Debate: Social Physics, or the Interpretative Turn? – *peer-reviewed*
Edited by Noortje Marres
https://doi.org/10.60923/issn.1971-8853/24341

Sociologica. V.20 N.2 (2026)
ISSN 1971-8853
https://sociologica.unibo.it/

Machine Demons

Juan Pablo Pardo-Guerra*

Department of Sociology, University of California San Diego (United States)

Submitted: March 2, 2026 – Accepted: June 28, 2026 – Published: August 6, 2026

Abstract

Machine learning has reinvigorated longstanding ambitions to develop a predictive science of society. This essay argues that enthusiasm for ML as a turning point rests on a category confusion between two distinct scientific imaginaries: Laplace's deterministic universe, in which sufficient data and computational power yield precise prediction, and Comte's social physics, which sought descriptive regularities rather than causal mechanisms. ML's insertion within the dominant causal frameworks that characterize much quantitative social science implies, in particular, that the Comtian promise is far from reality on the ground. The essay explores three reasons why ML is unlikely to deliver a fundamental epistemic break: the performative instability of social classifications, the irreducible indexicality of data and claims, and the irreducibility of collective knowledge to information underdetermine ML's overall capacity.

Keywords: Machine learning; Artificial intelligence; Epistemology; Computational Methods.

* jpardoguerra@ucsd.edu

1 Introduction

A little bit more than two centuries ago, the French polymath Pierre-Simon Laplace wrote a tantalizing provocation about the awesome power of the emerging knowledge about mechanical systems. In his *Essai philosophique sur les probabilités* (1825), Laplace pondered about a being who, equipped with the deterministic equations that govern the movement of objects, would be capable of knowing the future precisely if provided with the position and forces governing all particles in the universe. When plugged into the equations that describe the mechanics of physical systems, this enormous yet simultaneously simple dataset would allow this supernatural force to predict any future state of the world. Whether the exact time when a flower blooms or the number of votes achieved by a candidate in an election, this demon knows everything set to happen based solely on data taken from a single moment and processed through enormous calculative force.

Laplace's demon is, of course, nothing but a stimulating thought. Predicting the future of any system is both practically and physically impossible: in addition to the sheer complexity of the calculations (even deterministic equations can get tricky and chaotic once reality kicks in!), constraints given by quantum indeterminacy — under standard interpretations — mean that conjoined knowledge of position and velocity is intrinsically unattainable. Even without this restriction, precision itself is an impossible target: small perturbations or measurement errors propagate over time, particularly in complex and chaotic systems, making long-term predictions—even about the fate of a single, humble particle — highly uncertain.

This dream of a deterministic, mechanistic universe that can be fully known through a combination of very specific parameters and the equations that describe their relations nevertheless continues to percolate the imaginaries of many scholars and, indeed, the public at large. Strong determinism may be recognizably rare, but its softer forms subsist in our collective imagination. We see this across numerous domains, from readings of genetics that assume a one-to-one correspondence between hereditary traits and complex individual behaviors, to economistic theories that reduce choice and valuation to fixed, computable preference functions — as if agents were themselves calculating engines optimizing a utility formula.

While these stronger and more evident variants of Laplacian determinism are admittedly uncommon, a softer form has found fertile ground at the intersection of computing and the study of society. Social scientists aren't exactly lauded for their forecasting abilities (predicting financial crashes, consumer demand, and sudden political shifts are three illustrations; The Forecasting Collaborative, 2023), yet the image of a predictive demon that, equipped with proper forms of measurement, enough data, and sufficient processing power, can foretell even the most complex collective human behaviors lurks in the shadows of research designs and framings of empirical puzzles throughout. The promise extends even to multifactorial, complex, almost intractable outcomes such as educational attainment, employment trajectories, and individual health, apparently all predictable with enough data and the right model (Chen et al., 2021; Leitgöb et al., 2023; Liu & Salganik, 2019; Salganik et al., 2019). Failures in forecasting should only "increase our efforts to rigorously measure and understand our ability — and inability — to predict the future" (Salganik, 2023). If we haven't hit the target, it may merely be a result of missing data or poor measurement. Perhaps, like the clouds above us, all that is necessary is enough data resolution and computing power to solve society's very own version of the Navier-Stokes equations that govern the mechanics of fluids. After all, hasn't physics already modeled complex social systems, representing financial actors as nodes in a crystalline structure, or social transformations as phase transitions described through the equations of statistical mechanics (Mantegna & Stanley, 1999; Mullick & Sen, 2025; Sen & Chakrabarti, 2014)? Almost like an echo of the psychohistory imagined in Isaac Asimov's (1951) *Foundation* series, in which a fictional mathematician develops equations capable of predicting the long-run trajectory of a galactic civilization through statistical aggregation, the belief that through the sum of its individuals, even the most complex and emergent facets of society are computationally tractable problems, inspires at least some strands of social scientific research that use computational methods to make sense of the world.

