



Synthetic Data Real Impact



Syntho is recognized by global leaders

Winner Philips Innovation Award



Winner Unesco's Gender Bias Challenge



United Nations
Educational, Scientific and
Cultural Organization



Synthetic data
assessed and approved



Shortlisted as
Generative AI startup

50%

Of data is locked

What does the Dutch DPA say?



AUTORITEIT
PERSOONSgegevens

“Testing with personal data is difficult to reconcile with the GDPR”

Mag ik testen met persoonsgegevens bij de ontwikkeling van een systeem of applicatie?

Dat is niet aan te raden. Testen is een complex proces, waarvoor zorgvuldigheid en meerdere gescheiden omgevingen nodig zijn. Het testen met persoonsgegevens brengt namelijk risico's met zich mee.

Aparte grondslag

De mensen van wie u persoonsgegevens verwerkt, verwachten niet dat u hun gegevens ook voor testdoeleinden gaat gebruiken. Dat betekent onder meer dat u voor het testen **een aparte grondslag moet hebben**.

Niet noodzakelijk

Verder is **het vaak niet noodzakelijk om te testen met persoonsgegevens, omdat er meestal alternatieven mogelijk zijn.**

Dat is een van de redenen dat testen met persoonsgegevens moeilijk in overeenstemming te brengen is met de AVG.

Classic 'anonymization' is no solution

Deleting direct personal identifiers

Original data

Name	Age	Gender	Item	Price	Data
Olivia	26	Female	Shoes	€125	4 March
John	75	Male	Laptop	€695	5 March
George	41	Male	Beer	€4	7 March
...
George	41	Male	Shirt	€25	9 March

N=100k

Classic anonymization

Name	Age	Gender	Item	Price	Data
xxx					
xxx					
xxx					
...					
xxx					

N=100k

Classic 'anonymization' is no solution

Indirect information is aggregated

Original data

Name	Age	Gender	Item	Price	Data
Olivia	26	Female	Shoes	€125	4 March
John	75	Male	Laptop	€695	5 March
George	41	Male	Beer	€4	7 March
...
George	41	Male	Shirt	€25	9 March

N=100k



Classic anonymization

Name	Age	Gender	Item	Price	Data
xxx	25-30				
xxx	70-75				
xxx	40-45				
...	...				
xxx	40-45				

N=100k



Classic 'anonymization' is no solution

And continue...

Original data

Name	Age	Gender	Item	Price	Data
Olivia	26	Female	Shoes	€125	4 March
John	75	Male	Laptop	€695	5 March
George	41	Male	Beer	€4	7 March
...
George	41	Male	Shirt	€25	9 March

N=100k



Classic anonymization

Name	Age	Gender	Item	Price	Data
xxx	25-30	Female	Cloth	€100 - €200	March
xxx	70-75	Male	IT	€600 - €700	March
xxx	40-45	Male	Drink	<€5	March
...
xxx	40-45	Male	Cloth	€20 - €30	March

N=100k

Classic 'anonymization' is no solution

Original data

Name	Age	Gender	Item	Price	Data
Olivia	26	Female	Shoes	€125	4 March
John	75	Male	Laptop	€695	5 March
George	41	Male	Beer	€4	7 March
...
George	41	Male	Shirt	€25	9 March

N=100k

Classic anonymization



Name	Age	Gender	Item	Price	Data
xxx	25-30	Female	Cloth	€100 - €200	March
xxx	70-75	Male	IT	€600 - €700	March
xxx	40-45	Male	Drink	<€5	March
...
xxx	40-45	Male	Cloth	€20 - €30	March

N=100k

✗ Always a privacy risk

✗ Destroys data

✗ Time consuming

So what *is* allowed?



AUTORITEIT
PERSOONSgegevens

Welke gegevens kan ik wel gebruiken om testen uit te voeren?

U kunt bijvoorbeeld onderzoeken of er **synthetische gegevens** of testdata ('dummy data') beschikbaar zijn. Stel daarbij altijd vast dat de dataset die u wilt gebruiken niet alsnog persoonsgegevens bevat.

De Rijksdienst voor Identiteitsgegevens biedt bijvoorbeeld een reeks **test-burgerservicenummers** aan.

Wilt u testen of een nieuw systeem of een nieuwe applicatie dezelfde uitkomsten genereert als het oude systeem of de oude applicatie? Het kan zijn dat u in beide applicaties vaak testdata

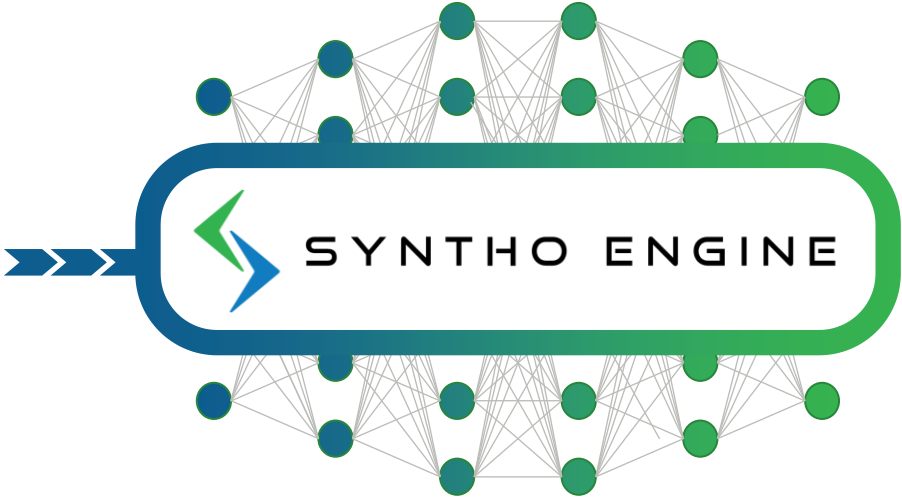
"You can explore the availability of **synthetic data or **mock data**"**

A synthetic data twin of your original data

Original data

Name	Age	Gender	Item	Price	Data
Olivia	26	Female	Shoes	€125	4 March
John	75	Male	Laptop	€695	5 March
George	41	Male	Beer	€4	7 March
...
George	41	Male	Shirt	€25	9 March

N=100k



Generate a synthetic data twin of your original data

Original data

Name	Age	Gender	Item	Price	Data
Olivia	26	Female	Shoes	€125	4 March
John	75	Male	Laptop	€695	5 March
George	41	Male	Beer	€4	7 March
...
George	41	Male	Shirt	€25	9 March

N=100k



Synthetic Data Twin

Name	Age	Gender	Item	Price	Data
NewID1	23	Male	Sofa	€790	1 March
NewID2	23	Female	Scarf	€40	3 March
NewID3	52	Male	Razor	€5	9 March
...
NewIDn	35	Male	Wine	€7	7 March

