Original research article

BIG DATA & SOCIETY

Big Data & Society
July–September: 1–14

DOI: 10.1177/20539517261481720
journals.sagepub.com/home/bds

Sage

Making up with parrots: A critical praise of machine learning agency

Florian Jaton[1,2] **and Marc Lenglet**[3]

Abstract

This paper aims to contribute to the reconciliation between radical critiques and endorsements of machine learning's capacity to act, intervene, and produce social effects (i.e., its agency). Building upon recent ethnographic studies of the mundane shaping of algorithmic systems, the paper begins by underlining three positive attributes of machine learning agency: "experience" (the ability to embody insights derived from referential datasets), "consistency" (the ability to remain aligned to encoded principles), and "intelligence" (the ability to link new elements with prior knowledge). To balance this reverential perspective, the paper then highlights the limitations of these attributes through the lens of philosophy. Chuang Tzu's concept of emptiness reveals the struggle of machine learning agency with openness and confusion; Bruno Latour's reflections on psycho-bearing entities highlight machine learning agency's imperviousness to external disruptions; and Henri Bergson's discussion of intelligence exposes the inability of machine learning agency to engage with lifeful phenomena. This exploration results in a form of "critical praise" that acknowledges both the constraints and capacities of machine learning agency, fostering diplomatic rapprochement between its strongest critics and advocates.

Keywords

Machine learning agency, critical praise, diplomacy, experience, consistency, intelligence

> I, and others, started out wanting a strong tool for deconstructing the truth claims of hostile science by showing the radical historical specificity, and so contestability, of every layer of the onion of scientific and technological constructions, and we end up with a kind of epistemological electroshock therapy, which far from ushering us into the high stakes tables of the game of contesting public truths, lays us out on the table with self-induced multiple personality disorder. (Haraway, 1988, 578)

Introduction

Since at least Hubert Dreyfus's pioneering work (1972), algorithms—loosely understood as computerized methods of calculation—have been the subject of sustained criticism from scholars in the humanities and social sciences (e.g., Hoffman and Novak, 1998; Introna and Nissenbaum, 2000; Introna and Wood, 2002; Zureik and Hindle, 2004). And in recent years, this criticism has justifiably intensified, driven by the combined rise of mobile computing—which has broadened the use of algorithm-based web applications (e.g., Bozdag, 2013; Bucher, 2012; Steiner, 2012)—and the development of advanced machine learning (ML) techniques capable of deriving new and powerful algorithms from formatted data (e.g., Jaton, 2021a; Bechmann and Bowker, 2019; Girard-Chanudet, 2025).

This critical scrutiny of the capacities attributed to ML algorithms to act, intervene, and produce social effects—what we refer to here under the synthetic term *agency*[1]—was salutary, if only to counterbalance the overenthusiasm of techno-solutionist artificial intelligence (AI) admirers. More generally, this critical approach has helped to problematize the notion of ML algorithm, somewhat distancing it from the modernist trope of pure objectivity. In short, thanks to what is now referred to as *critical algorithm studies* (Gillepsie and Seaver, 2015), ML algorithms have come to be understood publicly as they have long been privately:

[1]Institute of Social Sciences, UNIL, Lausanne, Switzerland
[2]College of Humanities, EPFL, Lausanne, Switzerland
[3]Strategy & Entrepreneurship Department, NEOMA Business School, Mont-Saint-Aignan, France

Corresponding author:
Florian Jaton, Institute of Social Sciences, UNIL, Lausanne, Switzerland.
Email: florian@florian-jaton.com

as technical-and-therefore-also-cultural devices (Akrich, 1992; Seaver, 2017) that inherently embody—and consequently *reproduce* and *reinforce*—aspects of the contingent social arrangements from which they originate (O'Neil, 2016). Mere stochastic parrots, in short, as Bender et al. (2021) famously (and provocatively) put it.

However, the backlash against this much-needed critical moment is very real. Beyond highly publicized events, such as the abrupt dismissal of ethical AI researcher Timnit Gebru by Google (Wong, 2020)[2] or infamous technopolitical supports for the authoritarian turn in U.S. politics (Coeckelbergh, 2026), there are subtler yet increasingly visible signs of normative tightening within computer science research and industry. From the downsizing of research units that lean towards social sciences and humanities (De Vynck and Oremus, 2023), to the removal of social science profiles from the boards of major companies (Mickle et al., 2023)—and even harsh statements questioning the legitimacy and seriousness of scholars in critical algorithms studies[3]—computer science practitioners and enthusiasts seem to be publicly expressing a growing reluctance to collaborate with social scientists, often viewing them as obstructive figures who merely "love mind games" (Moats and Seaver, 2019). There are, of course, notable exceptions (e.g., Choi and Park, 2025) that have succeeded in overcoming various forms of indifference, misalignment in publication venues (Domínguez Hernández and Owen, 2024), or, more prosaically, the compartmentalization of academic disciplines (Armstrong et al., 2012: 148). Yet such cases remain marginal, underscoring rather than reversing the growing difficulty—or at least stagnation—of stabilized interdisciplinary collaboration, even as the social consequences of ML algorithms and systems continue to be framed as increasingly massive.

This paradoxical tendency toward hermeticism is unfortunate, particularly in an academically inclined field where openness to critique is supposedly a fundamental value. But it also raises a harder question: if critique often encounters resistance, is that only due to (pervasive) industrial cynicism and institutional closure, or can some critical interventions also fail to engage constructively with the objects and practices they contest? In other words, as in the so-called "science wars" of the 1990s, could the current standoff over ML agency—that is, over the capacities attributed to ML algorithms and over the distribution of agency across humans, technical systems, and organizations—reflect, at least in part, an inability to respect essential (yet certainly arguable) *values* in computer science that are, ultimately, necessary to *compose* with (Latour, 2010a)?[4] This debatable hypothesis is the basis of our approach in this text, which aims to serve as a *diplomatic effort* (Janicka, 2023) to bridge computer science and critical algorithm studies, focusing on both the limitations and strengths of ML agency.[5] The idea is that of a small but real attempt at pacification, aimed at clarifying what ML algorithms can and cannot do, in order to foster new forms of engaged collaboration.

But how could one effectively initiate this effort to reconcile both radical critique and endorsement of ML agency? In this article, we propose to ground this effort empirically by drawing on a substantial corpus of fine-grained ethnographic inquiries into the construction of ML algorithms in various fields of applied computer science, which examine, for example, the everyday work of data preparation, model adjustment, evaluation, and deployment through which contemporary ML systems are made operational in concrete settings (e.g., Jaton, 2019, 2021b; Bechmann and Bowker, 2019; Engdahl, 2024; Henriksen and Bechmann, 2020; Kang, 2023; Muhr et al., 2023). By drawing on these descriptions of the everyday processes that shape ML algorithms, we first seek to highlight certain positive attributes of ML agency. To counterbalance this perspective, we then turn to its limitations, guided by philosophical reflections from classic yet unconventional thinkers. Our aim is to propose a nuanced characterization of ML agency, one grounded in both empirical observation and speculative thinking, and capable of resonating with both strong advocates and critics. We call this approach "critical praise" to signal its diplomatic ambition: "praise" because it respectfully underlies the specificity of ML agency, and "critical" because it refuses to treat this agency as innocent. In that sense, this article is addressed to two audiences at once. For critical scholars (of whom we are part), it argues that skepticism gains precision when it distinguishes carefully between what ML algorithms materially do and what is ideologically projected onto them. For practitioners and advocates, it argues that grand claims about the capabilities of ML or AI should be tempered by closer attention to the fundamental—almost logical—limits of these systems.

