



# Synthetic data Real impact

Unlocking Privacy-Sensitive Data with AI-Generated  
Synthetic Data in Healthcare & Pharma

*Wim Kees Janssen*  
*CEO & Founder Syntho*



# Wim Kees Janssen – Founder & CEO Syntho

- Contact: [kees@syntho.ai](mailto:kees@syntho.ai)



# About Syntho

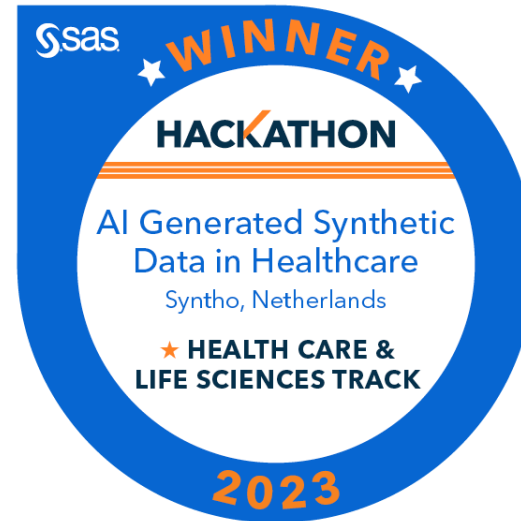


# Recognitions



Winner **Philips Innovation Award**

[Read more](#)



Winner of **SAS Global Hackathon** in healthcare & Lifescience

[Read more](#)



Shortlisted by **NVIDIA** as Generative AI startup to watch

[Read more](#)

---

# 50%

Of data is locked

---

**\$4T**

Missed data opportunities

# Turn data access into a competitive advantage

70%

**An increase in industry collaborations is expected with the use of privacy tools**

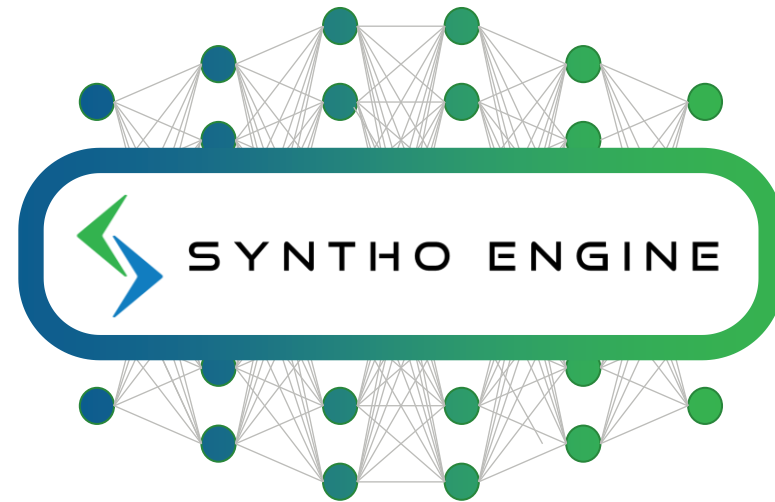
30%

**More revenue** for organizations that **earn and maintain digital trust** with customers and patients

50%

Of data will be **unlocked by privacy-enhancing techniques**

# Generative AI For Tabular Data

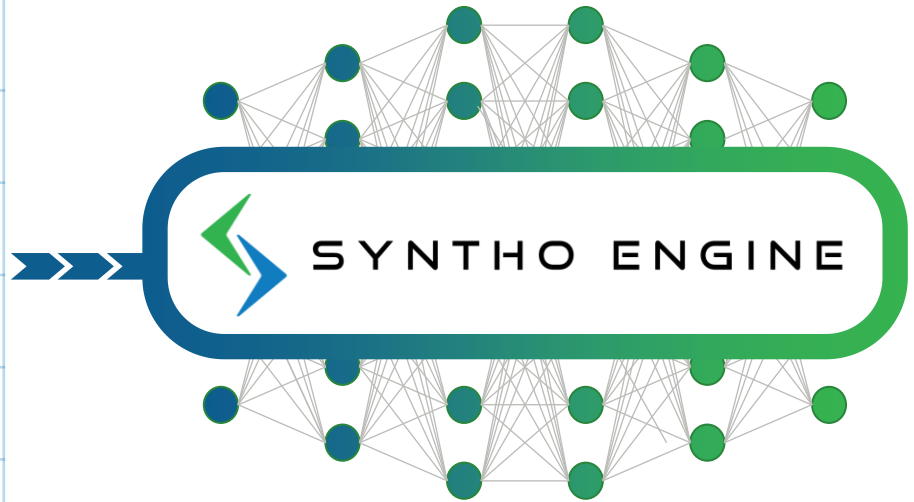




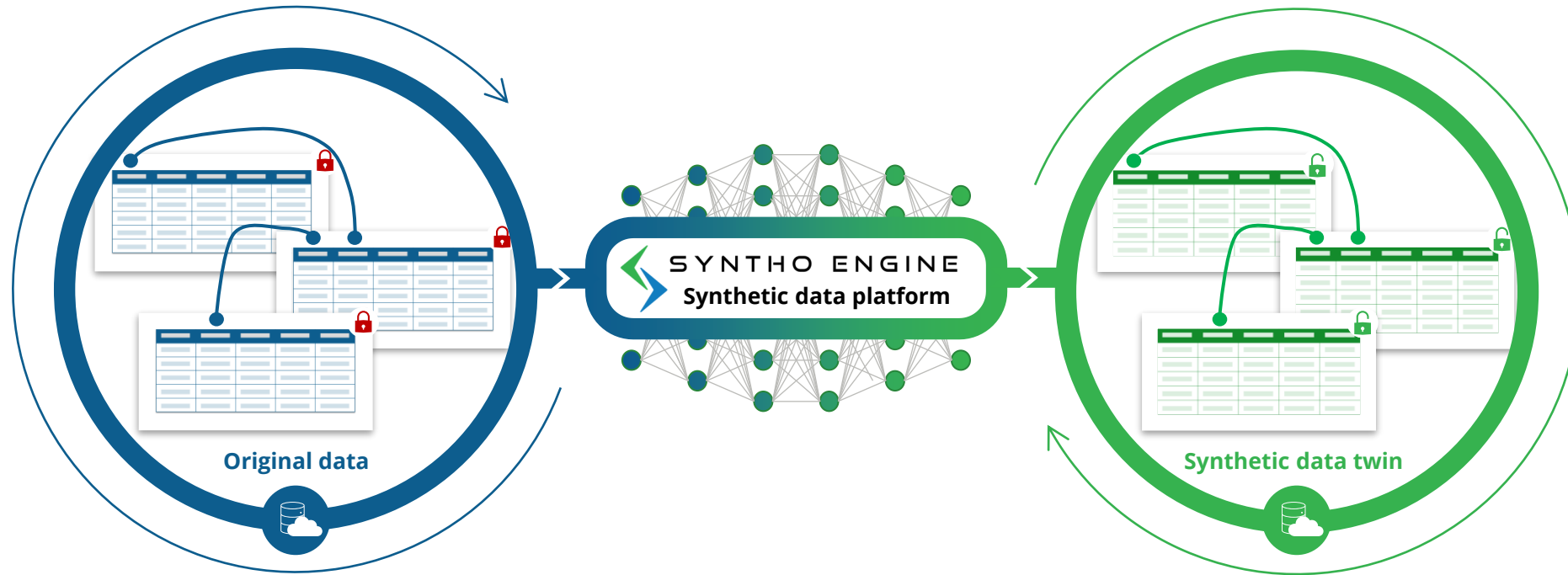
# A synthetic data twin of your original data

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**



# AI-generated synthetic twin of your original data



✓ Privacy by design

✓ High accuracy

✓ High scalability

# What do the privacy regulators say about synthetic data?

“ 

Under the **EU General Data Protection Regulation**, data synthesis would benefit from the same reasoning as the case for **not needing legal basis for anonymization**.

Article 29, EU Data Protection Work Party [IAPP 1 & 2 \(p.7\)](#)

“ 

Under the **U.S. Health Insurance Portability and Accountability Act**, synthesis would similarly be deemed to be a form of treatment, planning and operations and therefore does **not require additional authorization by a covered entity**.

US Department of Health & Services [HHS](#)

“ 

EU DPAs are open to **labeling synthetic data anonymous** under data protection standards.

[A guide to the EU's unclear anonymization standards \(2021\)](#)

# Quality evaluation of the synthetic data

## 1. Syntho's Quality Assurance (QA) Report

We provide a comprehensive quality assurance report for every synthetic data run, that demonstrates the accuracy of the synthetic data compared to the original data.

## 2. External assessment by the data experts of SAS

The data experts from SAS approved our AI-generated synthetic data for the use of model development as part of a real case study for a telecom customer.

