

Synthetic Data as the Future of AI Privacy, Explainability, and Fairness: An Introduction for Data Scientists and Data Executives

with Alexandra Ebert



Synthetic Data for Data Executives

What is synthetic data?

- A key enabler for privacy-preserving AI and data-driven innovation
- Allows your organization to anonymize data while retaining near-perfect data utility
- Can unlock the value of behavioral customer data

Three major strategic objectives synthetic data can support

1. Increase revenue.
2. Reduce costs.
3. Reduce risk.

Benefits of AI-generated privacy-preserving synthetic data

- Strong privacy protection
- Time to data
- Cost savings
- Data quality and utility

Top use cases for synthetic data

- AI training
- External data sharing
- Digital product development
- Cloud migration
- Responsible AI governance
- AI fairness
- Explainable AI
- Data democratization

Two approaches to providing synthetic data to your team

1. Data synthesis on-demand for specific projects and teams
2. Synthetic data democratization with internal marketplaces

Additional Resources

Harvard Business Review Report: <i>The Executives Guide to Accelerating Artificial Intelligence and Data Innovation with Synthetic Data</i>	Check out this report to learn more about how large organizations are using synthetic data and which benefits/ROI they gain from it. Note: this is behind a paywall.
Synthetic Data Report from the European Commission's Joint Research Centre (JRC)	You can read the full report here . I also wrote a blog post summarizing the 88 pages to provide a quick overview on why policy makers consider synthetic data not only highly relevant for privacy, but also for fairness, data democratization, and more.
Innovation Advisory Group from the UK's Financial Conduct Authority	The UK's Financial Conduct Authority established a Synthetic Data expert group (I'm also a member) to accelerate adoption of SD in the financial services industry. The group is currently working on a report on best practices across a range of different financial services SD use cases. The FCA in general is quite active around synthetic data. See this link for more details .
Tackling AI Bias At Its Source – With Fair Synthetic Data	To learn more about synthetic data for AI fairness check out this blog post .
A high-level introduction to synthetic data for Explainable AI	https://mostly.ai/blog/the-future-of-explainable-ai-rests-upon-synthetic-data
Publications on Privacy Enhancing Technologies, including Synthetic Data	https://royalsociety.org/topics-policy/projects/privacy-enhancing-technologies/
The <i>Data Democratization</i> Podcast	Check out this podcast to learn more about synthetic data, Responsible AI, AI Governance, and how to accelerate AI innovation across an enterprise organization. This is available on most podcast platforms.
	Episode 41: Synthetic data for impact: how to innovate in health insurance with Daniela Pak-Graf, Merkur Innovation Lab
	Episode 33: Fair synthetic data and ethical AI in healthcare with Humana
	Episode 31: How to get value out of PETs in banking with Ville Sointu