In what follows, I argue that the excitement around Machine Learning (ML) and Generative Artificial Intelligence (GenAI) as potential turning points for the social sciences rests on a category confusion — one that conflates the Laplacian imaginary of deterministic prediction with the Comtean reality of descriptive similitude (responding to the invitation that inspired this contribution). This confusion matters because it sets impossible benchmarks for what computational methods can deliver while obscuring the more modest but genuine contributions they afford. Three structural features of social scientific knowledge (the instability of its classifications, the indexicality of its claims, and the irreducibility of knowledge to information) set the epistemic limits within which ML/GenAI operate, regardless of how much data we collect, how powerful our algorithms become, or how impressive our transformers are.

Machine learning and Generative AI have grown into a new and exciting instrumental frontier capable of revealing patterns and relations unobservable otherwise — in a sense, fulfilling some of the promise of discerning the equations of society while solving them, a step closer to a social physics that, through empirical observation, can foretell the social world. I call these "instrumental" because, much like a mass spectrometer or a high-resolution scientific sensor on a telescope, they are an intermediary between a collection of inputs (data) and a series of outputs (classifications, clusters, graphs, or text).[1]

Much like other revolutions in instrumentation, the rise of ML/GenAI will invariably shift the social sciences in new directions. ML/GenAI aren't merely an extension of existing analytical practices but an expansion of capabilities beyond what previous approaches made possible. Sure, some of the social science research that uses machine learning boils down to a regression with a few independent variables produced through some computational reduction plugged into the design (cf. Boelart & Ollion, 2018). A growing number of researchers, however, has found in ML both a means for uncovering new patterns and data clusters of direct analytical relevance as well as extending theory in new directions (Abramson et al., 2018; Molina & Garip, 2019; Nelson, 2020).

Whether these computational innovations ushered a new phase of "maturity" for the social sciences is, nevertheless, a more difficult argument to make. New techniques of data analysis certainly invite an expansion of empirical fields and analytical capabilities, but it remains unclear whether they constitute a sufficiently stark epistemic break from longstanding social scientific paradigms.

1. Throughout this essay, "machine learning" and GenAI are used to cover both classical ML methods (decision trees, ensemble methods, support vector machines, operating primarily on tabular data; cf. Breiman, 2001) and contemporary deep learning, including transformer-based large language models. These differ substantially in architecture and interpretability: deep learning models operate in high-dimensional vector spaces, generating stochastic outputs that are semantically complex in ways that classical ML is not. The argument I make does not depend on that distinction. Both paradigms produce patterns of association rather than causal mechanisms, and what this essay addresses is the conflation of the two. Where the distinction becomes analytically relevant, particularly in the discussions of categorical stability and the information/knowledge problem, it is flagged in the text.

https://doi.org/10.60923/issn.1971-8853/24341

Some disciplines have, admittedly, radically changed their approach. The empirical turn in economics, largely based on the ability to analyze new sources and large amounts of data, substantively transformed training and scholarship, leading to a new wave of publications that are clearly incompatible with the standard *Econometrica* article from the 1970s. That much of this is framed by a very specific experimental logic, however, should make us pause in declaring an epistemic break — some of the ingredients of this empirical turn existed already in the labs of behavioral economists of the 1980s while data crunching has been a core practice of financial economics since its origins in the late 1960s and early 1970s (Banerjee, 2020; de Souza Leão & Eyal, 2019; Webber & Prouse, 2018).

In other fields, like sociology, machine learning and GenAI have expanded the scope and depth of analytical capabilities. In some cases, it unveiled associations that were invisible with previous theoretical and methodological lenses (Molina & Garip, 2019; Salganik et al., 2019). In others, it allowed testing theories at scale and in ways that were painstakingly difficult to conduct before — augmenting researchers' abilities to process textual and visual data that required individualized coding by a team of humans (Nelson, 2020). It has also created opportunities to lower data-collection costs by creating new "synthetic respondents and interlocutors" for surveys and interactive experiments (cf. Boelaert et al., 2025). And yet in others, it suggested new theoretical avenues to explore, developing the field in significantly new directions. And so, the question emerges: does this rising family of techniques bring us closer to the promise of a "social physics"?