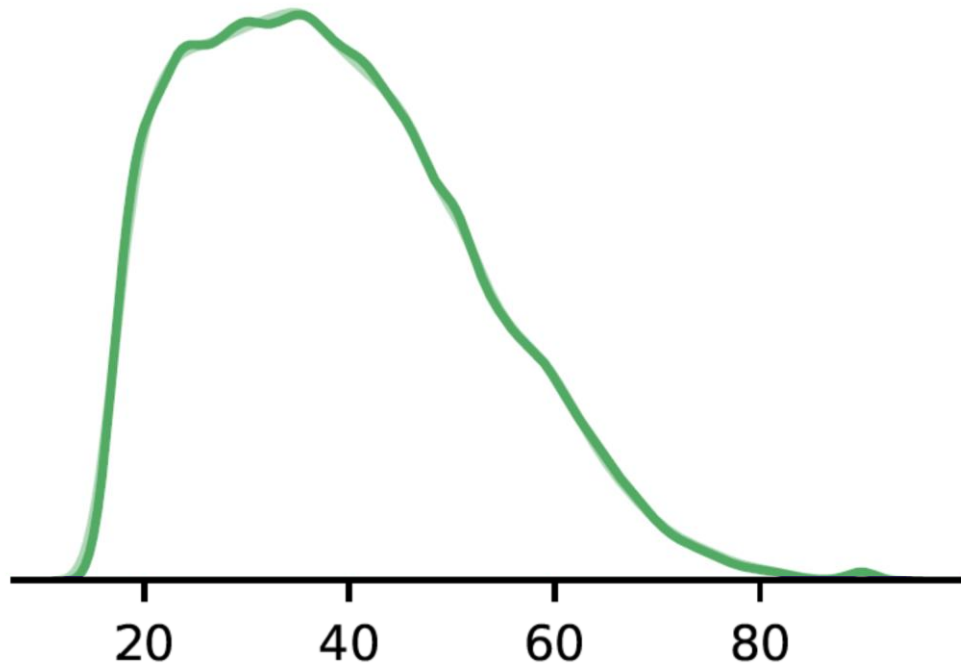


Our synthetic data is **approved**  
by the **data experts of SAS**

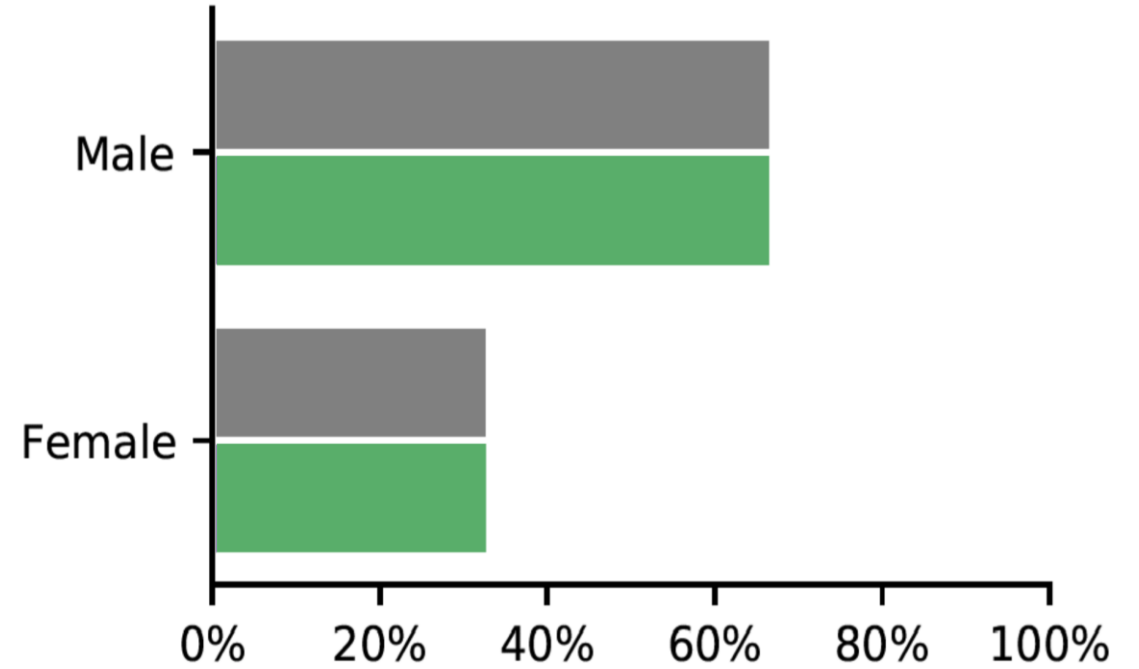


# Distributions are preserved

## Age



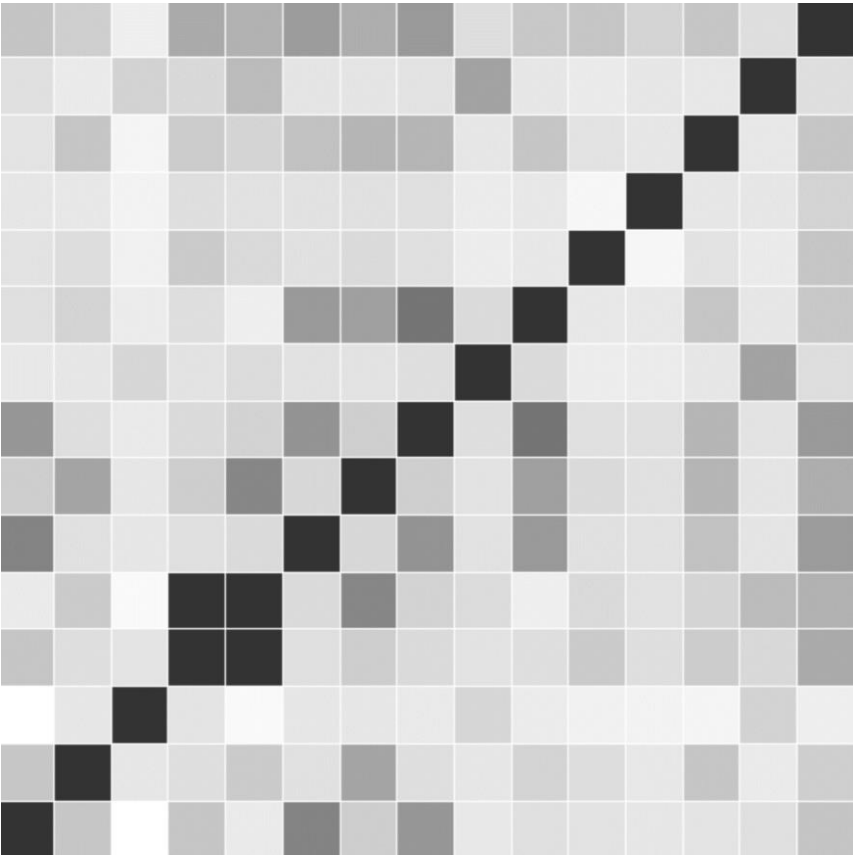
## Gender



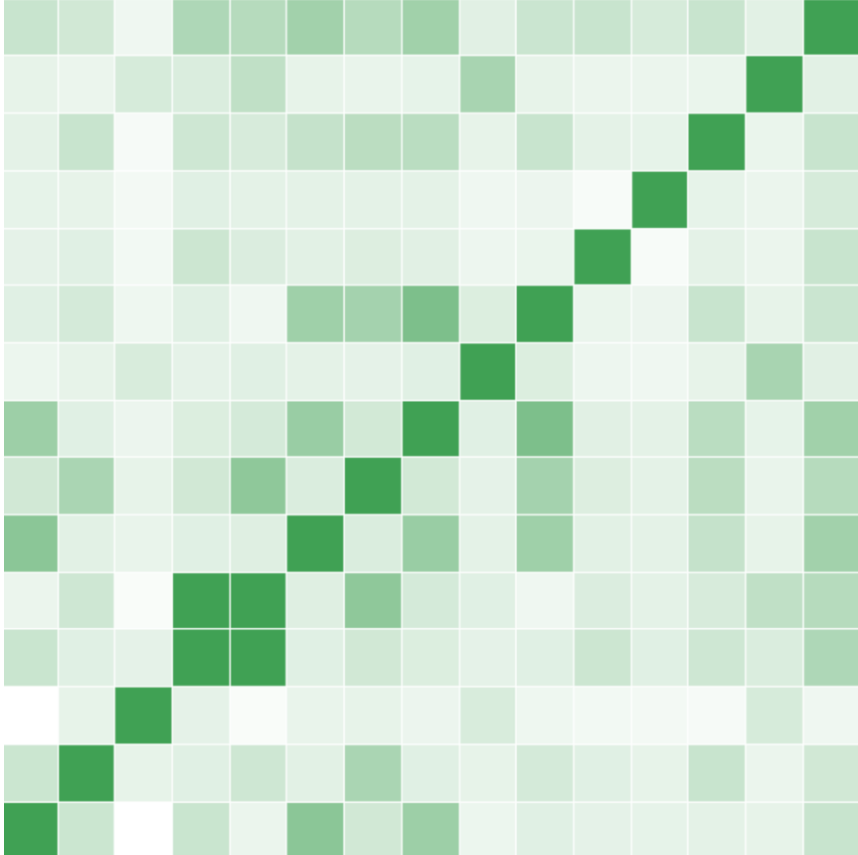
Original data

Synthetic data

# Correlations are preserved

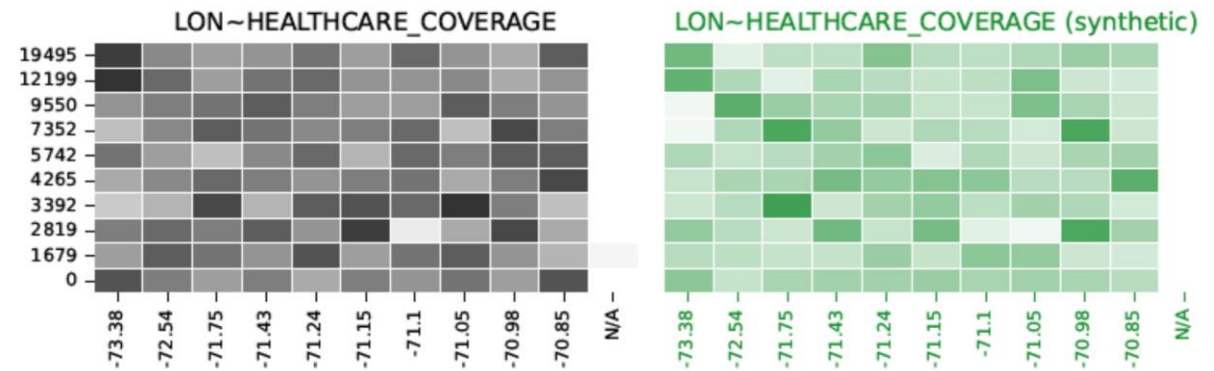
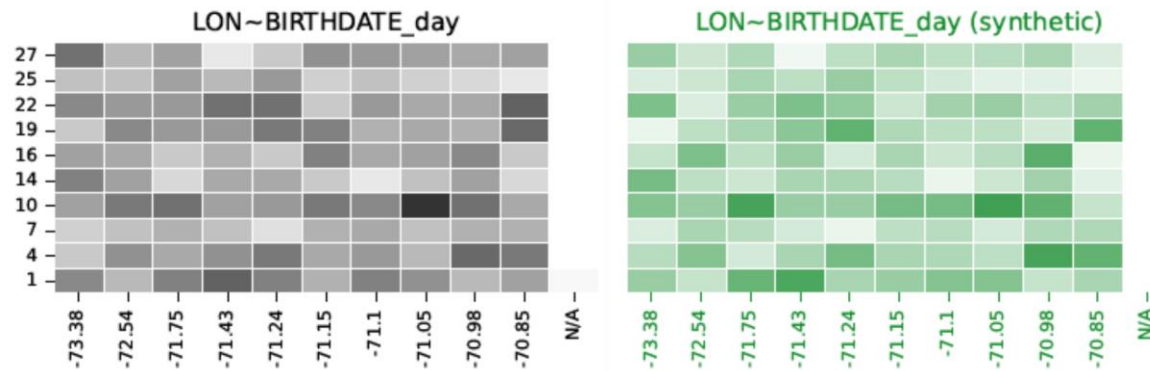
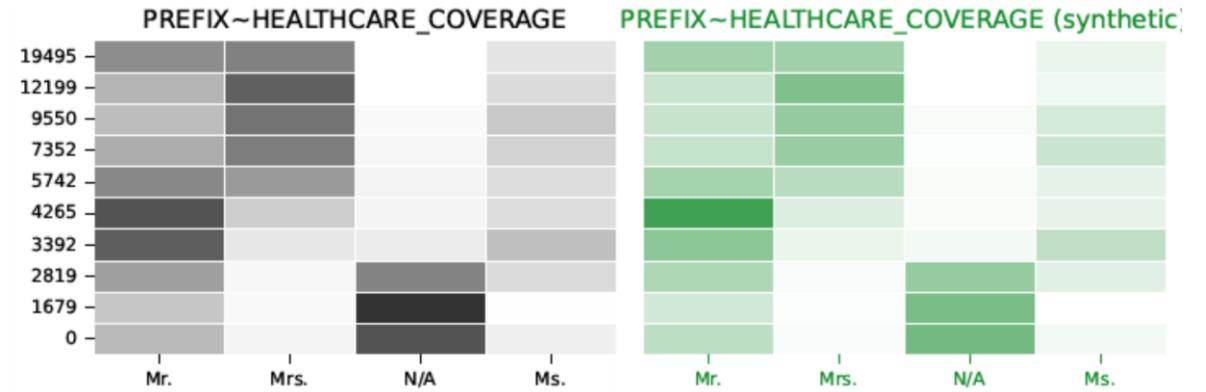
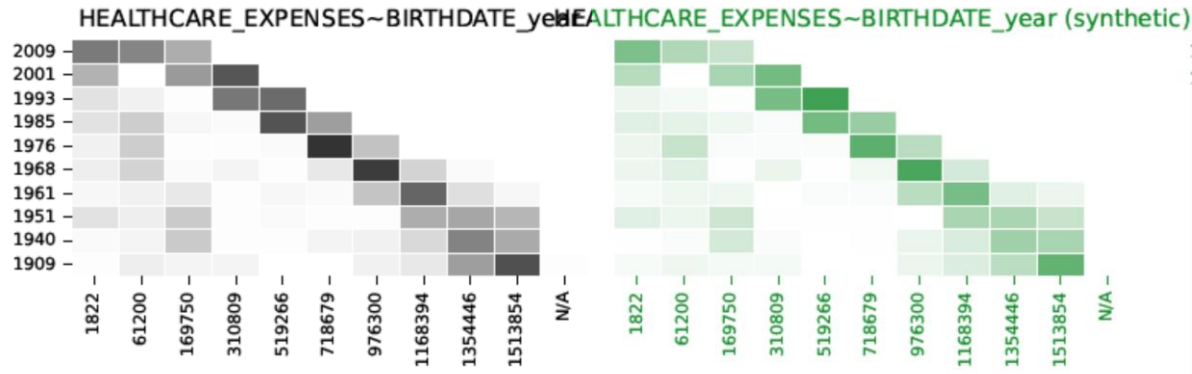


