Digital Society (2025) 4:44
https://doi.org/10.1007/s44206-025-00200-y

BRIEF COMMUNICATION

Unraveling the Regimes of Synthetic Data Metrics: Expectations, Ethics, and Politics

Louis Ravn[1] · Vassilis Galanos[2] · Matthew Archer[3] · Danielle Shanley[4]

Received: 30 October 2024 / Accepted: 29 April 2025

Abstract
Synthetic data - artificially produced data used for various data science tasks - have become the subject of intense scholarly interest, engendering both hope and hype in fields like machine learning (ML) and data privacy. In this commentary, we shed light on a little-studied facet of the emerging synthetic data landscape: their evaluation through the use of different quality measures, such as privacy, utility, and fidelity metrics. While these may seem highly technical, this commentary argues that evaluation metrics are inextricably linked to the expectations, ethics and politics of synthetic data. Situating synthetic data metrics within longer histories of data measurement in big data and ML discourses, we unfold a conceptualization of synthetic data metrics as *metrological regimes* which highlights the multifaceted ways in which they are implicitly and explicitly political. We put this concept to use by providing a three-fold preliminary analysis of metrics for the evaluation of synthetic tabular data: first, we outline the current *constitution* of synthetic data's metrological regimes around utility, privacy, and fidelity metrics; second, we highlight the *performativity* of these metrological regimes; that is, how they overshadow other crucial measures and enact quantifications of essentially contested concepts; and third, we emphasize the *fragility* of synthetic data's metrological regimes by pointing to the eruption of specific negotiations regarding which privacy metrics (not) to use for synthetic data evaluation. By foregrounding how metrics shape the expectations, ethics, and politics of synthetic data, this commentary underlines the need for their critical study.

Keywords Synthetic data · Evaluation metrics · Metrological regimes · Expectations · Data ethics · Data politics

Extended author information available on the last page of the article

Published online: 04 June 2025

1 Introduction

In recent years, there has been a growing interest in synthetic data, a concept denoting artificially produced data used for various data science tasks (Jordon et al., 2022). In addition to addressing privacy and confidentiality concerns, compelling arguments in favor of synthetic data include gains in cost efficiency, bias mitigation, and the ability to simulate rare or dangerous scenarios, all of which spur hopes for the use of synthetic data in such domains as finance (Assefa et al., 2020), healthcare (Vallevik et al., 2024), and even human rights research (IOM, 2024). By juxtaposing synthetic data with so-called "real-world" data (Ravn, 2025), proponents suggest that the rise of synthetic data signals a fundamental breakthrough with respect to the promises of data privacy and the optimization of machine learning (ML) systems. However, the imagined dichotomy between real-world data and synthetic data (see also Helm et al., 2024, p. 4) overlooks that so-called real-world data are only ever a highly mediated snapshot of what they represent; moreover, the "real world" is always present in the production of synthetic data (Ferrari & McKelvey, 2022; Steinhoff & Hind, 2025).

Crucially, then, synthetic data are more than a mere technical innovation. As a burgeoning critical literature on synthetic data attests, they are also performative of expectational, ethical, and political implications. Regarding expectations, synthetic data enact the view that the field of ML will be able to avert the issues posed by data scarcity in the age of data-intensive deep learning systems. This is associated with both an economic interest in cheap labeled data (Steinhoff, 2022) and a risk-related concern for non-personal yet diverse data (Jacobsen, 2023). These expectations, importantly, have ethical and political implications in the sense that they shape how synthetic data are used today: on the one hand, their inexpensiveness contributes to the automation of AI technologies (Steinhoff, 2022); on the other, their tailored producibility discursively renders ML systems risk-free (Jacobsen, 2023), foreclosing broader critical debates as a result (Helm et al., 2024). Building on this, our commentary offers a critical examination of synthetic data evaluation metrics, an often-overlooked component of the sprawling synthetic data landscape that is nevertheless central to the development of synthetic datasets and their expanding applications.

Various actors have devoted considerable efforts at the formulation of metrics to evaluate different qualities of synthetic data. This manifests in systematic review papers of synthetic data evaluation (e.g., Vallevik et al., 2024), quality reports made available by synthetic data firms (e.g., Gretel, n.d.), or even open-source toolboxes published on platforms like GitHub (e.g., Lautrup et al., 2024). These quality metrics are essential to the proliferation of synthetic data: while positive evaluations of the privacy of synthetic data undergird claims about the social, ethical, and political legitimacy of synthetic data, evaluations of fidelity corroborate claims regarding synthetic data's sufficient similarity to underlying "real-world" data. Utility metrics, moreover, are key to establishing the suitability of synthetic data for diverse downstream tasks. Taken together, these evaluation metrics have ethical repercussions in the sense that they contribute to determinations about which uses of synthetic data are "right" and "wrong;" they are political insofar as they help shape the ways in which synthetic datasets are assembled and deployed. While we do not question that the production and use of synthetic data should be accompanied by rigorous and suitable modes of evaluation, our ambition here is to advance a critical perspective on how current evaluation practices are already actively shaping the trajectories of synthetic data.

