FME AI FOR INDUSTRY JAAREVENT

Generative AI: Toepassing van synthetische data

7 december 2023

FME POWERED BY DUTCH TECHNOLOGY

SYNTHETISCHE DATA

AGENDA

- Synthetic data, Real impact
  Wim Kees Janssen - Syntho
- Demcon Synthetic Data
  Vincent Bos – Demcon

FME POWERED BY DUTCH TECHNOLOGY

SYNTHO

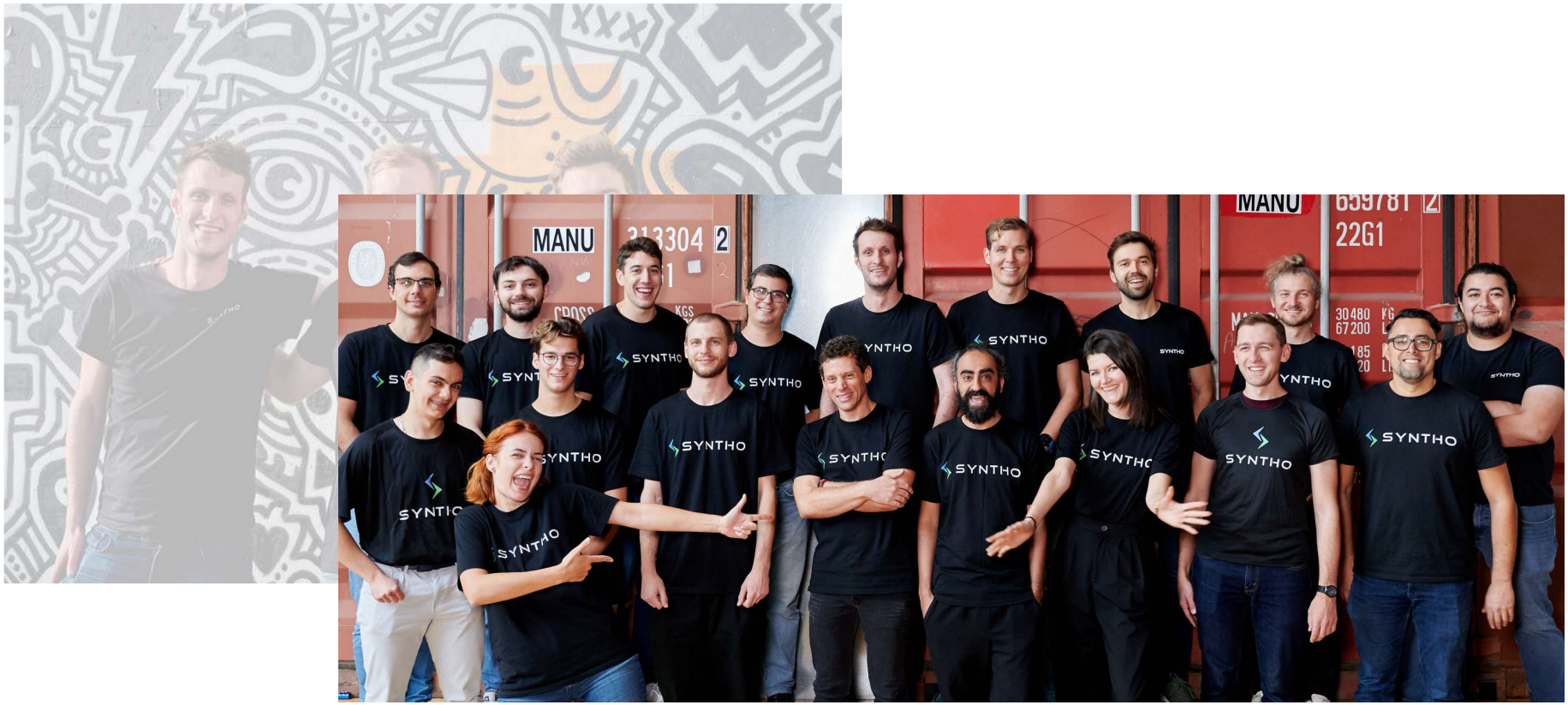
Synthetic data
Real impact

Self-service AI-generated synthetic data platform
to accelerate data-driven solutions.

kees@syntho.ai | www.syntho.ai/kees | www.syntho.ai/careers

About Syntho

About Syntho



Recognitions

Winner Philips Innovation Award

Read more

Winner SAS Global Hackathon in healthcare & Lifescience

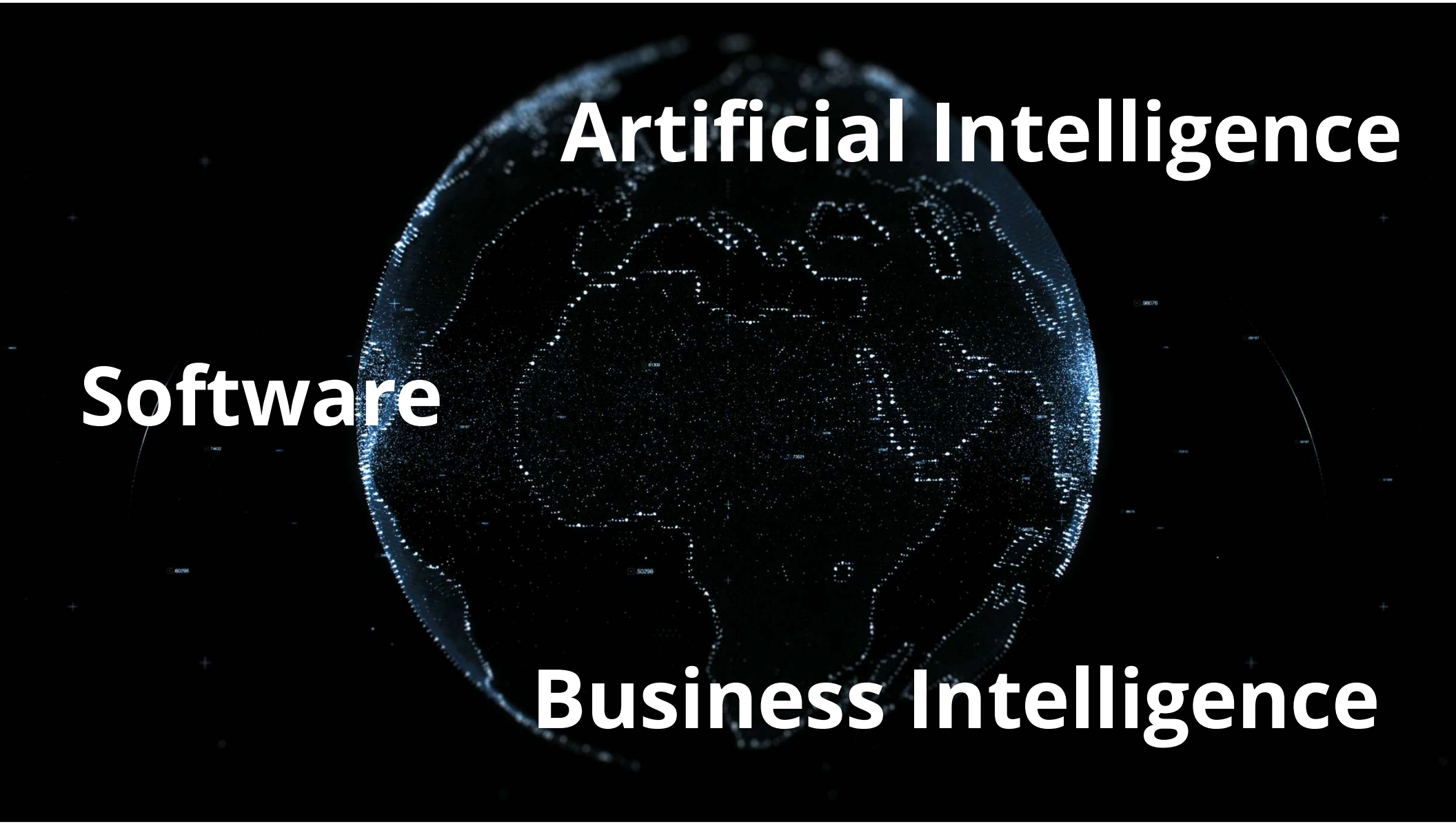
Read more

Shortlisted by NVIDIA as Generative AI startup to watch

Read more

Agenda

- Data challenges
- Anonymization
- AI Generated Synthetic Data (Tabular Data)
- Syntho Engine
- Use Cases
- Q&A


Artificial Intelligence
Software
Business Intelligence

50%

Of data is locked

$4T

Missed data opportunities

Synthetic data offers opportunities

70%

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

30%

More profits for companies that **earn and maintain digital trust** with customers

50%

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

Gartner®

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

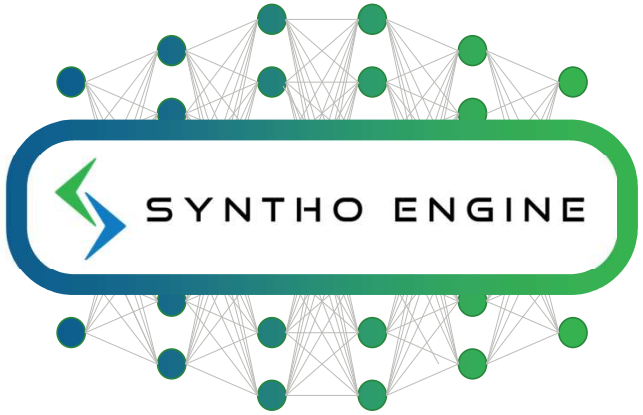
Classic anonymization					
Name	**Age**	**Gender**	**Item**	**Price**	**Data**
xxx	25-30	Female	Cloth	€100 - €200	March
xxx	70-75	Male	IT	€600 - €700	March
xxx	40-45	Male	Drink	<€5	March
…	…	…	…	…	…
xxx	40-45	Male	Cloth	€20 - €30	March