Original data



Synthetic data

# BI-variates are preserved



Original data

Synthetic data

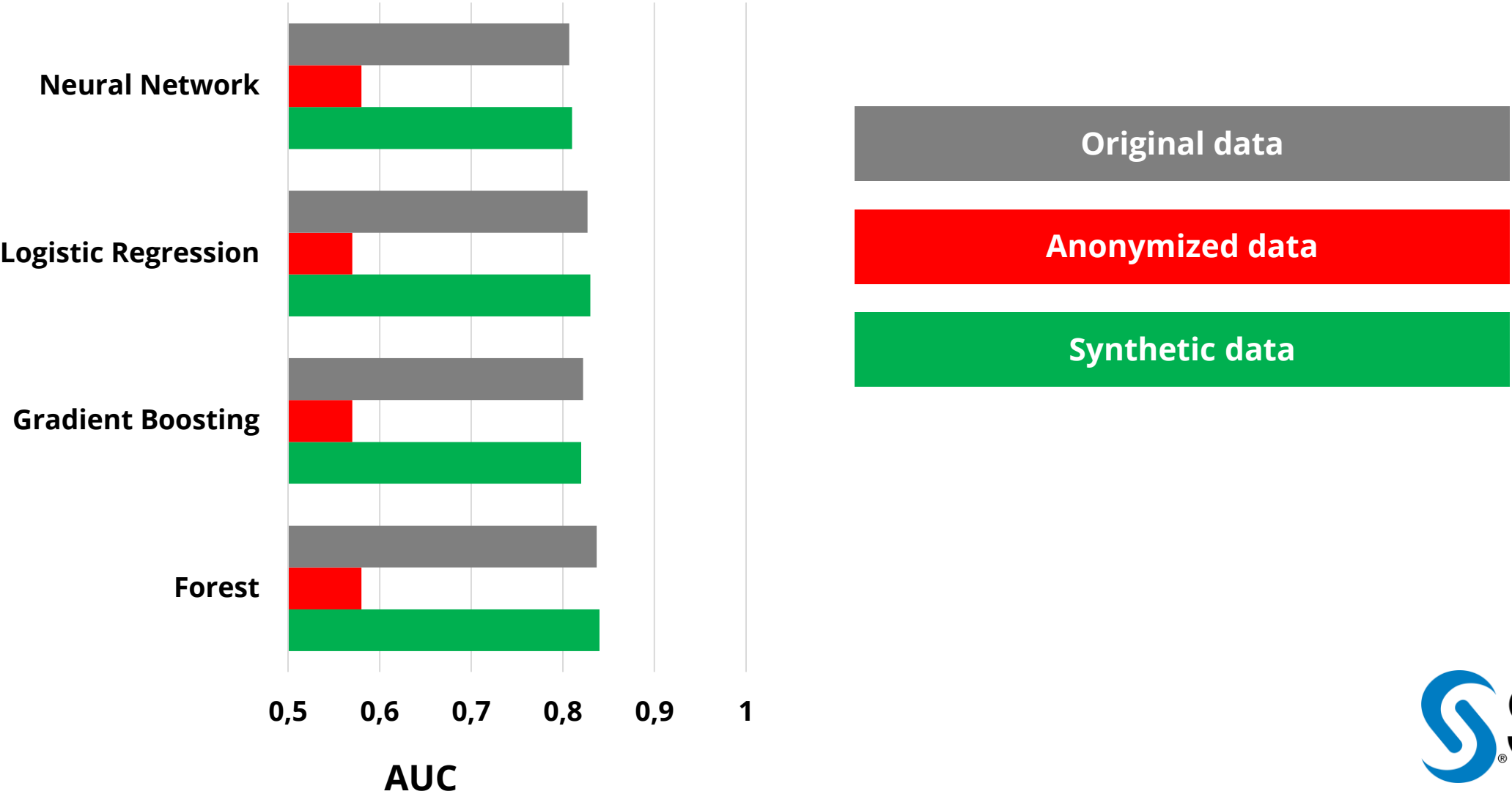
# Synthetic data **approved** by the data experts of **SAS**



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**



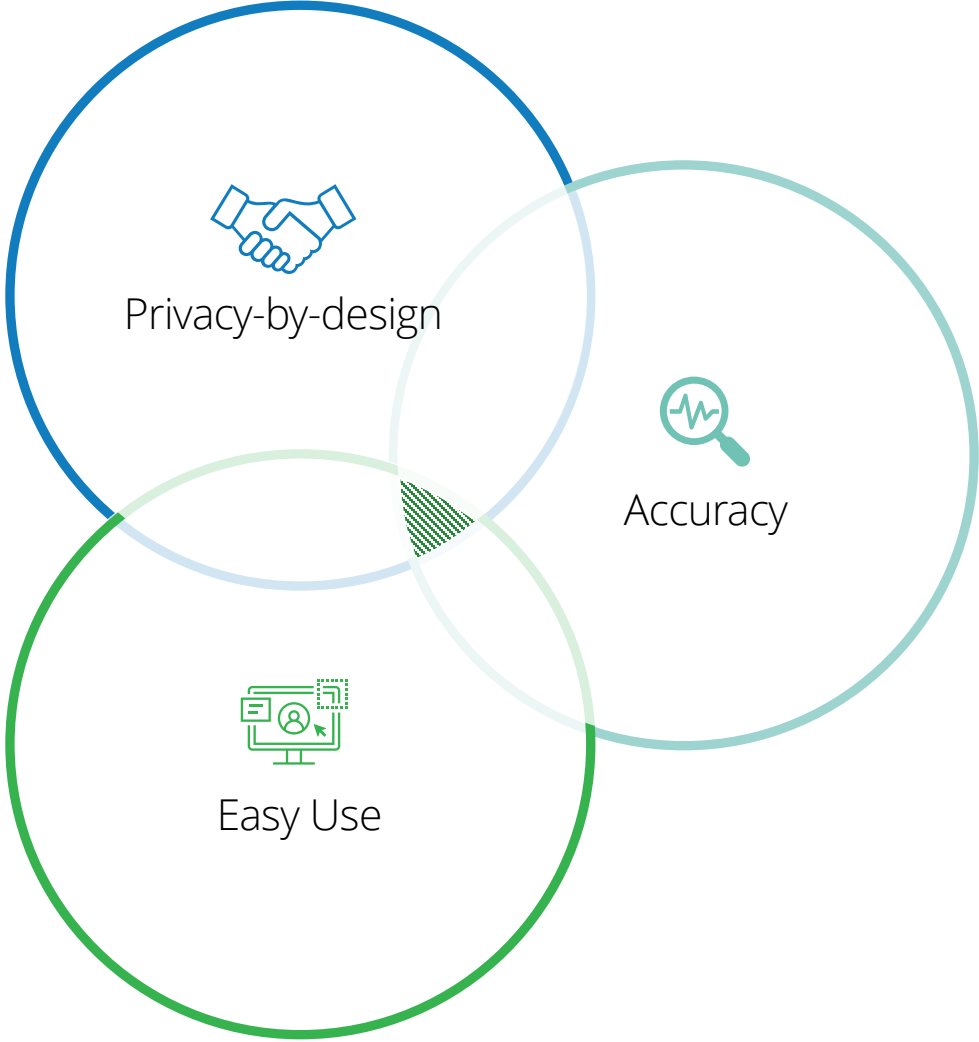
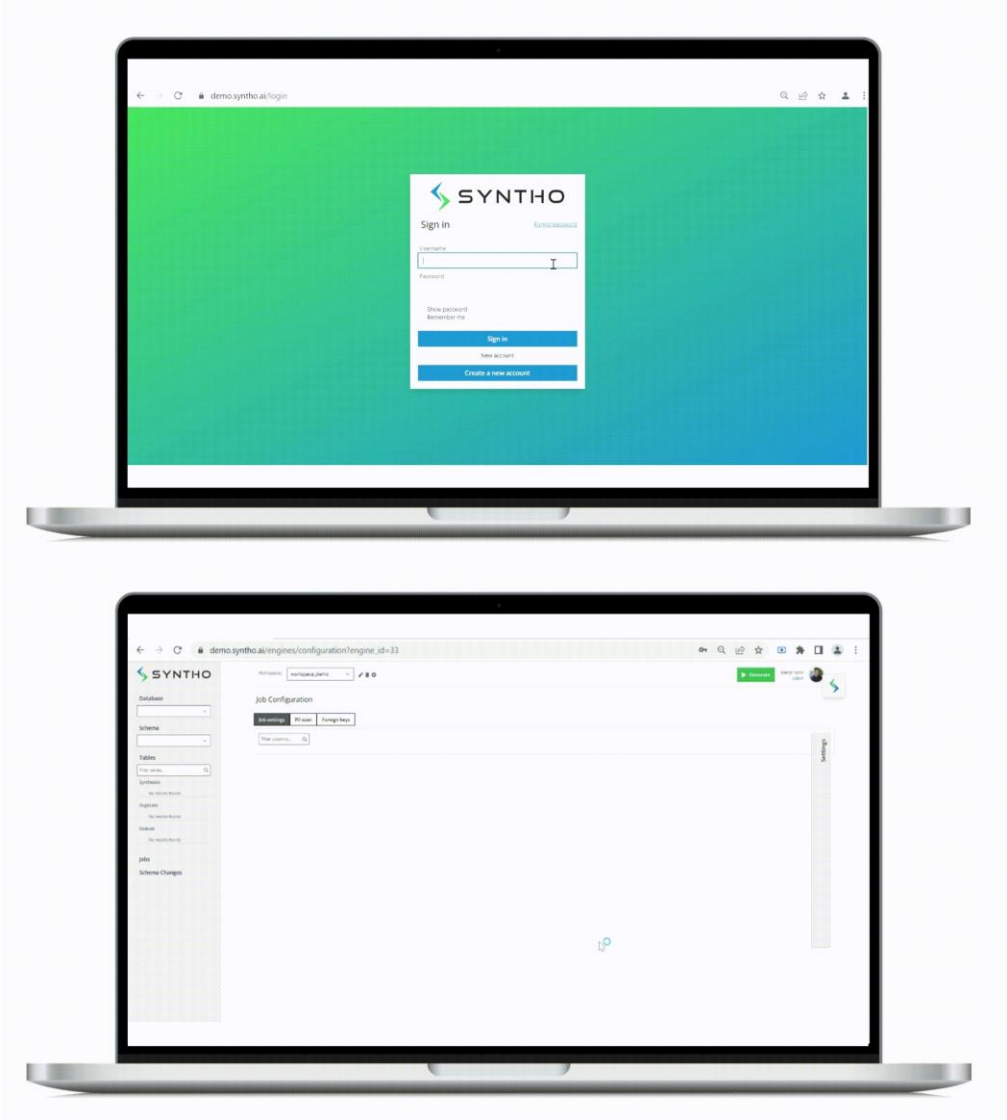
- **Synthetic data** is as-good-as the original data
- **Anonymized data** shows the worst performance
- **Easy, fast, and scalable**