The paper is organized as follows. We begin by noting a common observation across ethnographic studies of ML algorithm construction: most ML algorithms encountered in daily life derive from datasets that are, to some extent, arbitrarily assembled. From this observation, we define ML algorithms as *parameterized function approximations* designed to model relationships within these datasets as finely as possible. We then draw on this socio-ethnographic definition to identify three positive attributes of ML agency, treating them not as native categories given in the case studies, but as analytical consequences of our interpretation. Following—and therefore respecting—stakeholders in their propensity for anthropomorphizing (MacKenzie, 2019) and hyperbole (Kotliar, 2025), we name the first attribute a form of *experience*, as the ML algorithm—this parameterized function approximation—embodies accumulated representations from its referential dataset. The second attribute is a form of *consistency*, as the ML algorithm adheres to fixed principles enshrined in its defining parameters. The third attribute is a form of *intelligence*, in that the ML algorithm is effectively the expression of links between (*inter-legere*) a

multitude of elements gathered into sets. To add a critical dimension to this approach, we then highlight the limitations of these positive attributes through the lens of classical philosophy. The concept of emptiness as developed by the Chinese thinker Chuang Tzu, along with recent interpretations of his works (Billeter, 2004, 2010, 2014), helps us illustrate the limitation of the experience attribute, as ML agency can hardly revert to a state of openness and confusion. Bruno Latour's discussions of entities possessing a "psyche" (Latour, 2013: 181–206) underscore the limitations of the consistency attribute, as ML agency is resistant to disruptive external influences. Finally, Henri Bergson's differentiation between instinct and intelligence (Bergson, 1944 [1907]) sheds light on the limitations of the intelligence attribute, as ML agency is fundamentally extrinsic and therefore unable to engage with lifeful phenomena. In the conclusion, we expand on these insights and the broader notion of critical praise, with a view to clarifying a more balanced characterization of ML agency and examining the moral, political, and practical stakes of such a characterization for critique and interdisciplinary collaboration.

Machine learning algorithms as parameterized derivatives of referential datasets

A recent series of ethnographic studies has looked into the everyday work practices involved in constructing ML algorithms across applied fields, including image processing (Jaton, 2017, 2021b, 2022), computer audition (Kang, 2023; Seaver, 2022), video processing (Engdahl, 2024), algorithmic trading (Lange et al., 2019, 2024), text analysis (Bechmann and Bowker, 2019; Girard-Chanudet, 2025; Wu, 2024), digital journalism (Christin, 2020), and predictive medicine (Jaton, 2023; Henriksen and Bechmann, 2020; Muhr et al., 2023).[6] Taken together, the focus of these studies has been less on abstract models than on the concrete *work* through which ML systems become operative: coordinating heterogeneous actors, specifying targets through programming and validation practices, and negotiating what counts as a good result in particular domains. The cases are diverse—from image, video, and sound processing to trading, text analysis, journalism, and medicine—but they converge on one point that is central to our argument here: in all of them, ML work is inseparable from the preparation, formatting, linking, and evaluation of data. A recurring observation in this literature is thus that ML systems are tied to inscriptions collected/produced in more or less arbitrary ways and assembled in more or less structured *sets* (Amoore et al., 2024; Le Ludec et al., 2023).

These *referential datasets*—referred to as "ground truth,"[7] "benchmark," or simply "dataset," depending on the specific application field (Jaton, 2021a, 2025)—serve to organize, and thus relationally connect, at least two distinct subsets: (a) an input-data subset, which gathers the data the prospective ML algorithm will process, and (b) an output-targets subset which gathers the outcomes the ML algorithm is intended to produce (see Figure 1 and 2).

Far from peripheral, these referential datasets are central to the work of ML algorithm construction, for the fundamental purpose of ML algorithms is to reproduce, as finely as possible, the relationships present in those datasets, using mathematical formulations, specialized hardware, programming languages, evaluation mechanisms, and often *other* algorithms. In fact, the profound impact these datasets and their assemblage exert on both the development process and the content of ML algorithms even suggests the following socio-ethnographic definition: an ML algorithm can be viewed as a computer-compatible approximation of an ideal function that purportedly governs the relationship between the input-data and the output-targets of at least one referential dataset (Jaton and Lenglet, 2025; Mackenzie, 2017: 28; Rettberg et al., 2024: 85).[8]

There are at least two general methods for achieving this type of function approximation. The first is the "classical" approach, which involves assessing whether an established type of mathematical function (e.g., Fourier transform, discrete cosine transform) can express part of the relationships between input-data and output-targets. If successful, the function is then parameterized according to the dataset, with this parameterized function serving as the new algorithm. For many decades, this mathematical approach was the primary method of algorithm construction in applied computer science (e.g., early data compression schemes, path-finding methods). Over the past two decades, however, a second method has become more prevalent: using other algorithms specialized in function approximation. These algorithms, *delegated* the task of best fitting the constitutive relationships of referential datasets (Jaton, 2021b: 263–279), are typically rooted in statistics, probabilities, and—especially for large language models (LLMs)—linear algebra. Because they mechanically minimize errors, they are said to "learn," hence the label *machine learning* algorithms. This approach requires at least two algorithms: a first, "learning" algorithm, which mechanically defines the function approximation, and a second, "parameterized" algorithm, which constitutes the final approximated function.[9]

Due to various contingent factors, ML has recently become dominant within applied computer science. Ingenious methods, such as deep neural networks (Dhaliwal et al., 2024), have enabled the creation of ML algorithms with previously unimaginable performance levels. But regardless of whether they are massive, as in deep learning neural networks, or simpler, as in linear regression, ML algorithms ultimately revolve around approximating an ideal function that supposedly organizes the relationships between the data of a referential set (see Figure 3).

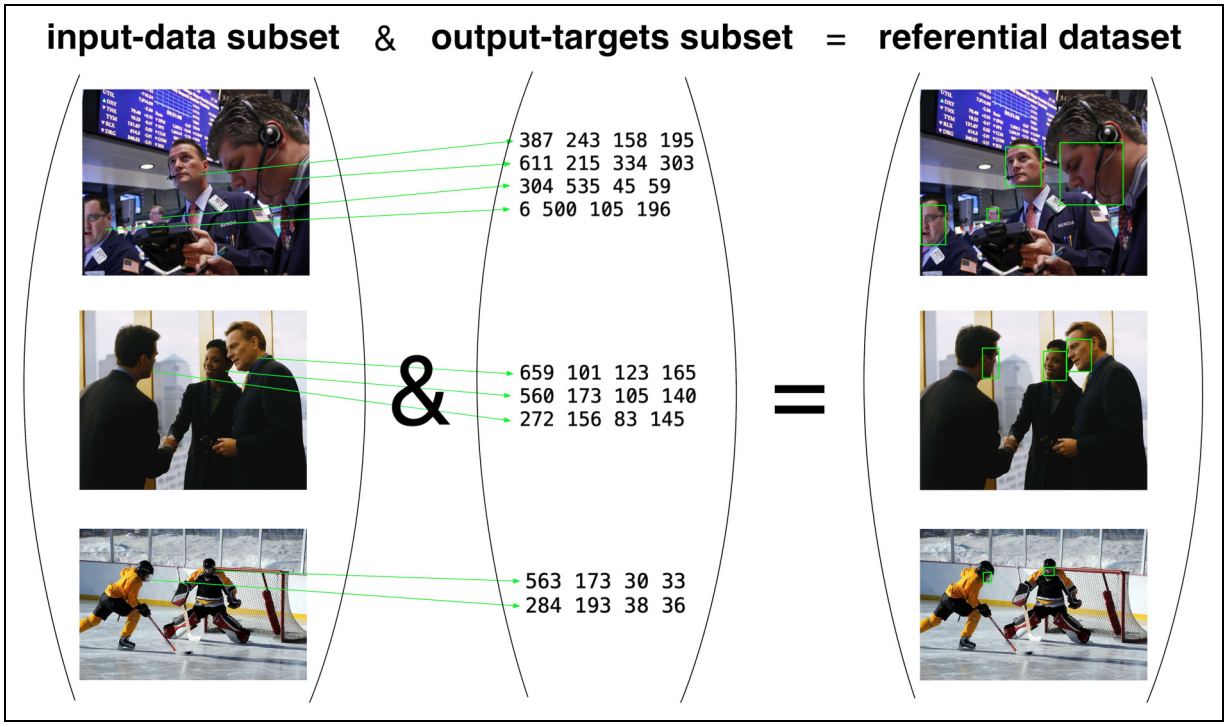

Figure 1. Taken from the WIDER FACE open dataset (Yang et al., 2016), this schematic illustrates a referential "ground truth" dataset for face detection. On the left are three examples from the dataset's 32,203 images, which are publicly available for face-detection research. In the middle, face annotations (expressed as numerical values) correspond to the human faces within these images (green arrows). Since each annotation belongs to the coordinate space of its associated "input" image, it can be represented as a set of four numerical values: the first two indicate the starting position of the label along the x and y axes, the third represents the width in pixels, and the fourth represents the height in pixels. This numerical information, corresponding to the bounding boxes around faces (as shown in the images on the right), was manually produced by one annotator and cross-checked by two others (Yang et al., 2016: 5527). Once linked, the images and labels are used to train and evaluate algorithms designed to detect faces in videos and photographs. The referential "ground truth" dataset operates as the optimal answer list for this task, that is, the task-specific reference against which outputs are judged. Source: adapted with permission (Jaton, 2021a).