N=100k

✕ **Privacy risks**

✕ **Low accuracy**

✕ **Low scalability**

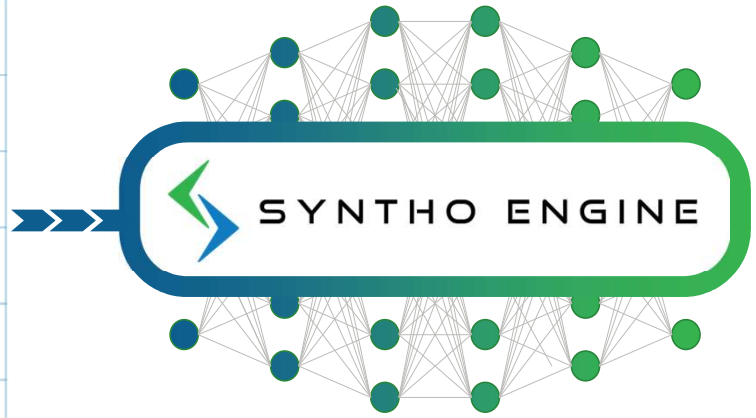


Generative AI
For Tabular Data

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



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

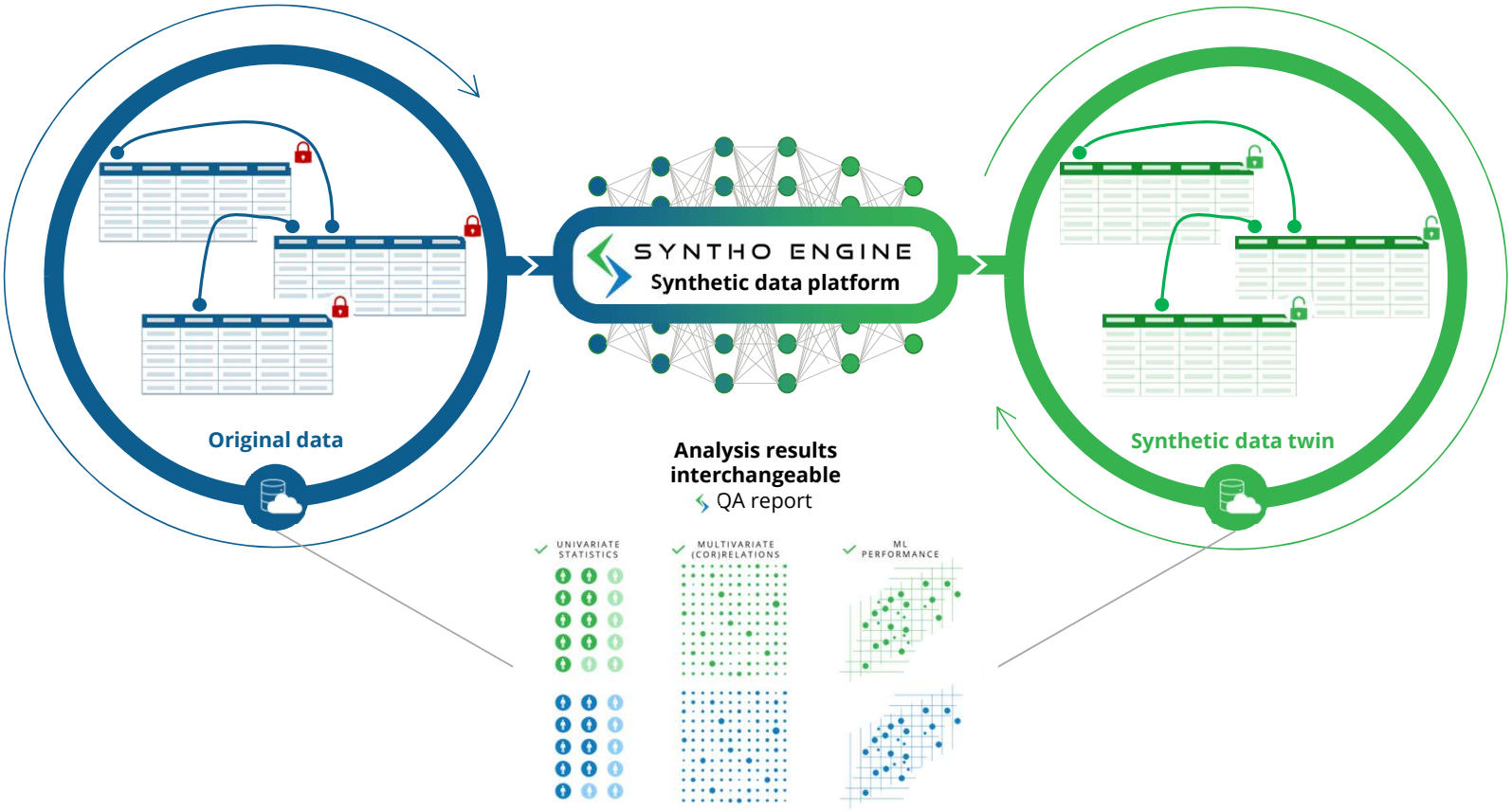
N=100k

✓ **Privacy by design**

✓ **High accuracy**

✓ **High scalability**

AI-generated synthetic data



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.

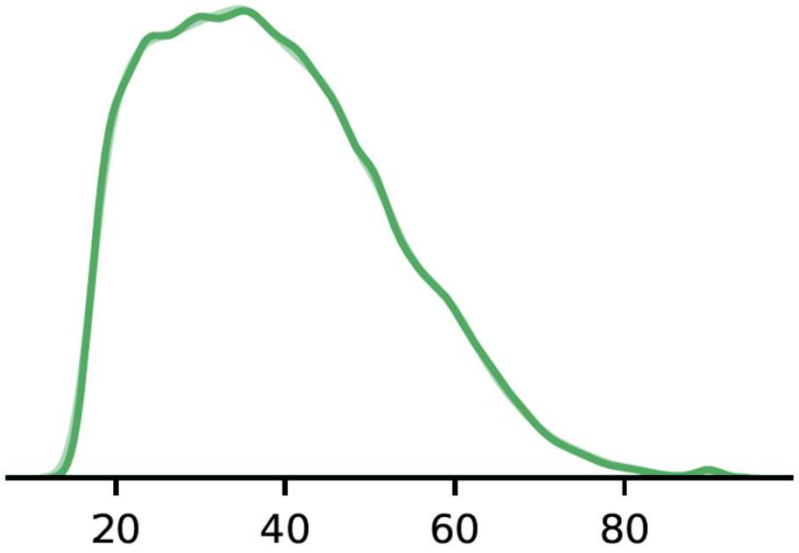
2. External assessment by the data experts of SAS

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.

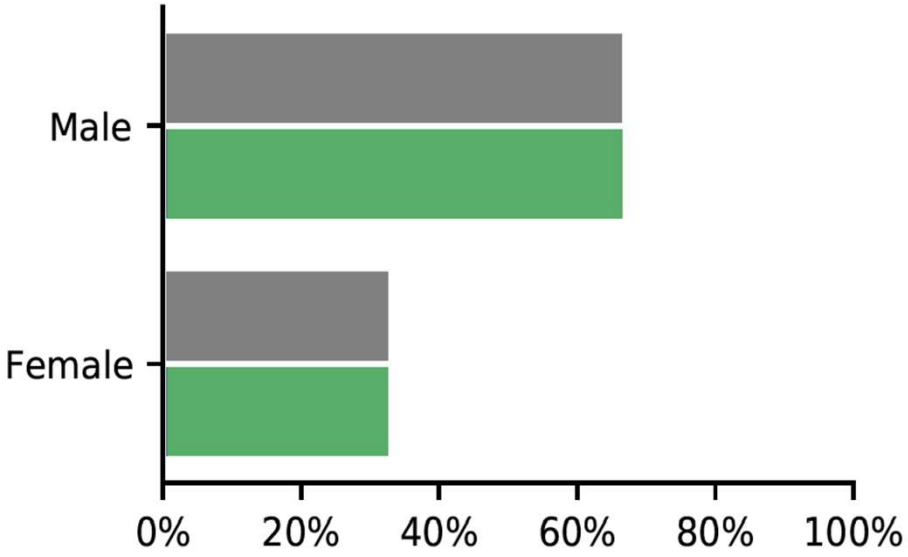
Our synthetic data is approved by the data experts of SAS