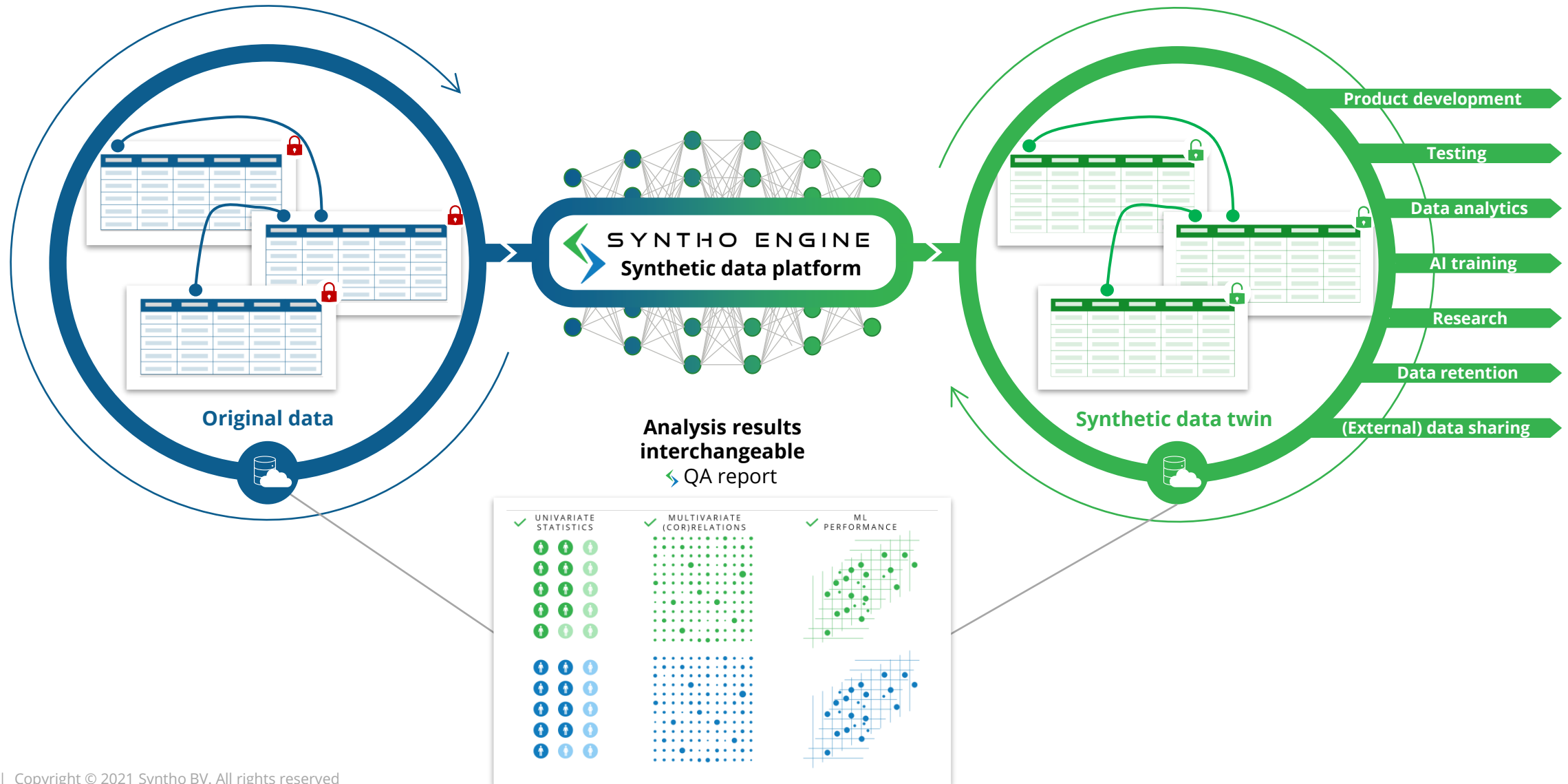
N=100k

✓ No privacy risk

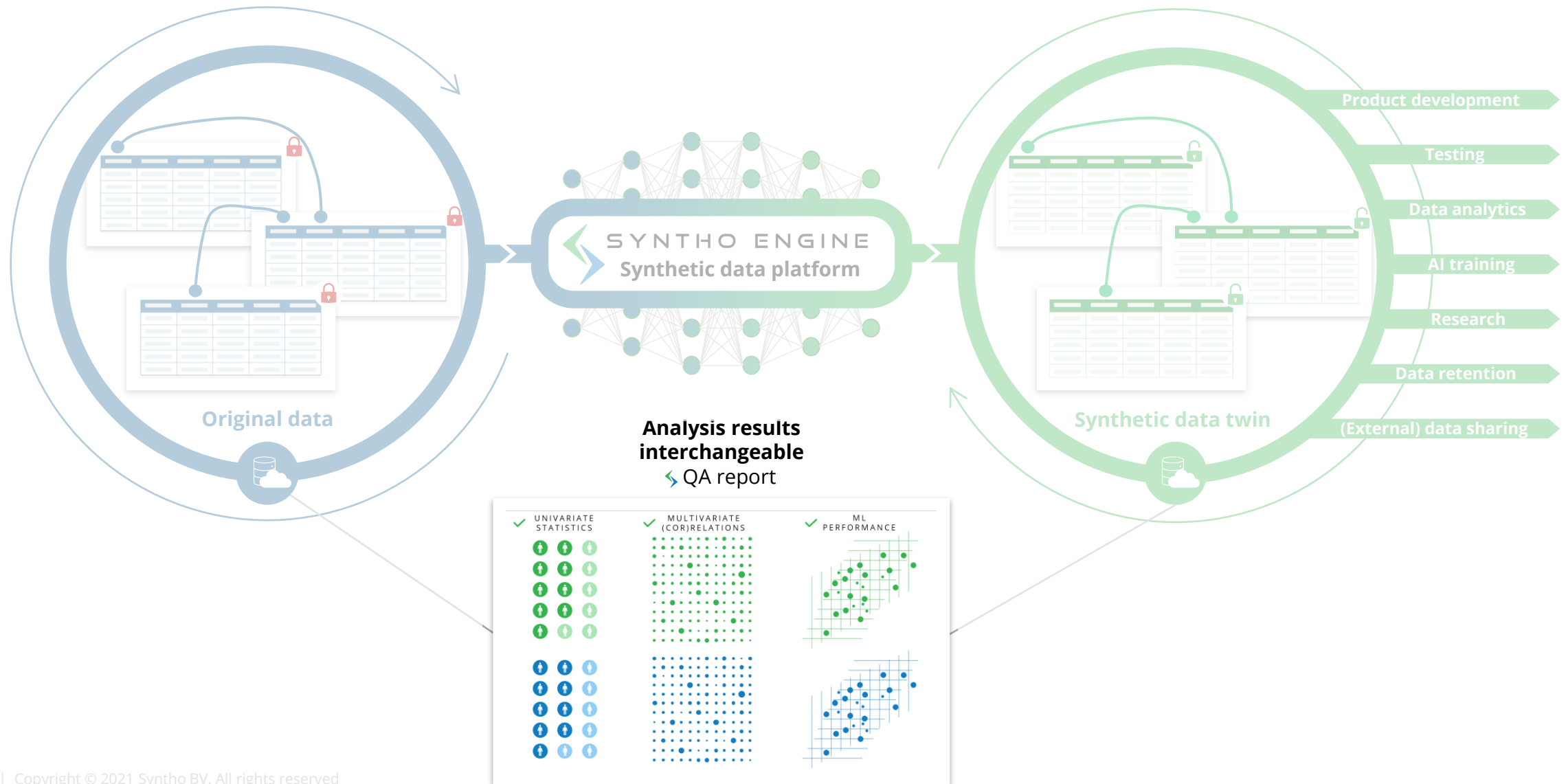
✓ As-good-as real

✓ Easy, fast, scalable

Generate a synthetic data twin of your original data



Generate a synthetic data twin of your original data