To that end, we approach synthetic data evaluation metrics as an emerging metrological regime, which Barry (2002, p. 281n14) defines as "a zone in which measurement has come to take relatively standardized forms." Although ongoing debates about the best way to measure the quality of synthetic data suggest that these metrics are at an early stage, the notion of metrological regimes nevertheless productively attunes us to the anti-political nature of metrics (Barry, 2002). Particularly, this underscores that debates about the technical aspects of metrics displace and can even foreclose critical discussions about the desirability and impacts of technoscientific practices more generally (Archer, 2024). Our contention is that the increasing focus on the technical aspects of synthetic data (centrally including utility, fidelity, and privacy metrics) forecloses more fundamental considerations of the social, environmental, and ethical implications of synthetic data's growing prominence. This critical investigation of the metrics used to evaluate synthetic data will bring more clarity to the central wager of this commentary: that these metrics are deeply entangled with the expectational, ethical and political dimensions of synthetic data, thereby speaking to this special issue's interest in the political dimensions of ML evaluation. Crucially, proposing the concept of metrological regimes as a lens for studies of the politics of ML evaluation has the central benefit of sensitizing us not only to those metrics as subjects of critical analysis in their own right, but also the central role these metrics play in the often-hidden politics of synthetic data and its numerous applications.

2 Histories of Data Measurement in Machine Learning

A growing body of literature in the past two decades has focused on the myriad ways in which uncritical faith in datasets, algorithms, and AI technologies has performative effects that shape decision-making. This is commonly based on the assumption that quantifying techniques, given voluminous amounts of data, can lead observers closer to a view from nowhere (Baumgartner et al., 2023; Boyd & Crawford, 2012; Gustavsson & Ljungberg, 2021; Striphas, 2015). The emergence of data- and algorithm-mediated life, this literature helps reveal, has been spurred by imaginaries of quantified selves and datafied ecologies, typically with a nod toward managerialist improvements of spaces, places, and the bodies that inhabit them.

The measurement of synthetic data's properties must therefore be seen in the light of a longer history of data measurement, one which foregrounds the relationship between metrics and power (Beer, 2016). In their review of data measurement practices, Mitchell et al. (2023, p. 3) observe that "the task of measuring data has been at the core of Machine Learning from the early days of the field;" for example, this has involved "goodness-of-data" (p. 3) metrics which serve the comparative evaluation of dataset quality or the measurement of bias within datasets. An important feature of dataset measurements in ML contexts, as Scheuermann et al. (2021, p. 13) show, is that they are both standardized and quantified, leading to their perception as "objective, reliable, and reproducible". This view of data metrics as objective representations, while widespread, fails to account for how they are embedded in socio-political dynamics that shape their materialization and use in decision-making processes.

While synthetic data evaluation metrics are connected to this genealogy, they also display idiosyncrasies. Most importantly, as Shanley et al. (2024) point out, synthetic data cannot be evaluated "in the same way that we would [evaluate] other forms of data" because synthetic data are "qualitatively different" in terms of risks: on one hand, advocates of synthetic data make optimistic claims about privacy protection that facilitate the expansion of ML applications; on the other hand, debates about the ethics of synthetic data tend to be lumped together with debates about the ethics of AI. This forecloses critical understandings of the important differences between diverse types of data and their uses. Indeed, the difference between synthetic data and other forms of data – although a question of degree rather than kind – is one of the core reasons for evaluating synthetic data with a unique set of quality metrics. As will be shown in more detail below, synthetic data metrics harbor close connections to their purposes of use. For example, while it may be the case that privacy has historically been a core motivation for using synthetic data (e.g., Drechsler & Haensch, 2024), privacy *metrics* are necessary from the perspective of developers and users of synthetic datasets in order to balance privacy concerns with functionality, efficiency, and other technical and financial considerations.

3 Synthetic Data as Additional Layer of Abstraction

Before rethinking synthetic data metrics, let us begin by visualizing our line of argument regarding the ontology of synthetic data in its complexity by way of a diagram. While sympathetic to the notion that synthetic data do indeed merit unique forms of evaluation, we crucially suggest understanding them as different in degree of computational manipulation relative to other forms of data, rather than as fundamentally different in kind. Specifically, synthetic data – despite constituting an additional layer of computational abstraction – remain grounded in the material world (Ferrari & McKelvey, 2022; Steinhoff & Hind, 2025).

