

Synthetic Location Data

for privacy-safe, granular-level data sharing



Dr. Michael Platzer

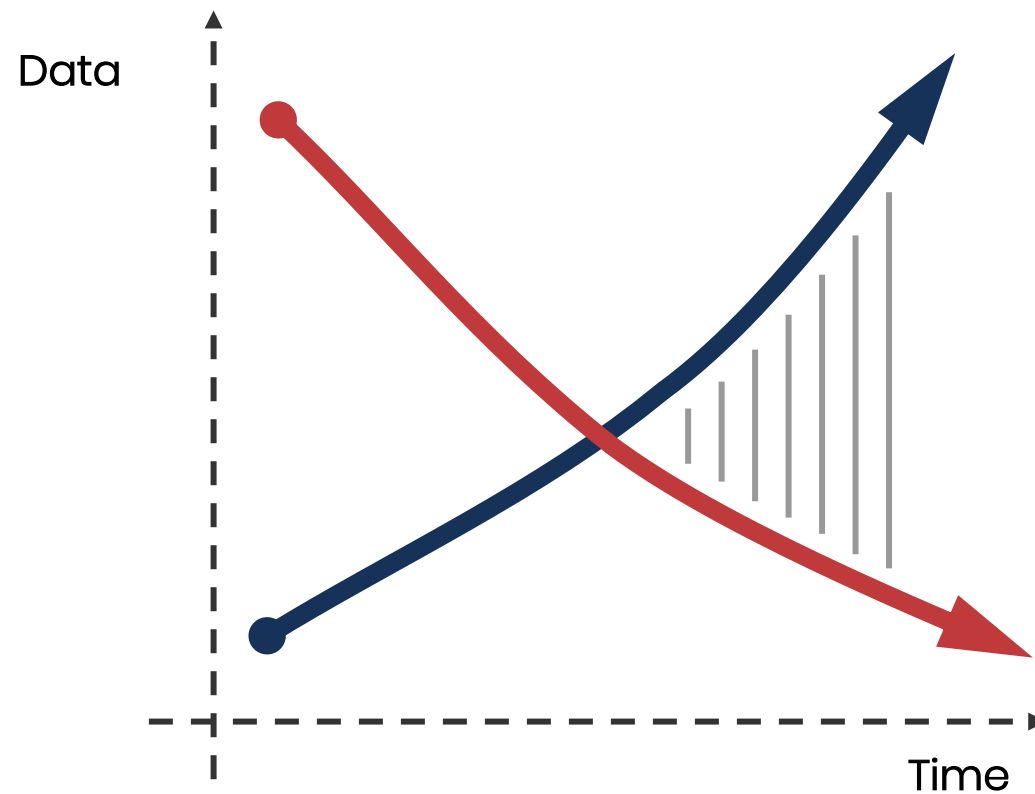
Co-Founder & Chief Strategy Officer

michael.platzer@mostly.ai

MOSTLY·AI

Structured Synthetic Data since 2017

The Data Fuel Crisis



AI, analytics & automation
drive demand for data

Regulation, risk & ethics
restrict availability of data

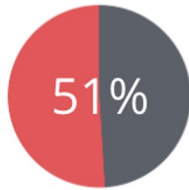
Why not “simply” anonymize your data?

truly anonymous data is not subject to regulation

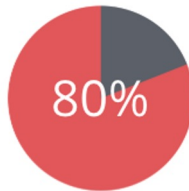
...because “anonymization” doesn’t work!



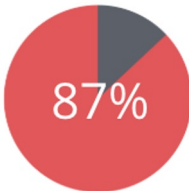
Academia knows it



of **mobile phone owners** are re-identified simply by **2 antenna signals**, even when coarsened to hour of day (Nature, 2013)



of **credit card owners** are re-identified by **3 transactions**, even when only merchant and the date of transaction is revealed (Science, 2015)



of **US citizens** are re-identified by **date-of-birth, gender and ZIP code** (Health, 2000)

Media knows it

AOL: "This was a screw up"

NetFlix Cancels Recommendation Contest After Privacy Lawsuit

The Guardian

Sticky data: Why even 'anonymized' information can still identify you

The New York Times

Researchers spotlight the lie of 'anonymous' data

Big data privacy is a bigger issue than you think

SCIENCE & TECHNOLOGY

You're not so anonymous

Sorry, your data can still be identified even if it's anonymized

Saying it's Anonymous Doesn't Make It So: Re-identifications of "anonymized" law school data