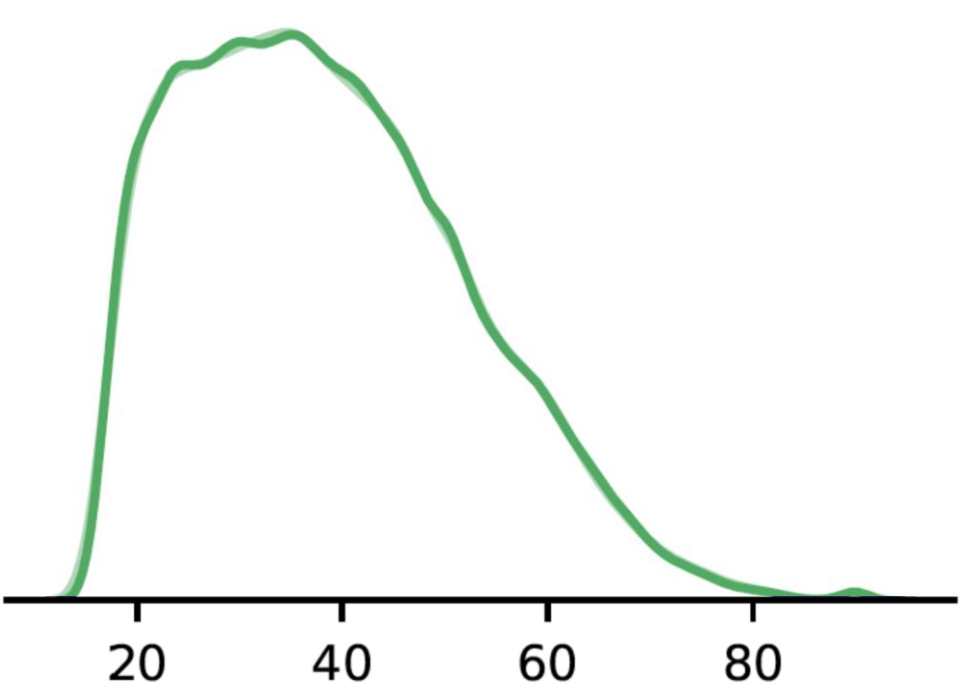


Automatic QA report that assess on Privacy, Accuracy and speed

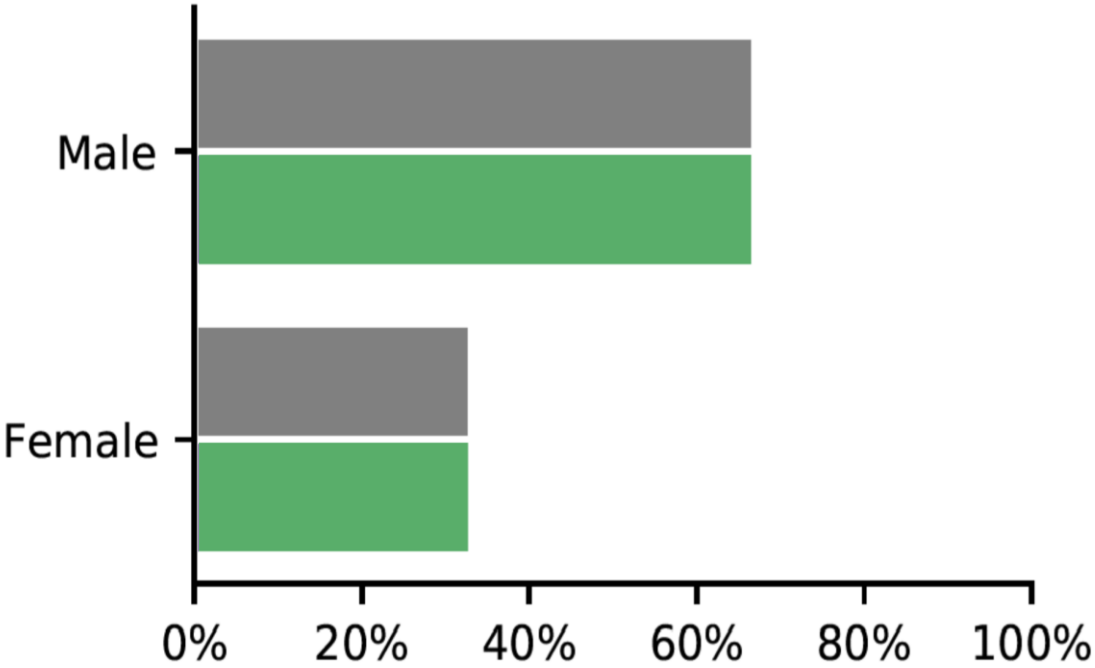


As-good-as-real synthetic data

Age



Gender

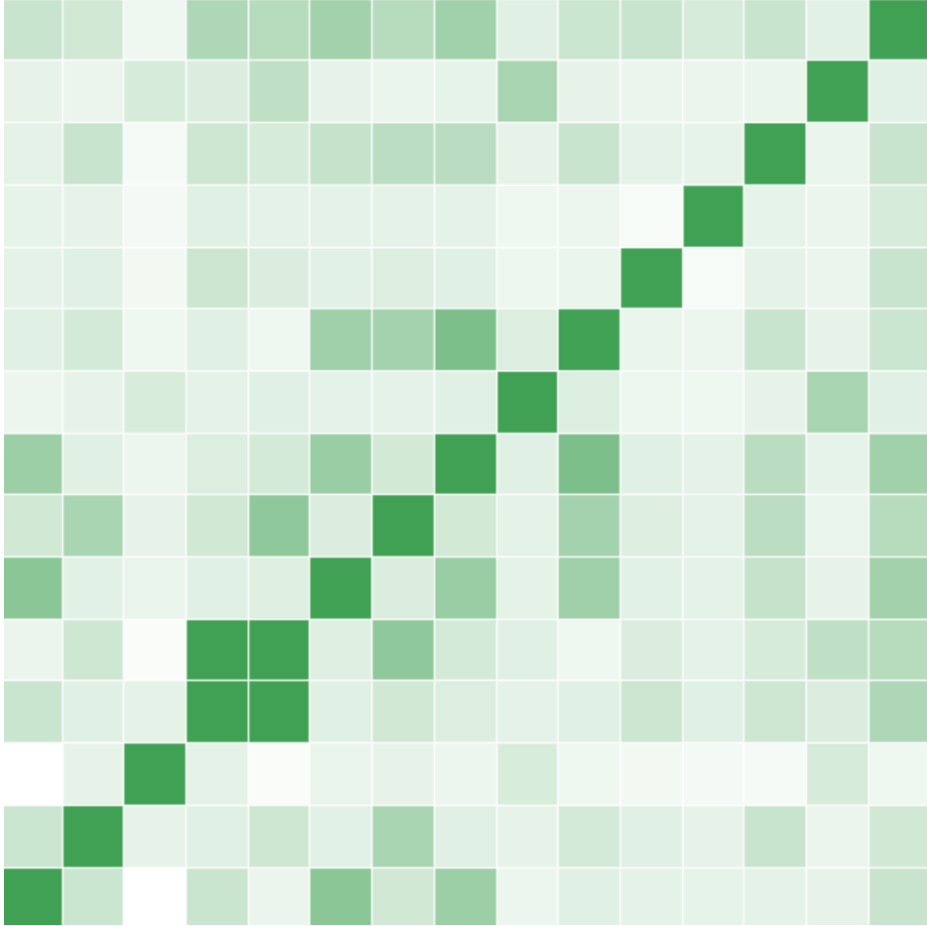


