ORGANISED BY

BDV BIG DATA VALUE ASSOCIATION

IN COLLABORATION WITH

IFE

NTNU

SINTEF

WESTERN NORWAY RESEARCH INSTITUTE
VESTLANDSFORSKING

DATAWEEK

JOIN.LEARN.SHARE.GET VALUE

DATA FJORDS: UNLOCKING AI FOR INDUSTRY & SOCIETY

5-6 May 2026 / Oslo

SUPPORTED BY

The Research Council of Norway

data-week.eu
#DataWeek26

DATAWEEK
JOIN.LEARN.SHARE.GET VALUE

Synthetic Data in Healthcare - Latest Successes, Challenges and Insights

Moderators:
Riske Meijer, Eindhoven University of Technology
Rob Smeets, Philips

ORGANISED BY BDV BIG DATA VALUE ASSOCIATION
SUPPORTED BY The Research Council of Norway
IN COLLABORATION WITH IFE NTNU SINTEF WESTERN NORWAY RESEARCH INSTITUTE VESTLANDSFORSKING

DATAWEEK
JOIN.LEARN.SHARE.GET VALUE

Synthetic Data in Healthcare

Brief Intro to Healthcare TF White Paper
(published September 2025)

Presenter: Mr. Rob Smeets, Philips

White Paper URL: https://bdva.eu/news/synthetic-data-in-healthcare/

ORGANISED BY BDV BIG DATA VALUE ASSOCIATION
SUPPORTED BY The Research Council of Norway
IN COLLABORATION WITH IFE NTNU SINTEF WESTERN NORWAY RESEARCH INSTITUTE VESTLANDSFORSKING

Synthetic Data: a Strategic Enabler

- Inherent Privacy-by-Design
- Bias Mitigation & Inclusion
- On Demand Generation & Fast Innovation
- Accessibility
  No need to share Real World Data

Prerequisites

- Ensuring Seed Data Quality
- Adoption of Standards, Interoperability
- Inclusive Governance
- Outcome-oriented Evaluation

Unlocking Value with Synthetic Data

virtual patient

Powers real-time monitoring and personalized care through privacy-safe, data-rich digital twins.

scarcity & bias

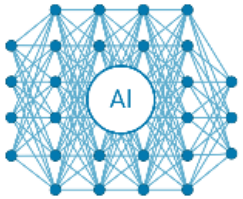
Fills data gaps and supports equity in rare disease and emergency scenarios.

data sharing & collaboration

Enables privacy-compliant, cross-border research with scalable, high-fidelity datasets.

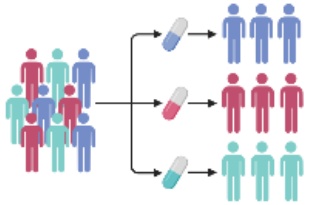
ethics, law & innovation

Reduces legal risk, promotes inclusive design, and accelerates time-to-market.



AI fairness & performance

Enhances model robustness, equity, and context specificity across domains.



clinical trials

Enables synthetic control arms, simulations, and earlier regulatory engagement.

AI system readiness

Provides safe, repeatable conditions for assessing AI reliability and risk.

training & workforce

Equips clinicians and developers with realistic, scenario-rich training tools.

Critical Gaps and Challenges in Synthetic Health Data Research → Recommendations

Clinical Data and AI Model Development for SD Generation

- Limited Access to Real-World Quality Seed Data
- Lack of Common Assessment and Benchmarking Frameworks
- Underdevelopment of Multimodal and Cross-Domain Synthesis
- Inadequate Model Accuracy, Explainability and Interpretability
- Lack of acceptance and trust in AI by clinical stakeholders

Governance and Legal Ambiguities

- Unclear Data Ownership and Exploitation Rights
- Absence of Fit-for-Purpose Data Governance Structures
- Gaps in Data Sharing Agreements

Infrastructure, Regulatory and Policy Gaps

- Lack of Regulatory Sandboxes and Testbeds
- Limited Support for Cross-Border Collaboration
- Insufficient Funding for Scalable Innovation

Scientific, Technical and Security Gaps

- Lack of Research on Privacy Guarantee Composition
- Gaps in Post-Quantum Security Readiness
- Usability Constraints

BDV BIG DATA VALUE ASSOCIATION

Home / Task Forces / Healthcare

BDVA Task Forces

Healthcare

Thank you!