Distributions

Age



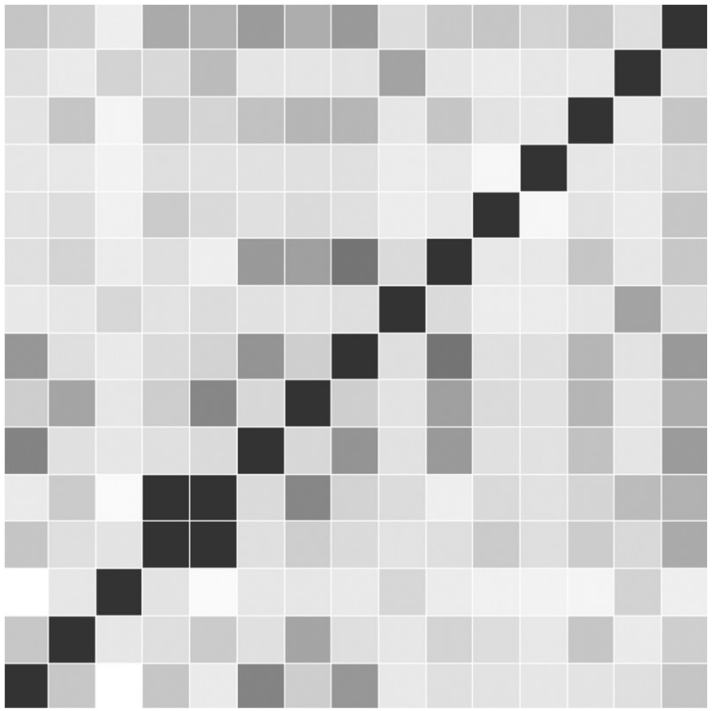
Gender



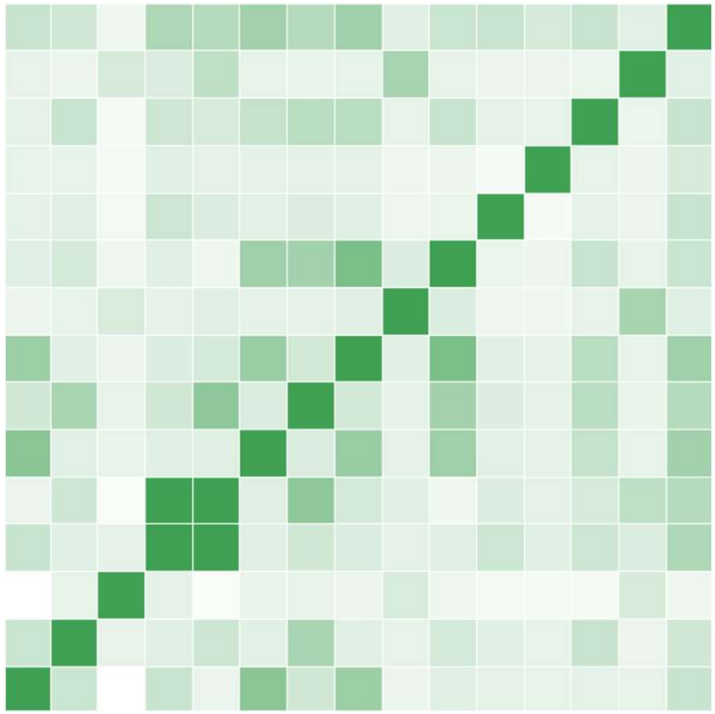
Original data

Synthetic data

Correlations

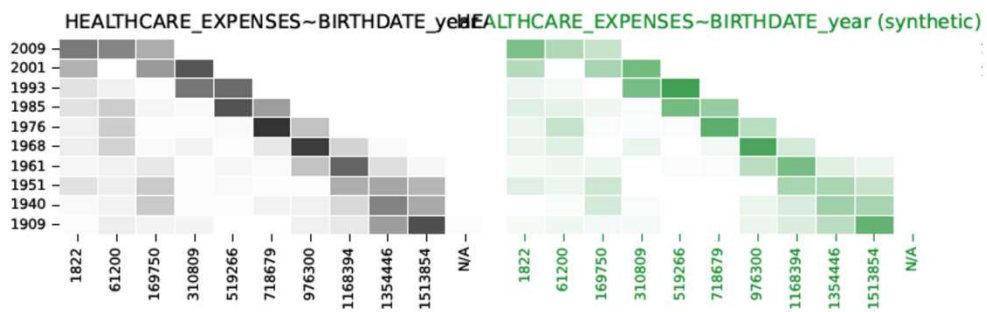


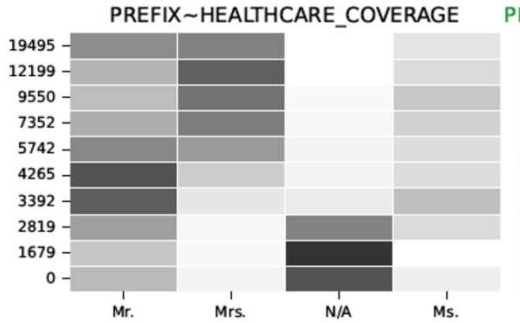
Original data

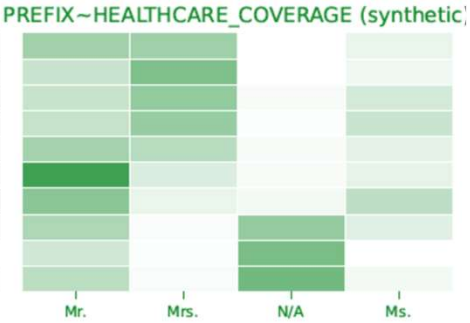


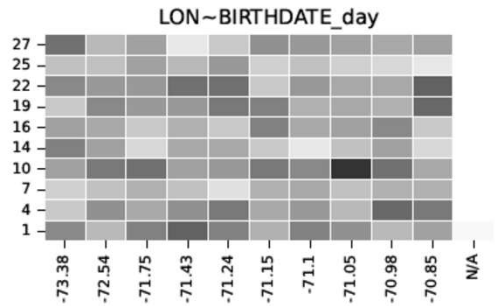
Synthetic data

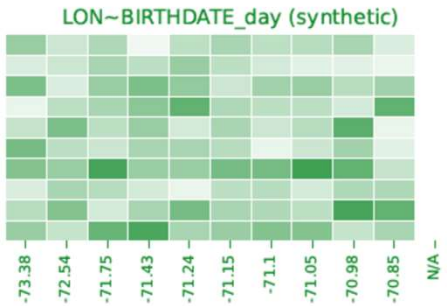
BI-variates

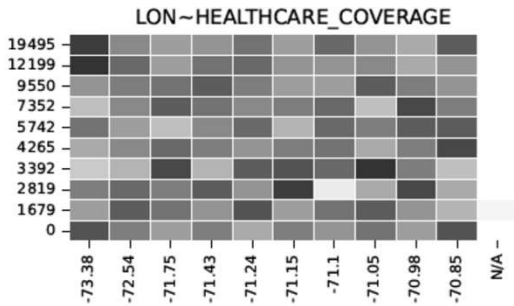


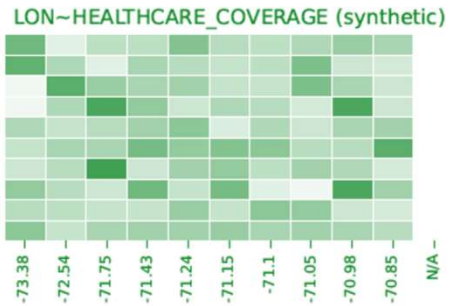












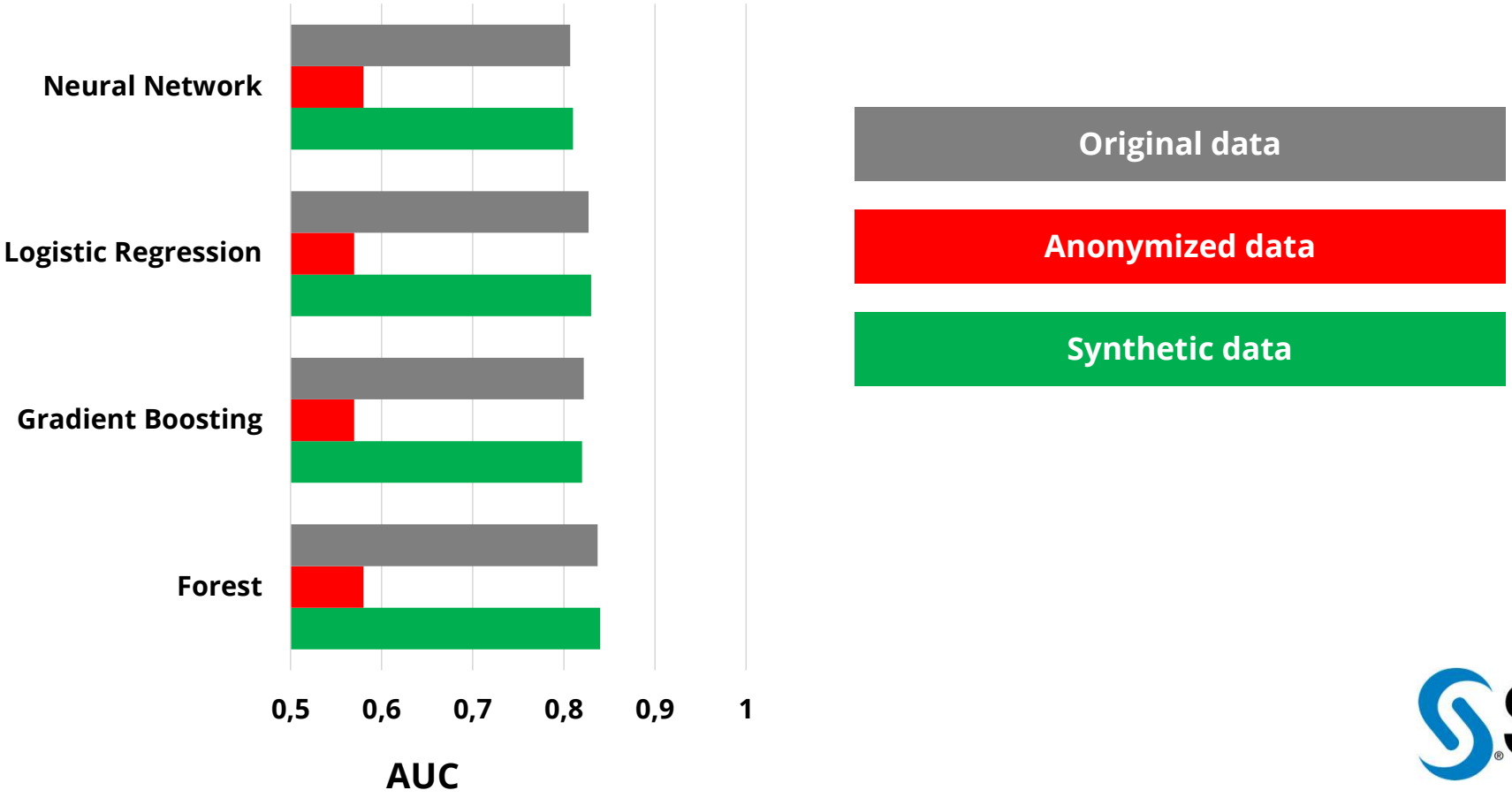
Original data

Synthetic data

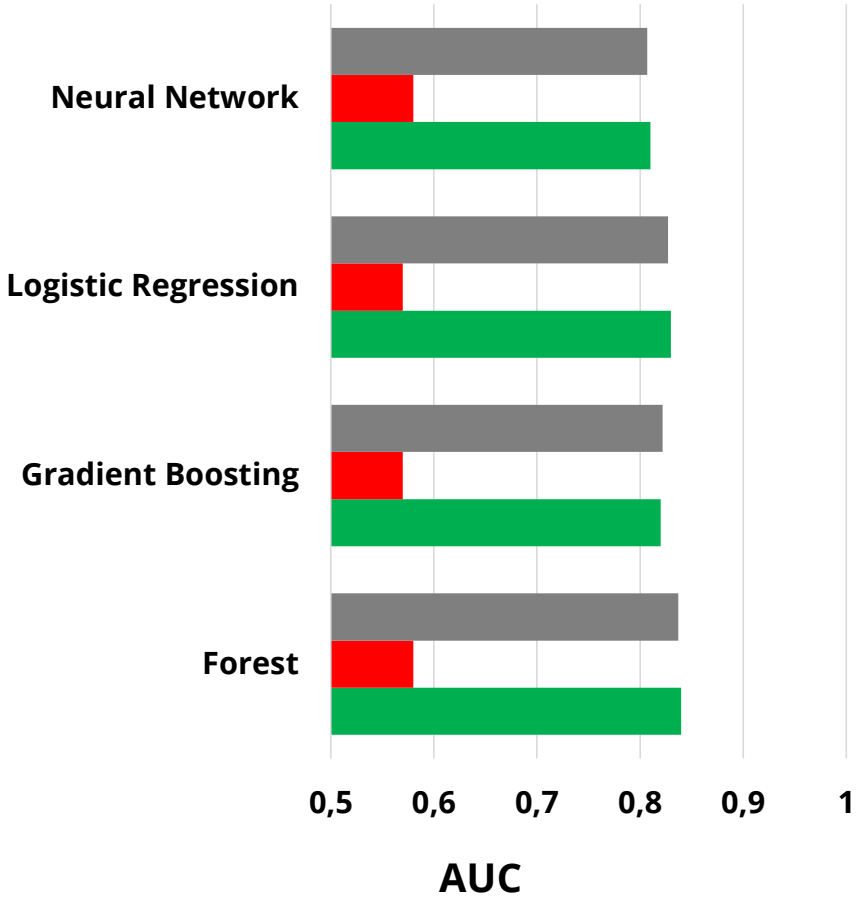
Synthetic data approved by the data experts of SAS

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

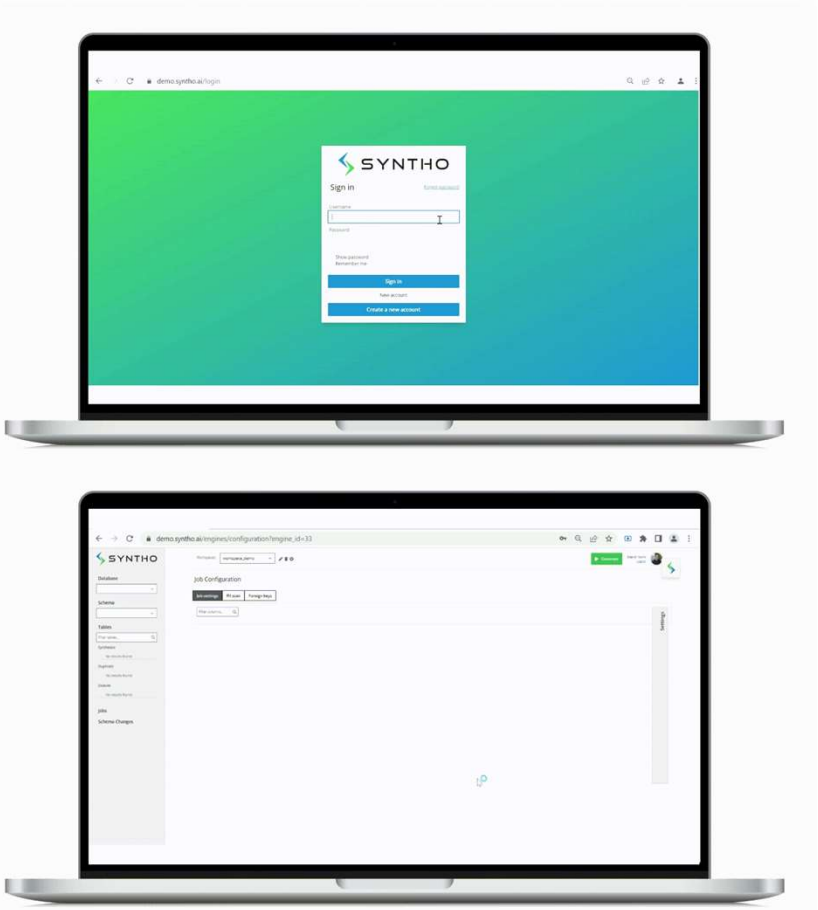


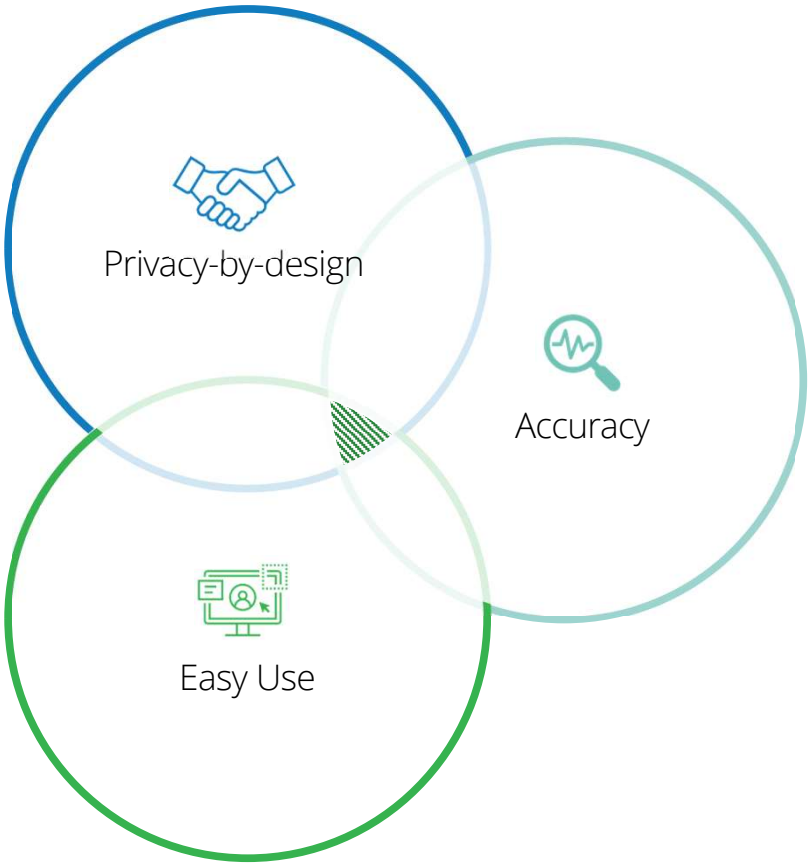
- Synthetic data is as-good-as the original data
- Anonymized data shows the worst performance
- Easy, fast, and scalable

