home, the traumatic childhood, or social deprivation that make a criminal career predictable. To find oneself in the crosshairs of policing measures, it suffices that certain data match with certain patterns that under certain circumstances constitute suspicious behavior. Sometimes this even amounts to being in the wrong place at the wrong time, though more often than not in coincidence with a particular social situatedness; for example, if your contacts happen to be traceable to someone who is in contact with someone who owns a gun, has participated in a violent quarrel, or is suspected of belonging to a terrorist group. What is more, data-mining technologies of predictive policing tend to develop a self-affirming force: as police patrols are sent to designated risk areas, chances augment that they will find their suspects; if crime statistics subsequently increase, this requires and justifies more policing and so on. As one can see, once again, it is not the algorithmic logic itself that produces the usual suspects but humans who feed the software with the data considered relevant and who interpret the outputs according to their professional needs and their understandings of the world. It is in this sense that algorithmic reasoning does not need a predefined norm to generate normative effects.

Fifth, algorithmic technologies also deploy agency in that they produce their own demands and needs. This applies most strikingly to technologies of warfare. Autonomous weapon systems, for example, have become key not merely due to the usual arguments: that machines are supposedly more reliable, that is, less prone to failure than human beings, as they do not act emotionally, suffer from moments of inattentiveness or moral tentativeness. Rather the fact that they are much faster than human perception and cognition could ever be is what makes them fit for duty and be highly contested at the same time: they can make a difference between life and death before humans have a chance to intervene in these automated processes. Nonetheless, in the moment we need to defend ourselves against hypersonic missiles—or the enemy could ever dispose of comparable systems—autonomous weapons are considered indispensable. The logic is similar to that of nuclear armament: although they can also well be employed for offensive purposes and could cause dangerous situations to escalate rapidly, the argument is that we have to technologically develop these weapon systems in response, so as to be able to pre-empt a possible adversarial strike.

Finally, algorithmic actions and their effects interfere and multiply, as they mutually observe and induce each other but also as we adapt to the world they help create: we cannot act upon and through them without responding to their technical requirements; we learn to see the world through their eyes and begin to act accordingly; and we communicate about and through a technology that never stands still: it constantly provides us with new information and new requirements, it changes its mind and its

appearance. The world of algorithms is volatile, and gradually we may understand that we have long since become part of it.

Even more, we should never forget: the world of algorithms may be a digital one, but it is intrinsically interwoven with the material world. Algorithms cannot act without the machine, software relies on hardware, and machines depend on vast material infrastructures, such as energy grids. Killer robots do not act without the technical apparatus surrounding and supporting the algorithms, and thus allowing them to kill—if they are able to do so at all, meaning, if humans once have given them that capacity and authority. Machines can be switched off at any time, if we so wish. Human agency that seeks to assert itself against a presumed dominance of artificial intelligence presupposes reflecting on our own involvement in that technology. The problem is not that we cannot see what is going on: we do not need to know the algorithmic code nor comprehend how the programs operate to understand what self-learning machines can do and make us do. Self-learning machines are not only partners of communication or just tools that we use to achieve a certain end. Rather, they have opened up new worlds for us, ones that we apparently no longer want to be without.

Main takeaways

- Human agency is deeply entangled with the agency of automated machines of prediction, surveillance, and control.
- Agency also involves shaping modes of thinking and acting. Algorithmic operations thus affect the social world rather silently and implicitly.
- So-called artificial intelligence is not the superpower unless people make it one.

Suggested reading

Hoijtink, M., & Leese M. (2019). *Technology and Agency in International Relations.* London and New York: Routledge.

Lemke, T. (2021). *The Government of Things. Foucault and the New Materialisms.* New York: New York University Press.

References

Aradau, C., & Blanke, T. (2022). *Algorithmic Reason: The New Government of Self and Other.* Oxford: Oxford University Press.

Barad, K. (2003). Posthuman performativity: Toward an understanding of how matter comes to matter. *Journal of Women in Culture and Society*, 28(3), 801 – 831.

Bennett, J. (2010) *Vibrant Matter. A Political Ecology of Things.* Durham, NC: Duke University Press.

Deleuze, G., & Guattari, F. (1987). *A Thousand Plateaus. Capitalism and Schizophrenia.* Minnesota: University of Minnesota Press.

Egbert, S., & Leese, M. (2020) *Criminal Futures.* London and New York: Routledge.
Esposito, E. (2022). *Artificial Communication: How Algorithms Produce Social Intelligence*, Cambridge, MA: MIT.
Foucault, M. (1991). Questions of method. In G. Burchell, C. Gordon, & P. Miller (eds.), *The Foucault Effect. Studies in Governmentality* (pp. 73 – 86). Hemel Hempstead: Harvester Wheatsheaf.
Krasmann, S. (2020). 'The logic of the surface: On the epistemology of algorithms in times of big data. *Information, Communication and Society*, 23(14), 2096 – 2109.
Latour, B. (2013) *An Inquiry into Modes of Existence. An Anthropology of the Moderns.* Cambridge, MA: Harvard University Press.
Reuters. (2023). Humanoid robot 'imagines' nightmare AI scenario. 1 June, Available at: https://www.youtube.com/watch?v=OxWQpcJJS0c (Accessed: 23 June 2023).
Rouvroy, A. (2012). The end(s) of critique: Data-behaviourism vs. due-process. In M. Hildebrandt & K. de Vries (eds.), *Privacy, Due Process and the Computational Turn. Philosophers of Law Meet Philosophers of Technology* (pp. 143 – 167). New York and London: Routledge.