Main editors

Dr. Sofia Tsekeridou (Netcompany SEE & EUI)
e-mail: sofia.tsekeridou@netcompany.com

Dr. Iñaki Fernández Pérez (CARTIF)
e-mail: inafer@cartif.es

Dr. Sinem Tas (Philips)
e-mail: sinem.tas@philips.com

Saurav Baidya (Philips)
e-mail: saurav.baidya@philips.com

DATAWEEK
JOIN.LEARN.SHARE.GET VALUE

Barriers and Solutions for Synthetic Healthcare Data

ORGANISED BY
BDV BIG DATA VALUE ASSOCIATION

SUPPORTED BY
The Research Council of Norway

IN COLLABORATION WITH
IFE
NTNU
SINTEF
WESTERN NORWAY RESEARCH INSTITUTE
VESTLANDSFORSKING

Gradiant (RTC ext Vigo, Spain)

Hi!
My name is
Jorge Amor Río

I lead the Data & AI infra teams:
Cloud Continuum, Data Quality, Big Data,
Data Ops, ML Ops, Data Spaces

Head of Data & AI Infrastructure

gradiant

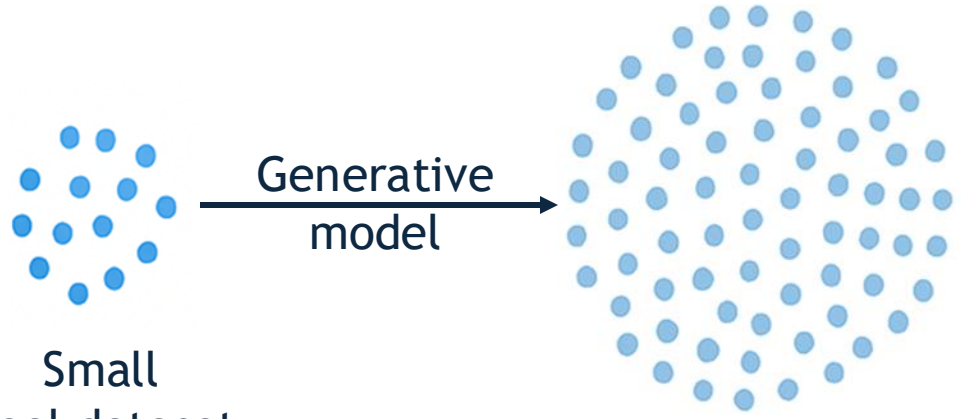
Innovation provider, with more than 15 years of experience in technology incubation

Vigo, Spain

The technical reality behind synthetic data

- Synthetic data is often seen as a solution to healthcare data scarcity
- Generative AI learns and samples from patterns in existing data distributions
- More synthetic data does not automatically mean more clinically relevant information for AI models