'Anonymised' data can never be totally anonymous, says study

REGULATION

There's No Such Thing as Anonymous Data

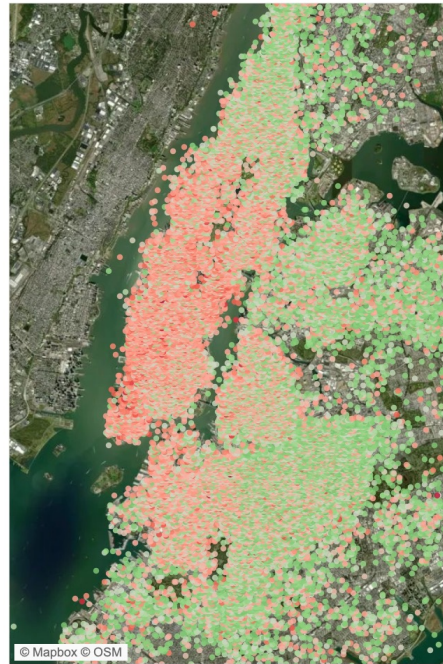
Harvard Business Review

Everyone knows it



Options ?

- ignore privacy ?



original
data

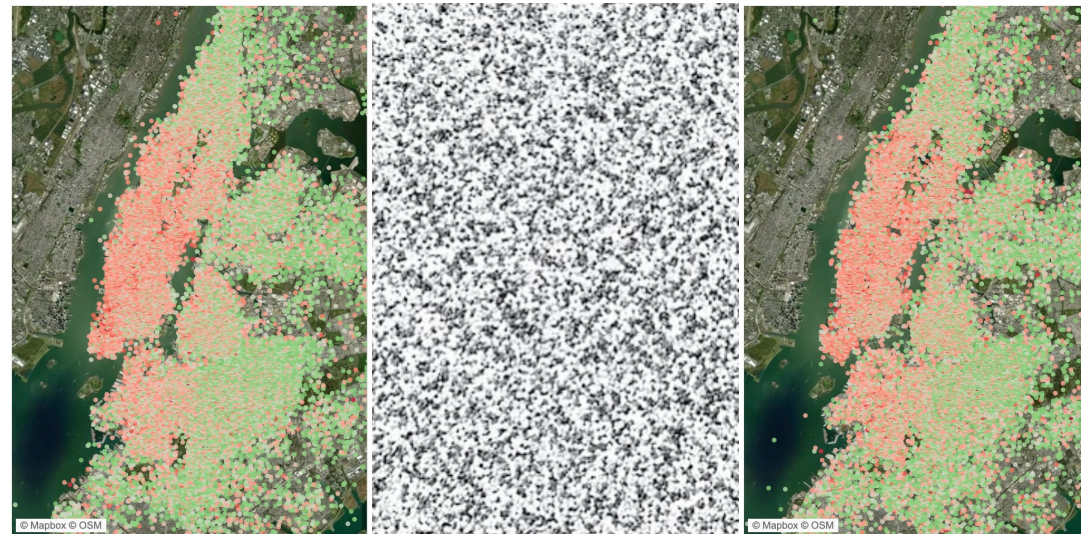
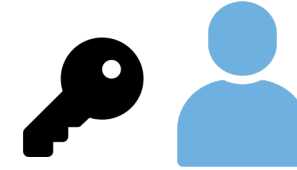
Options ?

- ~~ignore privacy~~
- not use data ?



Options ?

- ~~ignore privacy~~
- ~~not use data~~
- encryption



original
data



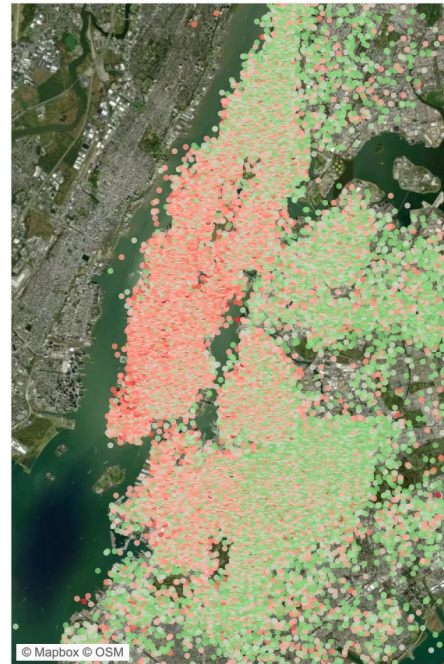
encrypted
data



decrypted
data

Options ?

