Surveillance & Society

Why Synthetic Data Can Never Be Ethical: A Lesson from Media Ethics

Andrew Fitzgerald

Volume 22, numéro 4, 2024

Open Issue

URI : https://id.erudit.org/iderudit/1115680ar
DOI : https://doi.org/10.24908/ss.v22i4.18324

Aller au sommaire du numéro

Éditeur(s)

Surveillance Studies Network

ISSN

1477-7487 (numérique)

Découvrir la revue

Citer ce document

Fitzgerald, A. (2024). Why Synthetic Data Can Never Be Ethical: A Lesson from Media Ethics. *Surveillance & Society*, *22*(4), 477–482. https://doi.org/10.24908/ss.v22i4.18324

Résumé de l'article

This Dialogue paper argues that the use of synthetic data can never be "ethical." My argument imports a normative stance from media ethics that "being-ethical-means-being-accountable" (Glasser and Ettema 2008). Building from discourse ethics, this stance positions such ethics as having "the facility to argue articulately and deliberate thoughtfully about moral dilemmas, which in the end means being able to justify, publicly and compellingly, their resolution" (Glasser and Ettema 2008: 512). Crucially, this approach is dialogical and social, necessitating a space open to all affected by relational practices and processes. While the use of synthetic data in commercial institutional contexts may offer workarounds to privacy concerns regarding personally identifiable information (PII) or unpaid user labor—or seem relatively innocuous, as in the case of training computer vision algorithms in video games—this facilitates, as others in this Dialogue section argue in their respective papers, a "fix" or "solutionist" framing that elides ethics and de-politicizes synthetic data. Synthetic data therefore intensifies a pre-existing lack of accountability inherent within automated systems more generally, and through this, entrenches and compounds surveillant practices. In some arenas, the stakes are quite literally life or death, such as in the development of medical AI, and more perniciously, the migration of models from commercial to state deployment in law enforcement and military contexts. Given the foreclosure of thoughtful, articulate, and reflexive inclusive deliberation on the significant moral implications of AI's vast and ever-growing assemblages, and synthetic data's role in further mystifying and legitimating its seemingly unbridled development and deployment, I argue that synthetic data can never meet the standard of "ethical" practice.

Surveillance & Society

Dialogue

Why Synthetic Data Can Never Be Ethical: A Lesson from Media Ethics

Andrew Fitzgerald

Rensselaer Polytechnic Institute, USA
Fitzga2@rpi.edu

Abstract

This Dialogue paper argues that the use of synthetic data can never be "ethical." My argument imports a normative stance from media ethics that "being-ethical-means-being-accountable" (Glasser and Ettema 2008). Building from discourse ethics, this stance positions such ethics as having "the facility to argue articulately and deliberate thoughtfully about moral dilemmas, which in the end means being able to justify, publicly and compellingly, their resolution" (Glasser and Ettema 2008: 512). Crucially, this approach is dialogical and social, necessitating a space open to all affected by relational practices and processes. While the use of synthetic data in commercial institutional contexts may offer workarounds to privacy concerns regarding personally identifiable information (PII) or unpaid user labor—or seem relatively innocuous, as in the case of training computer vision algorithms in video games—this facilitates, as others in this Dialogue section argue in their respective papers, a "fix" or "solutionist" framing that elides ethics and de-politicizes synthetic data. Synthetic data therefore intensifies a pre-existing lack of accountability inherent within automated systems more generally, and through this, entrenches and compounds surveillant practices. In some arenas, the stakes are quite literally life or death, such as in the development of medical AI, and more perniciously, the migration of models from commercial to state deployment in law enforcement and military contexts. Given the foreclosure of thoughtful, articulate, and reflexive inclusive deliberation on the significant moral implications of AI's vast and ever-growing assemblages, and synthetic data's role in further mystifying and legitimating its seemingly unbridled development and deployment, I argue that synthetic data can never meet the standard of "ethical" practice.

Introduction

With each wave of the artificial intelligence hype cycle, a slide from a 1979 presentation at IBM makes the rounds on social media. The slide states:

"A COMPUTER CAN NEVER BE HELD ACCOUNTABLE

THEREFORE A COMPUTER MUST NEVER MAKE A MANAGEMENT DECISION"

Ironically, IBM is one of the leading firms pushing artificial intelligence today, and is a major proponent of synthetic data—artificially produced data used to train machine learning models when the costs of real data are prohibitive, not enough real data exists, or there are concerns about privacy (Nikolenko 2021). While the use of synthetic data may offer workarounds to privacy concerns regarding personally identifiable information (PII) or unpaid user labor, synthetic data intensifies a pre-existing lack of accountability inherent within automated systems more generally. Importing a normative stance from media ethics that "being-ethical-means-being-accountable" (Glasser and Ettema 2008), this paper argues that the use of synthetic data can never be "ethical."