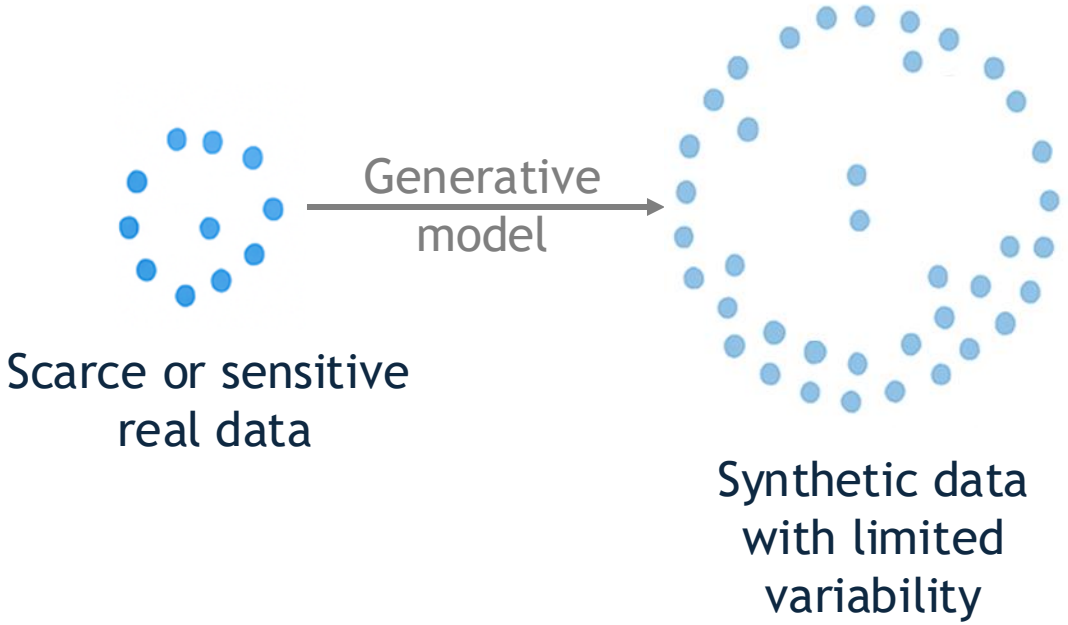
Volume can scale quickly, while clinically meaningful variability often remains constrained



The key challenge is generating meaningful variability

The data scarcity paradox in healthcare

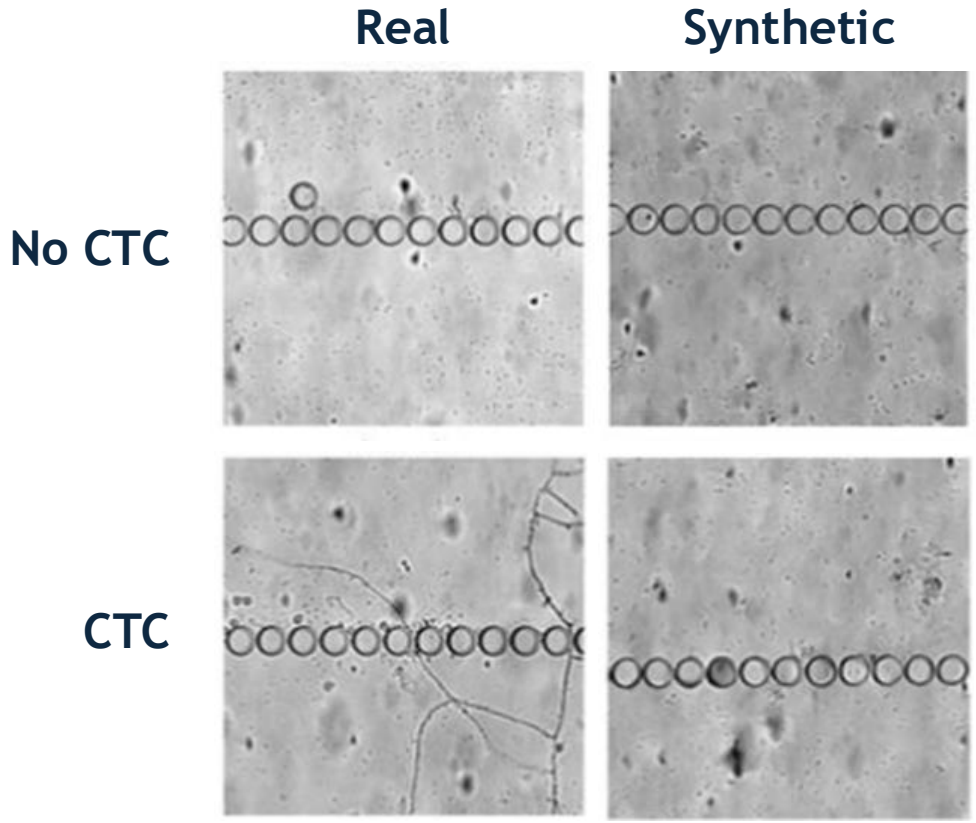
- Synthetic data is most valuable where healthcare data is scarce, sensitive or rare
- These settings often lack enough real data to train robust generative models
- Expertise and context become critical precisely in these situations