- ~~ignore privacy~~
- ~~not use data~~
- encryption
- sandboxes

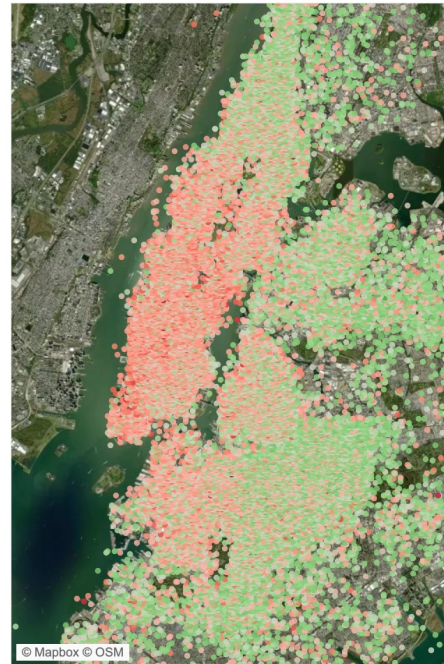


original
data

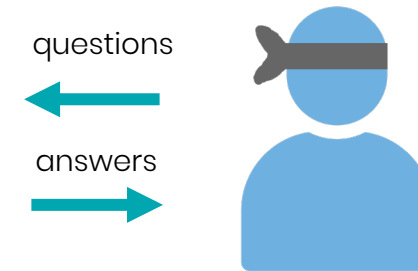


Options ?

- ~~ignore privacy~~
- ~~not use data~~
- encryption
- sandboxes
- query-based systems

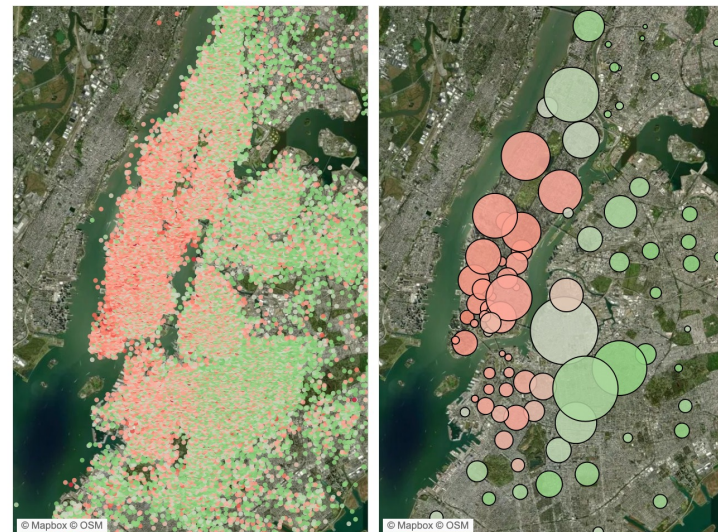


original
data



Options ?

- ~~ignore privacy~~
- ~~not use data~~
- encryption
- sandboxes
- query-based systems
- aggregation



original
data



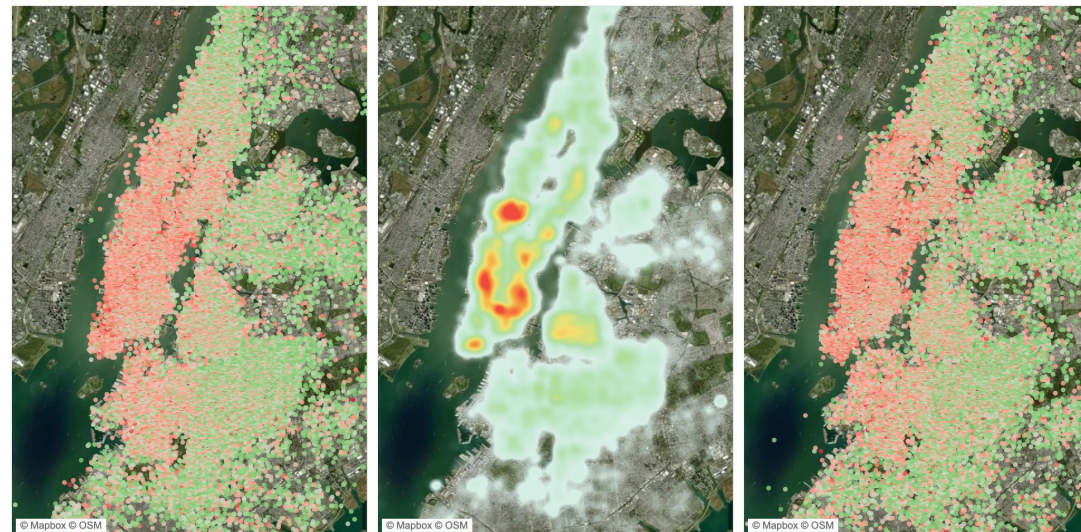
aggregated
data

High-level summary of original data, retaining information on few, selected attributes.