2 Social Physics, Revisited

For Auguste Comte, social physics emerges from the study of social phenomena which, much like natural phenomena, he assumed to be regulated by "invariable laws" that are the ultimate targets of knowledge. The foundational statement of this positivist program, running from the opening leçon of the *Cours de philosophie positive* (1830–1842) through the explicit construction of social physics in volumes four through six, rejects as inaccessible all inquiry into first or final causes. In the positive stage of knowledge, Comte writes, "the human mind, recognizing the impossibility of obtaining absolute notions, gives up seeking the origin and destiny of the universe and the intimate causes of phenomena, to attach itself uniquely to discovering, through the well-combined use of reasoning and observation, their effective laws, that is, their invariable relations of succession and similitude".[2] The epistemic goal, accordingly, is not to expose the generative causes of phenomena but "only to analyze with exactitude the circumstances of their production, and to connect them to one another through normal relations of succession and similitude".[3] As Eşanu (2019) characterizes this mathematical ontology, positive science operates through "magnitudes that were either derived from observation and experimentation or computed from other magnitudes that were themselves derived from

2. "l'esprit humain reconnaissant l'impossibilité d'obtenir des notions absolues, renonce à chercher l'origine et la destination de l'univers, et à connaître les causes intimes des phénomènes, pour s'attacher uniquement à découvrir, par l'usage bien combiné du raisonnement et de l'observation, leurs lois effectives, c'est-à-dire leurs relations invariables de succession et de similitude" (Comte, 1830–1842, leçon I).

3. "seulement d'analyser avec exactitude les circonstances de leur production, et de les rattacher les unes aux autres par des relations normales de succession et de similitude" (Comte, 1830–1842, leçon I.).

observation and experimentation" (p. 280), [4] organized into a theoretical structure, not merely aggregated.

It is instructive to recall that Comte's program was conceived in explicit opposition to an earlier, more reductive attempt at social physics developed by the Belgian statistician Adolphe Quételet. In his *Sur l'homme et le développement de ses facultés, ou Essai de physique sociale* (1835), Quételet had sought to derive social laws from statistical regularities: the remarkable consistency of crime rates, suicide rates, and marriage patterns across populations suggested that social phenomena obey something like mechanical laws. The constancy of these phenomena, what Quételet described as their "régularité effrayante" ("terrifying regularity") seen in the year-upon-year reproduction of suicide and crime statistics, pointed to deterministic-seeming forces operating beneath the apparent freedom of individual behavior. His concept of the "*homme moyen*" (the "average man") served as the statistical pivot of this enterprise: a parameterized abstraction capturing the central tendencies of human behavior from which deviations could be measured and classified. Society, Quételet argued, "prepares the crime, and the guilty person is only the instrument by which it is executed":[5] aggregate regularities were the primary data of social physics.

When Comte discovered Quételet had appropriated the term "physique sociale" (the label Comte had used for his own project), he was sufficiently exercised to coin an entirely new word, *sociologie*, precisely to mark his methodological distance (Pickering, 1993). Comte's disagreement with Quételet was not about the legitimacy of observation or the value of empirical regularity: it was about theoretical ambition. Statistical regularities, for Comte, did not yet constitute social science. In leçon 47 of the *Cours,* where he introduced the neologism of *sociologie*, Comte defined his new science as "this complementary part of natural philosophy which relates to the positive study of the ensemble of fundamental laws proper to social phenomena".[6] The architecture of this science was structured around two domains: the *statique sociale* (the laws of coexistence and social order) and the *dynamique sociale* (the laws of succession and historical development). What Quételet offered, in Comte's judgment, were descriptions of aggregate behavior lacking the theoretical framework capable of organizing them into genuine knowledge, a distinction that maps with uncomfortable precision onto what ML/GenAI produce when left uncoupled from a recognizable theoretical scaffolding.