Original data
Synthetic data

As-good-as-real synthetic data

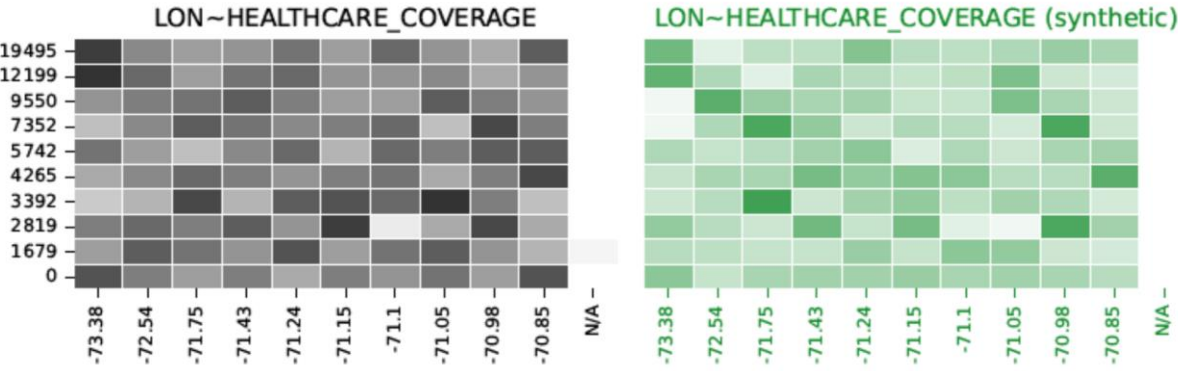
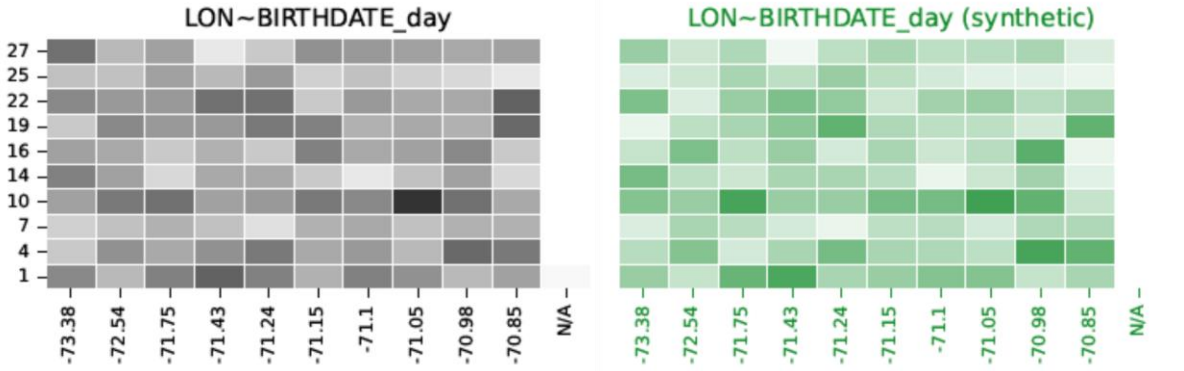
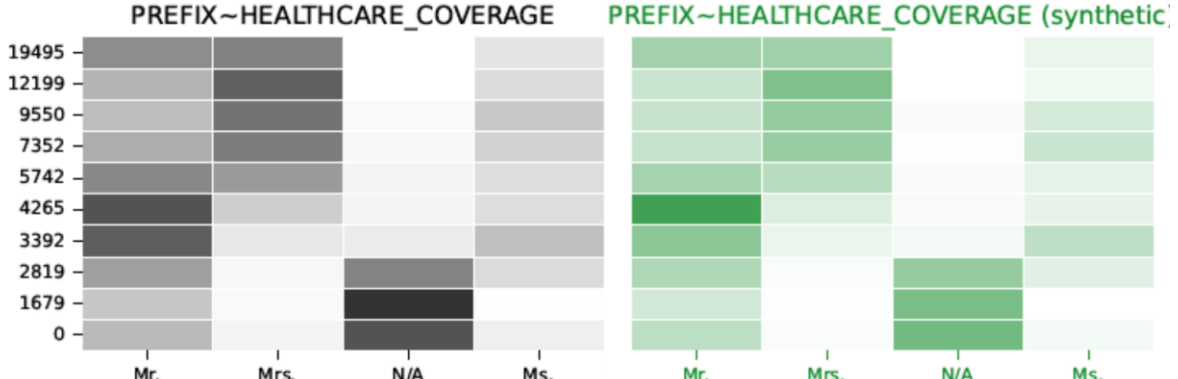
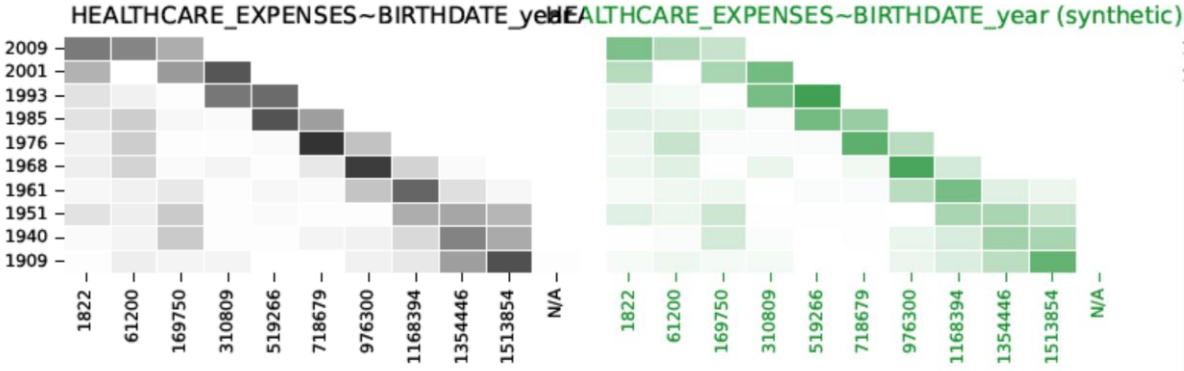