SYNTHO ENGINE 2.0

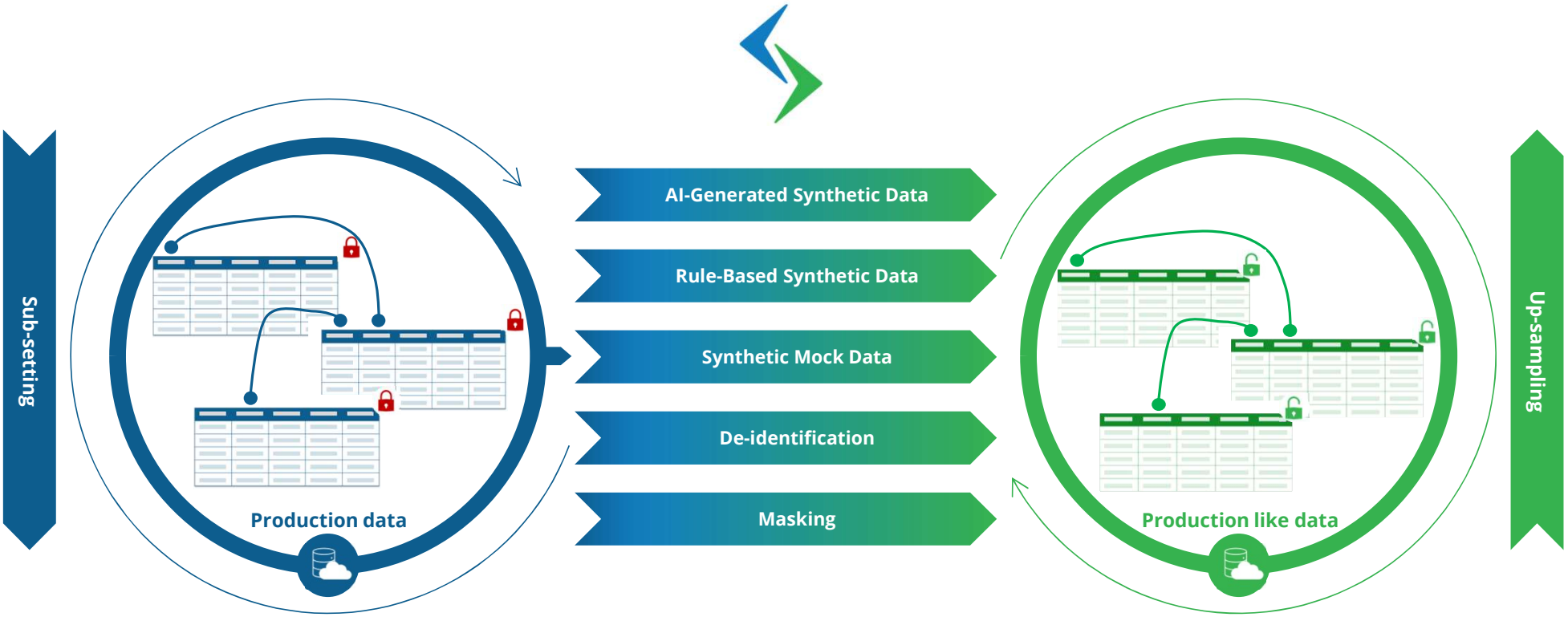
REVOLUTIONIZING AI GENERATED SYNTHETIC DATA

3 pillars of our Syntho Engine

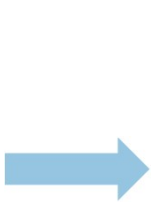




Smart Data Generation with Syntho



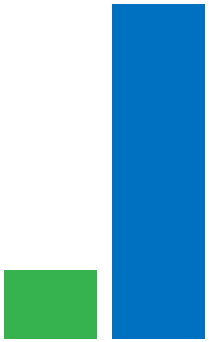
Smart up-sampling of underrepresented groups



Example:
10% FALSE
90% POSITIVE

SYNTHO ENGINE

Example:
50% FALSE
50% POSITIVE



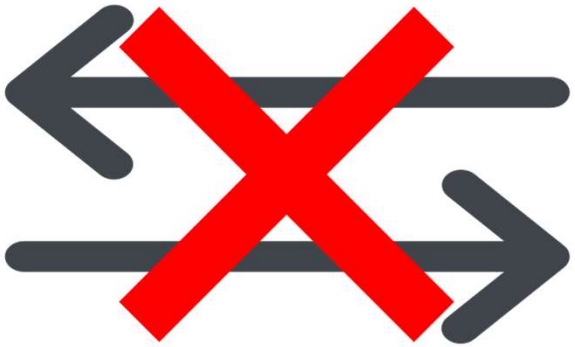
Synthetic data unlocks your data

Analytics

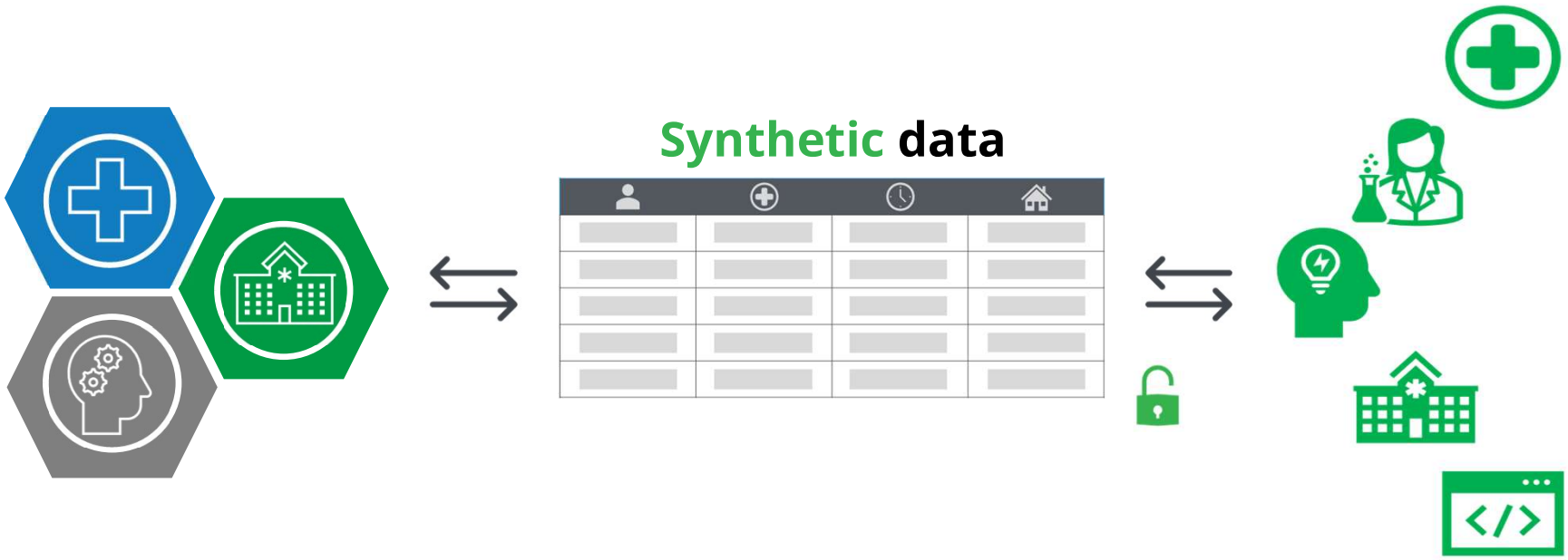
Test data

Data sharing

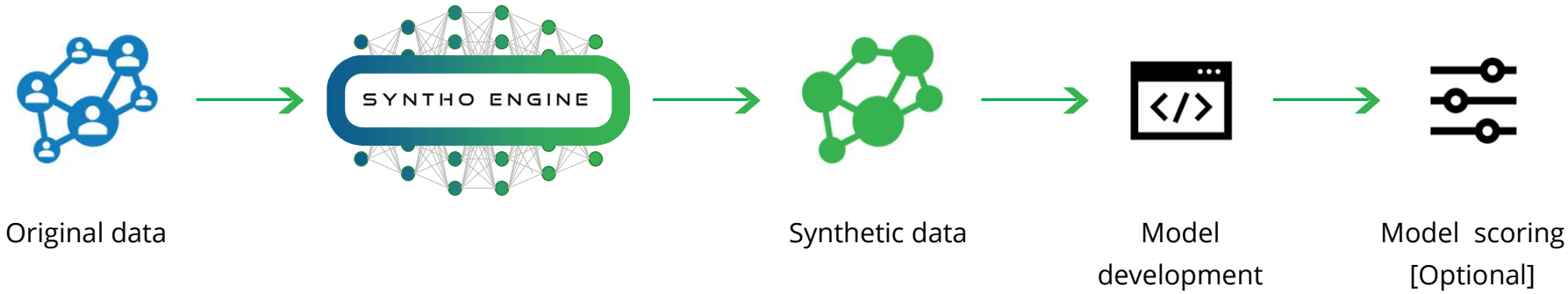
Data sharing: **using and sharing data is challenging**



Data sharing: **freely use and share synthetic data**

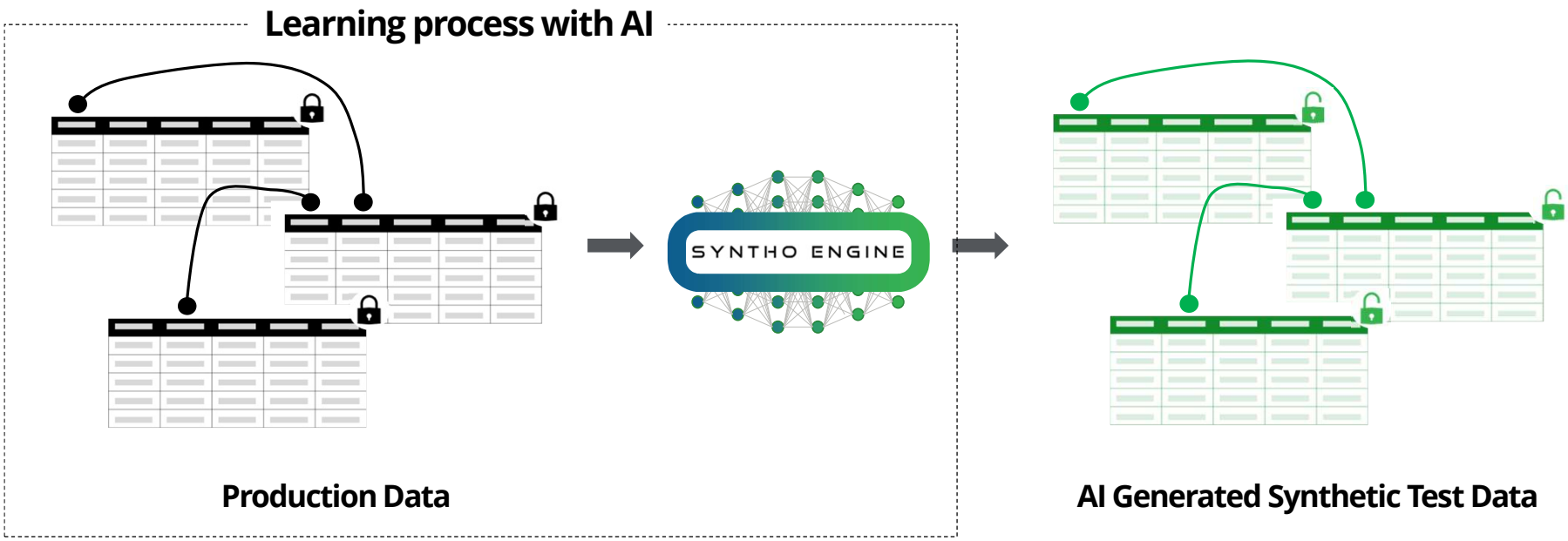


Analytics: Synthetic Data for Analytics



✓ Ranging from **business intelligence** to **artificial intelligence** solutions