Comte's relations were explicitly non-causal. In his discussion of mechanics, the paradigmatic case of a positive science, whose method he then extended systematically to sociology, Comte elaborates: "We have nothing to do here with the causes or modes of production of motion, but only with the motion itself. Thus, we are not treating of physics, but of mechanics, forces are only motions produced or tending to be produced" (Turner et al., 2011). Forces

4. The passage derives from leçon 15 of volume 1 of the *Cours* (Quinzième Leçon, p. 394 in the 1869 edition), where Comte is characterizing the method of classical mechanics rather than social physics directly. Its relevance is methodological: throughout the *Cours*, mechanics serves as the paradigmatic case of positive science, the discipline that most clearly demonstrates that forces are descriptive relations between observable phenomena rather than causal agents. This is a principle that Comte then extends systematically to sociology in volumes 4–6. The full title of leçon 50 is indicative: "Considérations préliminaires sur la statique sociale, ou théorie générale de l'ordre spontané des sociétés humaines". For Comte's own programmatic account of social physics, see leçons 46–60; for an authoritative secondary treatment, see Pickering (1993), Comte (2018), Eșanu (2019).

5. "prépare le crime, et le coupable n'est que l'instrument qui l'exécute" (Quételet, 1835).

6. "[...] cette partie complémentaire de la philosophie naturelle qui se rapporte à l'étude positive de l'ensemble des lois fondamentales propres aux phénomènes sociaux" (Comte, 1830–1842, leçon 47).

do not cause movement but describe it, and mathematical relations become relations of similitude rather than of cause.

In this respect, Comte's social physics operates through this logic of similitude rather than causation: the relationships it identifies are descriptive regularities, not mechanisms. This is, I would argue, precisely what machine learning and generative AI actually produce: patterns of co-occurrence, probabilistic associations, clusters in high-dimensional space, even when their practitioners frame their findings in the causal language of Laplacian or Queteletian determinism (Brand et al., 2023; Grimmer, 2015). This slippage is consequential. When a model predicts incarceration rates from neighborhood characteristics, or educational attainment from genomic markers, the output is a Comtean relation dressed in Laplacian clothing. The model describes a regularity; the interpretation imputes a force. Indeed, the analogy runs deeper than it might initially appear. Machine learning and GenAI in social science reproduce, in updated algorithmic form, precisely the move Comte rejected in Quételet, that is, the aggregation of observed regularities without the theoretical architecture that would give them explanatory weight. The statistical patterns ML/GenAI uncover are, structurally, a twenty-first-century *homme moyen*: stable, reproducible, and theoretically inert until embedded in a framework that can explain not just what co-varies with what, but why. Understanding this gap is essential because it explains why, despite their technical sophistication, ML and GenAI rarely dislodge the theoretical frameworks they encounter; they confirm their predictive relevance without explaining their mechanisms, leaving the underlying paradigms largely intact. The fact that much research with ML happens under a framework of "normal social science" that is strongly worded in causal terms suggests that a convergence with something akin to a social physics in Comte's sense is unlikely without some fundamental epistemological shift.

3 Teleologies in Silico

More broadly, and tied to the provocation that sparked this commentary, it is unclear, at least to me, whether "maturation" is the best descriptor for our current entanglements with computational methods. Besides the more difficult teleological question of what a mature social science would look like (is it closer to Comte's social physics or to Laplace's deterministic mechanics? Maybe something yet unimagined?), there is the fact that, despite all the findings and contributions provided by machine learning, it has yet to herald a Copernican revolution, a quantum jump, or a new central dogma that radically transforms our understanding of the social world. Longstanding categories of sociological analysis remain relevant predictors of collective and individual outcomes. Classical constructs including race, class, gender, and ability continue to describe the forces shaping lives and institutions; the paradigms surrounding these have hardly budged. Even predictive tasks have not become better in general — with some potential exceptions that, however promising, seem slightly atheoretical (see Savcisens et al., 2024).

I admit to being agnostic about the potential of the new techniques of ML and GenAI: I see their adoption as important methodological innovations but do not assume that they will, in and of themselves, shift the field in any specific direction — that some of the discussion has now shifted to Generative AI after previous forms of ML having yet to materialize fundamental breakthroughs certainly raises skepticism (Bail, 2024). This is even so with the more impressive, agential models that are now capable of single-handedly writing quantitative social science papers through a single prompt (Engzell & Wilmers, 2026): although certainly awe inspiring, these systems — like any other Large Language Model — are unlikely to be able to transcend the patterns and biases of their training data, limiting their ability to truly surprise (cf. Waight et al., 2026).