Original data



Synthetic data

As-good-as-real synthetic data



Original data

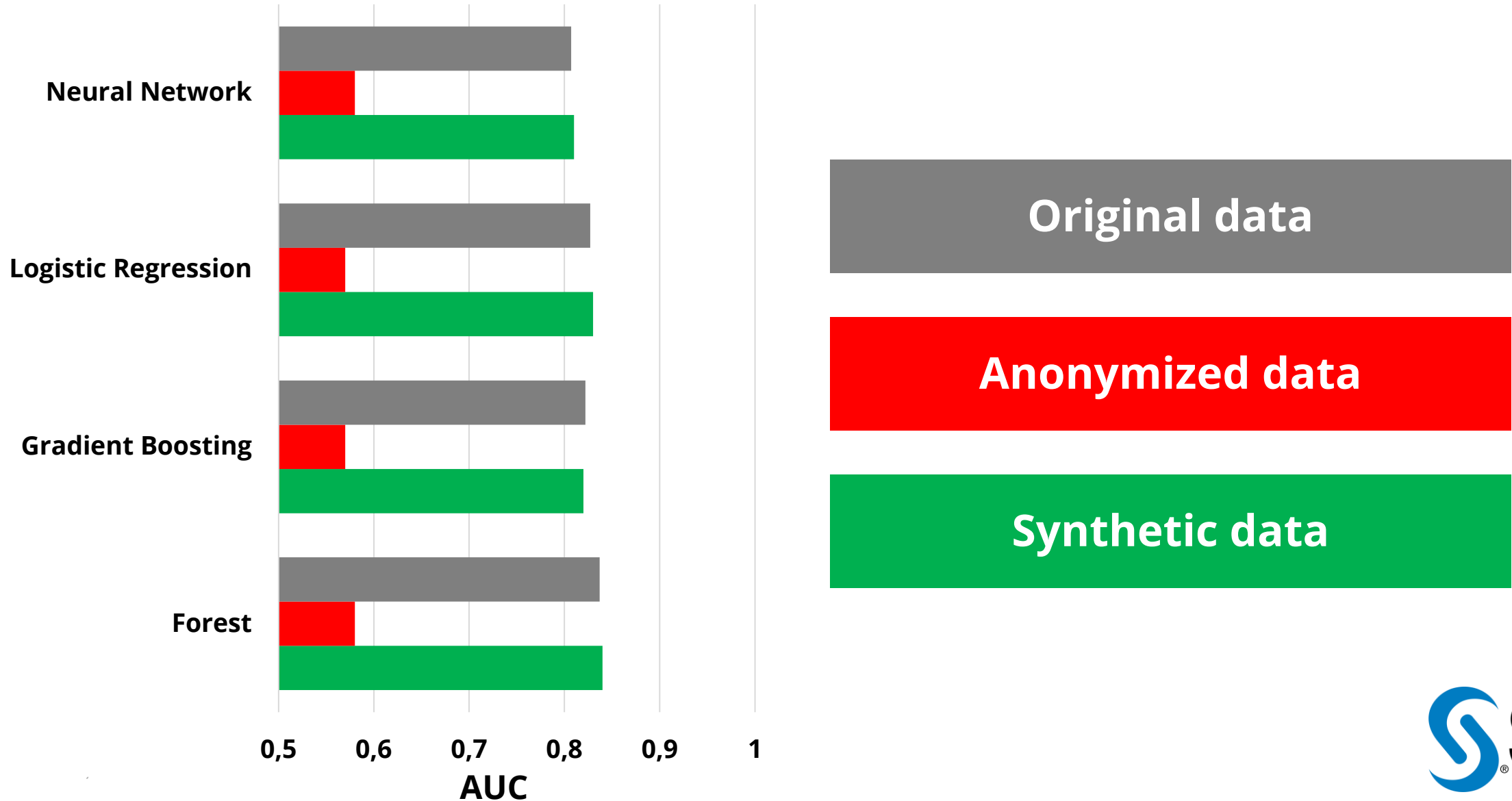
Synthetic data

Synthetic data **approved** by the data experts of **SAS**

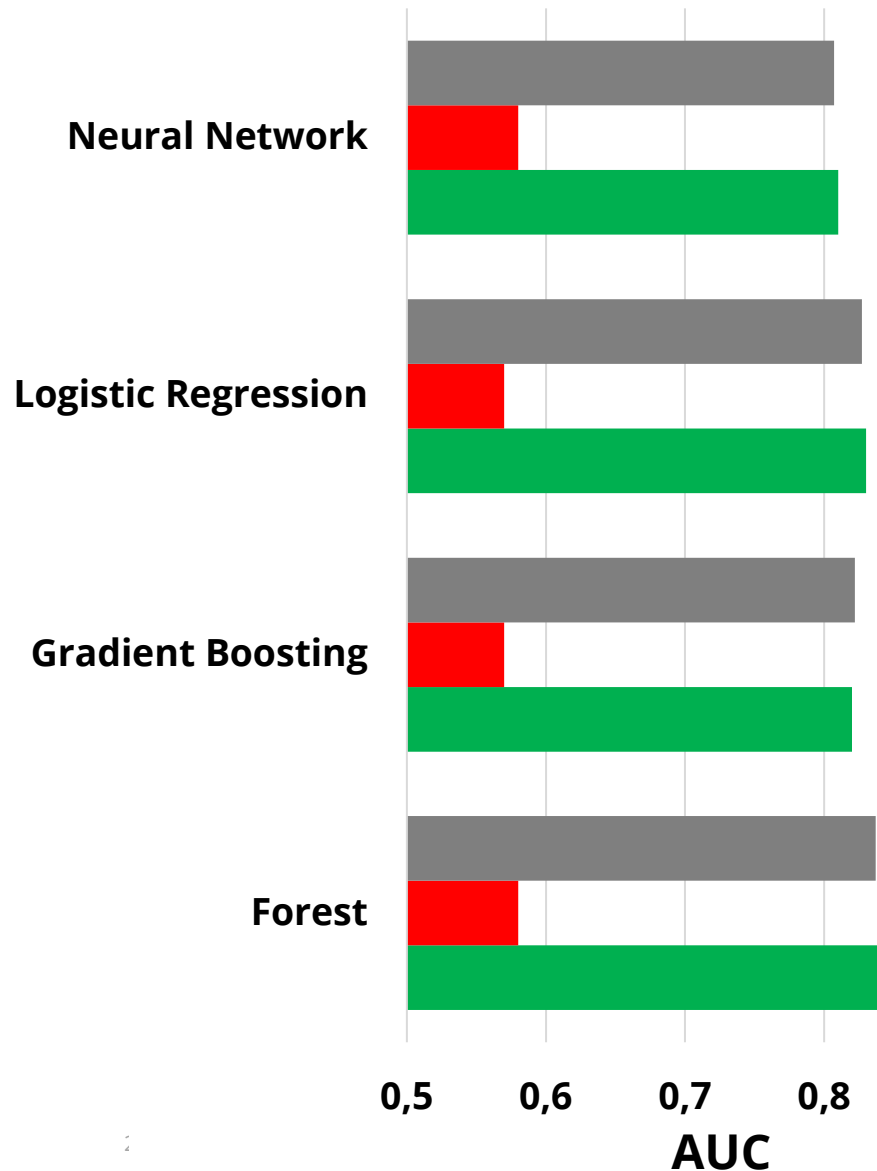


The **data experts**
from **SAS**
approved our AI
generated
synthetic data.