Options ?

- ~~ignore privacy~~
- ~~not use data~~
- encryption
- sandboxes
- query-based systems
- aggregation
- synthetic data



original
data

estimated
distributions

synthetic
data

Model learns how people look like, what they do, how they behave, etc. based on original data.

New, statistical representative personas are being sampled from distributions that retain the learned patterns.

AI-Generated Synthetic Data !



AI-Generated Synthetic Location Data

trips_id	STATE	ORIG	TIMESTAMP	latitude	longitude
05M1eO8a66YAfb0dlc	n/a	25	04/16/2014 12:01:17	41.14605	-8.61758
05M1eO8a66YAfb0dlc	n/a	25	04/16/2014 12:01:19	41.14604	-8.61759
05M1eO8a66YAfb0dlc	n/a	25	04/16/2014 12:01:22	41.14588	-8.61759
05M1eO8a66YAfb0dlc	A3	25	04/16/2014 12:01:27	41.14488	-8.61726
05M1eO8a66YAfb0dlc	A3	25	04/16/2014 12:01:44	41.14501	-8.61646
05M1eO8a66YAfb0dlc	B2	25	04/16/2014 12:01:59	41.14500	-8.61649
05M1eO8a66YAfb0dlc	B2	25	04/16/2014 12:02:08	41.14499	-8.61649
05M1eO8a66YAfb0dlc	B2	25	04/16/2014 12:02:43	41.14512	-8.61627
05M1eO8a66YAfb0dlc	B2	25	04/16/2014 12:02:50	41.14514	-8.61612
05M1eO8a66YAfb0dlc	A3	25	04/16/2014 12:03:09	41.14527	-8.61606
05M1eO8a66YAfb0dlc	A3	25	04/16/2014 12:03:30	41.14526	-8.61605