Fitzgerald, Andrew. 2024. Why Synthetic Data Can Never Be Ethical: A Lesson from Media Ethics. *Surveillance & Society* 22 (4): 477-482.
https://ojs.library.queensu.ca/index.php/surveillance-and-society/index | ISSN: 1477-7487

Being-Ethical-as-Being-Accountable vs. "Ethical AI"

This approach to media accountability contrasts with dominant trends (and critiques) within the rapidly emerging field of "Ethical AI." It views "ethics as a process, not merely an outcome; an argument, not merely a choice" (Glasser and Ettema 2008: 512), and builds on a concept of "discourse" or "communicative ethics" popularized by Jürgen Habermas (1993). In brief, this approach is dialogical and social—in contrast to say, Kant's ethics (Kant 2002) where an individual can solely justify an action—and necessarily opens a space "for public discussion open to everyone affected" by relational practices and processes (Glasser and Ettema 2008: 512).

In this view, "accountability" is not merely retributive. Instead, being-ethical-as-being-accountable positions ethics as having "the facility to argue articulately and deliberate thoughtfully about moral dilemmas, which in the end means being able to justify, publicly and compellingly, their resolution" (Glasser and Ettema 2008: 512). While this does not rule out the possibility of sanctions for one's actions resulting from such debate, accountability "means that one can be constrained to reveal what one has done and why one has done it; thus, the action and the reasons for it are open to a critique by strangers who have few inhibitions about demanding justification and reasonable grounds" (Gouldner 1976: 102; see also Glasser and Ettema 2008: 513). Ultimately, such ethics also demand reflexivity and intersubjectivity in relation to one's interlocutors—those asking the "what" and "why"—and a degree of epistemological accessibility to this information for affected parties.

Communicative ethics grounds Glasser and Ettema's (2008) approach to accountability in the case of journalism. This is rather distinct from "AI Ethics" to date, and my adoption here also varies productively from critiques of "AI Ethics" grounded in AI justice (Heilinger 2022) or human rights frameworks (Fukuda-Parr and Gibbons 2021), without ruling out their merits. In contrast to communicative media ethics and critical approaches to AI, applied AI ethics overwhelmingly operates "from an individualist framework of ethical responsibility" (Asaro 2023: 675). It often entails vague procedures for "stakeholder input" or voluntary "ethics guidelines" that disproportionately come from corporations themselves, aimed at mitigating harm and holding "responsible parties" accountable when (that is, after) harm has occurred (Fukuda-Parr and Gibbons 2021; Maas 2022). While machine learning literature on synthetic data tends to skirt robust discussion of ethics—an entire scholarly monograph (Nikolenko 2021) dedicated to the topic mentions "ethics" merely three times—the companies pitching synthetic data to their clients often appeal to virtue ethics (Vallor 2016), namely "trust" in automated systems trained using synthetic data.

The Pitch for "Trusting" Synthetic Data

As Ravn's Dialogue paper (in this issue) argues, the deployment of synthetic data enacts "differential implications depending on the surveillant assemblage (Haggerty and Ericson 2003) into which they become imbricated and which they reconfigure." Synthetic data is increasingly deployed in numerous professional fields and institutional contexts, particularly finance and medicine due to the sensitive nature of the data, or where the costs to acquire real data would be prohibitive. Synthetic data is also widely utilized in the development of "computer vision," where manual labeling for training machine learning algorithms is extremely difficult and labor intensive—computer vision models trained on synthetic data are often then used for applications like automated fraud detection or autonomous vehicles and robots (Nikolenko 2021). This extends and intensifies surveillant practices, but does so, as Ridgway and Malevé note in their essay (in this issue), by "modeling and predicting futures without the physical presence of objects." Evangelists tout benefits including cheap data that are easy to produce, are "pretty good," and allow for one to "perfectly label" the artificial data without the messiness of real life; such benefits will allegedly be so substantial and fiscally competitive that IBM predicts that, by 2025, AI will need 70% less real data than in 2023 (IBM Technology 2023). Ridgway and Malevé (in this issue) further highlight how this "fix" framing allows synthetic data's rapid uptake to elude ethical scrutiny while also legitimating a broader ontological shift

from representation and indexicality (within stabilizing assemblages of annotation and classification) to simulation—a shift affecting both the "output" of ML algorithms using synthetic data and the "human user who needs to appraise critically the results."

