Synthetic Customers

for Big Data Privacy

mostly AI

The Privacy vs. Innovation Clash

1 **Data privacy** restricts sharing of data and thus **hampers innovation**.

We demand highest standards for data protection, but also need to collaborate broadly on data in order to develop next-gen digital services and processes.

2 **Pseudonymization** offers **no safety**, while **Full Anonymization** falls short for big data.

Classic anonymization techniques need to destroy most of the available information to prevent re-identification of individuals (see appendix).

1 **Synthetic data is anonymous**.

Synthetic data is not restricted in its usage, and is free to store, to use, to explore, to experiment, to modify, and to share, within and outside of the organization.

2 **Generative AI** allows **highly accurate** synthetic data to be generated at scale.

Academic advances on deep generative neural networks have resulted in highly realistic synthetic images, near indistinguishable from real ones.

mostly AI

Pseudonymization

Pseudonymization offers **no safety**!

Pseudonymization

Pseudonymization offers **no safety**, while
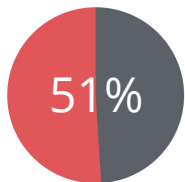**Classic Anonymization destroys** data **value**!

Pseudonymization

Pseudonymization offers **no safety**, while
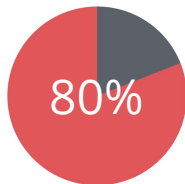**Classic Anonymization destroys** data **value**!
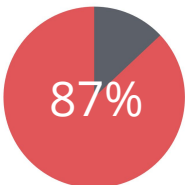
Classic Anonymization Fails for Big Data

i.e. for High−Dimensional, Highly Correlated Data

51% of **mobile phone owners** are re-identified simply by 2 antenna signals, even when coarsened to the hour of the day
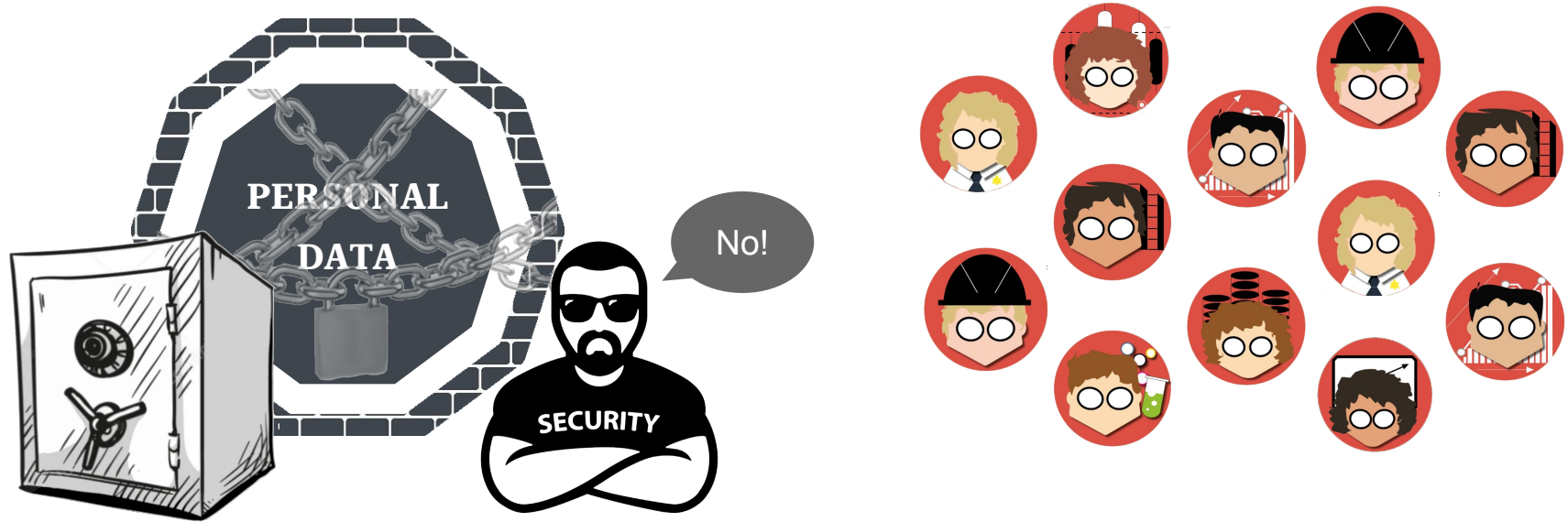
80% of **credit card owners** are re-identified by 3 transactions, even when only merchant and the date of transaction is revealed