Test data: AI Generated Synthetic Test Data



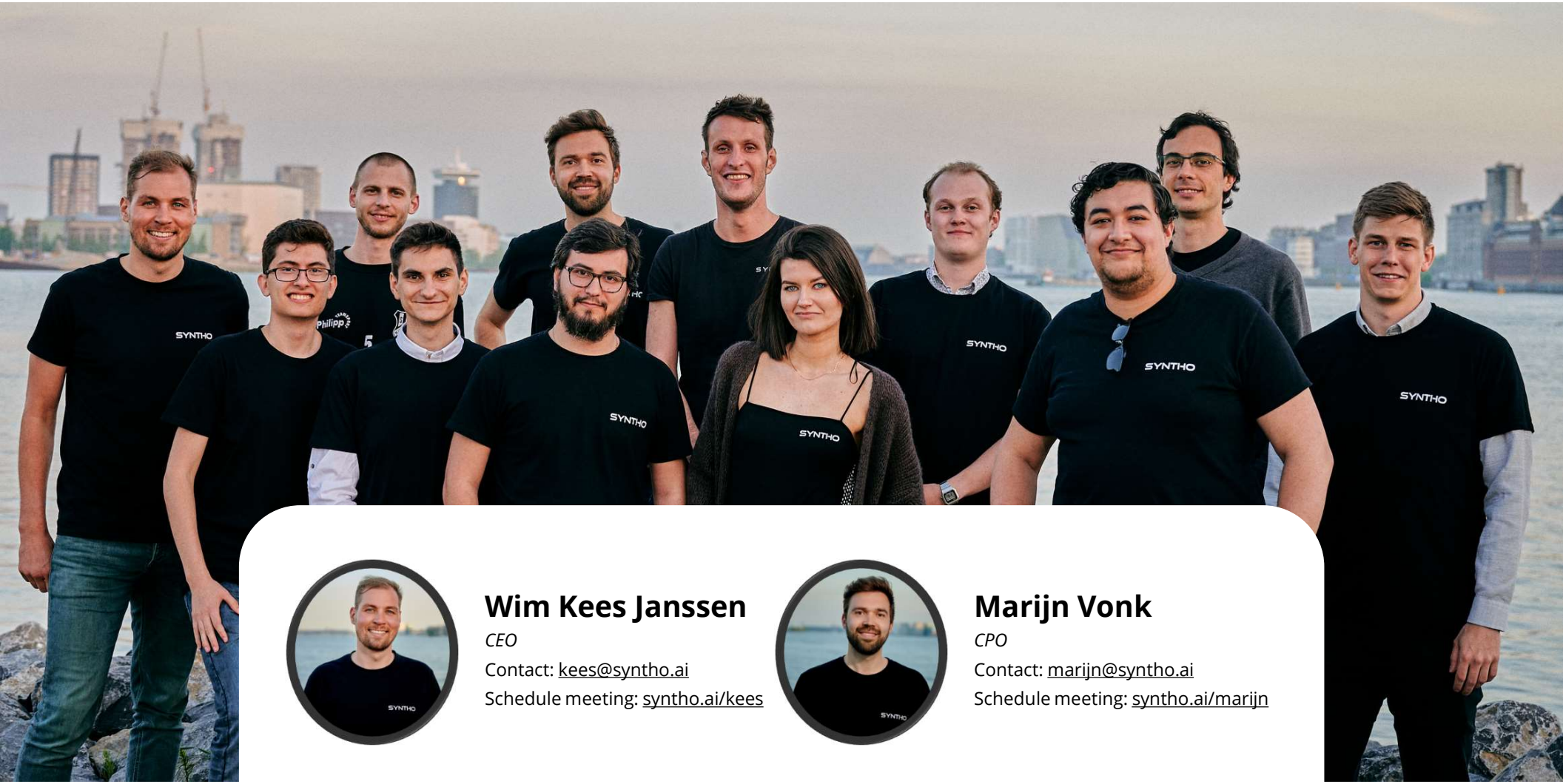
50%

Syntho unlocks data

$4T

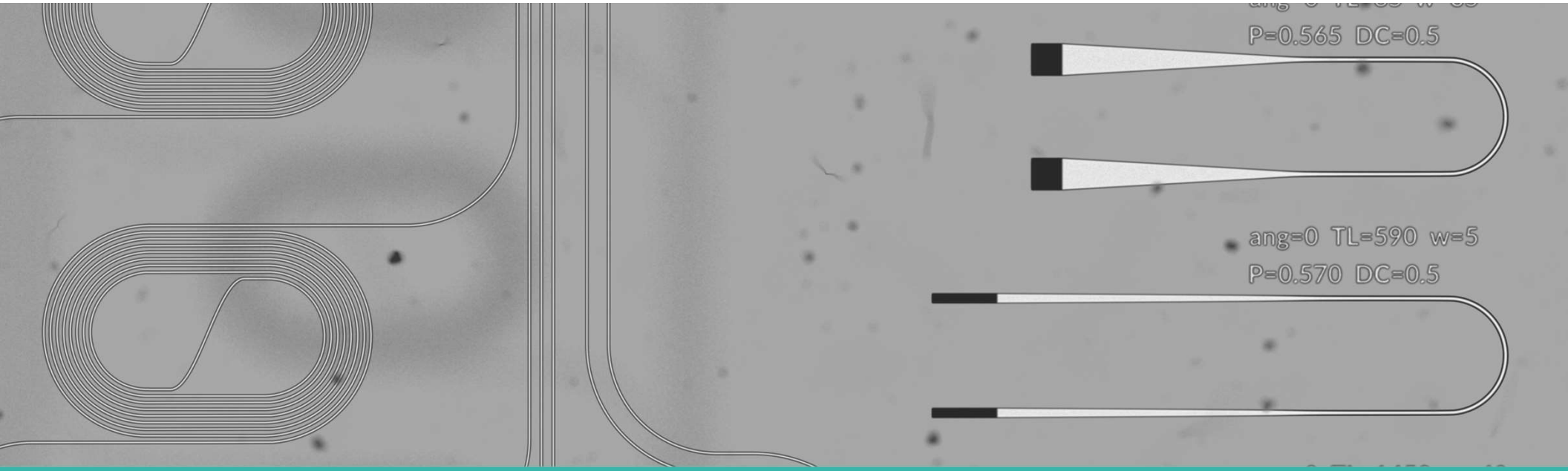
Realize data opportunities

Synthetic data — Real people!



Wim Kees Janssen
CEO
Contact: kees@syntho.ai
Schedule meeting: syntho.ai/kees

Marijn Vonk
CPO
Contact: marijn@syntho.ai
Schedule meeting: syntho.ai/marijn



Demcon Synthetic Data– FME

December 7th, 2023

VINCENT BOS

About Demcon

Design and Engineering

- Contract Engineering
  - High-Tech Mechatronic Systems
  - Life Sciences & Health
  - Optical Systems
  - Industrial Systems
- Contract Manufacturing
- Own Products
- Founded 1993
- 1200+ Employees
- ISO 13485, ISO 9001, CE, MDR, IVDR, FDA approval

DEMCON

Data Driven Solutions

Synthetic Data

Create data when there is none

Machine Vision

The system and the algorithms needed

Process Improvement

Data driven insights to improve processes and products

Time Series Analysis

Actionable knowledge from time series data

Reinforcement Learning

Optimization in a dynamic environment

DEMCON

What do we mean by Synthetic Data at Demcon?

DEMCON

“Simulated sensor data, and rich annotation of generated environment models”

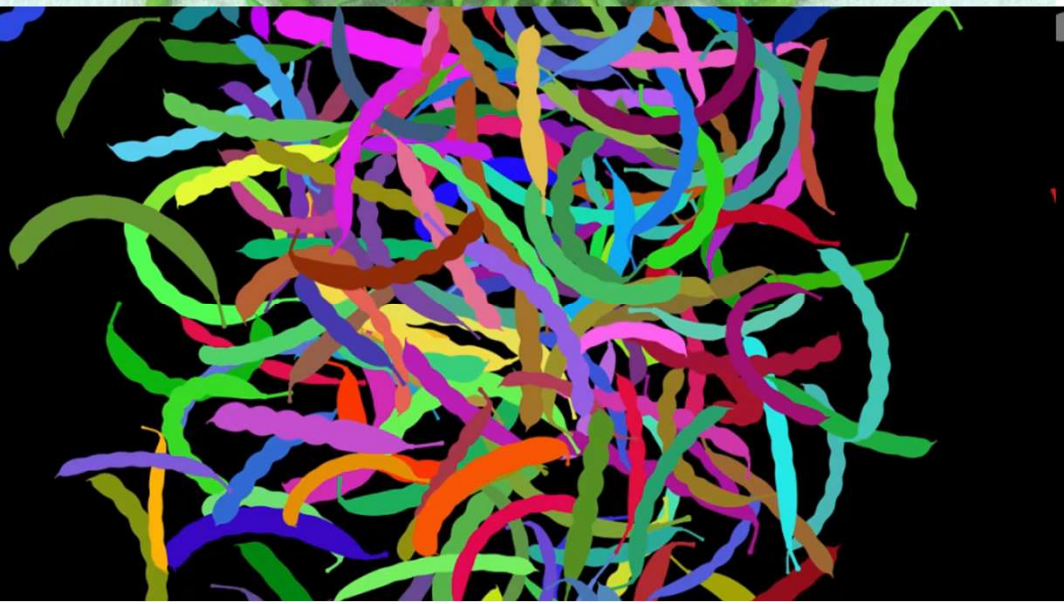
DEMCON

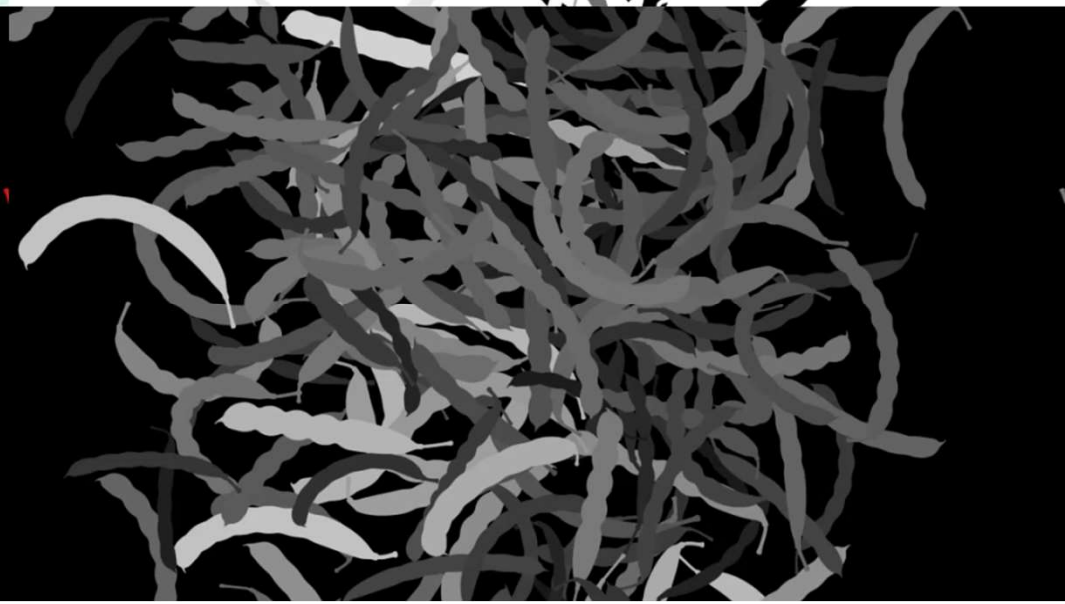
Synthetic Data Explainer - Demcon

Link explainer video Demcon Synthetic Data:

https://vimeo.com/810084583/7847b90c9e?share=copy

DEMCON

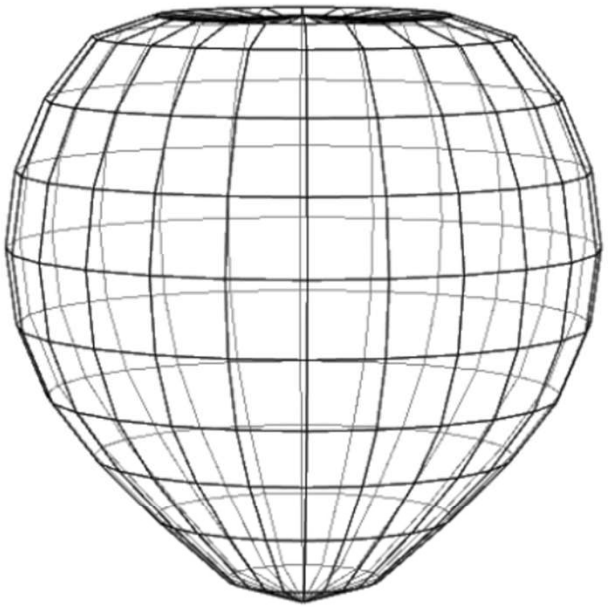




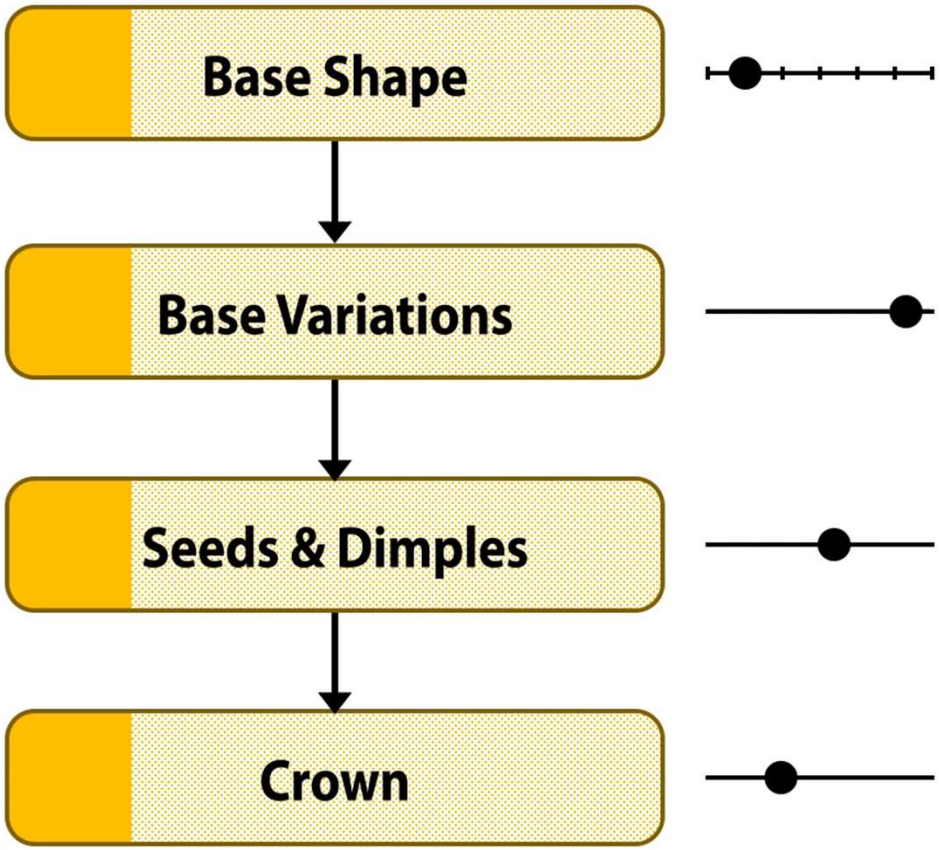
How..?

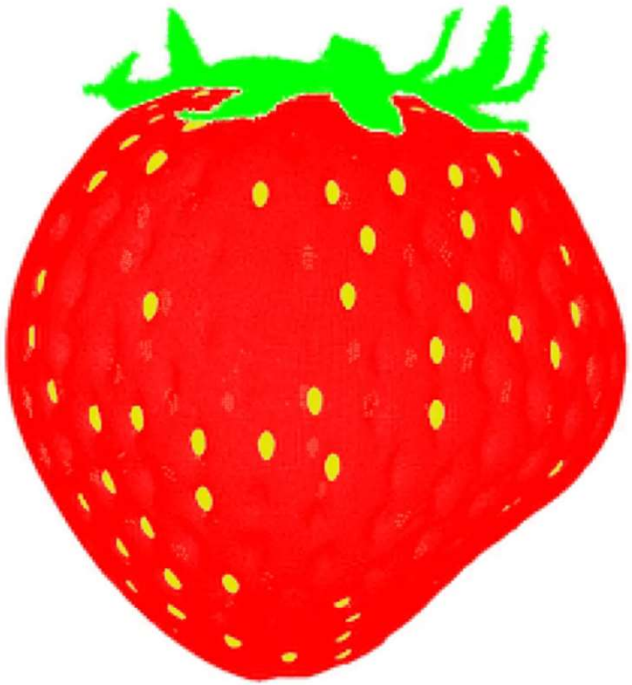
DEMCON

Strawberry Generator



Strawberry Generator





Why..?

DEMCON

Why Synthetic Data?

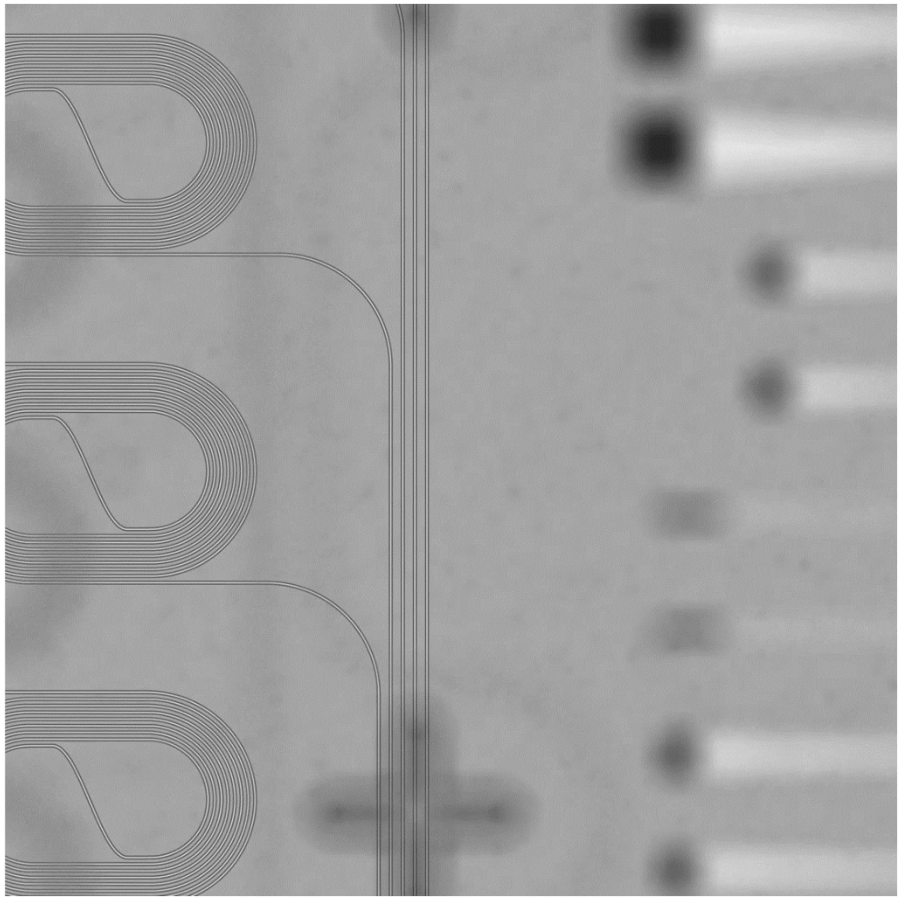
*Generate **data before the system is operational***

---

*Generate **highly specific data at** a **scale** and variation otherwise not feasible*

***Rich labeling**: enable new ML strategies*

***Full control over data**: resolve and address biases in simulation, add variation where needed*



DEMCON

Plastic Waste Sorting

Synthetic Data

DEMCON

Plastic Waste Sorting

Overview

- Identify valuable HDPE plastics using RGB cameras only: milk containers
- Highly varying stream of objects
- Data-centric approach: pipeline to generate synthetic data from CAD geometries

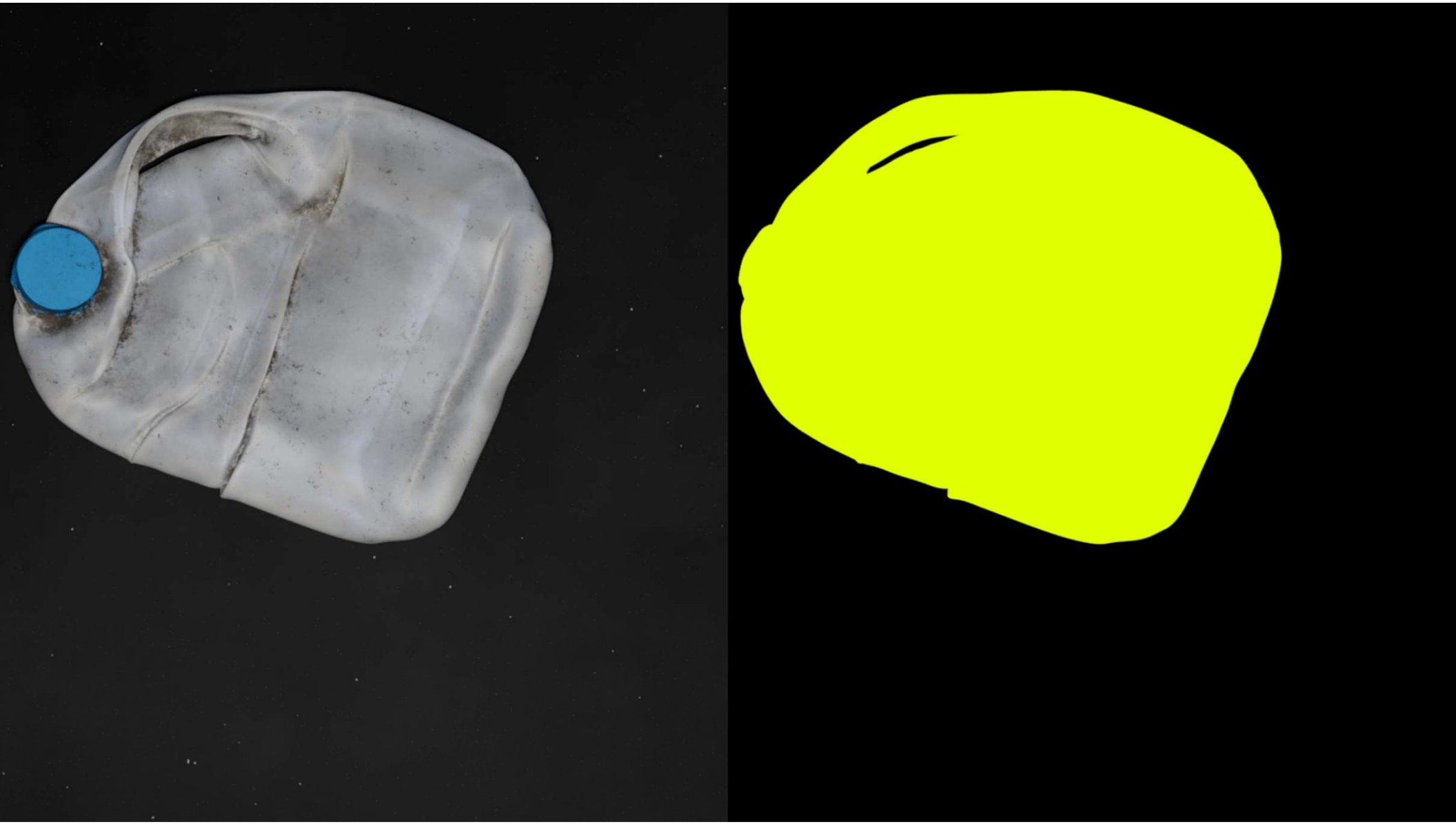
DEMCON

Plastic Waste Sorting

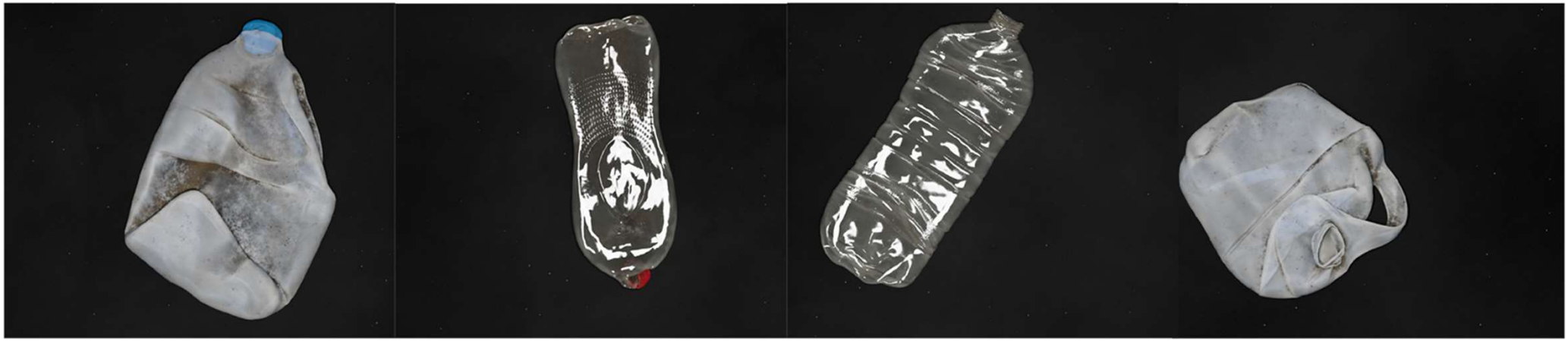
Procedural Methods

- Variation present in the real-world domain needs to be captured in the synthetic data
- Deformation, damage and dirt modules implemented
- Example on the right, shows deformation of a CAD model in simulation
- Examples of full quality renders on the next slide
- General, scalable architecture
- Automatic and pixel perfect labeling