The diagram below depicts what we perceive as degrees of abstraction between the material world (variously termed "reality," "world out there," "external world," etc.) and a sense-making observer. Between these two poles of reference there exist various intermediating nodes of distancing: (a) techniques for capturing data (more precisely, "capta" – i.e. "captured" units instead of "givens," as suggested by Jensen as cited in Becker, 1952, p. 278); (b) the datasets as units of analysis and the numerous methods employed for their analysis; (c) onto-epistemological frameworks (Barad, 2007), worldviews, and ideologies; and (d) institutional agendas and personal motivations that shape the process. All of these layers interact in multiple ways; for example, specific datasets may be selected to answer particular research questions influenced by political interests (Birhane et al., 2021). While all these dimensions add complexity to the observer's sense-making within the world, synthetic data – and the sociotechnical practices conditioning their production – constitute a further layer of abstraction from the material world. If one assumes the politics of data production, selection, and curation to be obscured by narratives of data objectivity, this obscura-

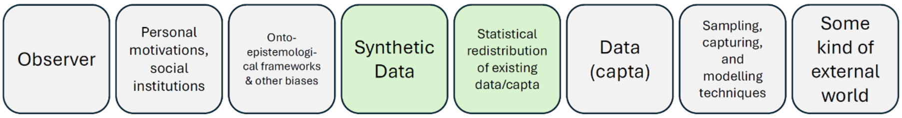


Fig. 1 Synthetic data as additional layer of abstraction

tion becomes even more pronounced through the abstractions performed by synthetic data (Fig. 1).

Following the aforementioned proposal for data as a misnomer for capta, it is worth considering Flusser's media-theoretical approach to "technical images" (as instances of the world captured and re-synethsized in computers). Thinking with Flusser (1985, pp. 42, 47–48), we suggest that data or datasets (as similar instances or snapshots of the world) represent a reversal of the traditional map-territory relationship, where our experiences in the world help us navigate the constructed realities of images or data rather than the other way around. In this framework, datasets are not mere representations of external realities, but projections that need to be understood for their intended (and very political) purpose, with their unintended biases, prompting us to decode their underlying programming rather than their apparently objective content. Instead of using the data to navigate and understand the world (a process already problematic in several ways), we now use our captured experiences in the world in order to navigate (and evaluate) synthetic data.

Understanding synthetic data as an additional layer of abstraction is conducive to the argument of this commentary for two main reasons. First, it allows us to conceive of synthetic data as different in degree rather than different in kind from other types of data; the main distinction lying in the extent to which they constitute an abstraction. Second, it highlights not only that synthetic data, like other forms of data, require forms of evaluation; most importantly, it helps explain why such metrics as utility, fidelity, and privacy dominate synthetic data evaluation (Vallevik et al., 2024): these measures, each in their own way, quantify synthetic data's levels of abstraction. While fidelity and utility metrics seek to ground synthetic data's abstractions as nevertheless corresponding to underlying realities (fidelity in relation to underlying data; utility in relation to 'downstream' ML tasks), privacy metrics conversely ensure that appropriate abstraction from underlying sensitive information thwarts any risk of re-identification.

4 Rethinking Synthetic Data Metrics as Metrological Regimes

Having conceptualized synthetic data as additional layers of mediation, we now articulate how the metrics for synthetic data evaluation may be seen as political. To that end, we suggest the notion of 'metrological regimes' (Archer, 2024; Barry, 2002) as a generative lens. Barry (2002, p. 276) follows Callon (2002) in observing a growing demand for the "politicization of the technological economy." The establishment of metrological regimes, zones "in which measurement has come to take relatively standardized forms" (Barry, 2002, p. 281n14), however, forestalls politicization by reducing the space of political debate. An example Barry provides is the measurement of exhaust fumes, which, although dependent on a range of actors (regulators, engineers, auto mechanics, etc.), leaves little room for debates about whether exhaust fumes are bad or should be regulated in the first place. Instead, what takes center stage is the improvement of methods for measuring these fumes, with politicians and other actors interested in depoliticizing the issue further pushing the institutionalization of these measures, for example through pricing mechanisms.

For Barry, metrological regimes are *inventive* in the sense that they create new objects of measurement; they are *fragile* in the sense that they are both always open to contestation and constantly being contested. Archer (2024) extends this through an analysis of corporate sustainability metrics, arguing that it is precisely the fragility of these regimes and the commensurative challenges they perpetually face that is, paradoxically, the main source of their resilience. The fragility of metrological regimes is the justification for never-ending attempts to improve measurement techniques – evaluation methods, indicator construction, etc. – which reinforces the idea that poor measurement techniques are the problem. This idea, however, leaves little room for more fundamental political negotiations around the desirability or sustainability of the things being measured in the first place.