# SYNTHO ENGINE 2.0

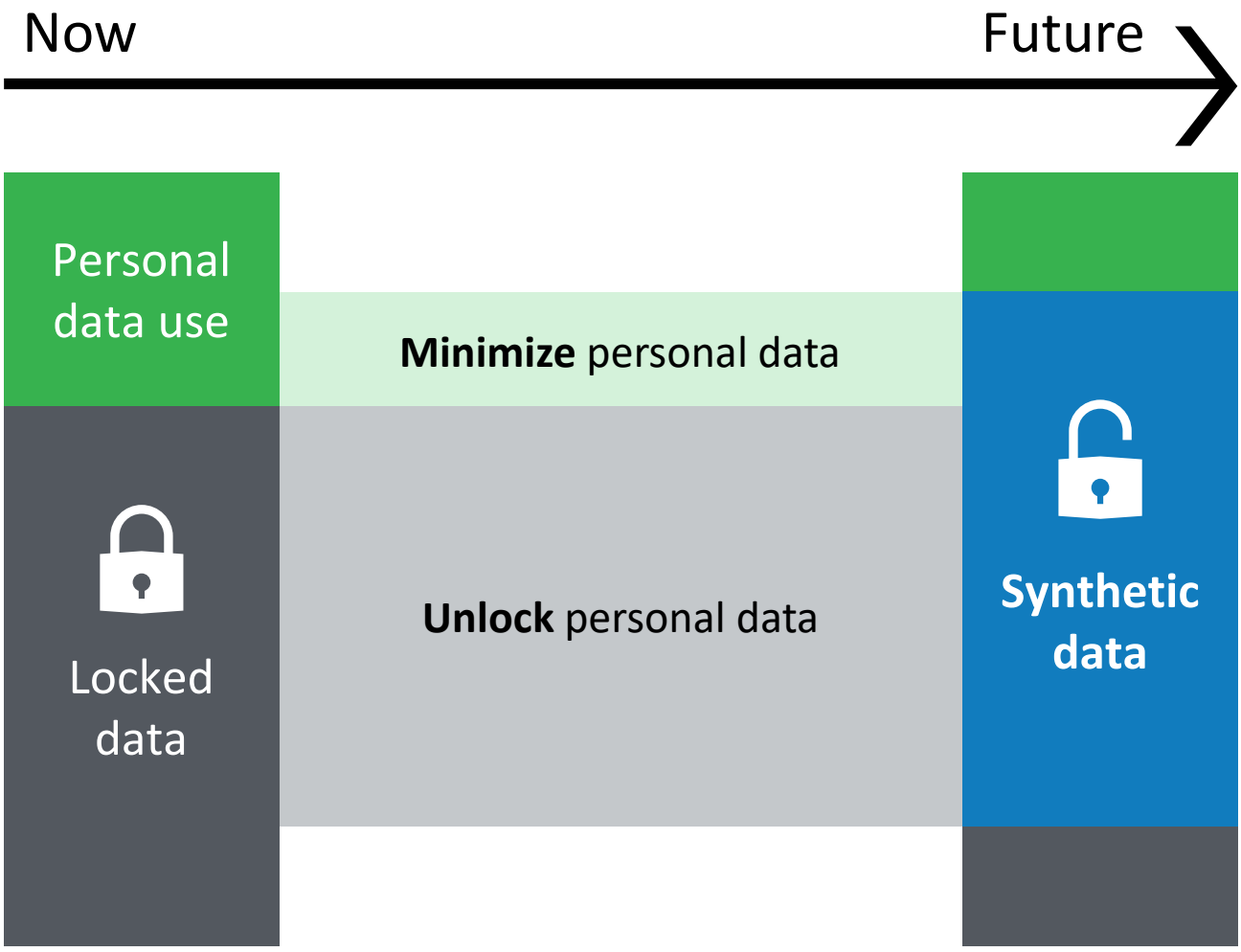
*REVOLUTIONIZING AI GENERATED SYNTHETIC DATA*