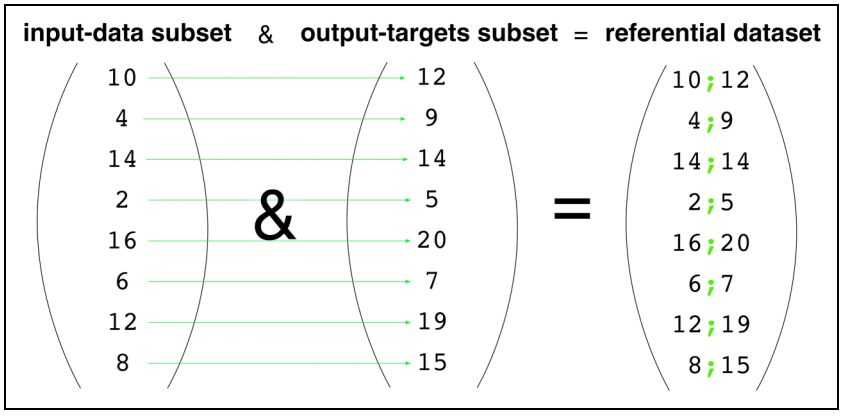

Figure 2. This diagram offers a hypothetical schematic of a simplistic referential "ground truth" dataset. Here, the input-data subset is a list of integers meant to be transformed (indicated by green arrows) into other integers in the output-targets subset. The referential "ground truth" dataset (shown on the right) "simply" aligns these two subsets to form a list of two-dimensional coordinates (indicated by the symbol ";"). Source: authors.

An algorithm of this kind can therefore be understood as a computer-compatible, parameterized function approximation derived from at least one referential dataset. Quite a down-to-earth thing, then. But does it imply that all the talk about AI, digital brains, thinking robots, or the imminent advent of artificial "general" intelligence is, above

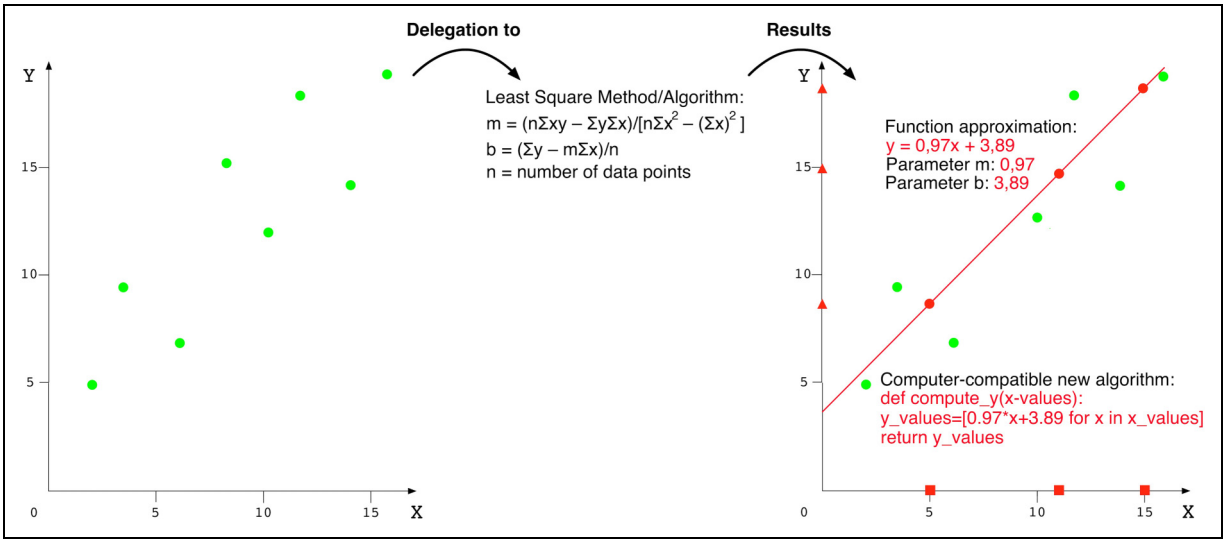

Figure 3. On the left, a graphical representation of the referential dataset from Figure 2 (green dots in the XY coordinate system) is shown. In the middle, details of the Least Squares statistical learning algorithm are provided, which has been tasked with fitting—and thus optimally expressing—the relationships between the dataset's data points. On the right are the results of this task: a function approximation with two parameters (m and b), which allows the prediction of new y-values (red triangles) based on previously "unseen" x-values (red squares) and the placement of these new xy pairs on the XY plane (red dots). At the bottom right, the computer-compatible version of this new, small, ML algorithm is presented in Python code. Source: authors.

all, a science-fictional pop fantasy? There have undoubtedly been—and remain—notable exaggerations, the most cynical of which must be openly opposed (Vertesi et al., 2026). Yet, in this text, we want to adopt a different, complementary perspective: instead of asserting that ML (or AI) algorithms are nothing more than parameterized derivatives of referential datasets, we suggest they are *nothing less*. For the painstaking work of function approximation has also been, and remains, a remarkable collective techno-scientific undertaking, endowing ML algorithms with a respectable agency, which we will now examine through a number of positive attributes.

Three positive attributes: Experience, consistency, intelligence

Before proceeding, one methodological clarification is necessary. The three positive attributes discussed in this section—experience, consistency, and intelligence—are not presented as native categories extracted directly from the ethnographic case studies, nor as terms that ML practitioners themselves would necessarily use in this form. Rather, they are analytic categories proposed here on the basis of the socio-ethnographic definition developed in the previous section. That definition, in turn, is grounded in our interpretation of ethnographic studies of ML algorithm construction work across several domains. In other words, the ethnographic material does not by itself present these three attributes as already established findings; instead, it provides the empirical basis for a definition of ML systems from which these attributes can then be interpretively derived. This is also why the attributes are not exhaustive and we do not claim that they are the only possible qualities one could emphasize. Our claim is narrower: given the definition advanced here, these three attributes are especially useful, we believe, for a diplomatic exercise in critical praise.

ML algorithms produce differences and, in this sense, they participate in forms of agency that are relational and distributed across design, training, deployment, monitoring, and use. The question, then, is how to characterize this specific form of agency, ideally first from a positive and slightly anthropomorphic viewpoint (if only to respect the values that seem to be central to those with whom we have to compose). Does the socio-ethnographic definition of ML algorithms as parameterized derivatives of referential datasets offer a sufficiently rich account of their capabilities?

Let us start with a ground-level observation. Although simplistic, the computer-compatible parameterized function approximation on the bottom right of Figure 3 was the product of a process—almost a short journey—within its limited referential dataset. It involved summing x, y, xy, and then x^2 values, and reorganizing them based on the number of data points (n). In that sense, even for such a simplistic "machine learning" linear regression, multiple *trials* were involved. And the reward for these training trials—overseen by another algorithm, in this case, the Least Squares Method—is the parameters m and b, which serve as numerical objectivations of some small but real acquired knowledge. In short, becoming a parameterized function approximation renders this algorithm—and others of this kind—an entity *with experience*.

Figure 4. The experience continuum, illustrated with three arbitrarily selected algorithms. Source: authors.