Synthetic data is most needed where it is hardest to generate properly

Synthetic data in practice: a healthcare example

- Microscopy images for automatic detection of circulating tumour cells (CTC)
- Limited and highly imbalanced real datasets, especially for rare phenotypes
- Synthetic image generation used to enrich the training data



Expert knowledge and validation are required to generate usable synthetic data

From isolated models to trusted ecosystems

- Synthetic data challenges cannot be addressed by models alone
- Healthcare requires shared expertise, validation, governance and trust
- Ecosystems enable collaboration across data, organisations and expertise

AI Factories
Model training and validation

Example
1HealthAI Factory

Data Spaces
Federated data and integrated AI services

Data Labs
Local prototyping and experimentation

TEF
Real-world AI testing and validation

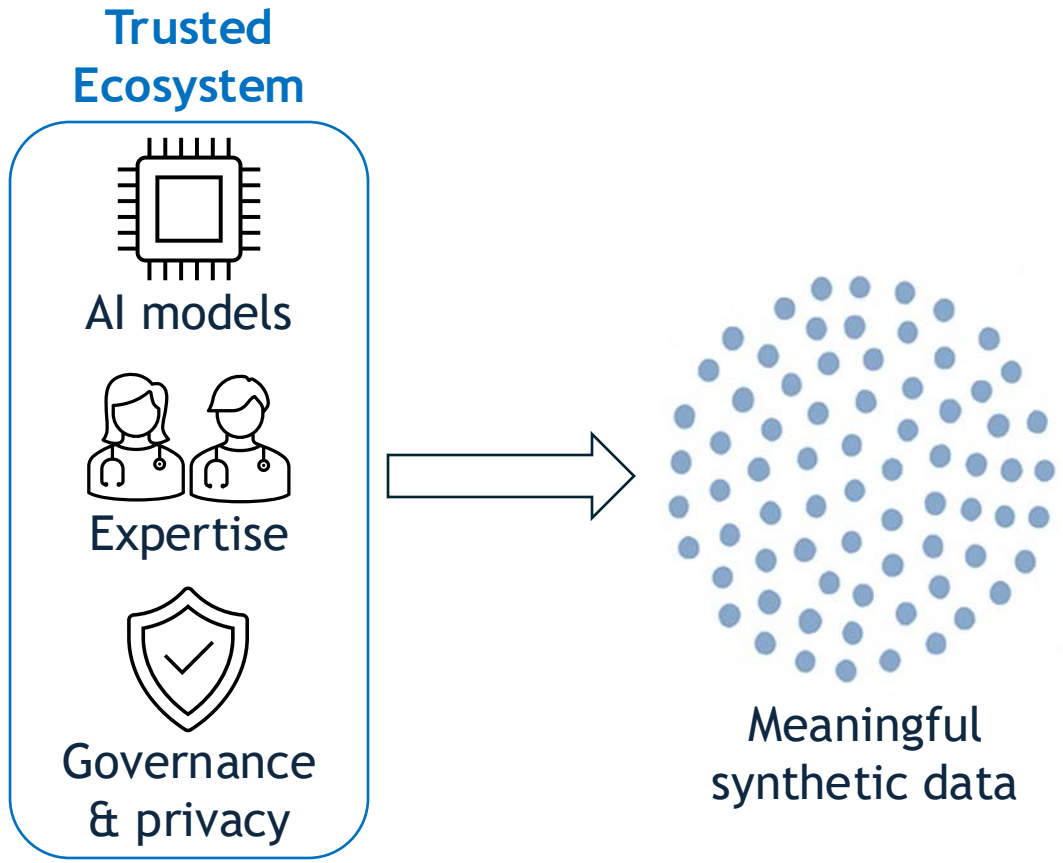
EDIH
Digital adoption and support

Antennas
Distributed access and AI adoption

Trusted ecosystems are needed to make synthetic data work in healthcare

Key takeaways

- Synthetic data can help, but it does not solve data scarcity by itself
- Meaningful variability requires data, expertise and validation
- Trusted ecosystems are key to scaling synthetic data in healthcare