DEMCON



Synthetic items



Feedstock pictures

Project

CT – Lung Nodules

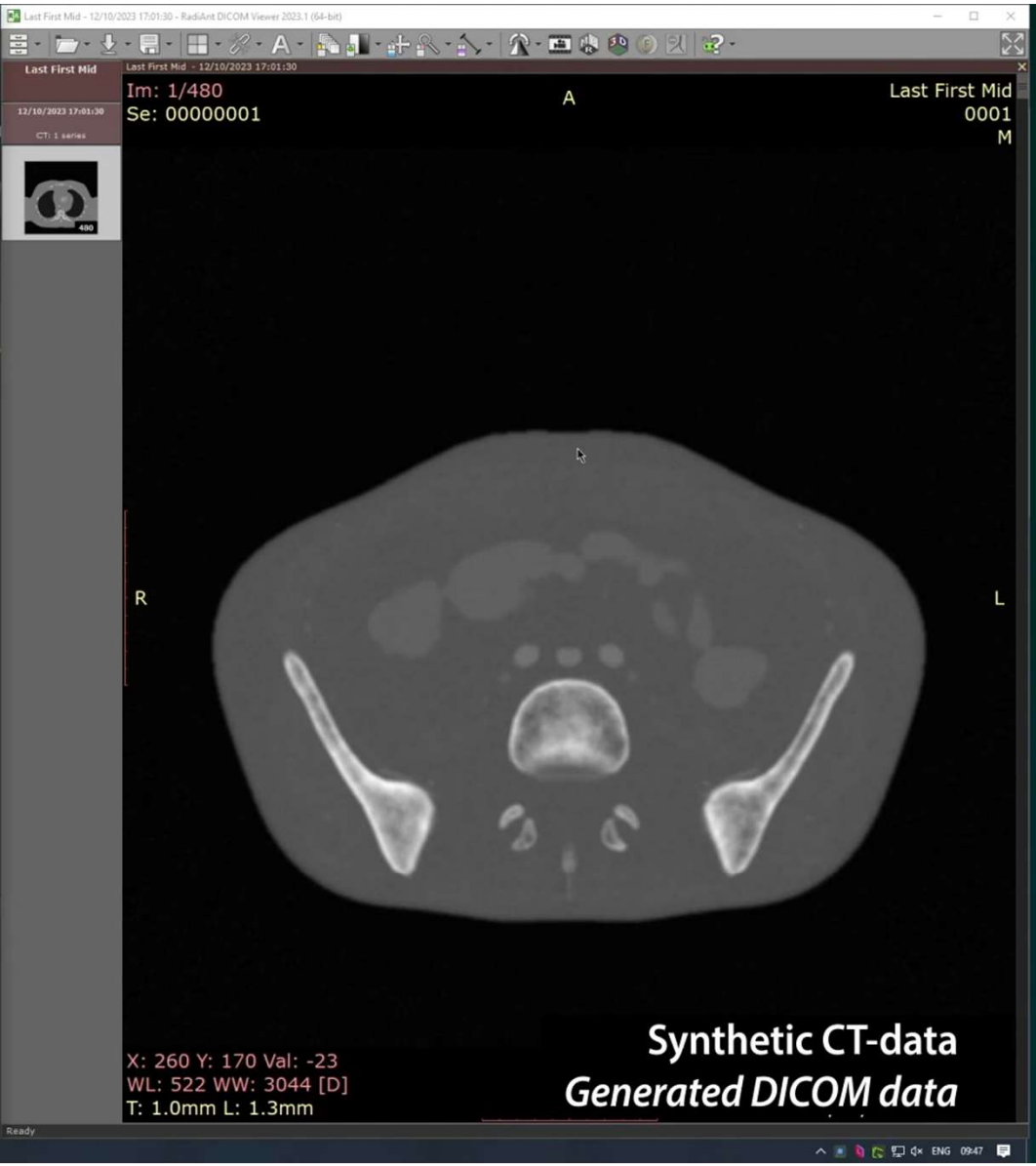
DEMCON

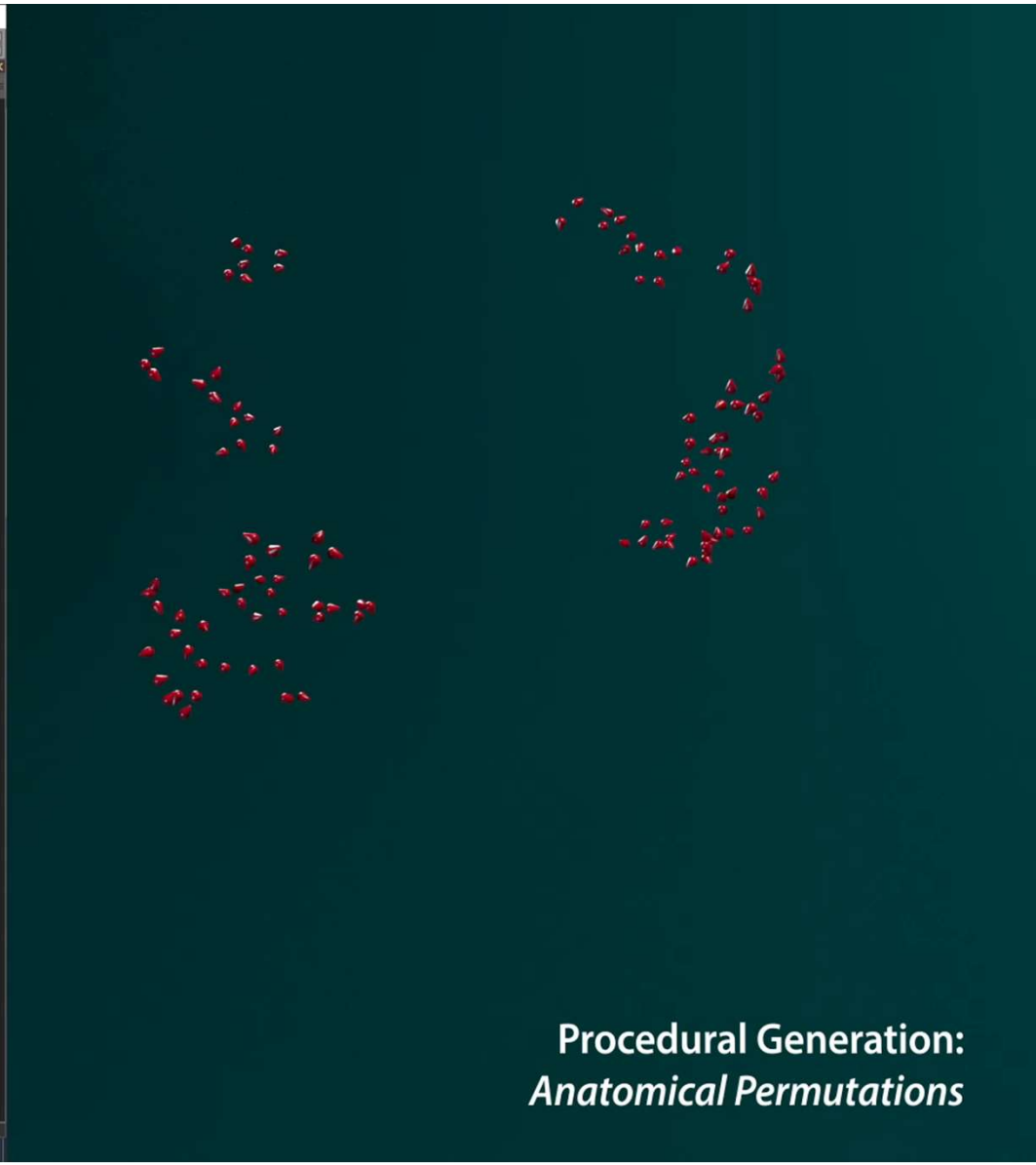
Synthetic Medical Data

Generated CT Data

- Generate permutations of anatomy
- Generate pulmonary nodules into the generated anatomy
- Embed nodule features from guidelines into synthetic data
- Transfer of domain knowledge to machine learning through the synthetic data model

DEMCON

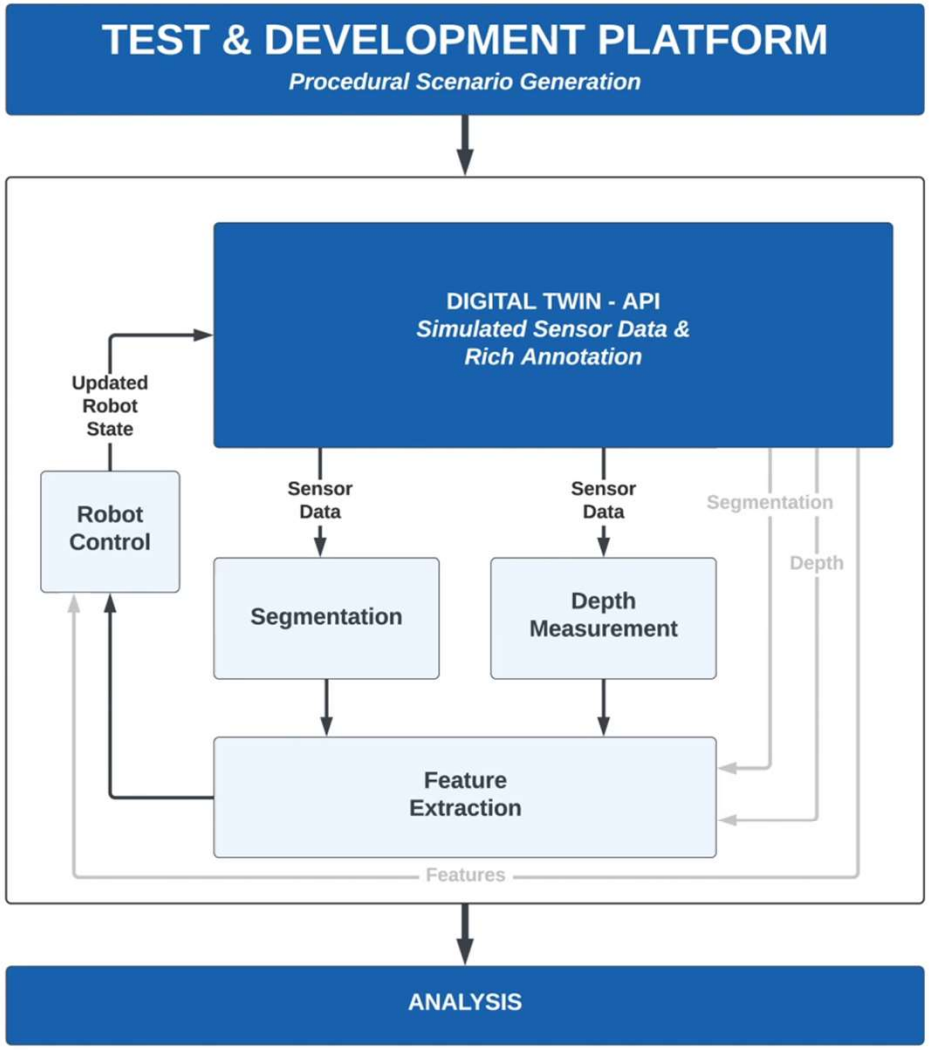

Last First Mid - 12/10/2023 17:01:30 - RadiAnt DICOM Viewer 2023.1 (64-bit)
Last First Mid
12/10/2023 17:01:30
CT: 1 series
480
Last First Mid - 12/10/2023 17:01:30
Im: 1/480
Se: 00000001
A
Last First Mid
0001
M
R
L
X: 260 Y: 170 Val: -23
WL: 522 WW: 3044 [D]
T: 1.0mm L: 1.3mm
Synthetic CT-data
Generated DICOM data
Ready
ENG
09:47