87% of **all people** are re-identified, merely by their date-of-birth, their gender and their ZIP code of residence

mostly AI

So, How to Utilize Big Data in the Era of Privacy?

a data-driven, customer-centric, innovative organization?

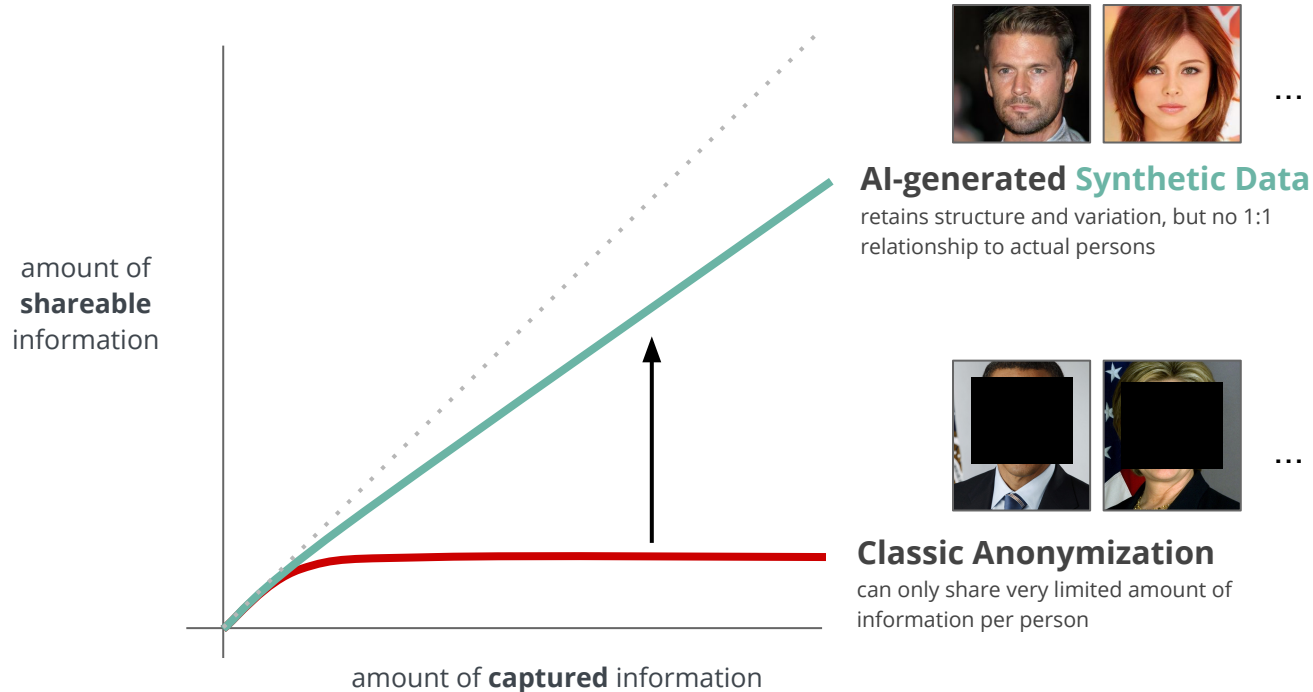# Dummy Data?