AI-Generated Synthetic Location Data

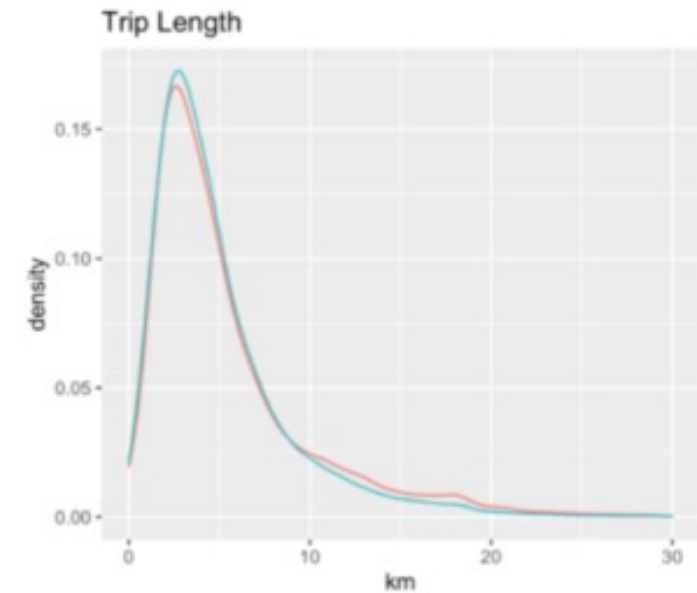
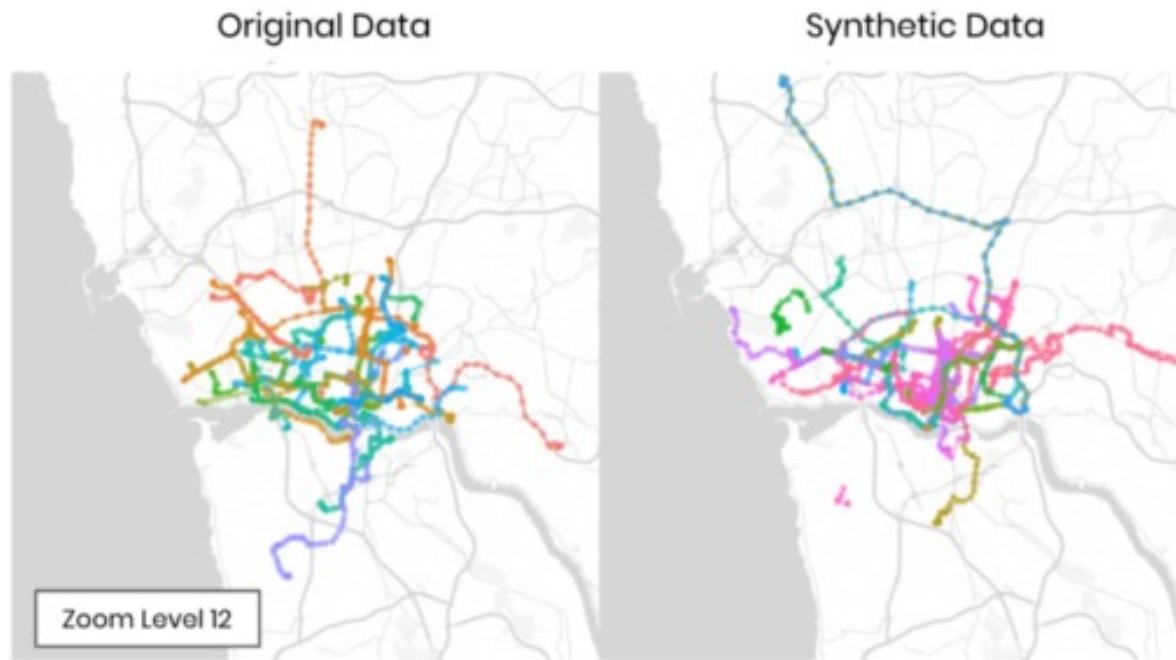
Original Data



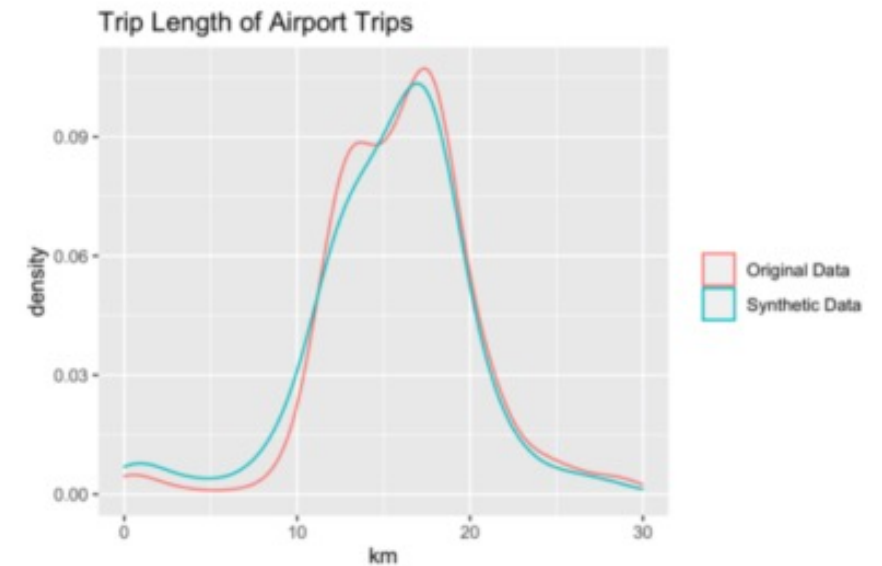
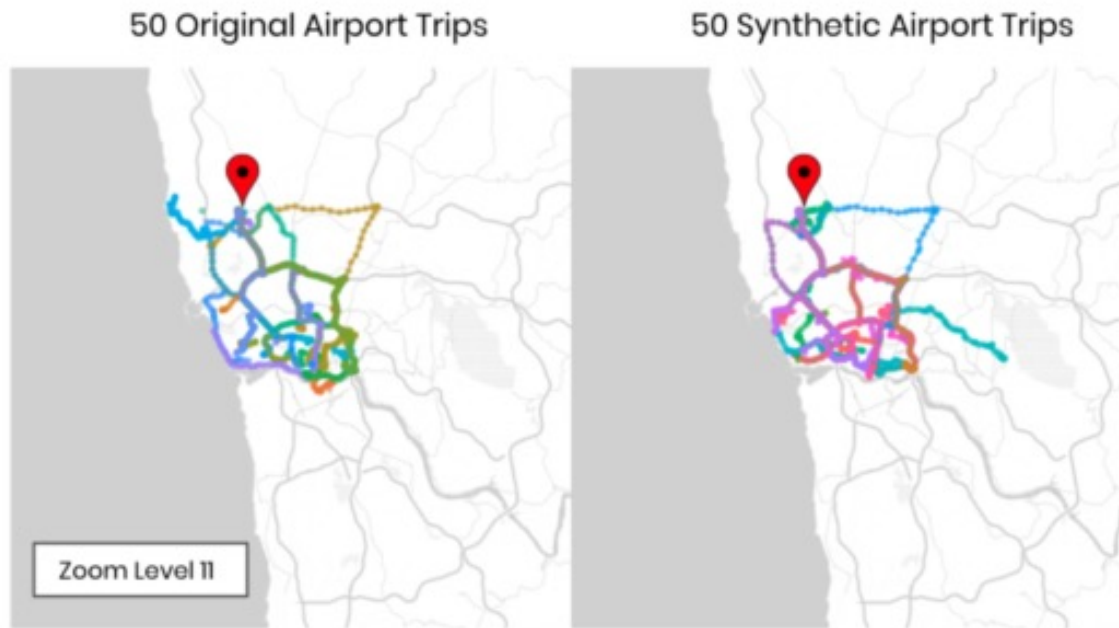
Synthetic Data