We use the term *experience* in a deliberately restricted sense: not sensation, embodiment, cognition, or emotion, but the incorporation of prior data relations into parameters. And compared with most recent Generative Pre-trained Transformers (GPT) whose trillions of parameters are derived from massive datasets, the linear regression shown in Figure 3, with only two parameters, sits at the very lower end of this experience continuum. Nonetheless, the difference is one of degree rather than kind, since even these two parameters capture a meaningful level of experience, grounded in the dataset from which this ML algorithm originates (see Figure 4).

To identify a second positive attribute of ML agency, we need to make a less intuitive observation: unlike what popular fictional narratives may sometimes suggest, the experience accumulated in an ML algorithm's parameters *remains constant* over its existence.[10] Returning to our simplistic linear regression in Figure 3, once learning and parameterization are complete, *this* small ML algorithm retains its parameterized experience (*m* and *b*), which it will then computationally apply to new, previously "unseen" input data.[11] These parameters thus act as constitutive elements: should they change, a different ML algorithm would, at least by virtue of the empirical socio-ethnographic definition we embrace here, come into existence.

We call this second attribute *consistency*: the degree to which ML outputs remain consistent with guiding parameters that are relatively stable over time, and the degree of effort required to modify them. Like experience, consistency exists along a continuum: for the linear regression in Figure 3, enriching the dataset and re-running the parametrization would be easy—a task achievable in two clicks in Excel. In contrast, refining a GPT foundation model involving bottomless layers of neural network and extensive hardware demands months of preparation (Riom and Tanferri, 2024), massive computational resources, and substantial investment. In that sense, GPT-based algorithms like GPT-4 (Achiam et al., 2024) rank highly on the consistency continuum, remaining relatively stable unless substantial effort is invested in changing them, while the simplistic linear regression of Figure 3 ranks much lower. Yet again, this difference is one of degree, rather than kind (see Figure 5).

Experience and consistency: these are already quite respectable attributes of the agency of computer-compatible function approximations. But a third attribute appears once we observe what happens when an ML algorithm processes new data.

Examining again our small linear regression in Figure 3, we find that the ML algorithm takes new input-data (values 5, 11, and 15 on the X axis), drawing on its acquired experience (parameters *m* and *b*). This operation produces new elements (values 8.74, 14.96, and 18.44, on the Y axis), enabling the placement of these *x*;*y* pairs within the coordinate system, on the regression line (red dots). In short, the ML algorithm *augments* new input data by *linking it* to its past experience, enabling the new element to become more *intelligible* (spatially situated on a plane, in this case). We term this quality *intelligence*, in the etymological sense of *inter-legere*: the capacity to link newly encountered elements to prior parameterized relations. We use the term cautiously; its colonial and patriarchal history (Cave, 2020; Weheliye, 2014) makes any celebratory use suspect. Here, it is meant in a restricted sense only: not consciousness or phenomenal awareness, but the capacity to render new inputs intelligible by *linking* them to prior data-based experience. Nothing more, but nothing less.

As with the previous attributes, intelligence also unfolds along a continuum. The agency of the linear regression in Figure 3 expresses, for example, a very low degree of intelligence, as it only permits one numerical value to be spatialized on a plane (which is already something). Conversely, the agency of the JPEG compression algorithm, ubiquitous in digital imaging, can be considered highly intelligent, as it facilitates the manipulation of large volumes of digital images, making them much more tractable to various entities, including other algorithms. But here too, the difference is one of degree, rather than kind (see Figure 6).

Chuang Tzu's confusion and emptiness

Experience, consistency, intelligence: this list of attributes may seem flattering, and it certainly is. Our point, however, is not to humanize algorithms excessively, but to identify—in a diplomatically attentive vocabulary—a restricted set of interesting capabilities, that yet also have their limits.

Starting with the experience attribute, we might ask: what *kind* of settled experience is this? Essentially, it is the result of multiplications, divisions, subtractions, and sums that eventually come to rest. Does this accumulation

of values constitute experience? Looking from a different angle, it could be seen as a mere *heap*. Can this heap grant ML agency true insight? Occasionally, for sure, but it can also *pull inward*, creating a tendency to revolve around the familiar. Has there ever been a metaphysics brave enough to deal with such a counter-intuitive pitfall of experience? To clarify this limit of experience, we turn to Chuang Tzu; more specifically, we draw on the interpretation proposed by Jean-François Billeter (2004, 2010, 2014), for whom Chuang Tzu's writings offer precise descriptions of ordinary regimes of activity.

What does Chuang Tzu tell us, then, and how could it help us identify a limit of experience, here understood as the insight acquired through trials happening in referential sets? A dialogue between Confucius—a recurring character in Chiang Tzu's texts—and one of his friends, Yen Hui, makes the point clearly:[12]

> Yen Hui said, "I'm improving!"
>
> Confucius said, "What do you mean by that?"
>
> "I've forgotten benevolence and righteousness!"
>
> "That's good. But you still haven't got it."
>
> Another day, the two met again and Yen Hui said, "I'm improving!"
>
> "What do you mean by that?"
>
> "I've forgotten rites and music!"
>
> "That's good. But you still haven't got it."
>
> Another day, the two met again and Yen Hui said, "I'm improving!"
>
> "What do you mean by that?"
>
> "I can sit down and forget everything!"
>
> Confucius looked very startled and said, "What do you mean, sit down and forget everything?"
>
> Yen Hui said, "I smash up my limbs and body, drive out perception and intellect, cast off form, do away with understanding, and make myself identical with the Great Thoroughfare. This is what I mean by sitting down and forgetting everything."
>
> Confucius said, "If you're identical with it, you must have no more likes! If you've been transformed, you must have no more constancy! So you really are a worthy man after all! With your permission, I'd like to become your follower. (Tzu, 1968, 6/h/89–93)

What does this passage—whose themes are echoed in other texts by Chuang Tzu (e.g., 21/d/24–27; 4/a/26–28; 7/f/31–33)—tell us? The first part of the dialogue presents a familiar image of learning: Yen Hui progressively internalizes morality, rites, and music to the point that he no longer needs consciously to attend to them. Confucius nonetheless replies each time that this is still insufficient. So far, Yen Hui is *learning*, accumulating what we have provisionally termed experience, yet Confucius indicates that Yen Hui's journey is far from complete.

The decisive turn comes when Yen Hui says that he can "sit down and forget everything." This does not mean that he has perfected mastery. It means, rather, that he can suspend the very drive to retain, categorize, and stabilize. He "drive[s] out perception and intellect," "do[es] away with understanding," and becomes identical with the "Great Thoroughfare." Confucius's response makes the implication explicit: to go through such transformation is to have "no more likes" and "no more constancy." Worthiness lies not in ever greater accumulation, but in the capacity to let prior acquisitions loosen their hold.

The lesson is straightforward, yet profound: accumulating experience and skills—and thereby gaining a form of mastery—is a virtue ("that's good," as Confucius remarks). Yet this accumulation has its limitations. Worthiness, which Confucius admires in Yen Hui at the end of the passage, arises precisely from the recognition of this asymptotic nature of learning, constrained by the ongoing transformations of the "Great Thoroughfare." But how does Yen Hui manage to transcend this tendency to accumulate mastery, a tendency he seemed so invested in at the beginning of the passage? First, it appears, by accepting a form of disorientation, and then by letting go of what he had accumulated. Indeed, it is through embracing a form of confusion (and, perhaps, a degree of anxiety) and then forgetting his prior acquisitions that Yen Hui achieves worthiness. Rarely has an apology for *dephasing*—that is, detaching oneself from a fixed tendency—been expressed so wittingly.

How, then, do these reflections help underline the limitations of ML agency's experience? Not because ML algorithms undergo an equivalent process of forgetting, embodiment, or self-transformation. They do not. The passage is useful precisely as a contrast. Yen Hui's worthiness depends on suspending prior acquisitions, letting go of established orientations, and inhabiting uncertainty. The ML systems discussed here, by contrast, retain and operationalize prior relations in parameterized form. In this limited sense, Chuang Tzu helps specify a boundary: such systems can accumulate and mobilize prior relations, but they cannot by themselves enact the practice of transformative forgetting described in the dialogue.