Synthetic data works best when combined with expertise and trusted ecosystems

gradiant

Jorge Amor Río

jamor@gradiant.org

+34 603 643 399

DATAWEEK
JOIN.LEARN.SHARE.GET VALUE

ORGANISED BY BDV BIG DATA VALUE ASSOCIATION

SUPPORTED BY The Research Council of Norway

IN COLLABORATION WITH IFE NTNU SINTEF VESTLANDSFORSKING

DATAWEEK
JOIN.LEARN.SHARE.GET VALUE

Controlled Synthetic Medical Data Generation and Evaluation

Fraunhofer PORTUGAL

ORGANISED BY BDV BIG DATA VALUE ASSOCIATION
SUPPORTED BY The Research Council of Norway
IN COLLABORATION WITH IFE NTNU SINTEF WESTERN NORWAY RESEARCH INSTITUTE VESTLANDSFORSKING

TEXT

¡Hello!
My name is Ricardo Santos

A Biomedical Engineer working on AI methods that address the challenges of real-world data, such as multimodality, missing data, and longitudinal data.
Interests span from representation learning to synthetic data and applications from health to industry settings.

Fraunhofer Portugal AICOS

Scientific Areas

HUMAN-CENTRED DESIGN

- Design Research
- Human-Computer Interaction
- Health Technology Assessment and Implementation

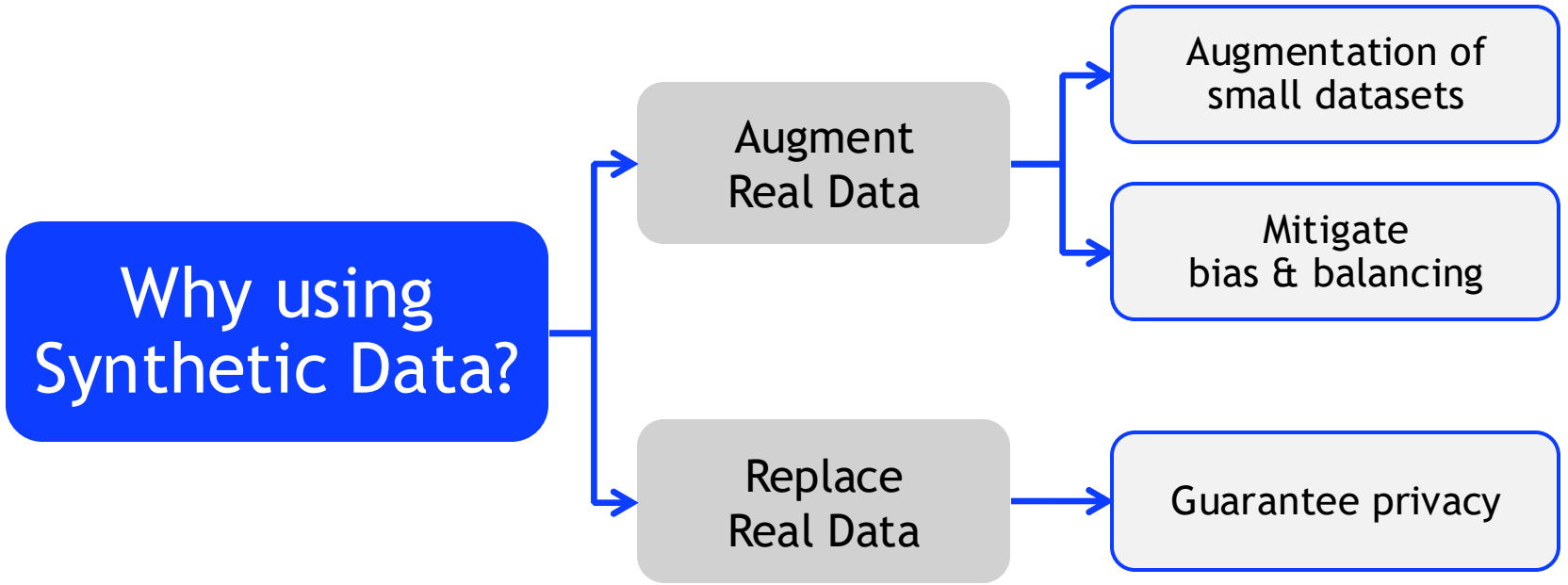
CONNECTED THINGS

- Embedded Electronics
- Edge and Cloud Computing
- Quality Assurance and Regulatory Affairs