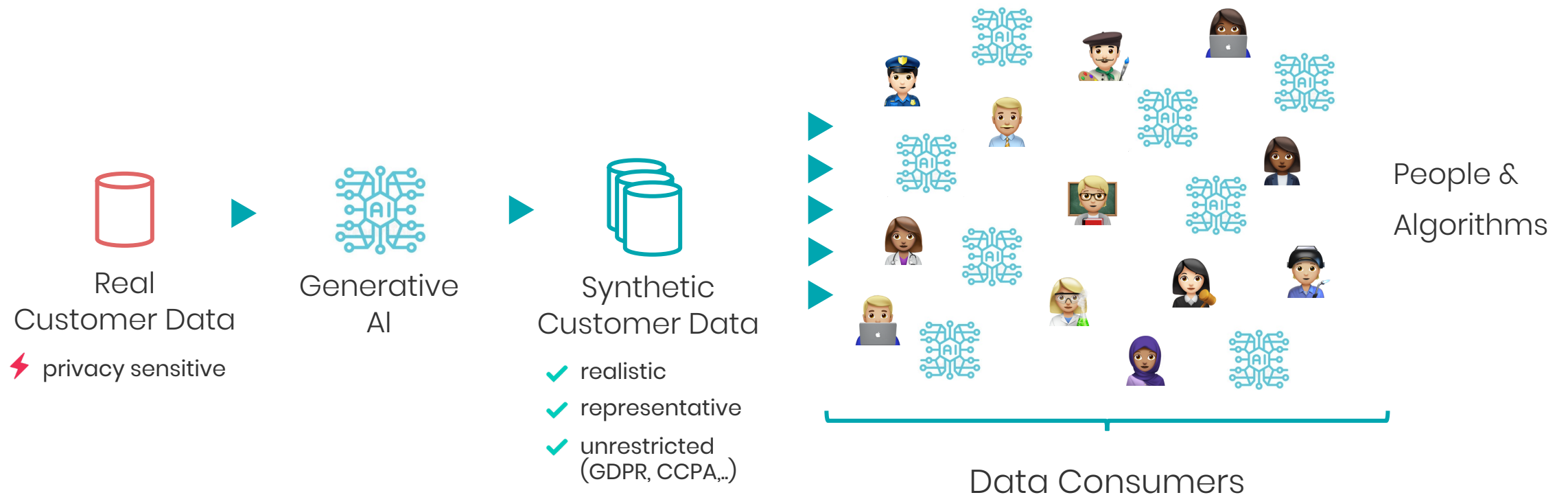
AI-Generated Synthetic Location Data



AI-Generated Synthetic Location Data



AI-Generated Synthetic Data



Generative AI distills the **insights** from real data into synthetic data, so that People as well as Algorithms can safely learn from.

MOSTLY·AI

Smarter Data for Smarter Organizations