Latour's beings of metamorphosis

It appears, then, that experience alone is not enough, for—as Chuang Tzu suggests—its constitutive phasing prevents

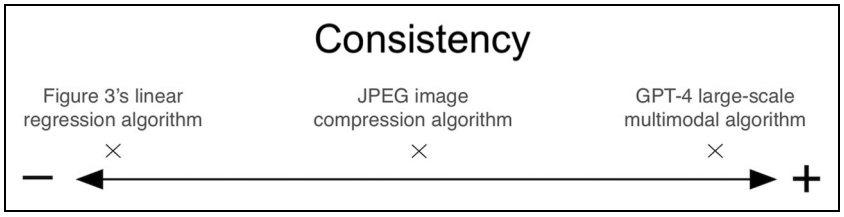

Figure 5. The consistency continuum, with three arbitrary selected algorithms. Source: authors.

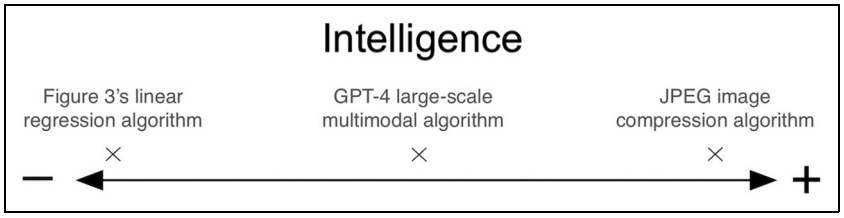

Figure 6. The intelligence continuum, with three arbitrary selected algorithms. Source: authors.

one from returning to the most fertile state of confusion and emptiness. This metaphysical insight carries implications for ML agency, firmly constraining its first attribute, though without entirely undermining its vitality. But what about the two remaining attributes? Let us set aside the sensitive issue of *intelligence* for now and focus instead on *consistency*, with an important clarification: the consistency we discuss here involves the preposition "with," signifying the propensity of ML agency to remain *consistent with* its guiding principles or parameters.

This clarification is especially relevant because the thinker we believe best suited to illuminate the limitations of this attribute is the sociologist and philosopher of science Bruno Latour, whose engagement with the concepts of truth, knowledge, and coherence has long been a source of debate and controversy. Fortunately, we don't need to resolve those broader Latourian questions here. For our purposes, a few ideas drawn from Latour's exploration of ethnopsychiatry (2010b) are enough, for it is in his attempt to reveal the networks that generate interiorities that we find a useful way of specifying a limit of ML agency's consistency.

So, what does Latour say, and how might it help us identify the limitations of the consistency attribute, understood here as a form of strong adherence to one's constitutive principles? Let us begin by examining a passage that describes a common experience rarely intellectualized—at least today—as something tied to external displacements:

> This experience is […] any kind of setting into motion—a hurtful word, a shocking attitude, an untoward gesture, sometimes even qualified by the term "evil eye"—that gives the impression that one is being taken over, overcome by an uncontrollable force—"I don't know what got into me." Let us call this experience a *crisis*. And since we are always "in the grip" of something in a crisis, let us call it a *gripping crisis*. (2013, 190, italics added)

These "gripping crises" are familiar, though sometimes feared, because they transport us into another state—sometimes comforting, often terrifying. Yet these external forces can also prove opportune, providing glimpses of salvation:

> [T]he *beings* that were deceiving us are on the contrary helping us to exist, finally, [we] can surf on them, through them, with them, thanks to them, by dint of skill, as on a wave that carries you away but that isn't targeting you personally. As if there were many beings that […] get you all *mixed up*. (2013, 192, italics added)

To underscore this transformative quality, Latour refers to these shifting entities—whose networks can be traced—as "beings of metamorphosis" (2013: 201–210) that shake, redirect, unsettle, but above all *engender*:

> What would we do without them? We would be *always and forever the same*. They trace […] paths of alteration that are at once terrifying (since they transform us), hesitant (since we can deceive them), and inventive (since we can allow ourselves to be transformed by them). (2013: 201, italics added)

These speculative beings of metamorphosis are traversing entities, capable of carrying, transporting, and altering. But how does this relate, at least on a metaphysical level, to our matters of ML agency? Precisely because the constitutive consistency of this agency, which adheres to the principles enshrined in its experience/parameters, has to remain *impervious* to these beings of metamorphosis. In other words, this agency must remain *isomorphic*, true to itself, unable to be disrupted by its own gripping crises. Beyond

its speculative tone, Latour's vocabulary helps specify ML agency's relative inability to be altered from within by transformative crises. But is this simply the nature of computers, with their stone-like core? Not necessarily, for as it is possible to design programs and systems that are permeable to the external forces that pass through them (though such systems may no longer be strictly "ML algorithms" by our definition)[13] just as there are people as tough as nails, whom nothing shakes or moves (Latour, 2013, 191), who thus express a form of ML agency as defined here.

But is it truly a limitation to be constitutively impervious to Latour's speculative beings of metamorphosis? There are, of course, advantages to consistently adhering to one's principles or parameters: a kind of staticity, for instance, that can be reassuring in turbulent times (much like those childhood friends who remain eternal teenagers). Yet to truly become something new—to innovate or even create—requires a deviation from one's established trajectory. This kind of genuine novelty—which, according to Latour, emerges not from deep within but from encounters with beings that turn us upside down—necessitates a rupture, a betrayal of the very principles that once defined us. Latour therefore helps us specify a limit of ML consistency: the agency of such systems can preserve continuity and stability, but must remain impervious to unsettling disruptive transformations through which genuinely new trajectories emerge.

Bergson's distinctions: Of intelligence, instinct, and intuition

It seems, then, that the consistency of ML agency also has a limit: it remains unresponsive to Latour's speculative beings of metamorphosis, entities that may bring chaos but also potential salvation—and, ultimately, novelty. But what about intelligence? Let us revisit, one last time, the definition of this attribute we proposed earlier in the text: ML agency can be considered intelligent insofar as it enriches newly encountered data-entities by *linking* them to previously acquired experience—thereby evoking the etymology of the term "*inter-legere*." If we accept this definition, the intelligence of ML agency is primarily determined by two factors: the richness of its referential dataset, and the sophistication of its function-approximating architecture. Consequently, intelligence emerges as an inherently extrinsic attribute—a quality that aligns with Henri Bergson's conception of intelligence as the capacity to manufacture artificial objects. Bergson argues that this capacity should be clearly distinguished from the abilities associated with instinct.

Indeed, when developing his ideas on intelligence, Bergson (1944 [1907]: 121) observes:

> "As regards human intelligence, it has not been sufficiently noted that mechanical invention has been from the first its essential feature, that even to-day our social life gravitates around the manufacture and use of artificial instruments, that the inventions which strew the road of progress have also traced its direction." (Bergson, 1944 [1907]: 152–153)

Intelligence is best understood as the capacity to manufacture artificial objects, and to extend this capacity by devising tools that serve in the making of other tools (Bergson, 1944 [1907]: 153–163). In setting intelligence apart from instinct, Bergson emphasizes a key difference: instinct concerns the use or even the construction of naturally organized instruments, whereas intelligence centers on the creation and use of instruments that are not organically given. From this perspective, instinct and intelligence embody two distinct modes of action, with innate knowledge in the case of instinct directed toward objects themselves, and in the case of intelligence directed toward relations among things.

This distinction between instinct and intelligence is further nuanced by Bergson's discussion of their respective limitations and strengths. Intelligence, he explains, is concerned with *relations*:

> "What is innate in intellect, therefore, is the tendency to establish relations, and this tendency implies the natural knowledge of certain very general relations, a kind of stuff that the activity of each particular intellect will cut up into more special relations. Where activity is directed toward manufacture, therefore, knowledge necessarily bears on relations." (Bergson, 1944 [1907]: 166)

While intelligence excels at making precise distinctions and establishing order, it is fundamentally constrained by its inability to engage with the fluid and indeterminate—the domain of instinct. Instinct, in contrast, provides a direct way of understanding that does not depend on external forms or representations. Whereas intelligence divides and categorizes, instinct moves seamlessly through continuities.