For commercial AI companies, public facing communication—for example, YouTube videos from IBM about synthetic data (IBM Technology 2023, 2024)—encourages a multistep procedure for building "trust" in one's synthetic data before "domain transfer," that is the redeployment of models trained on synthetic data to operate on real data, from which automated systems will take action in a "completely different domain" (Nikolenko 2021: vi). One suggested step is having a "preliminary deployment" stage, limiting the ability to generate and use such data to a select group of people that coordinate quality standards with "privacy and risk" teams before pushing the model to cloud for downstream use in real contexts by real clients. While this may assuage concerns about PII or first-order validity relative to existing non-synthetic datasets, the inability to trace the causal influence of such synthetic data once it is fed into an AI model leads to a further mystification of algorithmic decisions beyond the standard critique of the "black box," and intensifies anthropocentric illusions inherent to automated systems baked into the pipeline—adding another layer of non-reflexive, non-discursive process that can never achieve an ethics of accountability.

Mystification Beyond the Black Box: Anthropocentric Illusions of Accountability

In some ways, the debate appears to be a relitigation of algorithmic decision making more generally, including the critique of algorithmic "black boxes" in applied computational intelligence (Pasquale 2015). The term "black box" is often used as a catchall for what Burell (2016) notes are in fact three distinct forms of opacity—intentional corporate or state secrecy, technical literacy/illiteracy, and the nature of machine learning algorithms where the amount of data and the speed of processing is beyond human comprehension. The spread of these opacities has prompted calls for accountability mechanisms such as "algorithmic audits" (Costanza-Chock, Raji, and Buolamwini 2022; Raji et al. 2020), particularly in the context of algorithmic bias tied to racial (Benjamin 2019), gender (O'Connor and Liu 2023), and class inequalities (Eubanks 2018). Given the degree of intentional opacity around the deployment of algorithmic decision making, algorithmic "transparency" was also a much-touted approach, mainly from a virtue ethics perspective where, much like we might "trust" a company or algorithm, said company, team, or product might hold the virtue of being "transparent."

As Ananny and Crawford (2016) have convincingly argued, the transparency ideal for algorithmic accountability is insufficient at best and a red herring at worst. Such a focus "entrenches algorithmic opacity instead of lessening it" (Christin 2020: 900) by fixing the concept of "transparency"—and by extension, "accountability"—at limited stages in the process where it might be most impactful, such as allowing access to training data. The limits of the "black box" framing has led critical researchers, particularly ethnographers like Christin (2020) and Seaver (2017), to look "beyond the black box" to the relationships between human and non-human actors in "algorithmic assemblages" (Ananny and Crawford 2016).

Yet, at the practitioner level and in the broader discourse about computational intelligence and synthetic data that I briefly touched upon above, the commercial ideology that legitimates the adoption of these technologies hinges on the illusion of knowledge, control, and human-centered retributive accountability. "Accountability" in the context of automated decision making embraces the anthropocentric illusions (Kittler 1986; Packer and Reeves 2020) of ex post facto discipline for human "overseers" of AI agents in institutional and governmental contexts. While the "black box" may constitute a form of conceptual mystification, the third type of opacity Burell (2016) identified is inescapably at play: the nature of machine learning algorithms means that once the model is trained there is no unbaking the cake to remove "bad data," a whole new model would need to be trained. With the recent fixation on generative AI, and synthetic data in particular, another layer of obfuscation is added, one that requires not so much trust but faith in these systems—faith that the artificial data are "pretty good" and that domain transfers happen smoothly and

sufficiently. As Susser and Seema put it in their Dialogue paper (in this issue), "questions about how such 'ground truth' was established, and the function it serves, are put to the side." Such faith emphasizes the impossibility of reflexive and dialogical justification necessary for an ethics of accountability.

