

The Qualities of Quantities: Making Situated Chains of Representation

Position Statement for the Workshop on Synthetic Data: Representation and/vs Representativeness at Aarhus 2025

Francis Lee, Södertörn University, <francis@francislee.org>
Ericka Johnson, Linköping University, <ericka.johnson@liu.se>

We engage with the concept of representation in synthetic data within the context of scientific knowledge making practices. Our work employs theoretical concepts about the data-world relation drawn from STS (Latour & Woolgar 1979; Latour 1989; Knorr Cetina 2025), critical studies of AI (MacKenzie 2017; Amaro 2023) and feminist technoscience studies (Barad 2007; Suchman 2023; Klein & D'Ignazio 2024).

In many ways, we ask about possible “new” relationships between synthetic data and ground truths, but also what the relationship between synthetic data and “real” data is. Given our grounding (hahah, pun intended) in STS, we are interested in practices, particularly the way boundaries are drawn around and between real and synthetic data, and how scientists bridge between synthetic data and the world.

We would like to talk about how quantities have qualities. Data is often seen as quantitative and as simple to represent in numbers or in quantitative measures, but quantities always have qualities that need to be understood in order for the chain of representation to hold together. How processes of quantification and qualification relate to each other is one of the central topics that needs to be addressed in the generation and use of synthetic data. These qualities can be “things” like endogenous artifacts such as intersectional relations between different aspects of data, but it can also relate to the qualities of the original dataset, or even the qualities of the physical world where the original dataset was extracted from. At the workshop, we would like to think through how these qualities of synthetic data relate to the relationship between world, data, and synthetic data.

Perhaps of interest to others at the workshop, we are also concerned with how these qualities of quantities relate to paradata and metadata. We have recently been building on our current work on the complexities of synthetic data (Johnson & Hajisharif 2024; Lee, Hajisharif, Johnson 2025) and developed a particular interest in metadata and paradata for synthetic data, thinking about how scientists can make the qualities of their synthetic data more transparent (Johnson et al 2025). This work is sensitive to how the politics of categorisation are even more strongly obscured by the shimmer (and tarnish) of synthetic data and AI knowledge.

Integral to this discussion is also emerging practices that claim to assure provenance and pedigree – that the quantitative data represents that which it claims to represent, but this then also ties to the provenance and pedigree of qualities.

We feel the advent of synthetic data raises new and complex questions about the reuse and travel of data, like: What qualities of the original data are preserved in the synthetic dataset? How do you make the metadata and paradata of those qualities travel? How can you make visible both the process of production, the assumptions, and the data-relationalities visible to re-users of the datasets?

References:

- Amaro, R. (2023) *The Black Technical Object*. MIT Press
- Bowker GC and Star SL (1999) *Sorting Things out*. Cambridge: MIT Press.
- Johnson, E. & S. Hajisharif (2024) The Intersectional Hallucinations of Synthetic Data. *AI & Society*
- Johnson, E., Rayner, D., Kasmire, J. et al. (2025). Metadata/README elements for synthetic structured data made with GenAI. *AI Policy Exchange Forum (AIPEx)*.
- Klein L and D'Ignazio C (2024) Data Feminism for AI. *FAccT '24*. June 03-03, 2024 Rio De Janeiro.
- Knorr Cetina (2025) presentation at 'What is Knowledge', KVA Wennergren Symposium, Stockholm, 09-05-2025
- Latour, B. (1987). *Science in action*. Harvard UP.
- Latour, B. & S. Woolgar (1979) *Laboratory Life*.
- Lee, F., S. Hajisharif, & E. Johnson (2025) The Ontological Politics of Synthetic Data. *Big Data & Society*.
- MacKenzie A (2017) *Machine Learners*. Cambridge: MIT Press.
- Suchman L (2023) The uncontroversial 'thingness' of AI. *Big Data & Society*. 10(2).