Synthetic data **approved** by the data experts of **SAS**

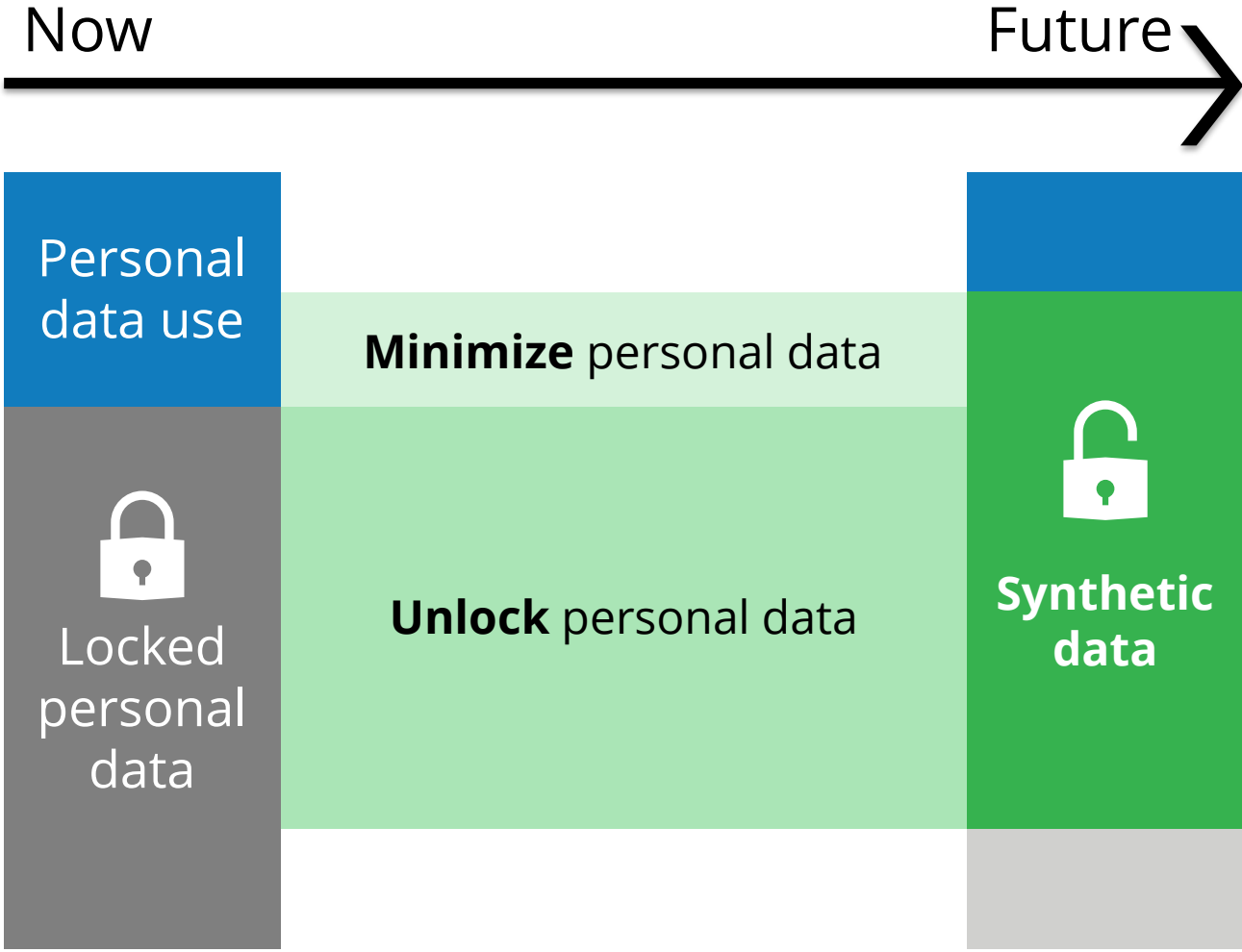


Synthetic data **approved** by the data experts of **SAS**



- **Synthetic data** is as-good-as the original
- **Anonymized data** shows worst performance
- **Generation is easy, fast and scalable**

Why use real data when you can use synthetic data?



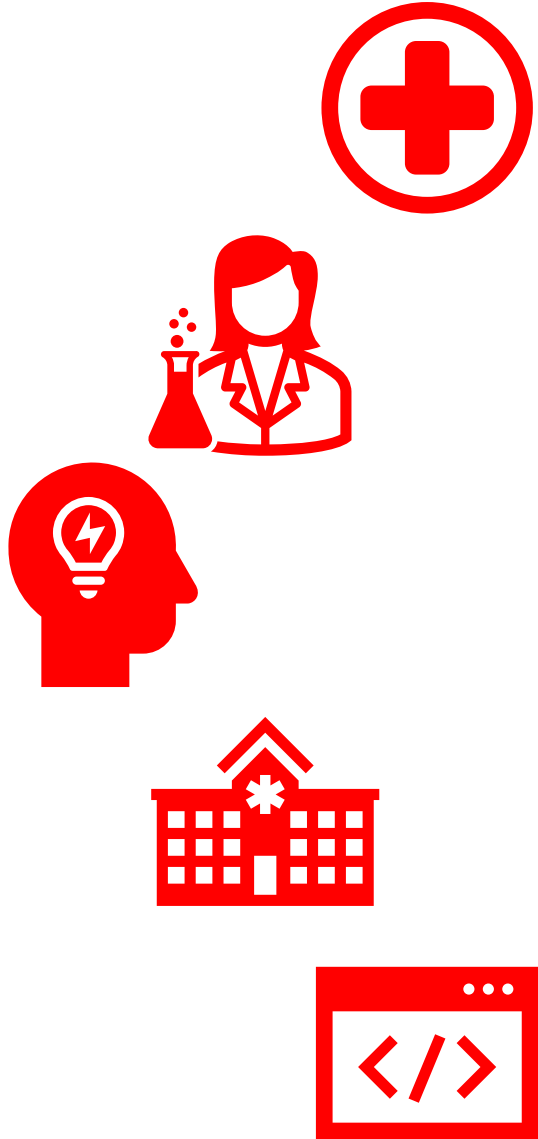
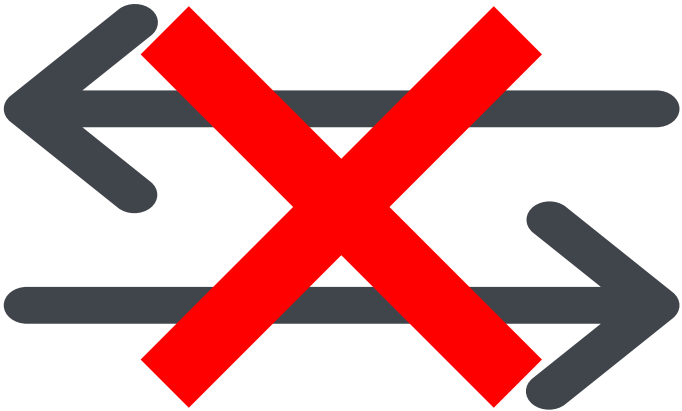
Key benefits

