

Position Paper for Workshop “Synthetic Data: Representation and/vs Representativeness” (Aarhus University, 18/08/2025)

Louis Ravn

University of Amsterdam (Institute for Logic, Language and Computation)

Synthetic populations and the politics of representation: Can the synthesized speak?

In lockstep with the growing uses of synthetic data in various societal domains (Offenhuber, 2024), critical scholarship has highlighted the multifaceted sociopolitical, ethical, and legal questions they engender (Jacobsen, 2023; Steinhoff, 2024; Beduschi, 2024). One particularly crucial insight of this body of scholarship is that synthetic data are complexly related to questions of representation. On the one hand, a central promise undergirding the rise of synthetic data is the notion that they facilitate the datafied representation of underrepresented and people (e.g., minorities; Jacobsen, 2025) and objects (i.e., “edge cases”; Jacobsen, 2023). On the other hand, scholarship has convincingly argued – drawing from a long lineage of STS research (Bowker & Star, 1999; Mol, 2002; Barad, 2007) – that synthetic data should be understood as performative rather than merely representative (Lee et al., 2025). Whether in the form of “synthetic supplements” (Jacobsen, 2024) or fully synthetic datasets, this perspective underscores that synthetic data shape the reality of that which is represented; in short, they engender questions of ontological politics (Mol, 2002): how are realities made in and through the representations of synthetic data and what are these realities’ implications?

I am interested in transposing this concept of synthetic data as performative rather than representative to an analysis of what we may term “synthetic populations”: representations of human populations constructed by means of synthetic data. For the sake of a relevant example, take the “Global Synthetic Dataset”: it constitutes the most comprehensive and fine-grained dataset containing information about survivors of human trafficking (International Organization for Migration, 2024). In my research on this dataset, a recurring narrative is that the representation of this group of people is facilitated thanks to the alleged privacy and safety afforded by synthetic data. And indeed, it may be noted here that “access to representation” has been identified as a core concern by data justice scholarship (Taylor, 2017). At the same time, the project is unclear about the extent to which the people – whose data form the basis of the synthetic datasets – are aware of these data reuses. Moreover, the Global Synthetic Datasets are performative: although individual rows of data are not directly linkable to distinct individuals (and hence “private”), the dataset – in its totality – makes legible the broader population of human trafficking survivors: individual and group privacy are, as we know, intertwined (Helm et al., 2024).

Against this backdrop, I would like to ask whether the synthesized can speak, thereby playing on Spivak’s (1988) famous essay: “Can the subaltern speak?”. Concretely, by raising this question, I am interested in collectively exploring the seeming paradox produced by *representation via synthesis*: marginalized populations may indeed gain access to representation via synthesis, but to what extent do they have a say in these synthesized representations? Moreover, how should we – as scholars examining synthetic data from various theoretical and methodological frameworks – encounter these synthetic populations? Should we try to speak to people represented within synthetic populations? But how do we know whether somebody feels “represented” by any given synthetic population? As already emerges in the reflections of this brief position paper, synthetic

populations seem to be deeply entangled with the performative politics of classification (Lee et al., 2025).

More importantly, I suggest that an examination of the relationship between synthetic populations and questions of representation/representativeness stands to gain from the sensibilities advanced by data justice scholarship (Taylor, 2017; Dencik, 2025). Particularly, while capabilities-based approaches (e.g., Taylor, 2017) emphasize concerns around privacy, autonomy, and non-discrimination, a recent special issue in *Information, Communication & Society* advances a data justice agenda focused on collectivity, specifically the “population-level effects of datafication” (Dencik, 2025, p. 13). This recent turn underscores the urgency of exploring populations as a particular object constructed by means of synthetic data – and how this relates to the politics of representation.

An additional challenge is constituted by the difficulty of theorizing synthetic populations. On one hand, they seem to belong to the long-established paradigm of biopolitics as the government of populations (Foucault, 1976). At the same time, of course, the growth of synthetic populations emerges at a moment when the very category of the “population” seems to dissolve, a consequence of deep learning epistemologies increasingly suffusing modes of contemporary governance (Amoore, 2022; Halpern & Mitchell, 2023; Johns, 2023; Törnberg et al., 2025). Are synthetic populations therefore situated at the border of biopolitics and the way in which “an ANN sees” (Törnberg et al., 2025), a hybrid of both, or does this depend on context?

To sum up, I would be very intrigued to explore questions of representation and representativeness as they materialize in the context of synthetic populations, broadly understood. Particularly, I would be interested in discussing these topics with an eye towards questions of data justice, biopolitics, and scholarly encounters of synthetic populations.

References

- Amoore, L. (2022). Machine learning political orders. *Review of International Studies*, 49(1).
- Barad, K. (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and meaning*. Durham, NC: Duke University Press.
- Beduschi, A. (2024). Synthetic data protection: Towards a paradigm change in data regulation? *Big Data & Society*, 11(1).
- Bowker, G. & Star, S. L. (2000). *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: The MIT Press.
- Dencik, L. (2025). ‘Rescuing’ data justice? Mobilising the collective in responses to datafication. *Information, Communication & Society*, 1-16.
- Foucault, M. (1978). *The History of Sexuality, Vol. 1: The Will to Knowledge*. London, UK: Penguin Random House.
- Halpern, O. & Mitchell, R. (2022). *The Smartness Mandate*. Cambridge, MA: The MIT Press.
- Helm P., Lipp B. and Pujadas R. (2024). Generating reality and silencing debate: Synthetic data as discursive device. *Big Data & Society*, 11(2).
- 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 23/05/2025].
- Jacobsen, B. N. (2023). Machine learning and the politics of synthetic data. *Big Data & Society*, 10(1): 1-12.
- Jacobsen, B. N. (2024). The Logic of the Synthetic Supplement in Algorithmic Societies. *Theory, Culture, and Society*, 41(4), 41-56.

- Jacobsen B (2025) Machine Learning, Synthetic Data, and the Politics of Difference. *Theory, Culture & Society*, 42(3), 41-57.
- Johns, F. (2023). *#Help: Digital Humanitarianism and the Remaking of International Order*. Oxford, UK: Oxford University Press.
- Lee, F., Hajisharif, S., Johnson, E. (2025). The ontological politics of synthetic data: Normalities, outliers, and intersectional hallucinations. *Big Data & Society*, 12(2).
- Mol, A. (2002). *The body multiple*. Durham, NC: Duke University Press.
- Offenhuber, D. (2024). Shapes and frictions of synthetic data. *Big Data & Society*, 11(2).
- Spivak, G. C. (1988). Can the Subaltern Speak? In C. Nelson & L. Grossberg (Eds.), *Marxism and the Interpretation of Culture*. (pp. 271-313). Basingstoke, UK: Macmillan Education.
- Steinhoff, J. (2024). Toward a political economy of synthetic data: A data-intensive capitalism that is not surveillance capitalism? *New Media & Society*, 26(6), 3290-3306.
- Taylor L (2017) What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society* 4(2).
- Törnberg, P., Söderström, O., Barella, J., Greyling, S. & Oldfield, S. (2025). Artificial intelligence and the state: Seeing like an artificial neural network. *Big Data & Society*, 12(2).