Mission Creep, or the Stakes of Normalizing Synthetic Data

In some scenarios, the deployment of synthetic data can appear relatively innocuous from a consequentialist perspective, such as training computer vision algorithms in video games. In another "domain," medical AI promises to revolutionize medical care with, proponents claim, the significant benefits outweighing the relatively low risk to patients if there is human oversight—if, for example, AI metrics aid medical providers as a "reliable indicator" in their broader deliberation about appropriate action and care (Durán and Jongsma 2021), offering at least a pretense of communicative ethics. Due to the stringent privacy concerns around medical data—and the expense of compensating clinical study participants—medicine is one field where synthetic data offers the greatest benefits for companies developing and deploying medical AI. A just-published study (Chouffani El Fassi et al. 2024) reveals that more than half of FDA approved AI health tools have not been clinically validated with real patient data, meaning they were either validated with "retrospective data"—data previously collected for different purposes other than validation of these specific medical devices—or they were missing clinical validation data and likely trained and validated using synthetic data. This lack of clinical validation with real patient data can pose risks such as misdiagnosis, overexposure in X-rays, and errors in radiation treatment, and "safety events" in ML-enabled medical devices have, in rare instances, even coincided with death (Lyell et al. 2023). But the biggest danger of synthetic data is the broader normalization of its use in automated or semi-automated systems, and the migration of models—and mindsets—from commercial to state deployment in law enforcement and military contexts.

Wiehn's dialogue paper (in this issue) underscores how despite (or because of) synthetic data's promise to remove "the issue of surveillance," its deployment in fact "adds a new layer to the data life cycle" while benefiting a growing datafied industry and de-politicizing it. At the same time, technologies frequently cross the porous boundaries of industry and state, particularly surveillant technologies. AI has led to a "platformitization of the military" with "new entanglements between the military and actors in the corporate domain, especially Big Tech, which play a key role in this context, as well as the open-source community that is organized around these platforms" (Hoijtink and Planqué-van Hardeveld 2022: 1). The institutional adoption and normalization of synthetic data comes with urgent moral stakes, particularly considering its possible deployment for large-scale warfare. Inaugurating a new era of militarized surveillance, AI is already used in targeting location and generation for airstrikes—the actual procedures of which only become knowable for public constraint and justification after they are leaked by whistleblowers. Evidence of synthetic data's downstream influence over several domain transfers in such applications seems likely. As Joseph Redmon (2020), the main author of the YOLO object detection architecture that revolutionized computer vision, noted about his decision to abandon computer vision research: "the military applications and privacy concerns eventually became impossible to ignore." Redmon's decision is not solely a consequentialist indictment, however, but speaks to the inability of even AI's leading practitioners—never mind the public—to constrain and demand answers to the "what?" and "why?" questions that professions and institutions like journalism might afford to meet the conditions of communicative ethics-as-accountability. With "synthetic data solutionism," to use Susser and Seema's (in this issue) words, we jettison additional institutionally born and constraining ethical guardrails associated with the use of real participant or patient data, further foreclosing the possibility of thoughtful, articulate, and reflexive deliberation on the significant moral implications of AI's vast and ever-growing assemblages. As the possibility for such dialogical justification of practices is essential to being accountable—and therefore being ethical—the use of synthetic data can never meet the standard of "ethical" practice.

References

Ananny, Mike, and Kate Crawford. 2016. Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability. *New Media & Society* 20 (3): 973–989.

Asaro, Peter M. 2023. Politicizing Data: AI Ethics as a Social Critique of Algorithms. *Social Research: An International Quarterly* 90 (4): 675–703.

Benjamin, Ruha. 2019. *Race after Technology: Abolitionist Tools for the New Jim Code*. New York: John Wiley & Sons.

Burrell, Jenna. 2016. How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms. *Big Data & Society* 3 (1): https://doi.org/10.1177/2053951715622512.

Chouffani El Fassi, Sammy, Adonis Abdullah, Ying Fang, Sarabesh Natarajan, Awab Bin Masroor, Naya Kayali, Simran Prakash, and Gail E. Henderson. 2024. Not All AI Health Tools with Regulatory Authorization Are Clinically Validated. *Nature Medicine* 30: 2718–2720.

Christin, Angèle. 2020. The Ethnographer and the Algorithm: Beyond the Black Box. *Theory and Society* 49 (5): 897–918.

Costanza-Chock, Sasha, Inioluwa Deborah Raji, and Joy Buolamwini. 2022. Who Audits the Auditors? Recommendations from a Field Scan of the Algorithmic Auditing Ecosystem. In *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Korea, June 21–24*, 1571–1583. New York: Association for Computing Machinary.