Mobilizing the concept of metrological regimes here allows us to highlight three aspects of the current landscape of synthetic data metrics. First, it allows us to point to the emergence of a set of relatively standardized metrics in the field of synthetic data, which we will call the *constitution* of synthetic data's metrological regimes in Sect. 5. Second, the concept's sensitivity to *performativity* highlights the ways in which metrics may be said to do more than just measure. Concretely, as soon as specific metrics become dominant and assume the status of metrological regimes, they sideline and prohibit alternative objects of measurement from surfacing. These two aspects of synthetic data metrics approximate what Barry (2002) describes as the "inventive" dimension of metrological regimes: while *constitution* indicates the emergence of particular metrics for synthetic data evaluation, *performativity* highlights their consequences. Finally, the concept of *fragility* of metrological regimes points to their being subject to contestations, revealing their provisional character. The following section unfolds this conceptual elaboration of metrological regimes by successively analyzing the constitution, performativity, and fragility of synthetic data metrics.

5 Metrological Regimes of Synthetic Data: Constitution, Performativity, and Fragility

In this section, we present an exploratory reading (Amoore et al., 2023) of a range of materials related to synthetic data's metrological regimes. We highlight three aspects: first, the constitution of particular metrics within synthetic data's emerging metrological regime; second, the performative consequences of these metrics; and finally, how these metrics exhibit fragility. Hence, the structure of this analysis follows our above exposition of the concept of metrological regimes. While synthetic data take on different shapes (Offenhuber, 2024), we focus here on synthetic tabular data for the sake of specificity. Accordingly, we do not claim this to be an exhaustive analysis of synthetic data metrics; rather, this section exemplifies the unique sensibilities of the concept of metrological regimes by means of an exploratory analysis.

5.1 Constitution: The Stabilization of Synthetic Data Metrics

While synthetic data and metrics for their evaluation are still “a relatively new domain” (Sarmin et al., 2024, p. 2), an exploratory reading of several review articles, industry-provided reports, and open-source tools suggests a convergence around groupings of metrics around the utility, privacy, fidelity, and – albeit considerably less often – the fairness of synthetic data. This early convergence is confirmed by overview articles (e.g., Jordon et al., 2022), industry-provided tools (YData, n.d.), and publicly available evaluation frameworks (Santangelo et al., 2024).

While the metrological regimes of utility, privacy, fidelity, and fairness contain an extensive range of more specific sub-metrics (e.g., Vallevik et al., 2024), it is still helpful to outline what each of these evaluate in relation to synthetic data. First, utility (also called ‘usability’) metrics measure the extent to which a given synthetic dataset (generator) is useful in relation to a given downstream task; this is split into ‘global utility’ metrics where the task is not known, and ‘outcome-specific’ metrics where tasks are specified (Drechsler & Haensch, 2024). Second, privacy metrics provide an estimate of the extent to which the synthetic dataset differs from the data upon whose basis it was generated as well as the risk of re-identification (Vallevik et al., 2024). This includes both measures of the synthetic data’s statistical distance to the data they are meant to resemble as well as metrics to measure the possibility of extracting information about the real data from the synthetic data. Fidelity (also called resemblance or similarity metrics), on the other hand, measures the extent to which the statistical properties of synthetic data match the statistical properties and distributions of “real” data (Belgodere et al., 2024). Finally, fairness metrics measure both the bias inherent to synthetic data as well as how downstream predictions may perform with respect to sensitive communities (Belgodere et al., 2024).

5.2 Performativity: The Hierarchies of Synthetic Data Metrics

This brings us to the performativity of synthetic data’s metrological regimes. Barry (2002, p. 277) explains: “[m]etrology creates new objects that make a difference in the world”, which then have “performative and regulative consequences”.

To begin with, synthetic data’s metrological regimes have instituted the dominance of certain evaluation metrics over others. For instance, in their review paper, Vallevik et al. (2024, p. 4) highlight: “Comprising over 80% of the recorded metrics, the categories similarity and usability clearly dominated over privacy, fairness, and carbon footprint.”[1]. A strikingly similar tendency is identified by Kaabachi et al. (2025, p. 2): “95% of the studies […] included utility evaluations while only 46% […] of the studies claiming to employ synthetic data for preserving privacy, i.e., those that should evaluate privacy, conducted any privacy evaluation.”. These examples show that the

[1] Note: ‘similarity’ and ‘usability’, respectively, here refer to what have above been called fidelity and utility metrics.

emergence of metrological regimes produce certain hierarchies of evaluation, here related to an overemphasis on utility evaluations and a neglect of such variables as carbon emissions. While measures of carbon emissions should be taken with a grain of salt since they enact neoliberal understandings of sustainability (Archer, 2024), questions around the environmental implications of synthetic data are nonetheless highly pertinent in light of their increasing production through large language models (Veselovsky et al., 2023) whose energy intensity is staggering (Hogan & Lepage-Richer, 2024).