Applying this distinction to ML agency, we may therefore say that its intelligence—though capable of expressing complex function approximations derived from referential datasets—is inherently bound to its structured and extrinsic nature. ML agency lacks the capacity to navigate the unstructured, the ambiguous, and fundamentally undefined aspects that make the fabric of life. One might then say that it operates most comfortably within the discontinuous, the static, and the lifeless. As Bergson observes, the intellect is marked by an intrinsic incapacity to grasp the essence of life (Bergson, 1944 [1907]: 182).

Following our reading of Bergson's reflections on intelligence, we thus uncover a compelling characterization of the third attribute of ML agency. For example, a carefully trained facial recognition ML algorithm may identify and categorize facial images very accurately, thanks to its acquired experience, thus expressing the ability to build relations. Yet the same ML algorithm, when presented with an

unfamiliar image—one that is partially distorted, or subject to unusual lighting conditions—may fail to process the image. This failure would not arise from coding errors or a lack of computational power: rather, it could result from its inherent dependence on predefined elements—the very elements that at the same time enable its success within its given referential dataset. When faced with domains where input data are less categorizable—areas where ambiguity or ambivalence are fundamental, such as interpreting regulations in a specific situation (Campbell-Verduyn and Lenglet, 2023, 2025; Lenglet, 2021) or tackling ethical dilemmas (Martin, 2019)—ML algorithms are likely to struggle to gain traction, and to establish their legitimacy convincingly.

From these considerations, it follows that ML agency can, in Bergsonian terms, legitimately be described as a form of intelligence. For Bergson, intelligence and instinct represent distinct orientations of knowledge: "the former toward inert matter, the latter toward life" (Bergson, 1944 [1907]: 194). On this basis, our expectations of ML algorithms—be they huge or small—should be tempered: the fact that they are capable of exhibiting intelligence is already significant. It is therefore misguided to fault ML agency for lacking *intuition*, which Bergson describes as "instinct that has become disinterested, self-conscious, capable of reflecting upon its object and of enlarging it indefinitely" (Bergson, 1944 [1907]: 194). The difficulty in much contemporary debate lies precisely here: by not sufficiently distinguishing between intelligence and intuition, such discussions risk overlooking the metaphysical dimension of knowledge production. Bergson thus helps specify the limit at issue here: ML agency's intelligence can establish and operationalize relations with great force, but it remains poorly equipped for the fluid, continuous, and indeterminate dimensions of life to which instinct and intuition remain more closely attuned.

Conclusion

In times marked by the rise of noxious forms of cybermodernism, to say the least, a direct and uncompromising critique of ML agency remains essential (Coeckelbergh, 2026; Vertesi et al., 2026). Yet alongside this agonistic stance, it is equally important, we believe, to lay the foundations for a diplomatic process, one that continues to seek to compose a common world, against all odds.

Such diplomacy, however, is not an invitation to reproduce the ideological uses of agency-related language nor does it ask critical scholars to forget how notions such as intelligence or autonomy may be mobilized to legitimize harmful deployments, concentrate power, or displace responsibility. Rather, it asks whether critique can remain uncompromising while also becoming more precise about what *exactly* is being attributed to ML algorithms, by whom, for what purposes, and with what consequences. It also depends on a deliberately limited premise: not that ML systems are inherently necessary or socially beneficial, but that many of them are already operative enough in collective life that critique must decide whether to reject, regulate, redescribe, or strategically engage with them.

If the paper has a practical ambition, it is this dual one: to make critique more indulgent, while making computer scientific discourse less prone to mystification. That ambition has led us to develop the argument throughout. Drawing on ethnographic studies of ML algorithm construction work, we proposed a restricted socio-ethnographic definition of the class of algorithms at issue here: contemporary ML algorithmic systems whose fundamental constitution derives from the preparation and processing of referential datasets. On that basis, we first identified three positive attributes of their agency—experience, consistency, and intelligence—before using Chuang Tzu, Latour, and Bergson to specify their corresponding limits—their inability to undergo transformative forgetting, their relative closure to metamorphic disruption, and their difficulty in engaging indeterminate, lifeful phenomena.

This is what we mean by critical praise: not a softening of critique, but an effort to describe ML fundamental capacities with sufficient precision that criticism does not remain external or mystified in reverse. Framing ML algorithms as parameterized function approximations that model relationships within referential datasets highlights the powerful yet constrained nature of their *experience*, derived from prior data; their respectable yet limiting *consistency*, rooted in relatively stable parameters; and their genuine yet constricted form of *intelligence*, which is effective in establishing and operationalizing relations but falters in fluid and noncategorizable situations. The capacities identified here are real, but they remain bounded by the sociomaterial conditions and epistemic logics from which they emerge.

Ultimately, a diplomatic engagement of this kind seeks a form of critique capable of confronting both the real capacities of ML systems and the discursive excesses that surround them. It aims to render claims about ML agency more empirically grounded, politically answerable, and institutionally contestable, thereby helping to create the conditions under which such systems can be criticized while being respected, and deployed while being governed.

Acknowledgments

The arguments developed in this text have been fortunate to benefit from the thoughtful scrutiny of many colleagues, whose incisive comments have helped make them better over the months (or so we hope!). Trying not to omit anyone, and in (roughly) chronological order of their involvement, we would like to express our sincere gratitude to Geoffrey C. Bowker, Kate Crawford, Fabian Muniesa, Philippe Sormani, Francis Lee, Basile Zimmermann, Dominique Vinck, the participants in the Machine Learning and

Inference workshop (organized by Louise Amoore, Alexander Campolo, and Luke Stark in November 2024), the participants in the STS-CH conference panel "Has Critique of Algorithms Run Out of Steam?" (in September 2025), the participants in the Futures & GenAI Symposium (organized by Noa Vaisman and Frederik Vejlin in November 2025), Isak Engdahl, the participants in the Digital STS Hub seminar (organized by Charlotte Högberg in February 2026), Björn Fischer, Donald MacKenzie, the three anonymous reviewers, and the editors of Big Data & Society. All remaining mistakes and oversights are, of course, ours alone.

ORCID iDs

Florian Jaton https://orcid.org/0000-0002-5001-9098
Marc Lenglet https://orcid.org/0000-0002-7267-9425

Ethical approval and informed consent statements

This does not apply to the present paper.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of conflicting interests

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

Data availability statement

This does not apply to the present paper.

Notes

1. This critique of what we provisionally call ML agency should not be understood as claiming that earlier critical scholarship was exclusively, or even always explicitly, about agency. The critical literature has also addressed technological determinism, bias, surveillance, power, and the broader social consequences of data-driven systems. Our emphasis on agency is therefore more limited: these critiques repeatedly raise the question of what ML algorithms are taken to do and how they intervene in situations. In this sense, "ML agency" is a synthetic term for controversies over the capacities attributed to ML algorithms and the consequences of those attributions. The literature also makes clear that such agency is neither autonomous nor self-sufficient, but embedded in wider sociotechnical arrangements (e.g., Klinger and Svensson, 2018; Lange et al., 2019; Neff and Nagy, 2016; Neff et al., 2017).
2. Dr. Timnit Gebru, an AI ethics researcher at Google, was dismissed in December 2020. She claimed her firing was due to her raising concerns about the ethical risks associated with large language models and how these technologies were being developed.
3. For instance, in a 2023 interview with *The Guardian*, when confronted with critiques of large language models emerging from the social sciences, Sam Altman dismissed them as outdated, doubling down with a derisive question: "Is that still a widely held view? I mean, is that considered—are there still a lot of serious people who think that way?" (Hern, 2024).
4. Following Latour (2010a: 474): "[C]ompositionism takes up the task of searching for universality but without believing that this universality is already there, waiting to be unveiled and discovered. It is thus as far from relativism (in the papal sense of the word) as it is from universalism (in the modernist meaning of the world […]). From universalism it takes up the task of building a common world; from relativism, the certainty that this common world has to be built from utterly heterogeneous parts that will never make a whole, but at best a fragile, revisable, and diverse composite material."
5. This text joins a recent body of work concerned with bracing reflexivity in computer science research. Recent contributions include Hirsbrunner et al. (2024), who introduced Reflexive Data Science (RDS)—building upon Agre's critical technical practice (1997)—to support reflexivity in AI practices; Domínguez Hernández and Owen (2024), who proposed collaborative ethnography as a method for bridging interdisciplinary gaps in responsible AI development; and John-Mathews et al. (2023), who used reflective interviews and re-enactments to bring moral inquiry into the everyday work of data scientists. The present text complements these efforts by approaching reflection from a more theoretical and foundational perspective.
6. There are many reasons for this recent interest in the ethnography of ML algorithm construction work (Jaton, 2019), but it is undoubtedly partly due to discomfort with the aporias of what Malte Ziewitz (2016) wittingly termed the *algorithmic drama*: a disarming and repetitive critical discourse—mainly originating from the humanities and social sciences—that argues ML algorithms derive their power from their abstract nature. This abstraction is said to make them even more powerful, creating a cyclical narrative: the more abstract ML algorithms become, the more powerful they are perceived, and vice versa. Similarly to the ethnographic studies of the late 1970s by scholars such as Bruno Latour and Steve Woolgar (1986), Michael Lynch (1985), and Karin Knorr-Cetina (1981), which aimed to move beyond the abstractions of (critical) epistemology by providing grounded perspectives on how scientific knowledge is constructed, these modern ethnographers of *computer* science attempt to move away from the abstractions of (critical) digital epistemology. Instead, they seek to develop a concrete and realistic understanding of ML algorithms in order to provide, ultimately, a more empowering perspective.
7. "Ground truth" is a vernacular technical term for the reference labels, annotations, or target values against which outputs are trained or evaluated. In that sense, "truth" does not name an absolute or philosophical truth claim, but a practical reference point internal to a given task. The sociopolitical ramifications of ground truth datasets are the focus of numerous publications in the field of science and technology studies. For a