Like with other instrumental revolutions, however, I recognize that the consolidation of computational social science *in general* may live up to ML/GenAI's promise of disciplinary renewal to some (modest) degree, although even this may take many years of organizational and practical change to bear fruit. Better measurement and new techniques of analysis have been foundational for epistemic breaks in the past — without advances in crystallography, knowledge about the double helix and the emergence of the central dogma in genetics would have occurred in very different ways, if at all. ML, GenAI, and new forms of computational analysis are also particularly well-suited to the deluge of machine-readable digital data that now permeates and shapes social life (Fourcade & Healy, 2024). I do not discount the chance that these technologies will shed new light on core questions and concepts (like better telescopes that provided greater access to the skies), perhaps even displacing or fundamentally altering the form of the regnant causal paradigm of social scientific explanation. And yet, certain features of social scientific knowledge, as well as of machine learning more generally, suggest that a rupture may require more than new techniques.

There are three specific characteristics of the social sciences and how they apprehend the world that, I believe, set an epistemic limit on what the new methods and technologies can perform. The first is the underlying importance of classifications for parsing and making sense of social and individual behaviors. Unlike colleagues in physics or chemistry, the constructs that social scientists deal with are particularly vexing. In Ian Hacking's terminology (Hacking, 1999), they are "interactive" kinds rather than "indifferent" kinds, prone to change by virtue of their use, and subject to what Hacking calls the looping effect, whereby the act of classification alters the classified. We may talk much about the mass of an object without it changing in physical characteristics; the same is not true of at least some social classifications. Unlike parameters that describe indifferent kinds, we have no "social classometers" to quantify an individual's social class, I often like to say, but just proxies of this broader conceptual entity that speak only partially to its critical, causal features. What kind of food we eat, drinks we consume, salary we make, and movies we watch are all expressions of this broader entity, but none is capable of fully standing in for the concept. These classifications and constructs are, too, highly variable across time and space. The way ethnicity and race are coded in a British census differs fundamentally from how they are approached in the United States. The racialized experiences of a person in Brazil are similarly quite different from those lived in Mexico, or Texas, or France. Race and racialization are, undoubtedly, global processes with a common history of violence and dispossession, but operate differently across countries, cultures, and times. The fact that ML/GenAI ultimately relies on classifications fixed in the data and tied to training sets that are partial to a subset of global experiences makes their revolutionary potential less plausible, at the very least for analyses at larger scales, given that capturing these multiple categorical realities is impractical. This is not merely a measurement problem amenable to better data collection. It is a constitutive feature of social categories that reflects their performative, self-referential character — the very thing that makes them sociologically interesting also makes them resistant to the kind of stable operationalization that CSS pipelines require. Trading zones between social and computational scientists regularly fracture on precisely this point: what counts as a valid operationalization of "class" or "race" is not a technical question but a theoretical and political one, and no amount of computational sophistication dissolves that.

A second constraint comes from the way data and claims are made sense of and given value. A slight variation of the Infinite Monkey Theorem provides some insight. Running for long enough time, a generator machine that prints out random sequences of alphanumeric characters will eventually produce text that is identical to a yet unwritten, paradigm-shifting publication (although this may take an admittedly inordinate amount of time). If an imagined social scientist were able to read and evaluate these outputs, they would likely be unable to recognize the significance of the text, no matter how revolutionary. It might seem *linguistically* reasonable. But paradigm-shifting? That evaluation requires a form of signification that emerges from a collective process of sense-making and designation (Barnes, 2003 & 2013; Barnes et al., 1996). More generally, data and the claims that are made with it are always indexical, tied to specific circumstances that grant meaning and value. However big or comprehensive the datasets may get, breaking with this indexicality is impossible. Claims and patterns are, likewise, the result of specific, embedded forms of collective sense-making. The output of an ML/GenAI system gains significance only after it is evaluated in relation to communally accepted knowledge — that is, framed in relation to a broader stock of claims that are shared by a community of practice. Indeed, even the most radical quantitative exercise is interpretative in the end. Machine learning and GenAI are not the exception.

What Science and Technology Studies scholars have begun to document is that this interpretive dimension is not merely a feature of how outputs are received but is structurally embedded in the training process itself: it's fundamentally structural. ML systems, for example, learn by minimizing error against labeled reference datasets, what practitioners call ground truths: collections of annotated cases that define what a correct answer looks like. Jaton (2021) shows that such datasets are the product of three constitutive, and decidedly non-neutral, choices: how the learning problem is defined, what data are collected and arranged, and how targets are labeled. Far from neutral benchmarks, ground truth datasets encode specific social judgments, from what categories matter and what counts as a correct classification to whose labeling practices carry authority; it is against these judgments that the algorithm's weights are adjusted, and its performance measured. Kang's (2023) concept of "ground truth tracings" extends this analysis by showing how qualitatively complex social phenomena (such as voice quality, emotional register, or professional fit) are ontologically flattened in the act of rendering them learnable: reduced, through successive translations, to the discrete, unambiguous labels that supervised learning requires. The ground truth is never given; it is made. And the social judgments embedded in its making are not subsequently available for correction by the algorithm; they are the very standard against which the algorithm's outputs are validated. In this sense, the indexicality of social knowledge penetrates ML/GenAI systems not only at the point of their interpretation but at the foundational moment of training.