# 3 pillars of our Syntho Engine



# Why use real data when you could use synthetic data?



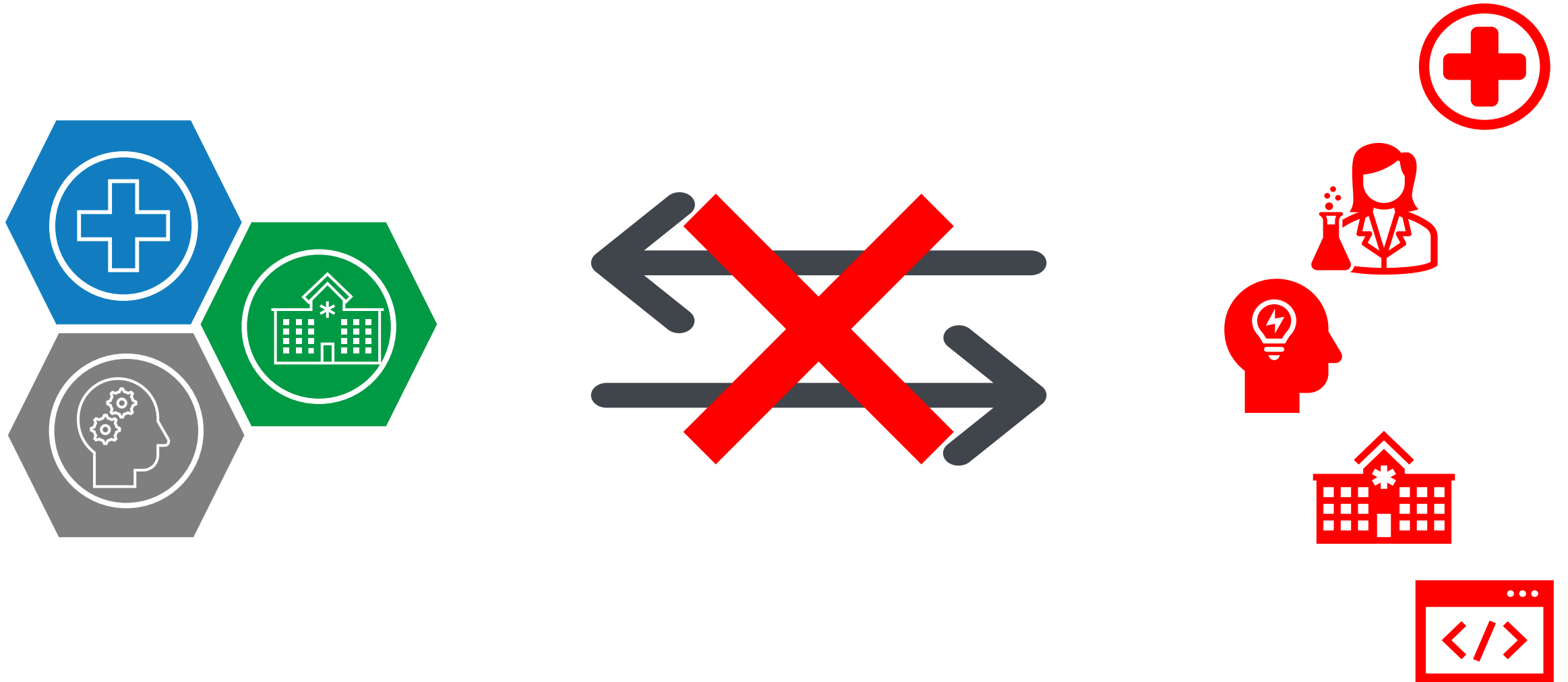
## Key benefits

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

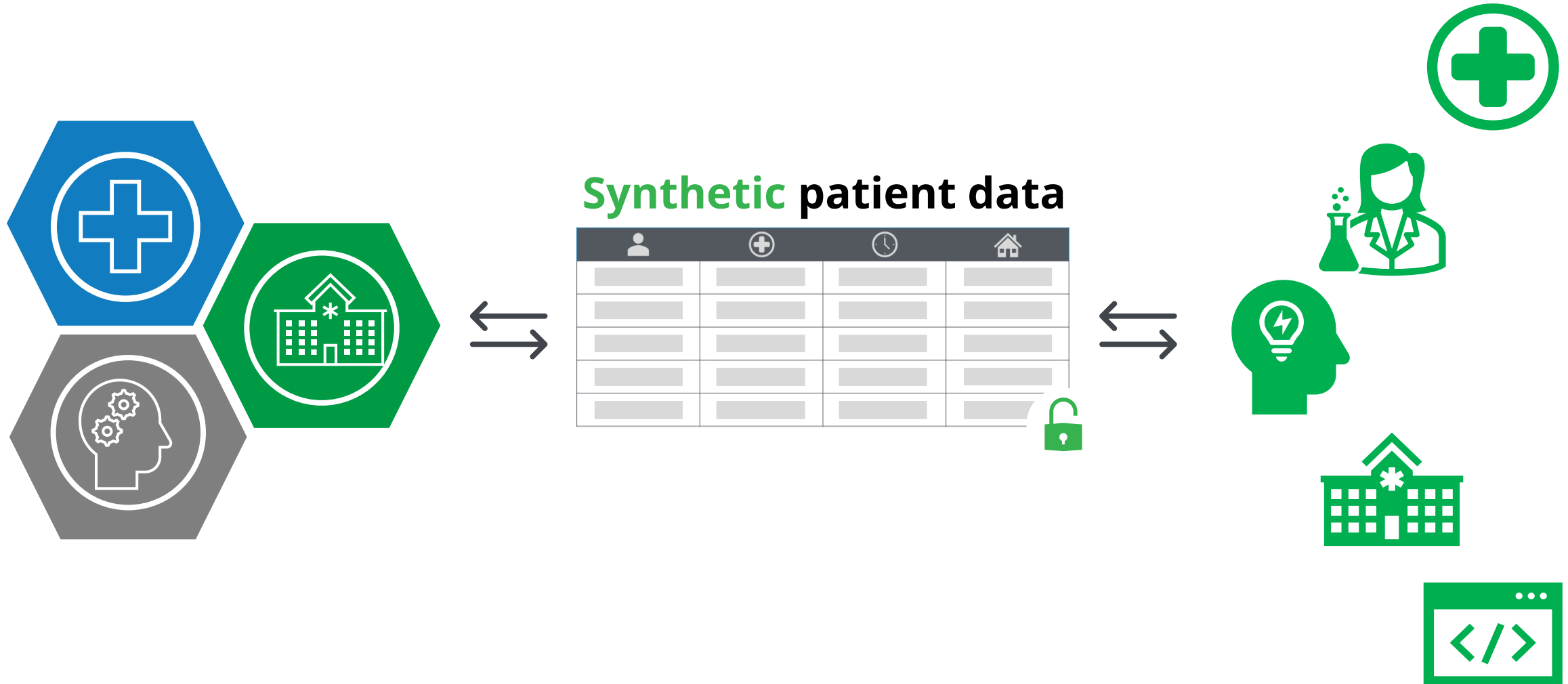


**Strong data foundation**

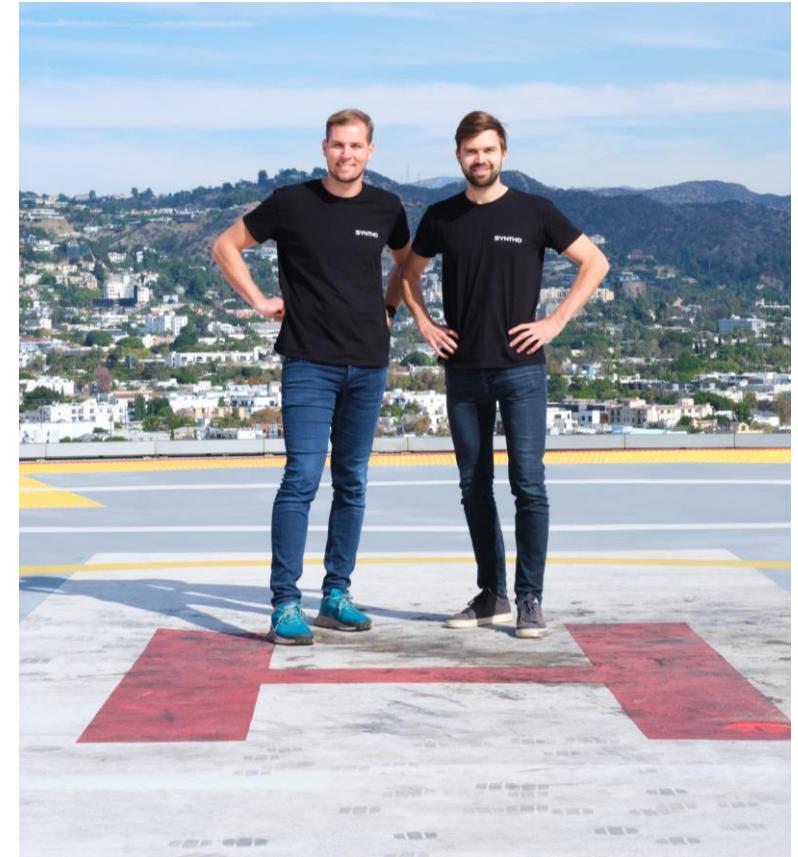
# Data sharing: using and sharing data is challenging



# Data sharing: freely use and share synthetic data



# Case Study: Synthetic healthcare data with Cedars Sinai Medical Center





# HEALTH CARE & LIFE SCIENCES **WINNER**

## AI Generated Synthetic Data in Healthcare

Syntho, Netherlands



Using AI-generated synthetic data in hospitals can be a powerful tool for utilizing data faster, without privacy concerns.