nonexhaustive list, see https://docs.google.com/document/d/1zDg396HuGU2ckSIQDQztIHe4XIQJwzVq/edit

8. This definition also usefully distinguishes between a "program" and an "algorithm:" while all effective ML algorithms are programs (they have no choice but to be), not all programs are ML algorithms. Broadly speaking, it is the process of referencing an external dataset to reproduce or approximate some of its constitutive relationships that—perhaps—defines the ML algorithmic form.
9. Boullier and El Mhamdi (2020) refer to "learning algorithms" and "execution algorithms," which essentially amount to the same thing.
10. The idea that ML algorithms can autonomously program themselves has caused significant confusion. This notion supports the misleading belief in the supposed autonomy of these systems, which are, in fact, highly dependent on external inputs and processes. This has also contributed to the abstract and misleading idea of intelligence in the Cartesian sense of consciousness. In practice, ML algorithms are parameterized based on training data, which allows them to define (or approximate) a function. So far, there is nothing particularly extraordinary about this. What is surprising, however, is that these functions—sometimes of immense complexity—are expressed solely in the form of computer code rather than in mathematical terms, as is the case with linear regression (maybe the simplest form of ML). This distinction has led some to claim that these algorithms program themselves, which is only true if one ignores all the programming work required to prepare the extraction and parametrisation of these functions. In this sense, it is indeed a misuse of language to claim that ML algorithms program themselves. This misuse of language can, however, be understood in a political sense, as it aligns with the modernist (therefore, abstract) vision of computer science. Another common misunderstanding is the belief that ML algorithms can reparameterize themselves. In reality, to parameterize a function mechanically (in a machine-like way), another algorithm is required. A single algorithm—at least according to the definition we adopt here—cannot reparameterize itself. However, the parameters defining an ML algorithm can be used to refine outputs by including previous outputs as new inputs (but this does not amount to reparameterization).
11. This is now often referred to as inference in the field of applied computer science.
12. Here, we use the translation by Watson (Tzu, 1968), which is among the most highly regarded in Anglo-Saxon studies of Chuang Tzu. In our references to the text, the first number indicates one of the thirty-three books, the letter corresponds to a subdivision of the book, and the subsequent numbers refer to the lines of text as they appear in the compiled corpus *A Concordance to Chuang Tzu* (Harvard-Yenching Institute, 1956).
13. See, for instance, the art installation "It's Normal for Some Things to Come to Your Attention" (Negativland, 2021), notably discussed in Muniesa (2023).

References

OpenAI, Achiam J, Adler S, Agarwal S, et al. (2024) GPT-4 Technical Report. *arXiv (Computing Research Repository)*.

Agre P, et al. (1997) Toward a critical technical practice: Lessons learned in trying to reform AI. In: Bowker G, Gasser L and Star L (eds) *Social Science, Technical Systems and Cooperative Work: Beyond the Great Divide*. Erlbaum, 131–157.

Akrich M (1992) The description of technical objects. In: Bijker WE and Law J (eds) *Shaping Technology / Building Society: Studies in Sociotechnical Change*. MIT Press, 205–224.

Amoore L, Campolo A, Jacobsen B, et al. (2024) A world model: On the political logics of generative AI. *Political Geography* 113: 103134.

Armstrong M, Cornut G, Delacôte S, et al. (2012) Towards a practical approach to responsible innovation in finance: New product committees revisited. *Journal of Financial Regulation and Compliance* 20(2): 147–168.

Bechmann A and Bowker GC (2019) Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society* 6(1): 205395171881956.

Bender EM, Gebru T, McMillan-Major A, et al. (2021) On the dangers of stochastic parrots: Can language models be too big? In: *In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, 610–623.

Bergson H (1944 [1907]) *Creative Evolution*. The Modern Library.

Billeter JF (2004) *Études sur Tchouang-tseu*. Allia.

Billeter JF (2010) *Notes sur Tchouang-tseu et la philosophie*. Allia.

Billeter JF (2014) *Leçons sur Tchouang-Tseu*. Allia.

Boullier D and El Mhamdi EM (2020) Le machine learning et les sciences sociales à l'épreuve des échelles de complexité algorithmique. *Revue d'anthropologie des connaissances* 14(1): 1–33.

Bozdag E (2013) Bias in algorithmic filtering and personalization. *Ethics and Information Technology* 15(3): 209–227.

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.

Campbell-Verduyn M and Lenglet M (2023) Imaginary failure: RegTech in finance. *New Political Economy* 28(3): 468–482.

Campbell-Verduyn M and Lenglet M (2025) Beyond the hype: In what sense are algorithmic technologies transforming regulation? *The Capco Journal of Financial Transformation* 61: 74–81.

Cave S (2020) The problem with intelligence: Its value-laden history and the future of AI. In: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. Association for Computing Machinery, 29–35.

Choi MA and Park BS (2025) More-than-human mediation: Social scientists in the making of environmental AI systems. *Science, Technology, & Human Values* 51(5): 907–933.

Christin A (2020) *Metrics at Work: Journalism and the Contested Meaning of Algorithms*. Princeton University Press.

Coeckelbergh M (2026) Technofascism: AI, Big Tech, and the new authoritarianism. *AI & Society* 41(5): 5163–5176.

De Vynck G and Oremus W (2023) As AI booms, tech firms are laying off their ethicists. Washington Post, 30 March, Available at: https://www.washingtonpost.com/ technology/ 2023/03/30/tech-companies-cut-ai-ethics/ (accessed 15 August 2024).

Dhaliwal RS, Lepage-richer T and Suchman L (2024) *Neural Networks*. University of Minnesota Press.

Domínguez Hernández A and Owen R (2024) We have opened a can of worms: Using collaborative ethnography to advance responsible artificial intelligence innovation. *Journal of Responsible Innovation* 11(1): 2331655.

Engdahl I (2024) Agreements "in the wild": Standards and alignment in machine learning benchmark dataset construction. *Big Data & Society* 11(2): 20539517241242457.

Gillepsie T and Seaver N (2015) Critical algorithm studies: A reading list. Available at: https://socialmediacollective.org/reading-lists/critical-algorithm-studies/ (accessed 15 August 2024).

Girard-Chanudet C (2025) Ground-truth is law: The invisible conceptual work behind AI. *Big Data & Society* 12(2): 20539517251352823.

Haraway D (1988) Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies* 14(3): 575–599.

Harvard -Yenching Institute (1956) *A Concordance to Chuang Tzu*. Second Printing edition. Harvard University Press.