overly **simplistic and biased**

AI-Generated **Synthetic Data!**

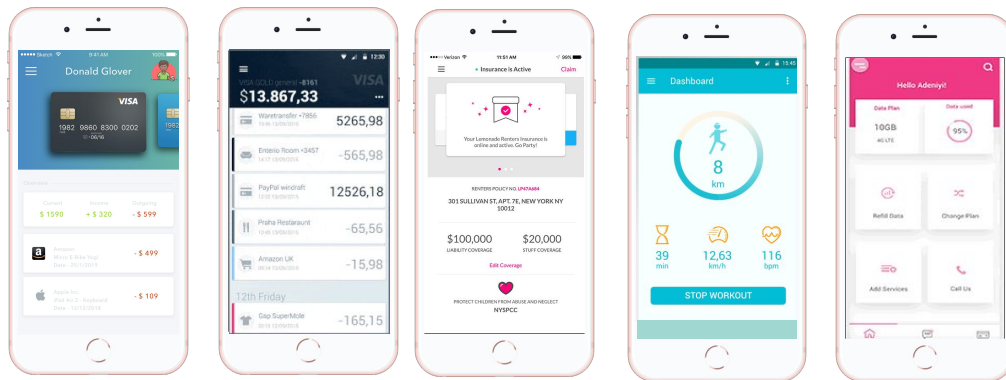highly realistic & representative, **as-good-as-real**, captures full richness of your actual data

NVIDIA.

Game Changer for Big Data Anonymization

amount of **shareable** information

amount of **captured** information

AI-generated Synthetic Data
retains structure and variation, but no 1:1 relationship to actual persons

Classic Anonymization
can only share very limited amount of information per person

mostly AI

AI-generated **synthetic populations** of customers and their behavior

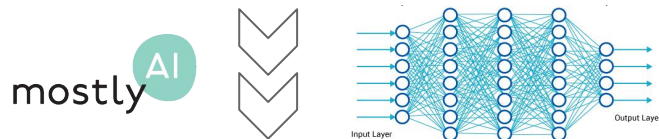| CRM/HR | Transaction | Insurance | Health | Mobility |

```
NAME   ZIP AGE GENDER     ITEM   EUR       LAT    LONG    DAY   TIME
Mary  1220 25y female     Book   12€   48.20N  16.78E    Fri  10:00
Mary  1220 25y female     Swim    8€   48.32N  16.82E    Fri  15:10
...
John  1210 39y   male     Lunch  13€   47.32N  16.02E    Mon   9:30
John  1210 39y   male     Food    8€   47.32N  16.02E    Mon  10:30
```

The **Synthetic Data Engine** by Mostly AI

AI-generated **synthetic populations** of customers and their behavior

retains unprecedented amount of **detail, structure and variation**

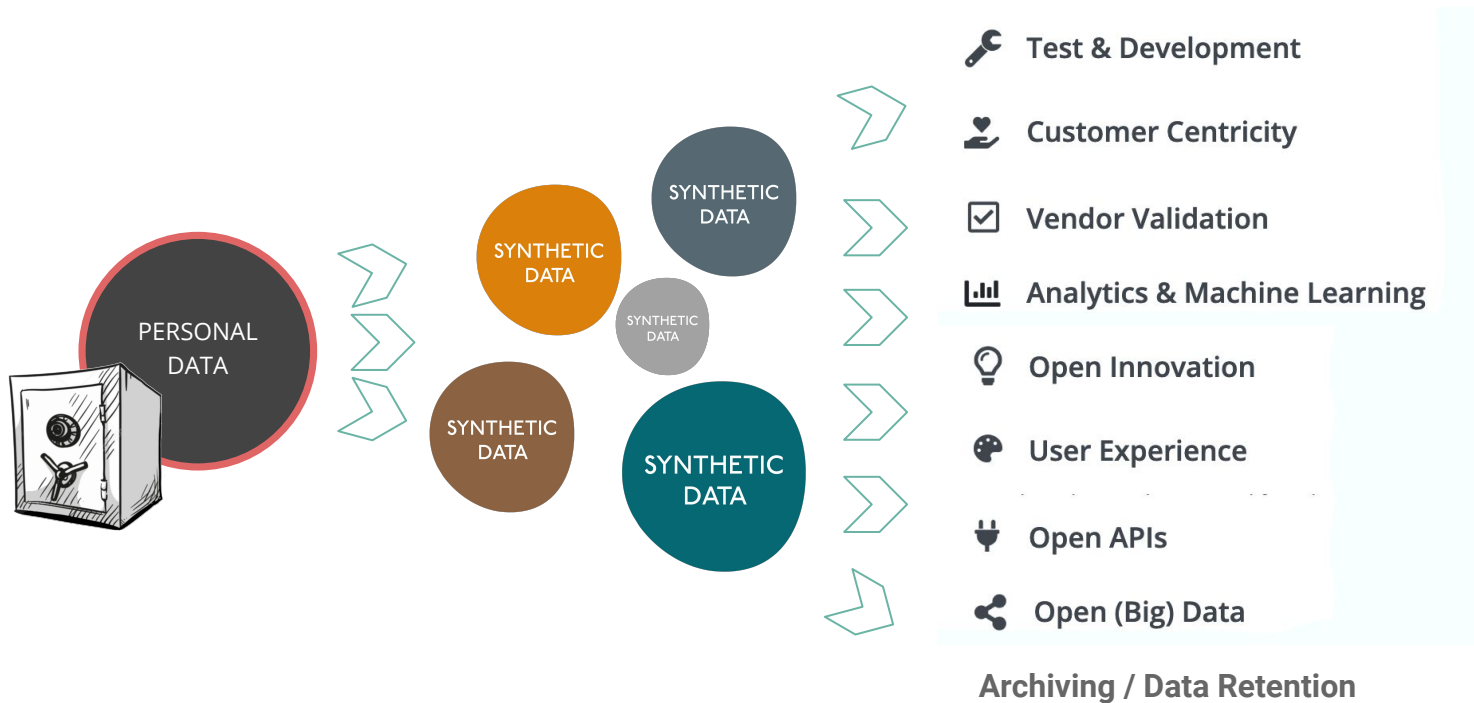**on-premise** software solution in protected environment