**SAS HACKATHON**  
#SASHackathon



# Correlations

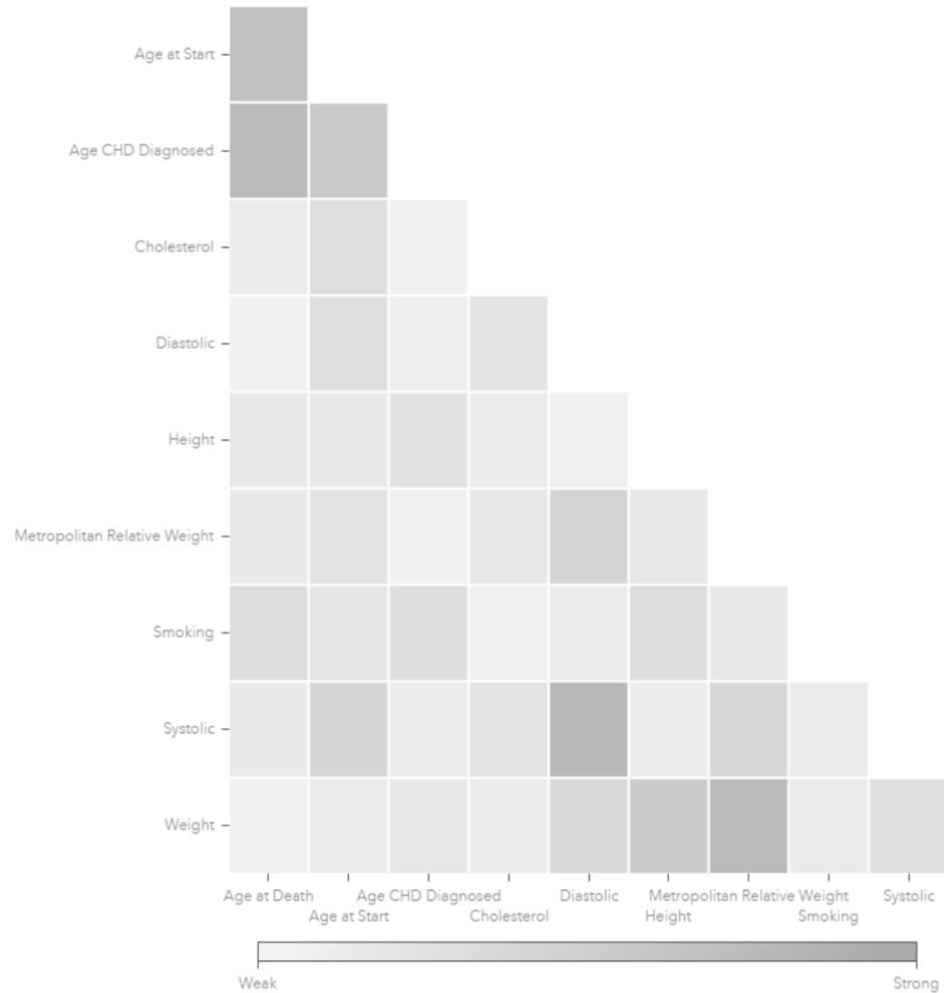


Original data

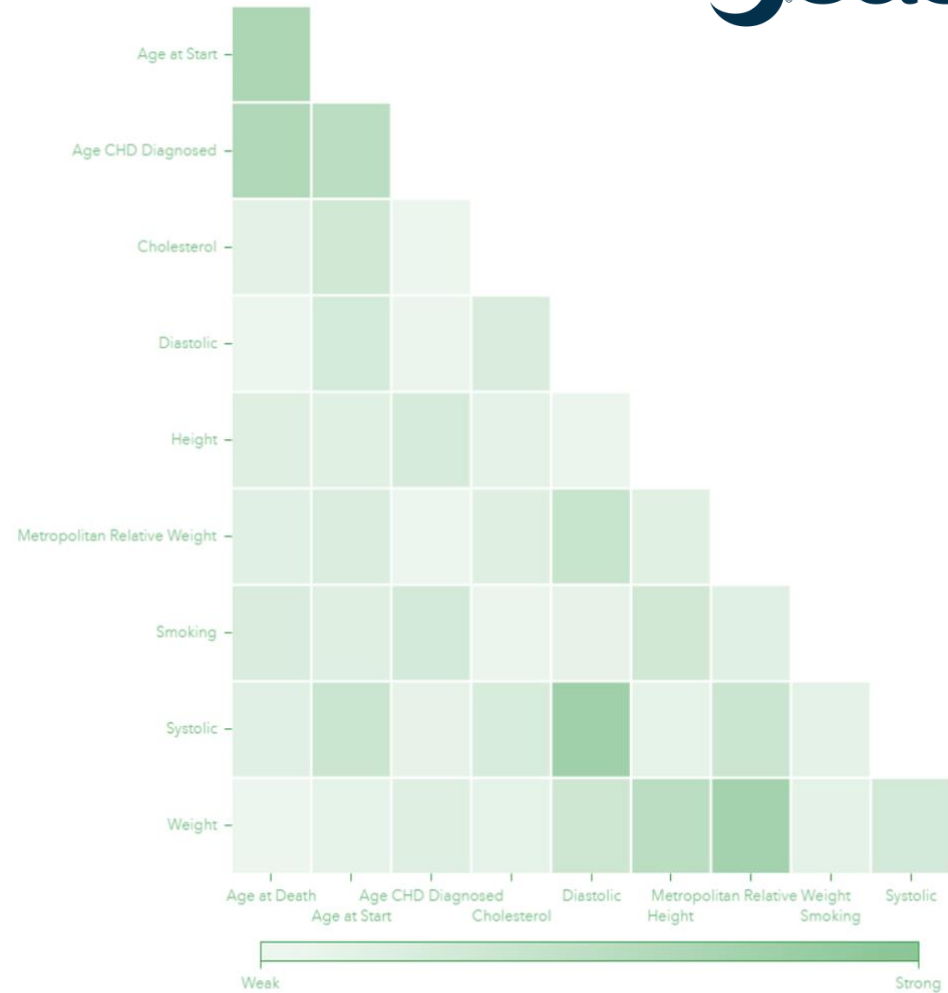
Synthetic data



Original data



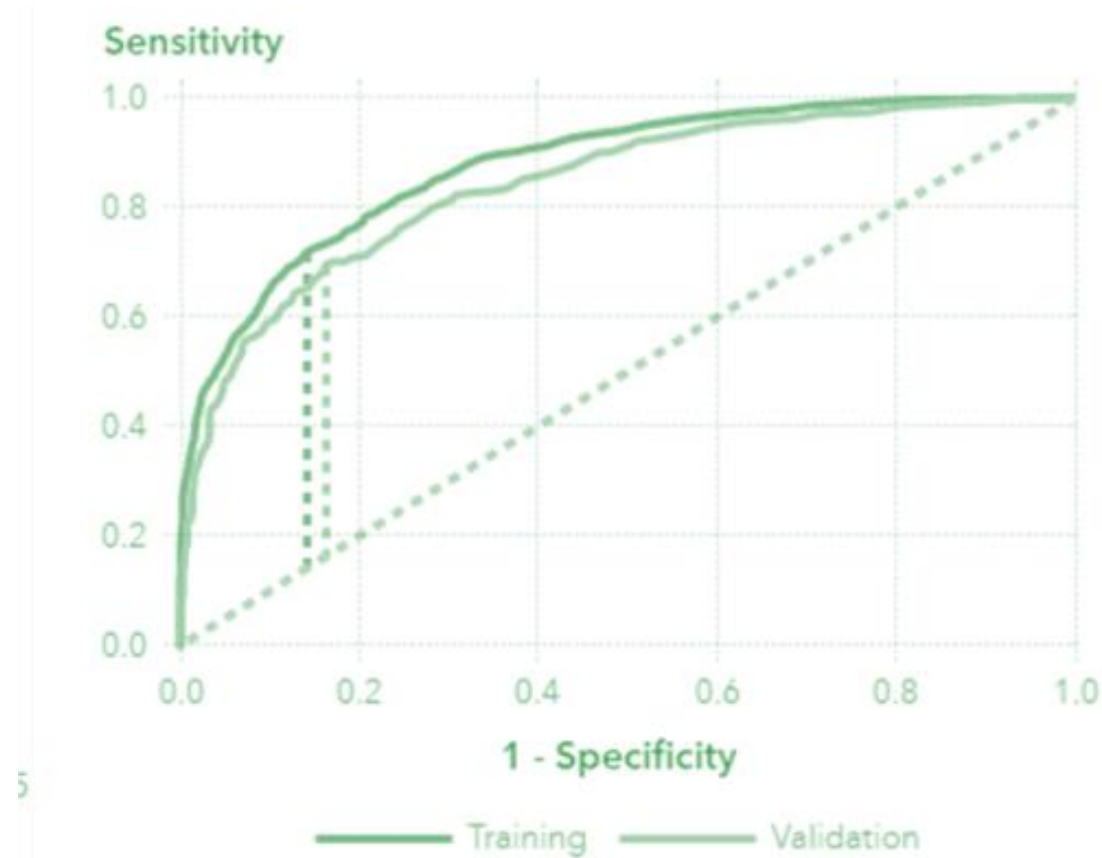
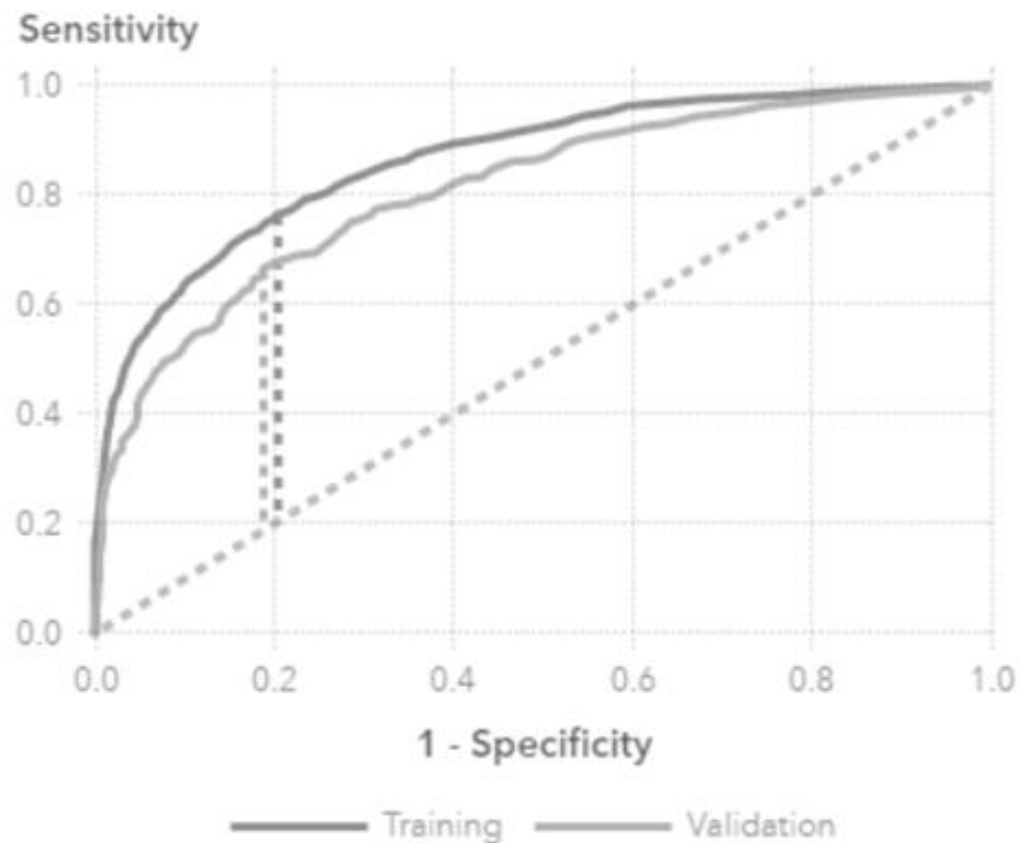
Synthetic Data



# ROC Curve

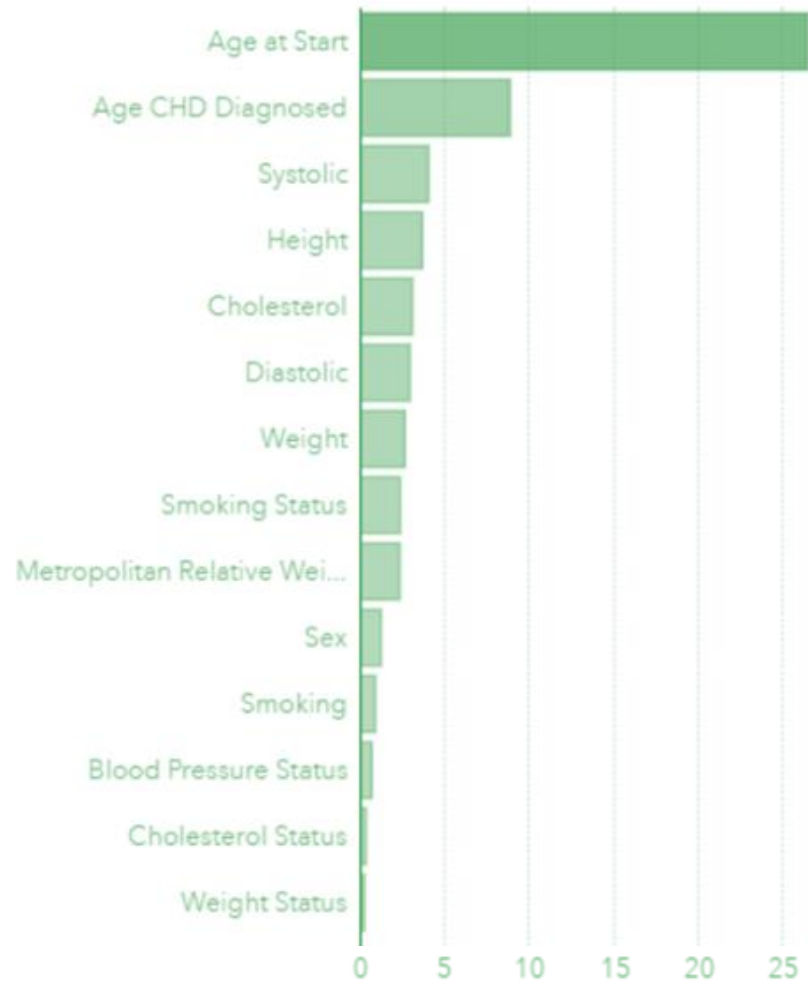
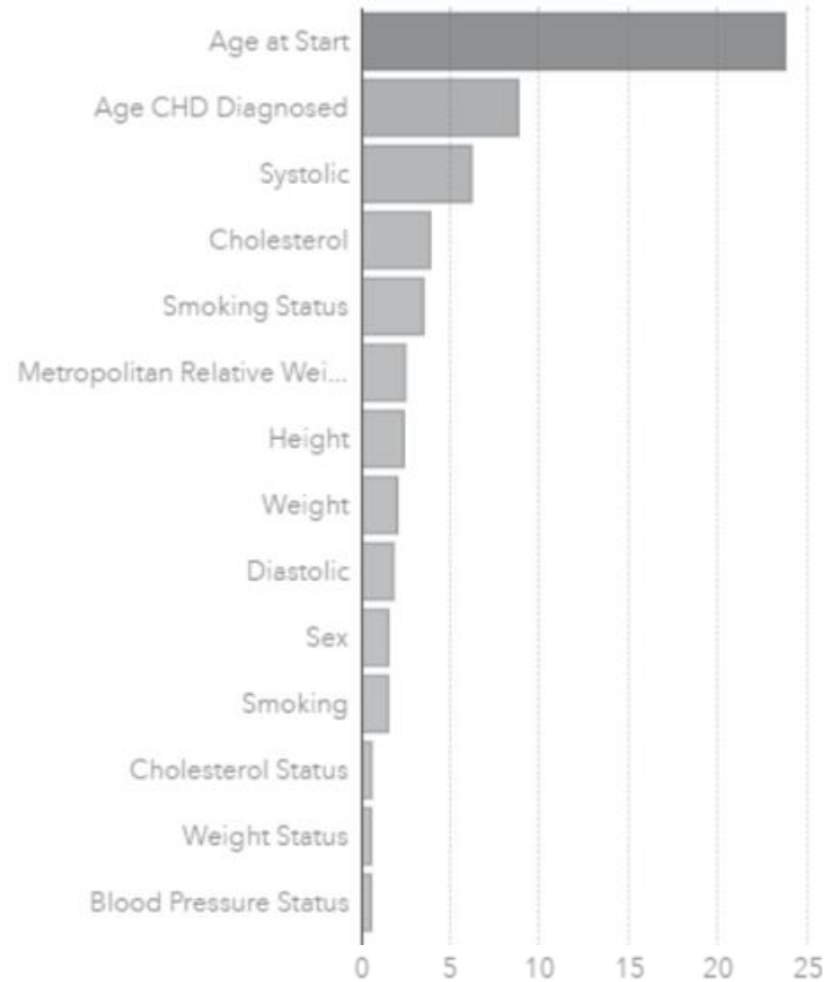


