What is Synthetic Data?

Mason Egger

@masonegger

Gretel

Common Developer Data Issues

▶ Access to usable testing data

▶ Limited Data Sets

▶ Biased Data

@masonegger

Synthetic Data to the Rescue

▶ Making private data accessible and shareable

▶ Generating more samples with limited data sets

▶ Reducing bias in machine learning datasets

　　▶ [Reducing AI Bias in Predicting Heart Disease with Synthetic Data](#)

@masonegger

Isn't that just "fake data"?

▶ Synthetic data is different from "fake" or "mock" data

▶ "Fake/mock" data has no accuracy. It is purely random

▶ Synthetic data can be nearly as accurate, or and in some cases even [improve on the accuracy of real-world data](#).

What can I use Synthetic Data for?

▶ **Automotive and Robotics** — leveraging synthetic data to create simulated environments for training robots, self-driving car software, and even testing safety and crash prevention technologies.

▶ **Financial Services** — creating synthetic time-series data to enable data sharing that doesn't compromise their customers' privacy

▶ **Cybersecurity and Infosec** — using synthetic data to train machine learning models to better detect rare events including fraud and cyber attacks

▶ **Healthcare and Life Sciences** — creating synthetic genomic data to fuel medical breakthroughs and encourage better medical care

▶ **Manufacturing** — using synthetic data to simulate complex supply chain operations and predict where failures may occur.

▶ And more!

How can I start using Synthetic Data?

▶ Gretel has open-sourced their models https://github.com/gretelai

▶ You can also use the Gretel Cloud at https://gretel.ai

   ▶ Try out the free tier

▶ Check out Gretel's blog What is Synthetic Data?

That's all for this time!

- ○ Follow me on Twitter [@masonegger](https://twitter.com/masonegger)
- ○ Join me for an Open Space on Synthetic Data!
  - ○ 250 AB at 4:00pm on Saturday

- ○ Slides on my website, https://mason.dev

gretel™