This has a specific implication for the promise of scale. Enlarging the dataset does not reduce indexicality — it multiplies it. A corpus of a billion social media posts is not more transparent about its conditions of production than a corpus of a thousand. If anything, the larger the dataset, the more heterogeneous and opaque the circumstances under which its constituent elements were generated. The fantasy of the view from nowhere, still present in some forms of quantitative social science, is not resolved by computational power.

This connects to a third and final constraint. Behind the promise of prediction and the power of these new technologies, scholars often invoke the underlying algorithms as capable of "extracting" valuable knowledge from data by identifying previously unseen patterns and relationships. Much of this capacity stems from a persistent, historically deep conflation of "information" with "knowledge". As Day (2001) has documented, the pre-Shannonian sense of "information" in English carried explicitly semantic weight — information *informed*; it reduced uncertainty by contributing to a state of knowing. Shannon's formalization decisively severed that connection, defining information as a purely syntactic measure indifferent to meaning (Shannon, 1948). Yet the older semantic valence persists in everyday usage and, consequently, in the imaginaries surrounding machine learning and GenAI.

Knowledge is a collective, partly self-referring, self-validating social institution (Barnes, 1983) that provides the context that says whether a sequence of signs called "information" reduces or not our uncertainty about an outcome. This prior, collective knowledge is not reducible to a sequence of signs. Machine learning and GenAI, like other computational techniques, deal with information and are part of a system where, as Claude Shannon canonically wrote, the "semantic aspects of communication are irrelevant to the engineering problem" (Shannon, 1948). Data mining and pattern recognition aren't meaning-making until humans are looped into the evaluation — Knowledge Discovery in Databases is, for example, an abysmal misnomer. This is, perhaps, the most striking constraint we face: the disconnect between what machines can process and how they process it, and how collective sign-making practices work. The latter is not reducible to the former. There is no knowledge in information systems, and, by extension, no knowledge that can be "extracted" by ML/GenAI, pinning any eventual significance to the work performed within a broader epistemic milieu where they are used. This constraint has particular relevance for current discussions of generative AI as a vehicle for social scientific interpretation. Large language models operate at the level of information (they are extraordinarily sophisticated at identifying and reproducing patterns in signs), but they do not possess the collective, self-validating epistemic infrastructure that constitutes knowledge in Barnes's sense. They are token machines, not knowledge makers. Treating their outputs as theoretical contributions rather than as prompts for theoretical work repeats, in more elaborate form, the misnomer that plagued earlier claims about Knowledge Discovery in Databases.

These three constraints imply that these new technologies, which promise so much, are, in and of themselves, not a turning point for the social sciences as much as an opportunity to, perhaps, reflect on and transform practices and understandings of the world in the way disciplines regularly do so. In one concrete aspect, though, these technologies *are* particularly promising as instrumental innovations: they provoke discussions about methods that often involve interdisciplinary conversations (Sutherland, 2018) that ask us to consider the boundaries of our field, the nature of data, the standing of evidence, and the construction of facts. That these conversations are frequently the product of mass collaborative work (Garip, 2020) merely makes of these emerging trading zones (Galison, 1997) spaces that may, in due time, change the way we see, study, and intervene in the social world. What would it take, then, for them to constitute a genuine epistemic break rather than instrumental innovations? Three conditions seem necessary, none of which are currently in place. First, it would require new classificatory schemes that emerge from computational analysis itself — categories that are not simply operationalizations of pre-existing theoretical constructs but that generate novel, empirically grounded concepts with genuine theoretical purchase. Second, it would require epistemic communities capable of validating claims produced through ML/GenAI in ways that are not merely extensions of existing paradigmatic commitments, trading zones that do more than translate existing disputes into new vocabularies. Third, it would require a theoretical integration that connects the patterns identified to mechanisms capable of explaining, not merely describing, social dynamics. Until these conditions are met, the excitement around computational social science is best understood as the productive noise of a field in emergence: significant, consequential, but not yet revolutionary. The demon, for now, remains theoretical.