scales to millions of individuals with billions of events

enables sharing and processing **without privacy restrictions**

The **Synthetic Data Engine** by Mostly AI

PERSONAL DATA

SYNTHETIC DATA

SYNTHETIC DATA

SYNTHETIC DATA

SYNTHETIC DATA

SYNTHETIC DATA

- Test & Development
- Customer Centricity
- Vendor Validation
- Analytics & Machine Learning
- Open Innovation
- User Experience
- Open APIs
- Open (Big) Data

Archiving / Data Retention

Big Data Anonymization for unlimited opportunities for your customer data !

mostly AI

Use Cases – Synthetic Data

for **Internal Data Sharing**

- Testing & Development
- Product Design / UX
- Insights & Analytics
- AI Training Data
- Cross-Border Sharing

for **External Data Sharing**

- Open APIs
- Open Innovation
- Research Collab
- Vendor Validation
- Sandboxing

for **External Data Monetization**

- Strategic Partnerships
- Data Marketplaces
- Data Resellers
- Market Research Intel

The more information is captured without risking privacy → the more value can be generated !
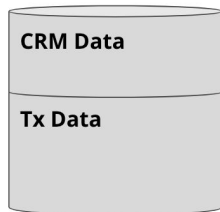
The higher the automation in anonymization process → the faster data can be shared & utilized !

mostly AI

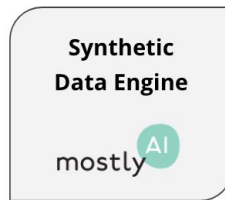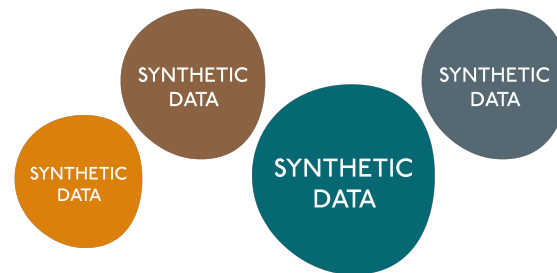# Go Synthetic Today!
to unlock your data assets

1. Data & Hardware Provisioning

CRM Data

Tx Data

2. Software Setup & Model Training

Synthetic Data Engine

mostly AI

3. Unlimited no. of realistic & representative synthetic data

SYNTHETIC DATA

SYNTHETIC DATA

SYNTHETIC DATA

SYNTHETIC DATA

1-3 month is the project time - delivery guaranteed

mostly AI