- Original data
- Synthetic data



5

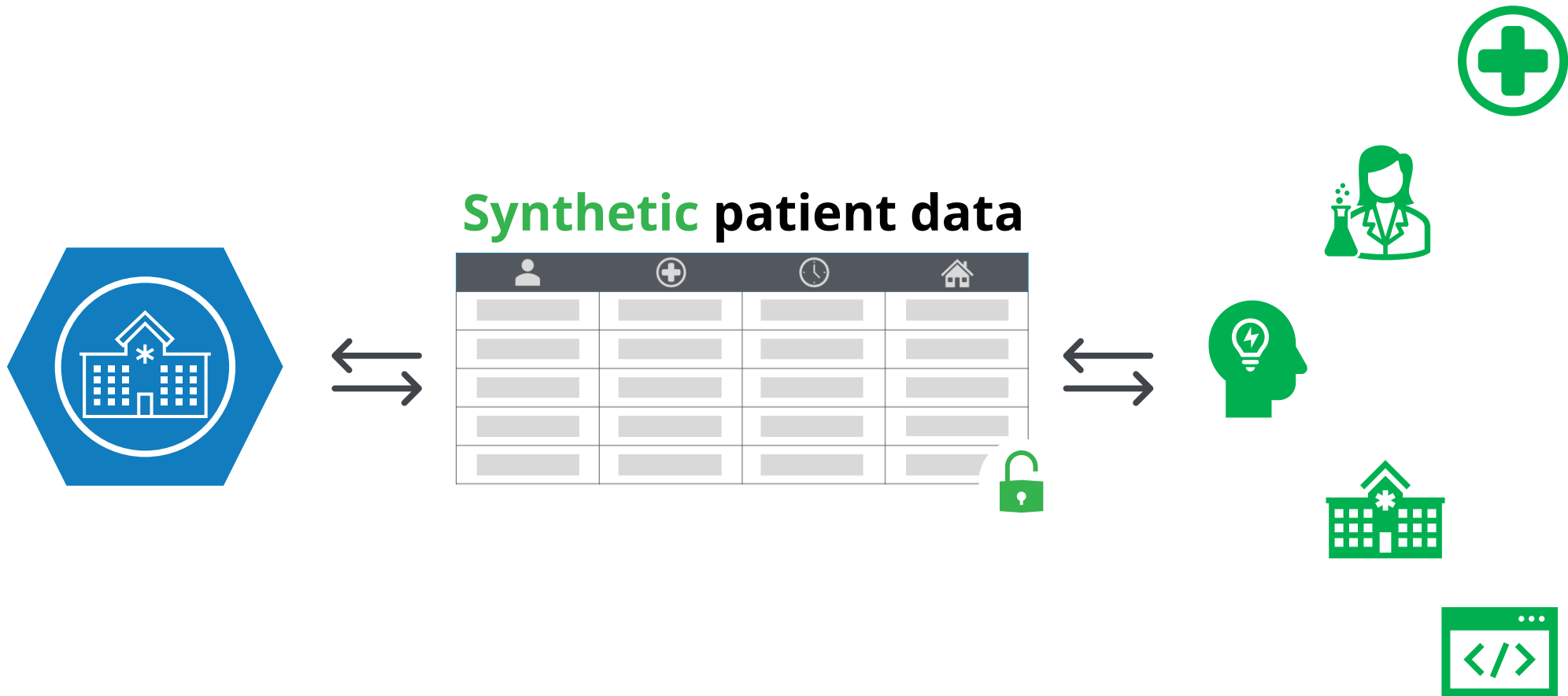
# Variable importance



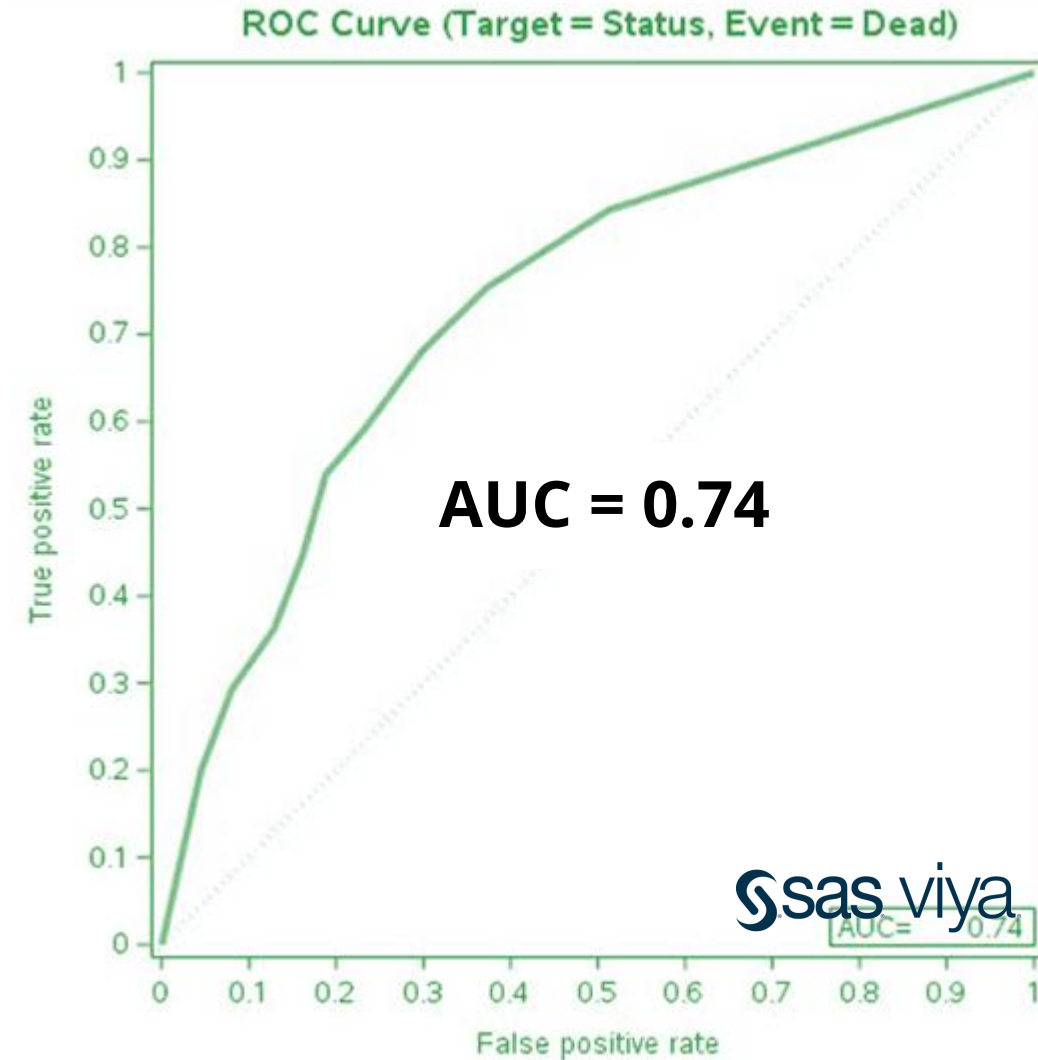
Original data

Synthetic data

# Case Study: Predicting deterioration and mortality as part of cancer research on synthetic data



# Case Study: Predicting deterioration and mortality as part of cancer research on synthetic data



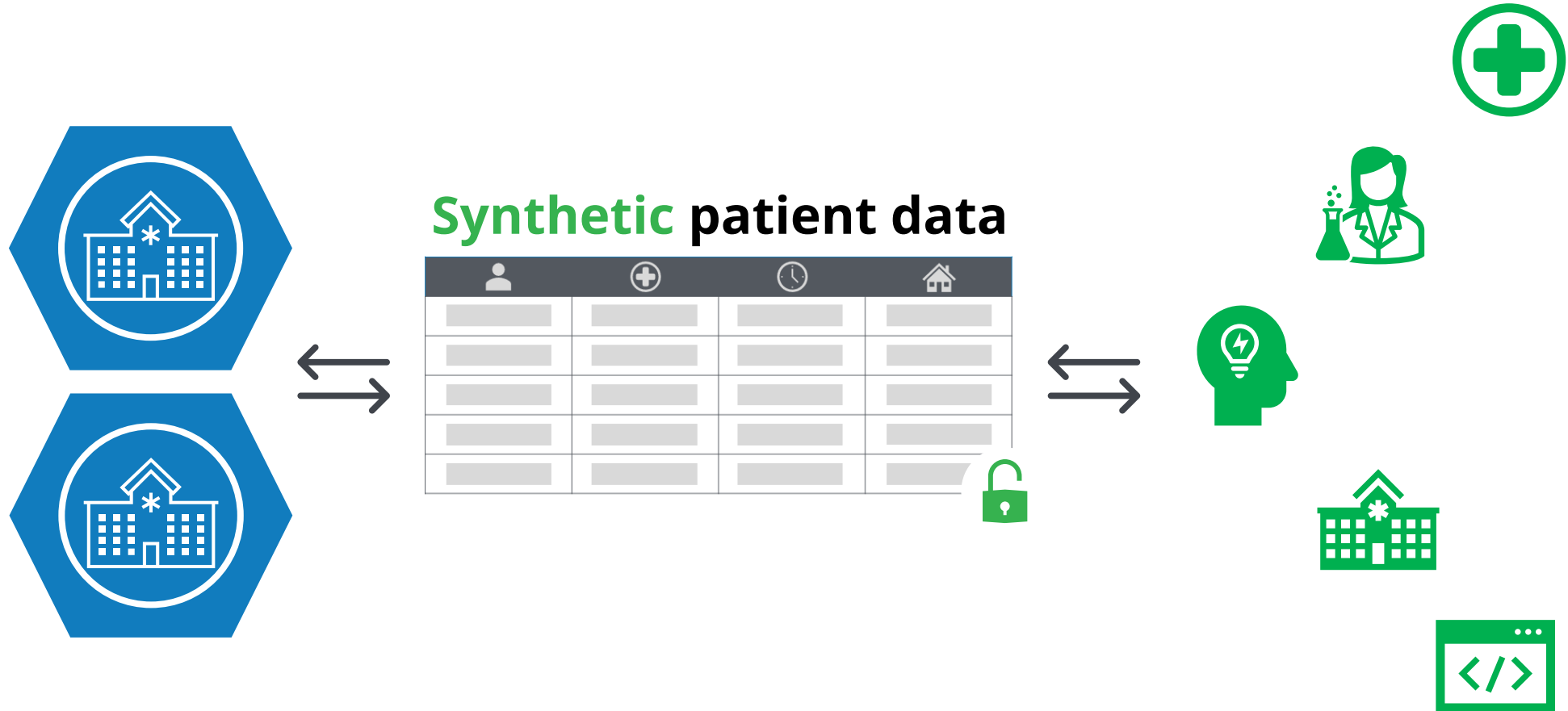
# **Case Study:** Predicting deterioration and mortality as part of cancer research on synthetic data

## **Conclusion**

*Privacy-sensitive healthcare data can be unlocked with synthetic data within the hospital to realize data-driven insights*