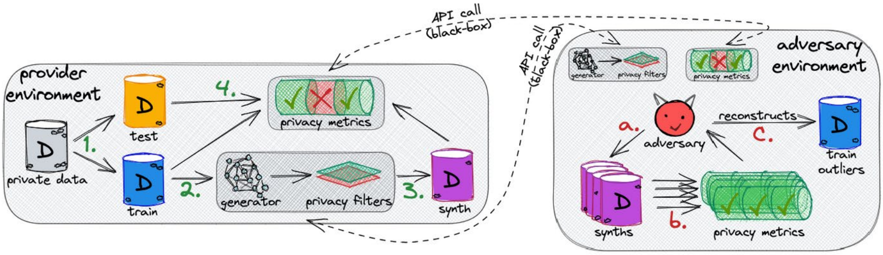
In addition to these processes of overshadowing, we may question the more fundamental ethicopolitics of evaluating some of these variables in their dominant quantified forms: what does it mean to measure the fairness of synthetic data quantitatively? While a growing willingness to consider fairness of synthetic data is laudable, what may be called into question is the way in which fairness is enacted through such metrics. More specifically, we argue that the risk with such quantitative evaluations is that they enact very particular understandings of 'essentially contested concepts' (Gallie, 1955; see also Orwat et al., 2024); in this case, an understanding of fairness as a one-off quantitative evaluation instead of a long-term engagement with the complex structural inequities often perpetuated through ML systems (e.g., Le Bui & Noble, 2020). Taken to its extreme, the complex ethicopolitical question of synthetic data's fairness here risks becoming a tick-box exercise under the assumption that the future of data is paved with syntheticity.

Moreover, one may raise the question if the imaginary of synthetic data as somehow inherently private shapes researchers' perceived need for privacy evaluation: "We found that most of the studies have utilized synthetic data "as is", assuming inherent privacy benefits without empirical verification" (Kaabachi et al., 2025, p. 2). This finding invites further study of the nexus between data expectations and data evaluation, particularly to the extent that methods for measuring these qualities of synthetic data relate to performative assessments in terms of those particular qualities.

5.3 Fragility: The Contestability of Synthetic Data Metrics

Finally, we would like to call attention to the fragility of synthetic data's metrological regimes. One particularly telling example is the case of privacy metrics, which we focus on here. As outlined above, privacy metrics constitute one crucial metrological regime of synthetic data evaluation. As part of the overarching category of privacy metrics there exist so-called "Similarity-Based Privacy Metrics" (SBPMs), a subcategory of metrics which, to simplify, aim at measuring "how close synthetic records are to their nearest neighbors in the train data", the idea being that "synthetic data should be similar and representative of the train data but not too close" (Ganev & De Cristofaro, 2023, p. 1). As the authors moreover detail, such SBPMs are the main set of metrics being utilized by several synthetic data startups as well as, sporadically, by researchers (Ganev & De Cristofaro, 2023).

Significantly, the authors demonstrate that SBPMs are deeply flawed in several ways. Among other problems, SBPMs do not provide theoretical guarantees, treat privacy as a binary rather than continuous variable, and approach privacy as a prop-



2: Overview of *ReconSyn*. The provider **1.** splits the real data into train/test, **2.** fits a generative model on the train data, **3.** generates synthetic data (privacy filters are applied), **4.** runs the privacy metrics on the synthetic data. The adversary can make API calls (they have black-box access) to the fitted generative model and privacy metrics. They **a.** generate synthetic datasets, **b.** run them through the privacy metrics to observe the pass/fail tests and scores (if tests pass), **c.** reconstruct underrepresented train records (outliers) through *SampleAttack* and *SearchAttack* (introduced in Algorithm 1).

Fig. 2 Ganev and De Cristofaro (2023) exposing fragility of SBPMs

erty of data rather than the generation process. Ganev and De Cristofaro (2023, p. 8) moreover provide a practical demonstration of the flaws of SBPMs, designing a successful reconstruction attack (see Fig. 2) that manages to recover some training data outliers as well as “to single individuals out and enable their identification or link them to the real data”. Further articulating the risks of SBPMs, Kaabachi et al. (2025, p. 3) argue that these “offer a false sense of security regarding the privacy-preserving capabilities of synthetic data”. Ganev and De Cristofaro (2023) go as far as suggesting that SBPMs to evaluate synthetic data do not suffice to consider them GDPR-compliant.