References

Abramson, C.M., Joslyn, J., Rendle, K.A., Garrett, S.B., & Dohan, D. (2018). The Promises of Computational Ethnography: Improving Transparency, Replicability, and Validity for Realist Approaches to Ethnographic Analysis. *Ethnography*, *19*(2), 254–284. https://doi.org/10.1177/1466138117725340

Asimov, I. (1951). *Foundation*. New York: Gnome Press.

Bail, C.A. (2024). Can Generative AI Improve Social Science? *Proceedings of the National Academy of Sciences*, *121*(21), e2314021121. https://doi.org/10.1073/pnas.2314021121

Banerjee, A.V. (2020). Field Experiments and the Practice of Economics. *American Economic Review*, *110*(7), 1937–1951. https:/doi.org/10.1257/aer.110.7.1937

Barnes, B. (1983). Social Life as Bootstrapped Induction. *Sociology*, *17*(4), 524–545. https://doi.org/10.1177/0038038583017004004

Barnes, B. (2003). Thomas Kuhn and the Problem of Social Order in Science. In T. Nickles (Ed.), *Thomas Kuhn* (pp. 122–141). Cambridge, UK: Cambridge University Press.

Barnes, B. (2013). *Scientific Knowledge and Sociological Theory*. London: Routledge.

Barnes, B., Bloor, D., & Henry, J. (1996). *Scientific Knowledge: A Sociological Analysis*. Chicago: The University of Chicago Press.

Boelart, J., & Ollion, E. (2018). The Great Regression: Machine Learning, Econometrics and the Future of Quantitative Social Sciences. *Revue française de sociologie*, *59*(3), 475–506. https://www.jstor.org/stable/48740414

Brand, J.E., Zhou, X., & Xie, Y. (2023). Recent Developments in Causal Inference and Machine Learning. *Annual Review of Sociology*, *49*, 81–110. https://doi.org/10.1146/annurev-soc-030420-015345

Chen, Y., Wu, X., Hu, A., He, G., & Ju, G. (2021). Social Prediction: A New Research Paradigm Based on Machine Learning. *The Journal of Chinese Sociology*, *8*(1). https://doi.org/10.1186/s40711-021-00152-z

Comte, A. (1830–1842). *Cours de philosophie positive* (6 vols.). Paris: Bachelier.

Comte, A. (1856). *Social Physics: From the Positive Philosophy of Auguste Comte* (H. Martineau, Trans.). New York: C. Blanchard. (Original work published 1830–1842)

Comte, A. (2018). *Écrits de jeunesse 1816–1828: Suivis du Mémoire sur la 'Cosmogonie'de Laplace, 1835* (Vol. 5). Berlin: Walter de Gruyter GmbH & Co KG.

Day, R.E. (2001). *The Modern Invention of Information: Discourse, History, and Power*. Carbondale, IL: Southern Illinois University Press.

de Souza Leão, L., & Eyal, G. (2019). The Rise of Randomized Controlled Trials (RCTs) in International Development in Historical Perspective. *Theory and Society*, *48*(3), 383–418. https://doi.org/10.1007/s11186-019-09352-6

Engzell, P., & Wilmers, N. (2026). The Paper Factory. [SocArXiv preprint]. https://doi.org/10.31235/osf.io/24xfq_v1

Eşanu, A. (2019). Auguste Comte and JS Mill on Physical Causes: The Case of Joseph Fourier's Analytical Theory of Heat. *Hopos: The Journal of the International Society for the History of Philosophy of Science*, *9*(2), 275–295. https://doi.org/10.1086/704373

Fourcade, M., & Healy, K. (2024). *The Ordinal Society*. Cambridge, MA: Harvard University Press.

Galison, P. (1997). *Image and Logic: A Material Culture of Microphysics*. Chicago: The University of Chicago Press.

Garip, F. (2020). What Failure to Predict Life Outcomes Can Teach Us. *Proceedings of the National Academy of Sciences*, *117*(15), 8234–8235. https://doi.org/10.1073/pnas.2003390117

Grimmer, J. (2015). We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together. *PS: Political Science & Politics*, *48*(1), 80–83. https://doi.org/10.1017/S1049096514001784

Hacking, I. (1999). *The Social Construction of What?* Cambridge, MA: Harvard University Press.