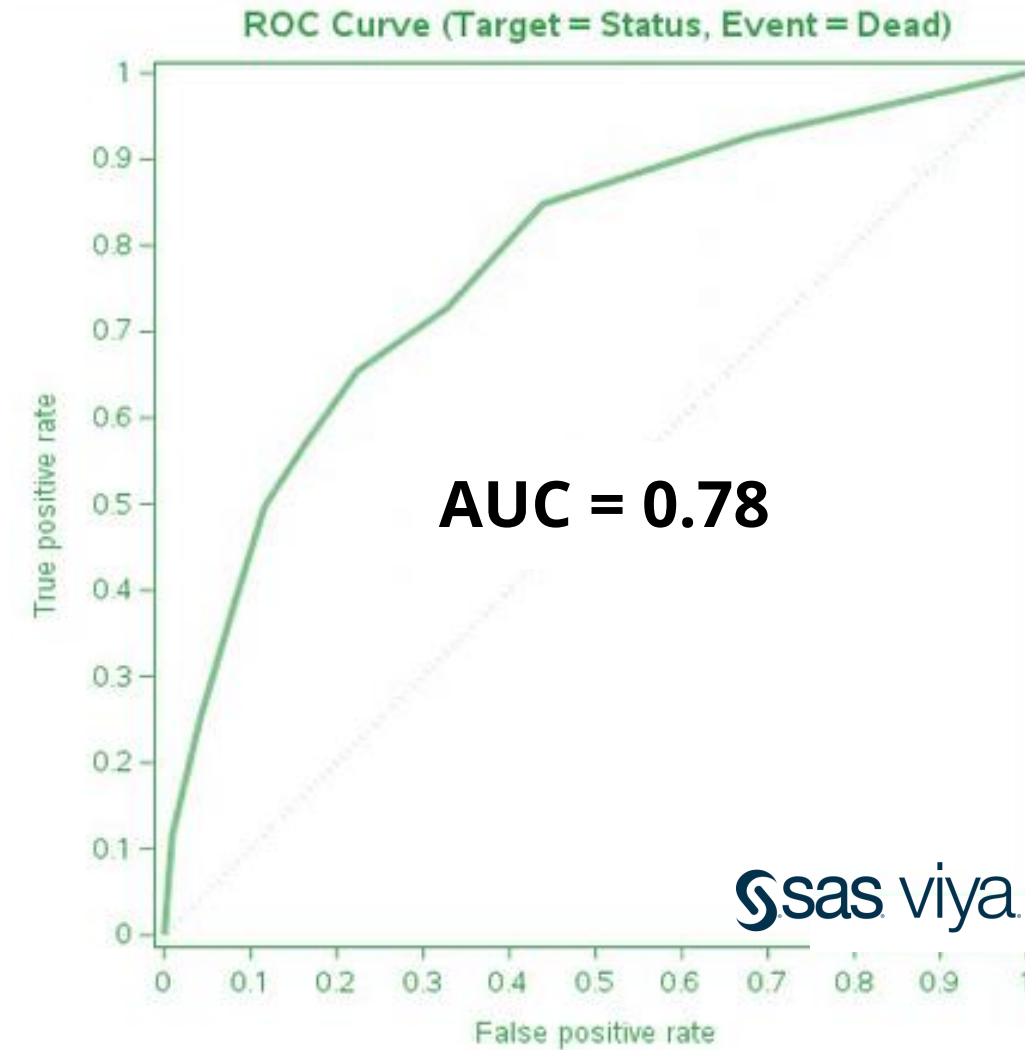
- ✓ Less risk
- ✓ More data
- ✓ Faster data access

# Case Study: Combine synthetic data from *two* hospitals

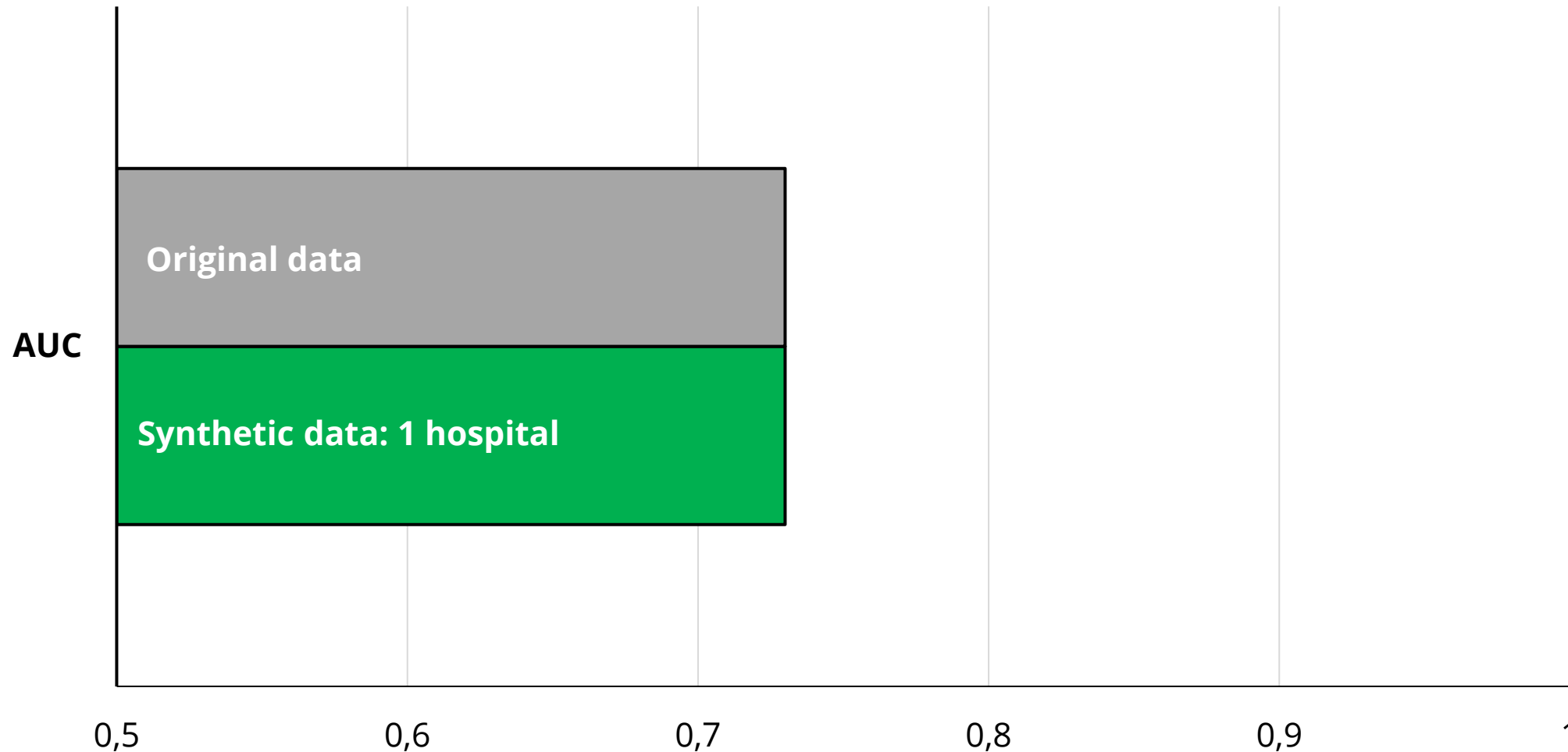




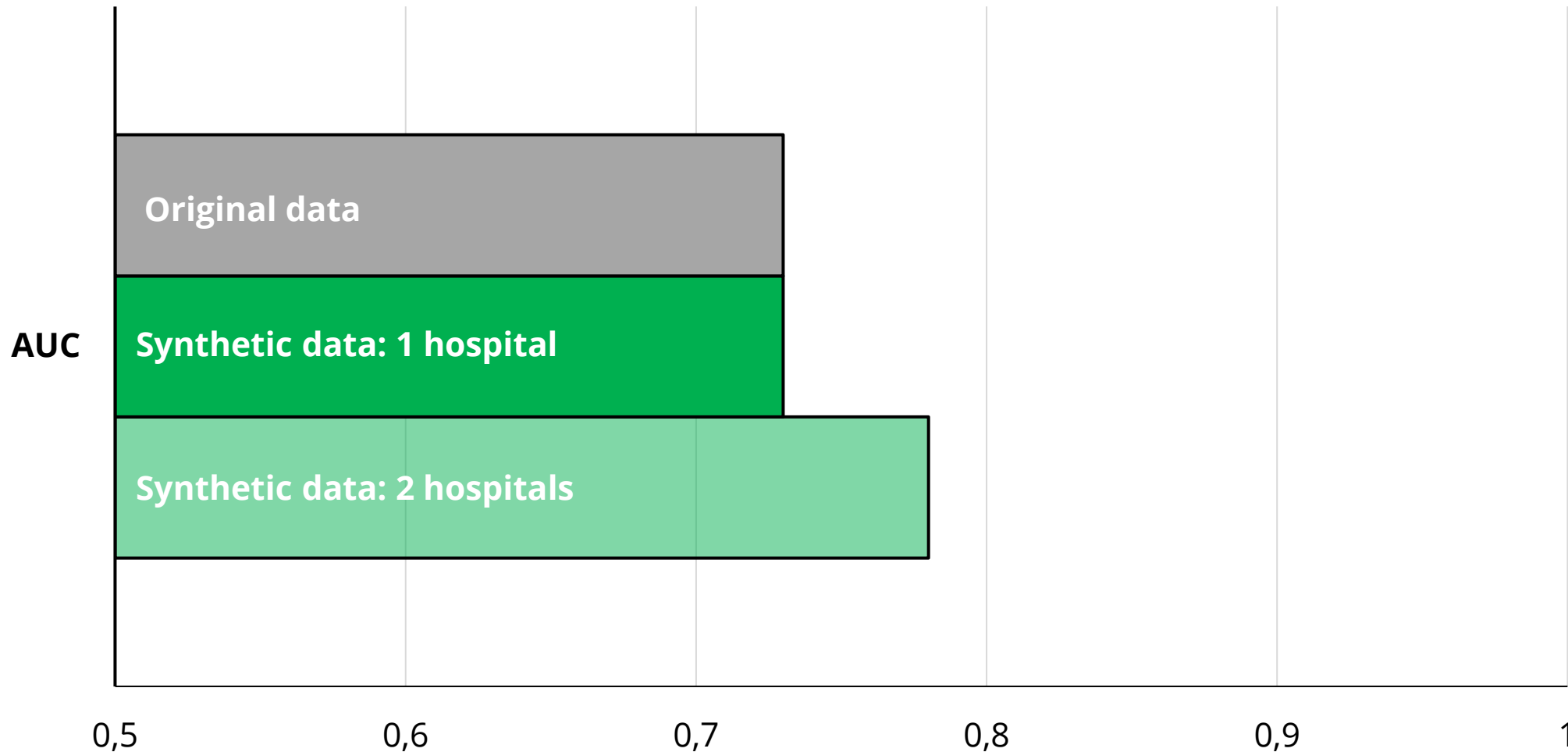
# Case Study: Combine synthetic data from *two* hospitals



# Case Study: Synthetic data is as-good-as-real and can be used for clinical research as alternative for sensitive real data



# Case Study: Synthetic data can easily be shared. Access to more data could improve the accuracy of predictive models.



---

**50%**

Syntho unlocks data

---

**\$4T**

Realize data opportunities

# Turn data access into a **competitive advantage**

# 70%

**An increase in industry collaborations is expected with the use of privacy tools**

# 30%

**More revenue for organizations that earn and maintain digital trust with customers and patients**

# 50%

**Of data will be unlocked by privacy-enhancing techniques**

# Synthetic data – Real people!



**Wim Kees Janssen**

*Founder & CEO*

Contact: [kees@syntho.ai](mailto:kees@syntho.ai)