While this dedicated reconstruction attack aptly illustrates the *fragility* of synthetic data’s metrological regimes, one should not stop here: the crucial point is that individual privacy protection does not guarantee collective privacy – which in turn shapes individual privacy. As Orwat et al. (2024, p. 10) argue: “recent approaches that involve quantifying and measuring privacy by means of privacy metrics […] such as k-anonymity or differential privacy are based on a narrow understanding of privacy as anonymity, secrecy or the confidentiality of systems”; in short, they are “an inappropriate simplification” of privacy (see also Helm et al., 2024). Put simply, even when privacy measures seem robust, it remains vital to scrutinize whether and how the release of seemingly private synthetic data about specific collectives interacts with the privacy of the individuals thus represented.

Even further, the fragility of synthetic data can be understood as a “double bind,” to borrow a term from Bateson (1972). A double bind is a set of contradictory commands (or promises) that given sufficient repetition and the condition that both have to be executed or believed can lead to schizophrenic traits. Galanos (2024) has argued that public AI discourse (simultaneously utopian and dystopian) acts as a double bind at the collective level, resulting in the overall generalized confusion, existential alarmism, and simultaneous mandates for increased investment in AI. At a micro-scale, this approach can be seen as conducive to synthetic data promissory work: simultaneous fidelity and robust privacy – the miraculous bridging of an inscrutable trade-off as the ultimate selling point of simulated realities that support corporate interests but do not prevent the stereotyping of individuals as members of collectively assessed groups.

While these insights into the fragility of particular synthetic data metrics are essential, a response focusing on proposing new and "better" metrics alone – although laudable – would necessarily be insufficient. Specifically, it would paradoxically run the risk of precisely fueling the resilience of synthetic data's emerging metrological regimes, analogous to Archer's (2024) analysis of the stabilization of neoliberal sustainability metrics. Nevertheless, highlighting the fragility of particular synthetic data metrics productively reveals their revisability. This serves as a critical reminder that metrological regimes, although seemingly solid, are in a continual process of becoming.

6 Conclusion

In this commentary, we have provided a critical perspective on how synthetic data metrics are entangled with the broader expectations, ethics, and politics of synthetic data. We argued that synthetic data metrics, rather than exclusively conform to their seemingly technical character, can productively be seen as metrological regimes. This allowed us to be analytically sensitive to the constitution (dominance of specific metrics), performativity (the effects of dominant metrics), and fragility (contestablity of metrics) of synthetic data's metrological regimes. Drawing on this conceptual apparatus, our exploratory analysis highlighted three aspects of the emerging metrological regimes of synthetic tabular data: that there is an early convergence around utility, privacy, and fidelity metrics; that a focus on these metrics performatively overshadows other questions and enacts highly quantified notions of essentially contested concepts; and that these metrics remain open to contestation and negotiation. Crucially, these metrological regimes significantly shape the trajectories of synthetic data: they induce expectations regarding the legitimacy of synthetic data use cases, reconfigure ethics as a question of quantified measures, and depoliticize debates about the adequacy and sustainability of synthetic data.

In outlining this approach, we hope to have contributed a new angle to both social studies of synthetic data (de Vries, 2020; Steinhoff, 2022; Jacobsen, 2023; Helm et al., 2024; Offenhuber, 2024) as well as research into the politics of ML evaluation (Raji et al., 2021; Hutchinson et al., 2022; Luitse et al., 2024). We also aim to bring together these two currently disjointed lines of inquiry: as a still nascent data technology, the case of synthetic data provides unique insights into how evaluation metrics shape the early stages of a technology's trajectory in expectational, ethical, and political ways. Most importantly, we suggest that the lens of metrological regimes can inspire future studies into the multi-faceted politics of ML evaluation both by foregrounding which metrics assume dominance in particular domains as well as by inviting a crucial question: which social, ethical, and political debates become overlooked, sidelined, or excluded as legitimate matters of concern and care?

Author Contributions LR: Conceptualization, Writing, Visualization – Original Draft & Review/Editing; VG: Conceptualization, Writing, Visualization – Original Draft & Review/Editing; MA: Conceptualization, Writing – Original Draft & Review&/Editing; DS: Conceptualization, Writing – Original Draft & Review/Editing.

Data Availability Not applicable since this is a "brief communication" article for which no original data were produced.

Declarations

Conflict of Interest On behalf of all authors, the corresponding author declares that there is no conflict of interest.

References

Amoore, L., Campolo, A., Jacobsen, B., & Rella, L. (2023). Machine learning, meaning making: On reading computer science texts. *Big Data & Society*, *10*(1). https://doi.org/10.1177/20539517231166887

Archer, M. (2024). *Unsustainable: Measurement, reporting, and the limits of corporate sustainability.* New York University Press.