Henriksen A and Bechmann A (2020) Building truths in AI: Making predictive algorithms doable in healthcare. *Information, Communication & Society* 23(6): 802–816.

Hern A (2024) Why AI's Tom Cruise problem means it is "doomed to fail." *The Guardian*, 6 August. Available at: https://www.theguardian.com/technology/article/2024/aug/06/ai-llms (accessed 15 August 2024).

Hirsbrunner SD, Tebbe M and Müller-Birn C (2024) From critical technical practice to reflexive data science. *Convergence: The International Journal of Research into New Media Technologies* 30(1): 190–215.

Hoffman DL and Novak TP (1998) Bridging the racial divide on the internet. *Science* 280(5362): 390–391.

Introna LD and Nissenbaum H (2000) Shaping the web: Why the politics of search engines matters. *The Information Society* 16(3): 169–185.

Introna LD and Wood D (2002) Picturing algorithmic surveillance: The politics of facial recognition systems. *Surveillance & Society* 2(2/3): 177–198.

Janicka I (2023) Reinventing the diplomat: Isabelle Stengers, Bruno Latour and Baptiste Morizot. *Theory, Culture & Society* 40(3): 23–40.

Jaton F (2017) We get the algorithms of our ground truths: Designing referential databases in digital image processing. *Social Studies of Science* 47(6): 811–840.

Jaton F (2019) « Pardonnez cette platitude » : De l'intérêt des ethnographies de laboratoire pour l'étude des processus algorithmiques. *Zilsel* 5(1): 315–339.

Jaton F (2021a) Assessing biases, relaxing moralism: On ground-truthing practices in machine learning design and application. *Big Data & Society* 8(1): 20539517211013569.

Jaton F (2021b) *The Constitution of Algorithms: Ground-Truthing, Programming, Formulating*. MIT Press.

Jaton F (2022) Éléments pour une sociologie de l'activité de programmation: Inscriptions provisoires, chaînes de référence et indexations. *RESET – Recherches en Sciences Sociales sur Internet* 11(1): 1–33.

Jaton F (2023) Groundwork for AI: Enforcing a benchmark for neoantigen prediction in personalized cancer immunotherapy. *Social Studies of Science* 53(5): 787–810.

Jaton F (2025) Examining algorithms in the light of their ground truth datasets: Results, objections, and avenues of reflection. *Digital Society* 4(2): 1–11.

Jaton F and Lenglet M (2025) Pour une apologie critique de l'agentivité algorithmique : Une perspective Bergsonienne. *Socio-anthropologie* 52: 51–67.

John-Mathews JM, De Mourat R, Ricci D, et al. (2023) Re-enacting machine learning practices to enquire into the moral issues they pose. *Convergence: The International Journal of Research into New Media Technologies* 30(1): 66–93.

Kang EB (2023) Ground truth tracings (GTT): On the epistemic limits of machine learning. *Big Data & Society* 10(1): 20539517221146122.

Klinger U and Svensson J (2018) The end of media logics? On algorithms and agency. *New Media & Society* 20(12): 4653–4670.

Knorr-Cetina K (1981) *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Pergamon Press.

Kotliar DM (2025) Can't stop the hype: Scrutinizing AI realities. *Information, Communication & Society* 29(3): 828–849.

Lange A-C, Lenglet M and Seyfert R (2019) On studying algorithms ethnographically: Making sense of objects of ignorance. *Organization* 26(4): 598–617.

Lange A-C, Lenglet M and Seyfert R (2024) High-frequency trading, spoofing and conflicting epistemic regimes: Accounting for market abuse in the age of algorithms. *Economy & Society* 53(4): 603–626.

Latour B (2010a) An attempt at a "compositionist manifesto." *New Literary History* 41(3): 471–490.

Latour B (2010b) Coming out as a philosopher. *Social Studies of Science* 40(4): 599–608.

Latour B (2013) *An Inquiry into Modes of Existence*. Harvard University Press.

Latour B and Woolgar S (1986) *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press.

Le Ludec C, Cornet M and Casilli A (2023) The problem with annotation. Human labour and outsourcing between France and Madagascar. *Big Data & Society* 10(2): 20539517231188723.

Lenglet M (2021) Algorithmic finance, its regulation, and Deleuzean jurisprudence: A few remarks on a necessary paradigm shift. *Topoi. An International Review of Philosophy* 40(4): 811–819.

Lynch M (1985) *Art and Artifact in Laboratory Science: A Study of Shop Work and Shop Talk in a Research Laboratory*. Routledge Kegan & Paul.

Mackenzie A (2017) *Machine Learners: Archaeology of a Data Practice*. MIT Press.

MacKenzie D (2019) How algorithms interact: Goffman's "interaction order" in automated trading. *Theory, Culture & Society* 36(2): 39-–359.

Martin K (2019) Ethical implications and accountability of algorithms. *Journal of Business Ethics* 160(4): 835–850.

Mickle T, Metz C, Isaac M, et al. (2023) Inside OpenAI's Crisis Over the Future of Artificial Intelligence. The New York Times, December 9. Available at: https://www.nytimes.com/2023/12/09/technology/openai-altman-inside-crisis.html.

Moats D and Seaver N (2019) You social scientists love mind games: Experimenting in the "divide" between data science and critical algorithm studies. *Big Data & Society* 6(1): 2053951719833404.

Muhr P, et al. (2023) The unobserved anatomy: Negotiating the plausibility of AI-based reconstructions of missing brain structures in clinical MRI scans. In: Flüchter A and Förster B (eds) *Plausibilisierung und Evidenz: Dynamiken und Praktiken von der Antike bis zur Gegenwart*. Bielefeld University Press, 169–192.

Muniesa F (2023) A science of stereotypes: Paranoiac-critical forays within the medium of information. *Distinktion: Journal of Social Theory* 24(2): 283–296.

Neff G and Nagy P (2016) Talking to bots: Symbiotic agency and the case of Tay. *International Journal of Communication* 10: 4915–4931.

Neff G, Tanweer A, Fiore-Gartland B, et al. (2017) Critique and contribute: A practice-based framework for improving critical data studies and data science. *Big Data* 5(2): 85–97.

Negativland (2021) *It's Normal for Some Things to Come to your Attention*. Negativland.

O'Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group.

Rettberg JW, Crawford K, Schultz J, et al. (2024) An AI society: Artificial intelligence is reshaping society, but human forces shape AI. *Issues in Science & Technology* 40(2): 76–88.

Riom L and Tanferri M (2024) On the trail of preparations. *Revue d'anthropologie des connaissances* 18(2): 1–23.

Seaver N (2017) Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society* 4(2): 2053951717738104.

Seaver N (2022) *Computing Taste: Algorithms and the Makers of Music Recommendation*. University of Chicago Press.

Steiner C (2012) *Automate This: How Algorithms Came to Rule Our World*. Portfolio Hardcover.

Tzu C (1968) *The Complete Works of Chuang-Tzu (trans. B Watson)*. Columbia University Press.

Vertesi J, Danah B, Taylor A, et al. (2026) Reckoning with the political economy of AI: Avoiding decoys in pursuit of accountability. *arXiv:2604.16106*. Available at: http://arxiv.org/abs/2604.16106 (accessed 14 June 2026).

Weheliye AG (2014) *Habeas Viscus: Racializing Assemblages, Biopolitics, and Black Feminist Theories of the Human*. Duke University Press.

Wong JC (2020) More than 1,200 Google workers condemn firing of AI scientist Timnit Gebru. The Guardian, 4 December, Available at: https://www.theguardian.com/ technology/2020/dec/04/timnit-gebru-google-ai-fired-diversity-ethics (accessed 15 August 2024).

Wu YH (2024) Capturing the unobservable in AI development: Proposal to account for AI developer practices with ethnographic audit trails (EATs). *AI and Ethics* 5: 1–14.

Yang S, Luo P, Loy CC, et al. (2016) Wider face: A face detection benchmark. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* June: 5525–5533.

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

Zureik E and Hindle K (2004) Governance, security and technology: The case of biometrics. *Studies in Political Economy* 73(1): 113–137.