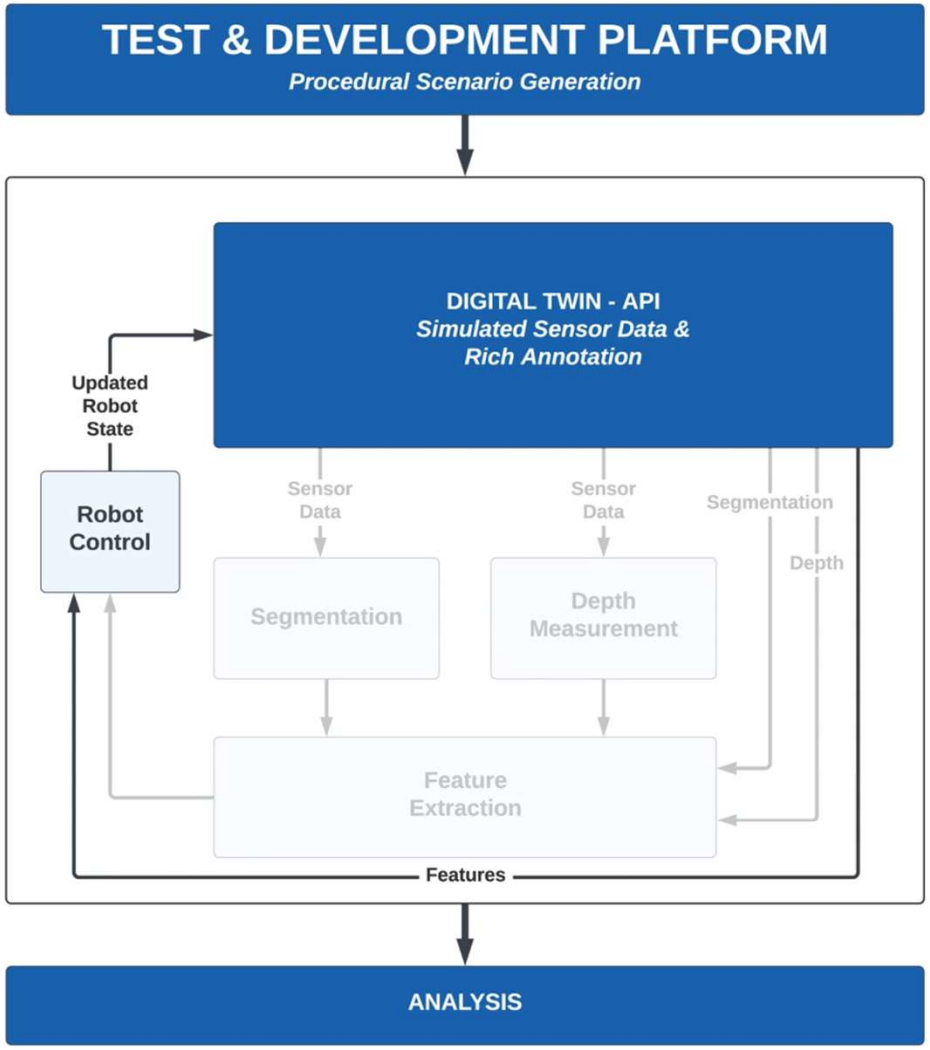

Procedural Generation:
Anatomical Permutations

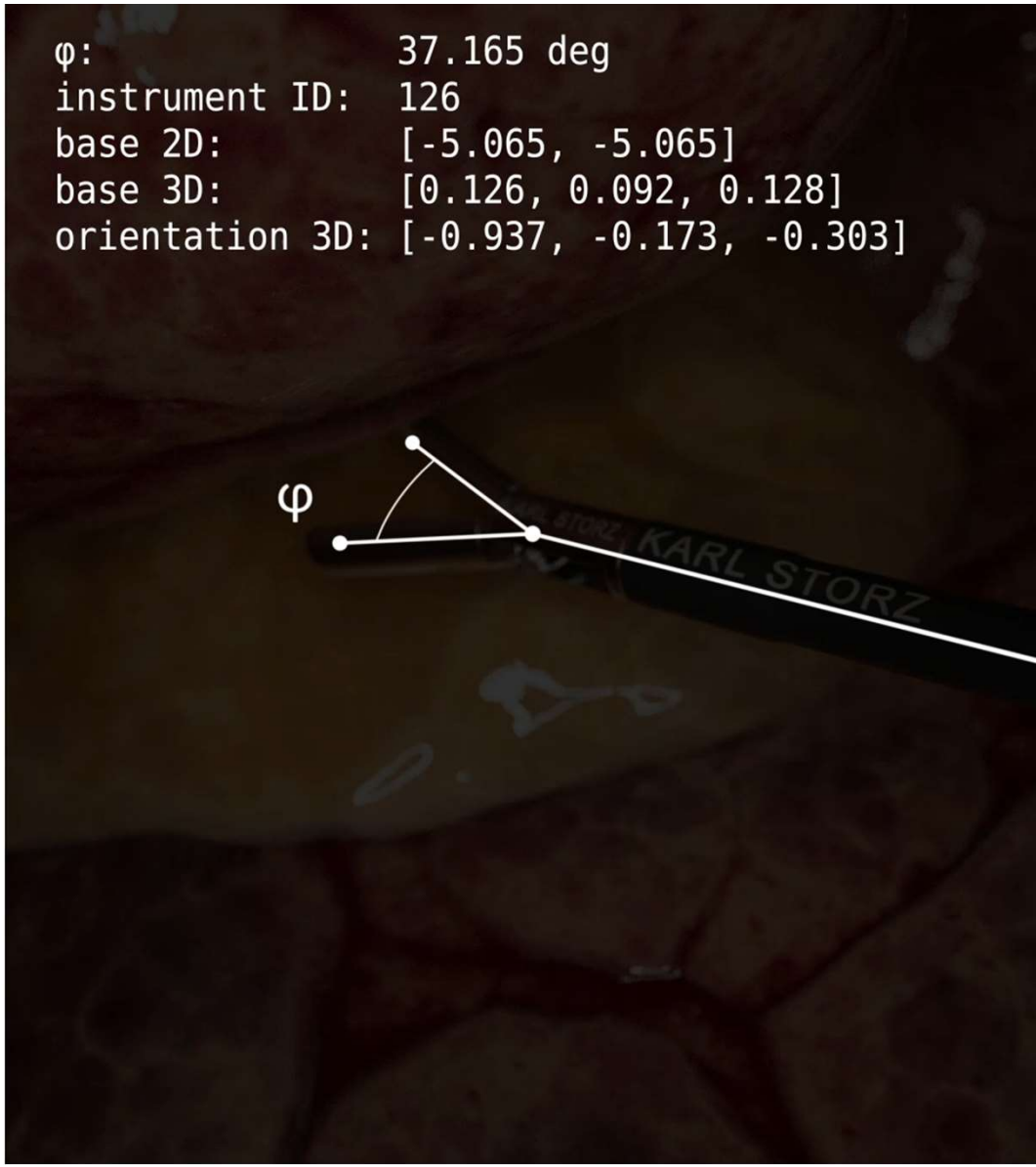
Autonomous Robotics

Test & Development Platform

DEMCON


TEST & DEVELOPMENT PLATFORM
Procedural Scenario Generation
DIGITAL TWIN - API
Simulated Sensor Data &
Rich Annotation
Updated
Robot
State
Robot
Control
Sensor
Data
Sensor
Data
Segmentation
Depth
Segmentation
Depth
Measurement
Feature
Extraction
Features
ANALYSIS


TEST & DEVELOPMENT PLATFORM
Procedural Scenario Generation
DIGITAL TWIN - API
Simulated Sensor Data &
Rich Annotation
Updated
Robot
State
Robot
Control
Sensor
Data
Sensor
Data
Segmentation
Depth
Segmentation
Depth
Measurement
Feature
Extraction
Features
ANALYSIS


φ: 37.165 deg
instrument ID: 126
base 2D: [-5.065, -5.065]
base 3D: [0.126, 0.092, 0.128]
orientation 3D: [-0.937, -0.173, -0.303]
φ
KARL STORZ

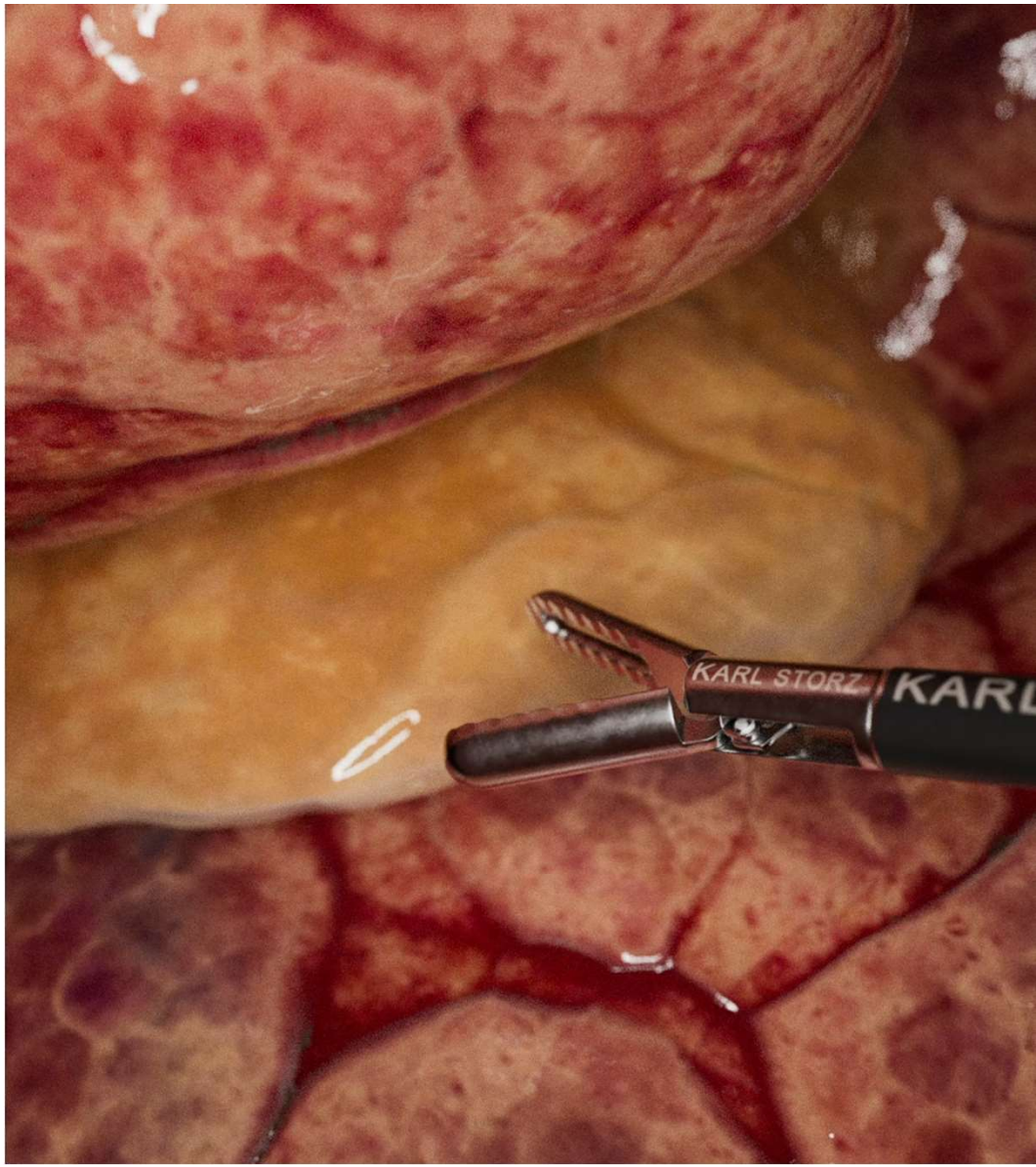
TEST & DEVELOPMENT PLATFORM

Permutations of Scenarios:

- Hardware failures
- Procedural model permutations
- Procedural animation permutations
- Edge cases
- **User scenarios**
- ...

Full robot-environment interactions in simulation

Procedural Scenario Generation



How do you validate these synthetic data sets?

If you could annotate at will,
what avenues could this open for you?

How about my use case?

Would you like to get some coffee?

Digital twinning: evaluate and
iterate design through simulation

Procedural geometry creation:
What about topology permutations?

Questions?

Can I link my specific simulation
tool to this pipeline?

vincent.bos@demcon.com

How well does a problem need to be
understood to be successfully modelled?

I have a funky idea on how to use
this technique, would you like to hear it?
(yes)

When does synthetic data makes sense
over real data?

DEMCON

FME AI FOR INDUSTRY JAAREVENT

Bedankt voor je aandacht!

FME POWERED BY DUTCH TECHNOLOGY