Assefa, S. A., Dervovic, D., Mahfouz, M., Tillman, R. E., Reddy, P., & Veloso, M. (2020). Generating synthetic data in finance: Opportunities, challenges and pitfalls. *ACM International Conference on AI in Finance*, 1–8. https://doi.org/10.1145/3383455.3422554

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.

Barry, A. (2002). The anti-political economy. *Economy and Society*, *31*(2), 268–284.

Bateson, G. (2000). *Steps to an ecology of mind*. University of Chicago Press. (Original work published 1972)

Baumgartner, R., Arora, P., Bath, C., Burljaev, D., Ciereszko, K., Custers, B., & Williams, R. (2023). Fair and equitable AI in biomedical research and healthcare: Social science perspectives. *Artificial Intelligence in Medicine*, *144*, 1–9.

Becker, H. (1952). Science, culture, and society. *Philosophy of Science*, *19*(4), 273–287.

Beer, D. (2016). *Metric power*. Palgrave Macmillan. https://doi.org/10.1057/978-1-137-55649-3

Belgodere, B., Dognin, P., Ivankay, A., Melnyk, I., Mroeuh, Y., Mojsilovic, A., Navratil, J., Nitsure, A., Padhi, I., Rigotti, M., Ross, J., Schiff, Y., Vedpathak, R., & Young, R. A. (2024). Auditing and generating synthetic data with controllable trust trade-offs. *arXiv*.

Birhane, A., Prabhu, V. U., & Kahembwe, E. (2021). Multimodal datasets: Misogyny, pornography, and malignant stereotypes. *arXiv*.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication & Society*, *15*(5), 662–679.

Callon, M. (2002). Technology, politics And the market: An interview with Michel callon. *Economy and Society*, *31*(2), 285–306.

de Vries, K. (2020). You never fake alone. Creative AI in action. *Information, Communication & Society*, *23*(14), 2110–2127.

Drechsler, J., & Haensch, A-C. (2024). 30 Years of synthetic data. *Statistical Science*, *39*(2), 221–242.

Ferrari, F., & McKelvey, F. (2022). Hyperproduction: A social theory of deep generative models. *Distinktion: Journal of Social Theory*, *24*(2), 338–360.

Flusser, V. (2011). *Into the universe of technical images* (N. A. Roth, Trans.). University of Minnesota Press. (Original work published 1985).

Galanos, V. (2024). Between monstrous dystopias and policy utopias in artificial intelligence: Reporting on a ten-year journey into the AI-topia. In V. A. Bruno, A. Campati, P. Caerlli, & A. Sfardini (Eds.), *Dystopian worlds beyond storytelling: Representations of dehumanized societies in literature, media, and political discourses: Multidisciplinary perspectives* (pp. 51–68). ibidem-Verlag.

Gallie, W. B. (1955). Essentially Contested Concepts. *Proceedings of the Aristotelian Society, 56*(1), 167–198.

Ganev, G., & De Cristofaro, E. (2023). On the inadequacy of similarity-based privacy metrics: Reconstruction Attacks against "truly anonymous synthetic data." *arXiv*.

Gretel. (n.d.). Synthetic data quality. Available at: https://docs.gretel.ai/optimize-synthetic-data/evaluate/synthetic-data-quality-report. [Accessed: 09/09/2024].

Gustavsson, M., & Ljungberg, J. (2021). Algorithms and their work: A performativity perspective. *12th Scandinavian Conference on Information Systems, 7.*

Helm, P., Lipp, B., & Pujadas, R. (2024). Generating reality and Silencing debate: Synthetic data as discursive device. *Big Data & Society*, *11*(2). https://doi.org/10.1177/20539517241249447

Hogan, M., & Lepage-Richer, T. (2024). *Extractive AI*. Centre for Media, Technology, and Democracy. Available at: https://static1.squarespace.com/static/5ea874746663b45e14a384a4/t/6664bcdf2cbd2d1d11ba6c03/1717877983870/CJT_Hogan+%26+Lepage-Richer.pdf. [Accessed: 11/09/2024].

Hutchinson, B., Rostamzadeh, N., Greer, C., Heller, K., & Prabhakaran, V. (2022). Evaluation gaps in machine learning practice. *arXiv*.

International Organization for Migration (IOM). (2024, February 27). IOM releases the Global Synthetic Dataset. Available at: https://migrantprotection.iom.int/en/spotlight/articles/publication/iom-releases-global-synthetic-dataset. [Accessed: 09/09/2024].

Jacobsen, B. N. (2023). Machine learning and the politics of synthetic data. *Big Data & Society*, *10*(1), 1–12.

Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., Cohen, S. N., & Weller, A. (2022). Synthetic Data – what, why and how? *arXiv*.

Kaabachi, B., Despraz, J., Meurers, T., Otte, K., Halilovic, M., Kulynych, B., Prasser, F., & Raisaro, J. L. (2025). A scoping review of privacy and utility metrics in medical synthetic data. *npj Digital Medicine*, *8*(1), 1–9.

Lautrup, A. D., Hyrup, T., Zimek, A., & Schneider-Kamp, P. (2024). SynthEval: A framework for detailed utility and privacy evaluation of tabular synthetic data. *arXiv*.

Le Bui, M., & Noble, S. U. (2020). We're missing a moral framework of justice in artificial intelligence: On the limits, failings, and ethics of fairness. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics of AI* (pp. 162–179). Oxford University Press.

Luitse, D., Blanke, T., & Poell, T. (2024). AI competitions as infrastructures of power in medical imaging. *Information, Communication & Society*, 1–22. https://doi.org/10.1080/1369118X.2024.2334393

Mitchell, M., Luccioni, A. S., Lambert, N., Gerchick, M., McMillan-Major, A., Ozoani, E., Rajani, N., Thrush, T., Jernite, Y., & Kiela, D. (2023). Measuring data. *arXiv*.

Offenhuber, D. (2024). Shapes and frictions of synthetic data. *Big Data & Society*, *11*(2). https://doi.org/10.1177/20539517241249390

Orwat, C., Bareis, J., Folberth, A., Jahnel, J., & Wadephul, C. (2024). Normative challenges of risk regulation of artificial intelligence. *NanoEthics*, *18*(11), 1–29.

Raji, I. D., Bender, E. M., Paullada, A., Denton, E., & Hanna, A. (2021). AI and the everything in the whole wide world benchmark. *arXiv*.

Ravn, L. (2025). The fabrication of synthetic data promises: Tracing emerging arenas of expectations and boundary work. *Big Data & Society*, *12*(1). https://doi.org/10.1177/20539517241307915

Santangelo, G., Nicora, G., Bellazzi, R., & Dagliati, A. (2024). SynthCheck: A dashboard for synthetic data quality assessment. *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technology*, *2*(1), 246–256.

Sarmin, F. J., Sarkar, A. R., Wang, Y., & Mohammed, N. (2024). Synthetic data: Revisiting the privacy-utility trade-off. *arXiv*.

Scheuermann, M. K., Denton, E., & Hanna, A. (2021). Do datasets have politics? Disciplinary values in computer vision dataset development. *Proceedings of the ACM on Human-Computer Interaction, 5*, 1–37.

Shanley, D., Hoogenboom, J., Lysen, F., Wee, L., Gomes, A. L., Dekker, A., & Meacham, D. (2024). Getting real about synthetic data ethics: Are AI ethics principles a good starting point for synthetic data ethics? *EMBO Reports*, *25*(5), 2152–2155. https://doi.org/10.1038/s44319-024-00101-0

Steinhoff, J. (2022). Toward a political economy of synthetic data: A data-intensive capitalism that is not surveillance capitalism? *New Media & Society*, *26*(6), 3290–3306.

Steinhoff, J., & Hind, S. (2025). Simulation and the reality gap: Moments in a prehistory of synthetic data. *Big Data & Society*, *12*(1).

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

Vallevik, V. B., Babic, A., Marshall, S. E., Elvatun, S., Brøgger, B., Alagaratnam, S., Edwin, B., Veeraragavan, N. R., Befring, A. K., & Nygård, J. F. (2024). Can I trust my fake data - A comprehensive quality assessment framework for synthetic tabular data in healthcare. *International Journal of Medical Informatics*, *185*, 1–20.

Veselovsky, V., Ribeiro, M. H., & West, R. (2023). Artificial artificial artificial intelligence: Crowd workers widely use large language models for text production tasks. *arXiv*.

YData. (n.d.). Synthetic data quality metrics. Available at: https://ydata.ai/synthetic-data-quality-metrics. [Accessed: 23/08/2024].

Authors and Affiliations

Louis Ravn[1] · Vassilis Galanos[2] · Matthew Archer[3] · Danielle Shanley[4]

✉ Louis Ravn
l.ravn@uva.nl

Vassilis Galanos
vassilis.galanos@stir.ac.uk

Matthew Archer
m.archer@maastrichtuniversity.nl

Danielle Shanley
danielle.shanley@maastrichtuniversity.nl

1 Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, Netherlands

2 Stirling Management School, University of Stirling, Stirling, Scotland

3 Department of Society Studies, University of Maastricht, Maastricht, Netherlands

4 Department of Philosophy, University of Maastricht, Maastricht, Netherlands