INTELLIGENT SYSTEMS

- Data Driven Intelligence
- Learning in Low-Data Scenarios
- Trustworthy and Operational AI

Synthetic Data Generation

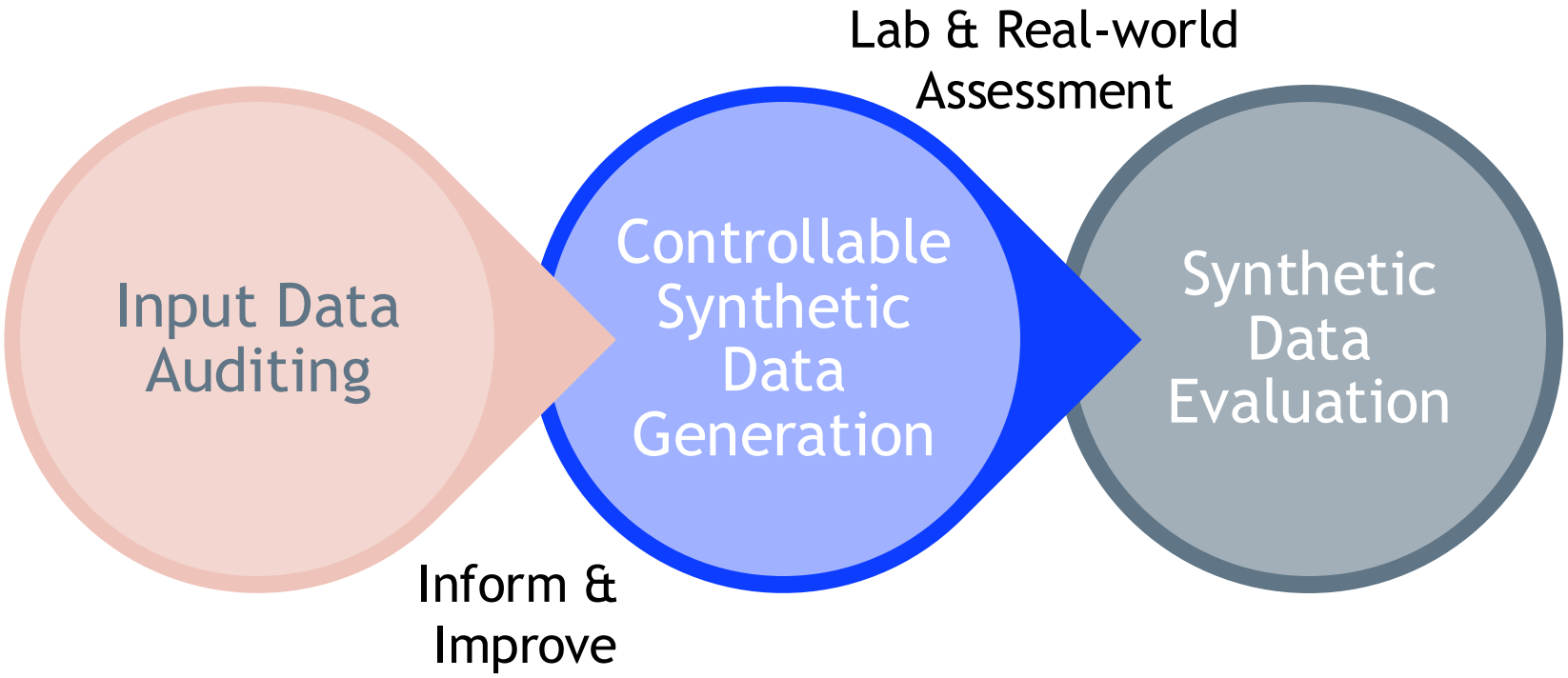


- Time-consuming or Expensive Acquisition
- Low Representativeness
- Poor Annotation