- 1. Less risk
- 2. More data
- 3. Faster data access

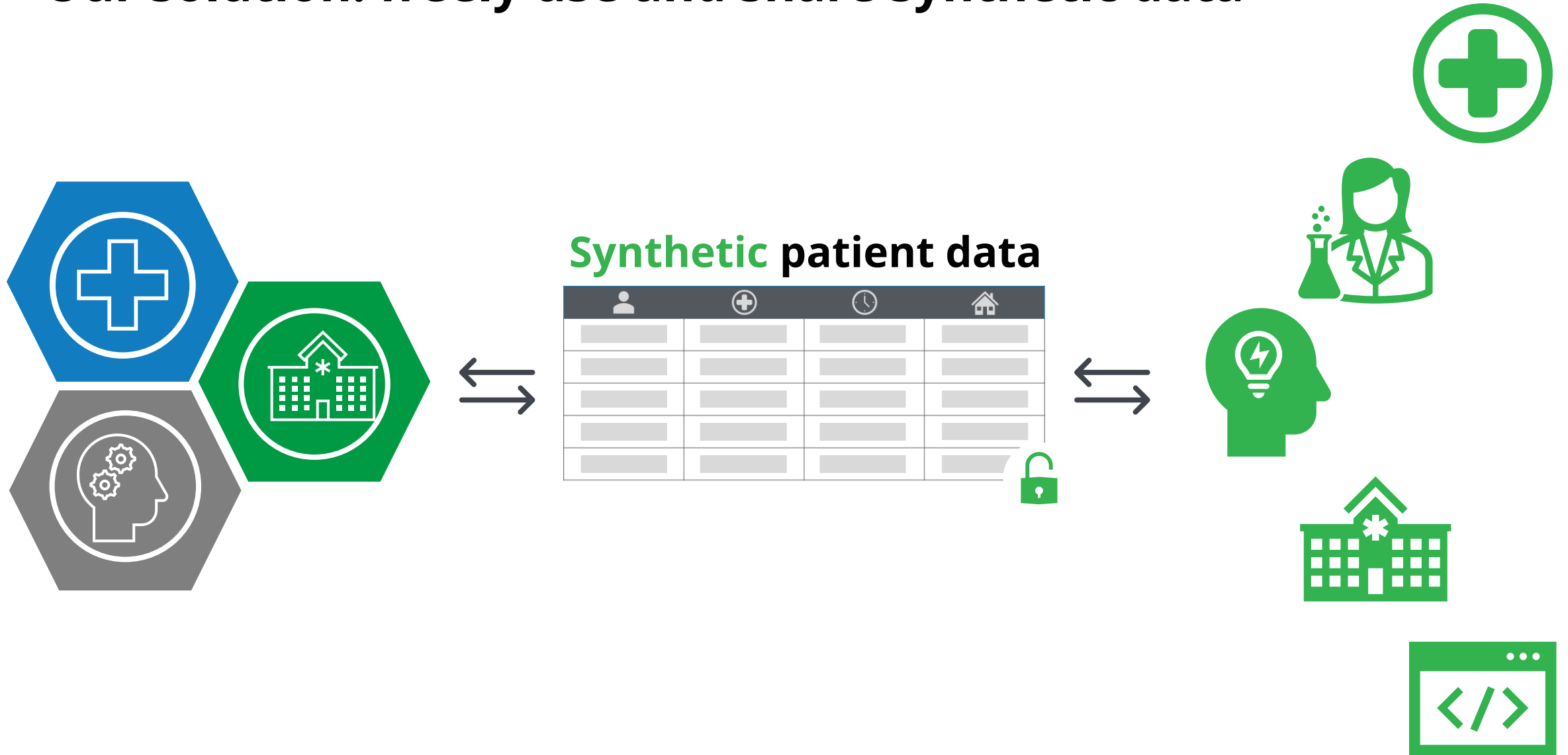


Boost Innovation

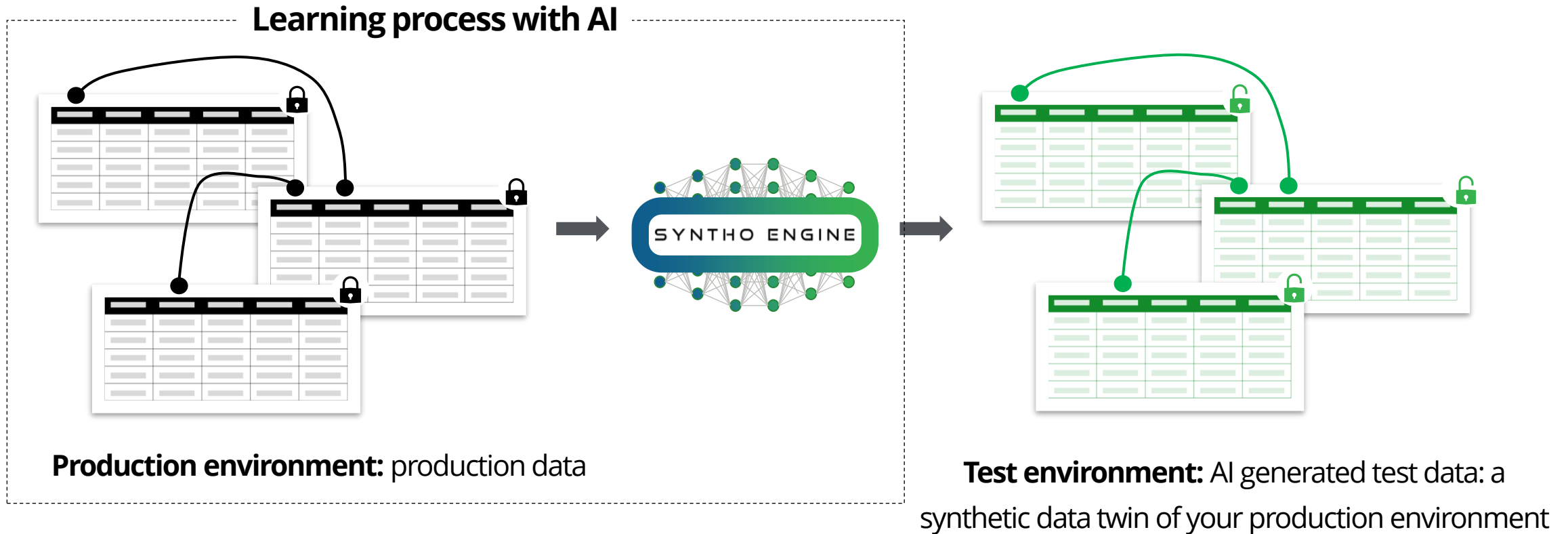
Using and sharing data is challenging



Our solution: freely use and share synthetic data



Synthetic Data as test for a large bank

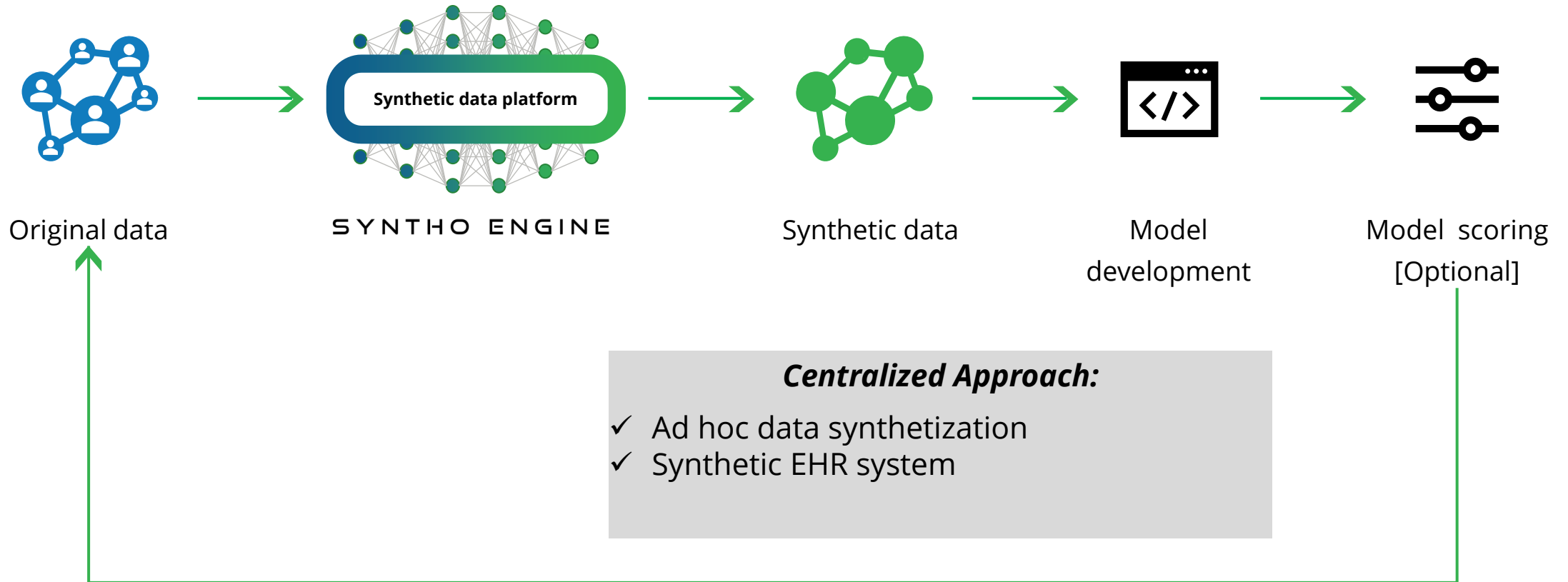


✓ Privacy-by-design

✓ Production-like data

✓ Easy, fast, scalable

Synthetic data for analytics at a leading hospital

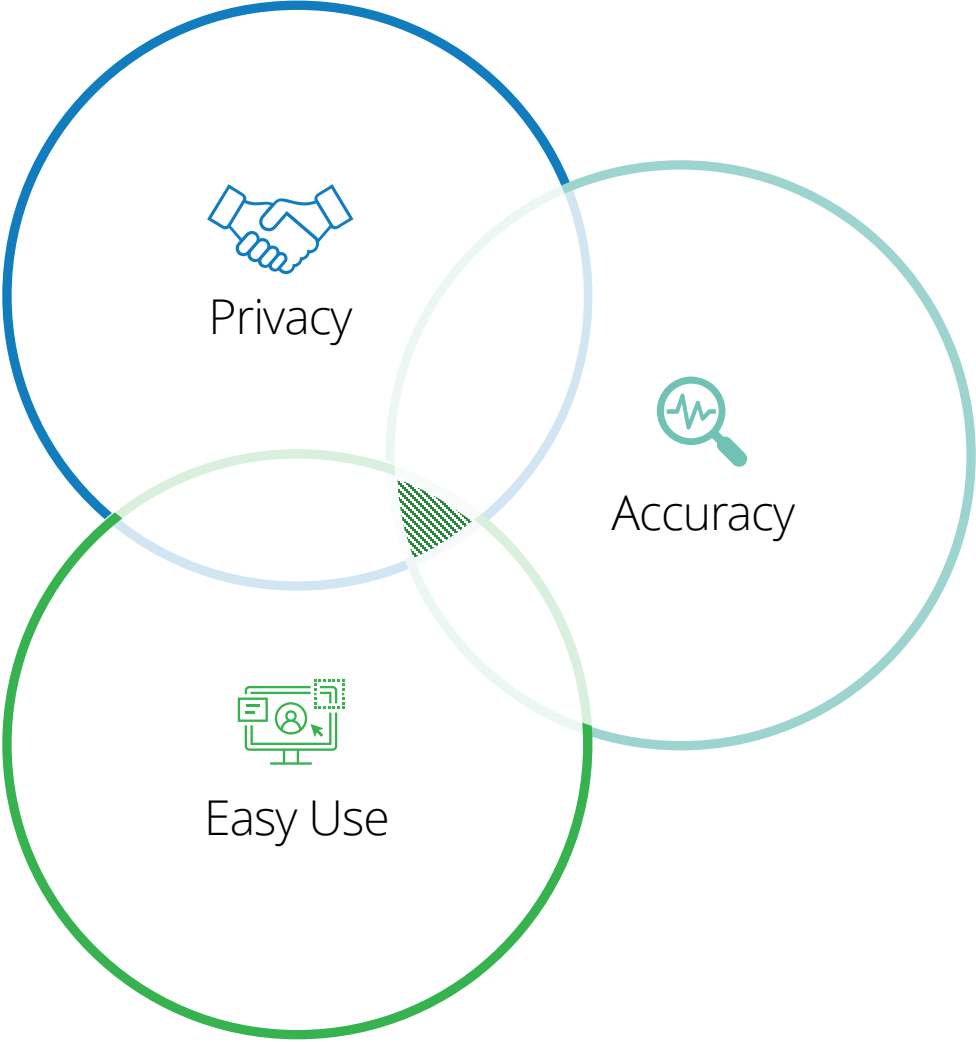
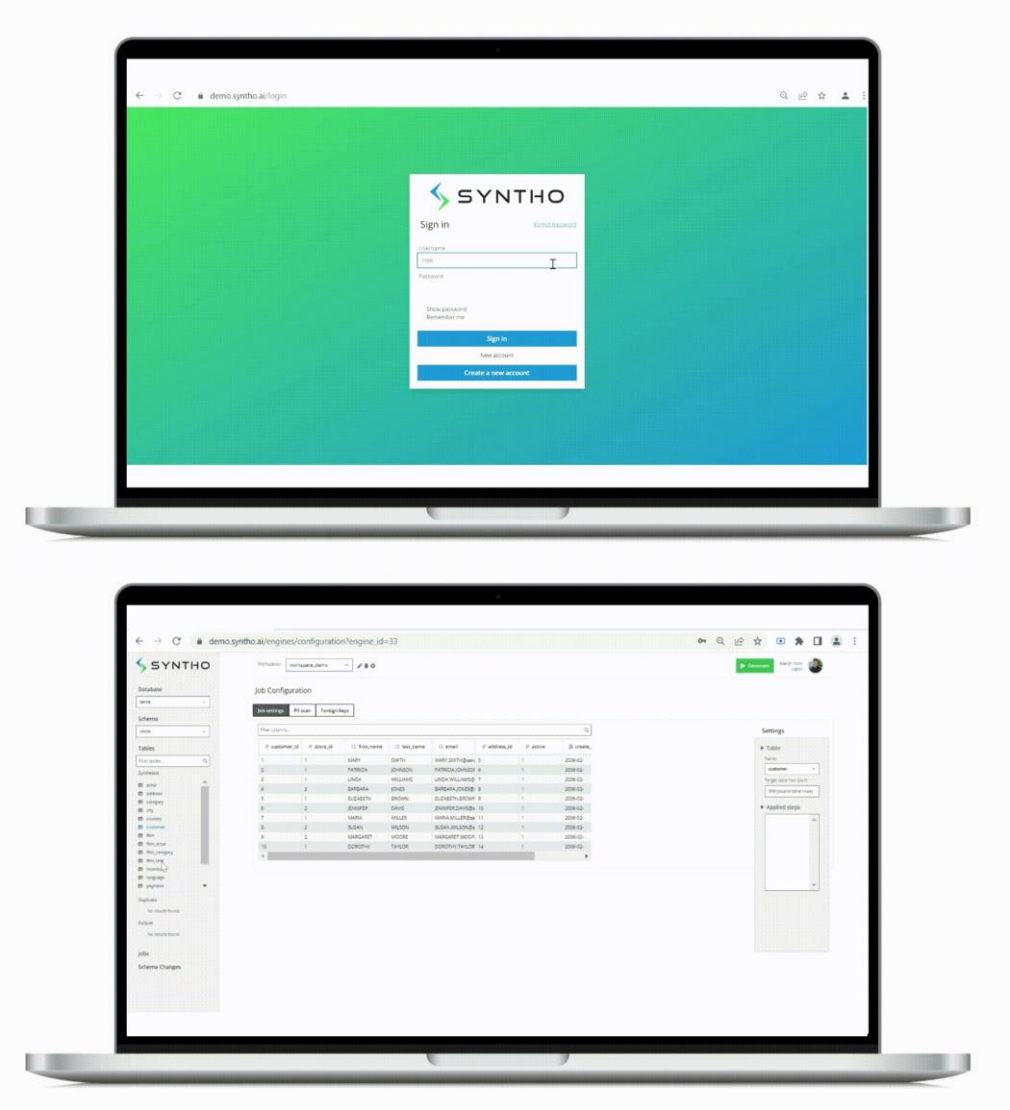


✓ **Unlock data**

✓ **As-good-as-real data**

✓ **Agile research**

AI Generated Synthetic Data at Scale



SYNTHO ENGINE 2.0

REVOLUTIONIZING AI GENERATED SYNTHETIC DATA



Syntho Engine 2.0: **Key differentiators**

Easy to use

AI-Generated Synthetic data in just 3 steps with our user-friendly self-service platform

Fastest

Fastest synthetic data generation (benchmarked against open source and commercial)

Massive data

Support for massive data with minimal computing resources

Time-series

Enhanced support for time series data / longitudinal data

Multi-table

Enhanced support for multi-table databases

Extended features

Extended features, such as: PII scanner and mockers, PII scanner in open texts, Advanced mocker, New connectors, Subsetting, Workspace sharing

50%

Syntho unlocks data

Synthetic data – Real people!



Wim Kees Janssen

Founder & CEO

Contact: kees@syntho.ai