Durán, Juan Manuel, and Karin Rolanda Jongsma. 2021. Who Is Afraid of Black Box Algorithms? On the Epistemological and Ethical Basis of Trust in Medical AI. *Journal of Medical Ethics* 47 (5): 329–335.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Fukuda-Parr, Sakiko, and Elizabeth Gibbons. 2021. Emerging Consensus on "Ethical AI": Human Rights Critique of Stakeholder Guidelines." *Global Policy* 12: 32–44.

Glasser, Theodore, and James Ettema. 2008. Ethics and Eloquence in Journalism: An Approach to Press Accountability. *Journalism Studies* 9 (4): 512–534.

Gouldner, Alvin W. 1976. *The Dialectic of Ideology and Technology*. New York: Seabury Press.

Habermas, Jürgen. 1993. *Justification and Application: Remarks on Discourse Ethics*. Translated by Ciaran Cronin. Cambridge, MA: MIT Press.

Haggerty, Kevin D., and Richard V. Ericson. 2003. The Surveillant Assemblage. *The British Journal of Sociology* 51 (4): 605–622.

Heilinger, Jan-Christoph. 2022. The Ethics of AI Ethics. A Constructive Critique. *Philosophy & Technology* 35 (3): https://doi.org/10.1007/s13347-022-00557-9.

Hoijtink, Marijn, and Anneroos Planqué-van Hardeveld. 2022. Machine Learning and the Platformization of the Military: A Study of Google's Machine Learning Platform TensorFlow. *International Political Sociology* 16 (2): https://doi.org/10.1093/ips/olab036.

IBM Technology. 2023. What Is Synthetic Data? No, It's Not "Fake" Data. YouTube, March 29. https://www.youtube.com/watch?v=HIusawrGBN4 [accessed August 27, 2024].

———. 2024. Can You Trust Synthetic Data? YouTube, March 29. https://www.youtube.com/watch?v=QQtSa9ngqQk [accessed August 27, 2024].

Kant, Immanuel. 2002. *Groundwork for the Metaphysics of Morals*. Edited by Allen W. Wood. Hartford, CT: Yale University Press.

Kittler, Friedrich. 1986. A Discourse on Discourse. *Stanford Literature Review* 3 (1): 157–166.

Lyell, David, Ying Wang, Enrico Coiera, and Farah Magrabi. 2023. More than Algorithms: An Analysis of Safety Events Involving ML-Enabled Medical Devices Reported to the FDA. *Journal of the American Medical Informatics Association* 30 (7): 1227–1236.

Maas, Jonne. 2022. A Neo-Republican Critique of AI Ethics. *Journal of Responsible Technology* 9 (April): https://doi.org/10.1016/j.jrt.2021.100022.

Nikolenko, Sergey I. 2021. *Synthetic Data for Deep Learning*. New York: Springer.

O'Connor, Sinead, and Helen Liu. 2023. Gender Bias Perpetuation and Mitigation in AI Technologies: Challenges and Opportunities. *AI & SOCIETY* 39: 2045–2057.

Packer, Jeremy, and Joshua Reeves. 2020. *Killer Apps: War, Media, Machine*. Durham, NC: Duke University Press.

Pasquale, Frank. 2015. *The Black Box Society*. Cambridge, MA: Harvard University Press.

Raji, Inioluwa Deborah, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, January 27–30*, 33–44. New York: Association for Computing Machinery.

Ravn, Louis. 2024. Synthetic Training Data and the Reconfiguration of Surveillant Assemblages. *Surveillance & Society* 22 (4): 460–465.

Redmon, Joseph [@pjreddie]. 2020. "I Stopped Doing CV Research Because I Saw the Impact My Work Was Having. I Loved the Work but the Military Applications and Privacy Concerns Eventually Became Impossible to Ignore." Twitter, February 20, 11:09 am. https://x.com/pjreddie/status/1230524770350817280.

Ridgway, Reneé, and Nicolas Malevé. 2024. Synthetic Data and Reverse Image Search: Constructing New Surveillant Indexicalities. *Surveillance & Society* 22 (4): 466–471.

Seaver, Nick. 2017. Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems. *Big Data & Society* 4 (2): https://doi.org/10.1177/2053951717738104.

Susser, Daniel, and Jeremy Seeman. 2024. Critical Provocations for Synthetic Data. *Surveillance & Society* 22 (4): 453–459.
Vallor, Shannon. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford, UK: Oxford University Press.
Wiehn, Tanja. 2024. Synthetic Data: From Data Scarcity to Data Pollution. *Surveillance & Society* 22 (4): 472–476.