Jaton, F. (2021). Assessing Biases, Relaxing Moralism: On Ground-Truthing Practices in Machine Learning Design and Application. *Big Data & Society*, *8*(1). https://doi.org/10.1177/20539517211013569

Jaton, F. (2025). Examining Algorithms in the Light of their Ground Truth Datasets: Results, Objections, and Avenues of Reflection. *Digital Society*, *4*. https://doi.org/10.1007/s44206-025-00197-4

Kang, E. (2023). Ground Truth Tracings (GTT): On the Epistemic Limits of Machine Learning. *Big Data & Society*, *10*(1). https://doi.org/10.1177/20539517221146122

Leitgöb, H., Prandner, D., & Wolbring, T. (2023). Big Data and Machine Learning in Sociology. *Frontiers in Sociology*, *8*. https://doi.org/10.3389/fsoc.2023.1173155

Liu, D.M., & Salganik, M.J. (2019). Successes and Struggles with Computational Reproducibility: Lessons from the Fragile Families Challenge. *Socius: Sociological Research for a Dynamic World*, *5*. https://doi.org/10.1177/2378023119849803

Mantegna, R.N., & Stanley, H.E. (1999). *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge, UK: Cambridge University Press.

Laplace, P.S. (1825). *Essai philosophique sur les probabilités*. Paris: Bachelier.

Molina, M., & Garip, F. (2019). Machine Learning for Sociology. *Annual Review of Sociology*, *45*, 27–45. https://doi.org/10.1146/annurev-soc-073117-041106

Mullick, P., & Sen, P. (2025). Sociophysics Models Inspired by the Ising Model. *The European Physical Journal B*, *98*(9). https://doi.org/10.1140/epjb/s10051-025-01053-7

Nelson, L.K. (2020). Computational Grounded Theory: A Methodological Framework. *Sociological Methods & Research*, *49*(1), 3–42. https://doi.org/10.1177/0049124117729703

Pickering, M. (1993). *Auguste Comte: An Intellectual Biography, Vol. 1*. Cambridge, UK: Cambridge University Press.

Salganik, M.J. (2023). Predicting the Future of Society. *Nature Human Behaviour*, *7*(4), 478–479. https://doi.org/10.1038/s41562-023-01535-7

Salganik, M.J., Lundberg, I., Kindel, A.T., & McLanahan, S. (2019). Introduction to the Special Collection on the Fragile Families Challenge. *Socius: Sociological Research for a Dynamic World*, *5*. https://doi.org/10.1177/2378023119871580

Savcisens, G., Eliassi-Rad, T., Hansen, L.K., Mortensen, L.H., Lilleholt, L., Rogers, A., Zettler, I., & Lehmann, S. (2024). Using Sequences of Life-events to Predict Human Lives. *Nature Computational Science*, *4*(1), 43–56. https://doi.org/10.1038/s43588-023-00573-5

Quételet, A. (1835). *Sur l'homme et le développement de ses facultés, ou Essai de physique sociale*. Paris: Bachelier.

Sen, P., & Chakrabarti, B.K. (2014). *Sociophysics: An Introduction*. Oxford, UK: Oxford University Press.

Shannon, C.E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sutherland, M.E. (2018). Computational Social Science Heralds the Age of Interdisciplinary Science. *Springer Nature Research Communities*, October 1 Retrieved from: https://communities.springernature.com/posts/computational-social-science-heralds-the-age-of-interdisciplinary-science

The Forecasting Collaborative. (2023). Insights into the Accuracy of Social Scientists' Forecasts of Societal Change. *Nature Human Behaviour*, *7*(4), 484–501. https://doi.org/10.1038/s41562-022-01517-1

Turner, J.H., Beeghley, L., & Powers, C.H. (2011). *The Emergence of Sociological Theory*. London: Sage.

Waight, H., Yang, E., Yuan, Y. *et al.* (2026). State Media Control Influences Large Language Models. *Nature*. https://doi.org/10.1038/s41586-026-10506-7

Webber, S., & Prouse, C. (2018). The New Gold Standard: The Rise of Randomized Control Trials and Experimental Development. *Economic Geography*, *94*(2), 166–187. https://doi.org/10.1080/00130095.2017.1392235

Juan Pablo Pardo-Guerra – Department of Sociology, University of California San Diego (United States)
https://orcid.org/0000-0003-3192-1104 | jpardoguerra@ucsd.edu
https://pardoguerra.org/
Juan Pablo Pardo-Guerra is an Associate Professor in Sociology at the University of California San Diego (United States). Previously, he was an Assistant Professor in Sociology at the London School of Economics and Political Science (United Kingdom). Trained in Science and Technology Studies (University of Edinburgh, United Kingdom), his work focuses on infrastructures, knowledge, and markets.