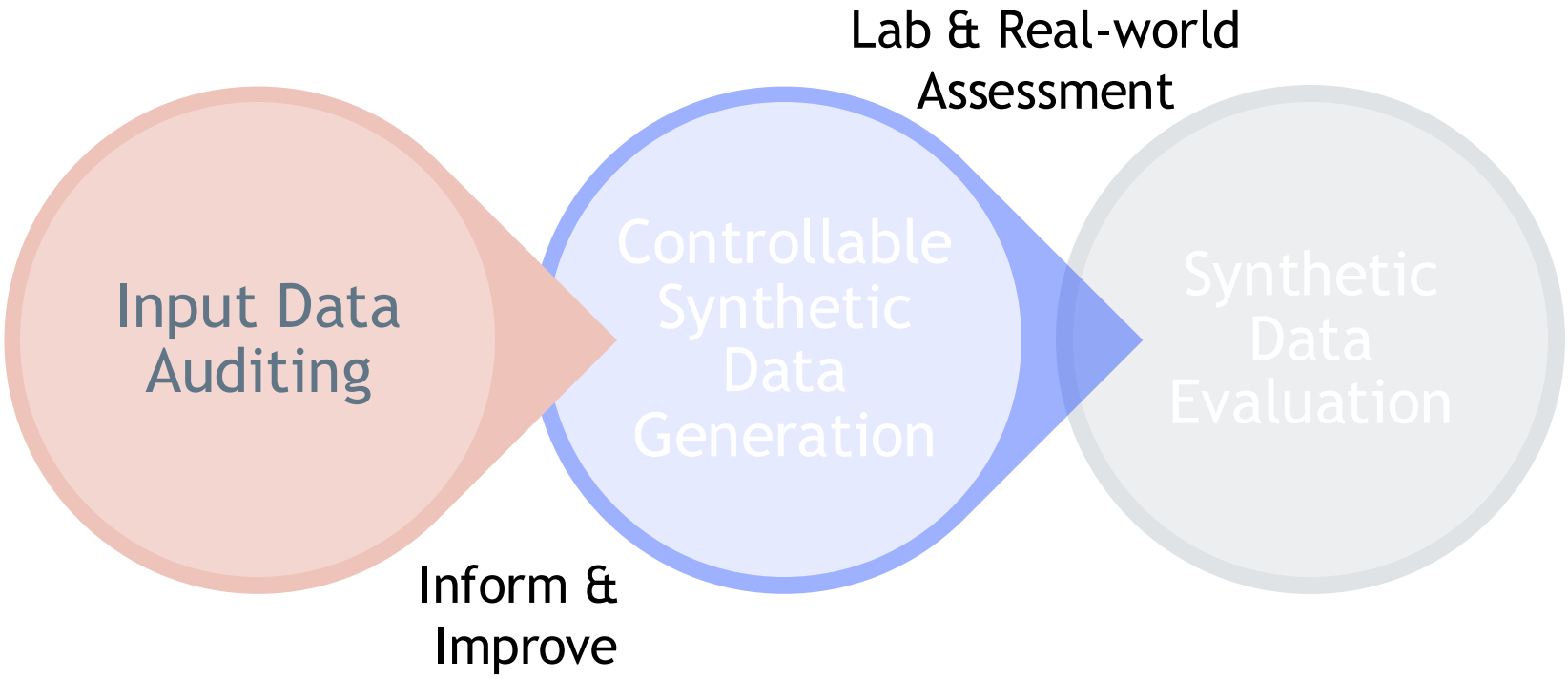
AISYM4MED

Co-funded by the European Union
UK Research and Innovation
Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

Synthetic Data Generation Pipeline



Synthetic Data Generation Pipeline



pyMDMA Library

Multimodal Data Metrics for Auditing Real and Synthetic Data

1 **Open-source Python library** with standardized metric implementations based on **NumPy** and **PyTorch**

2 Comprehensive suite of metrics for assessing **real and synthetic datasets** as **easily** and **fast** as possible

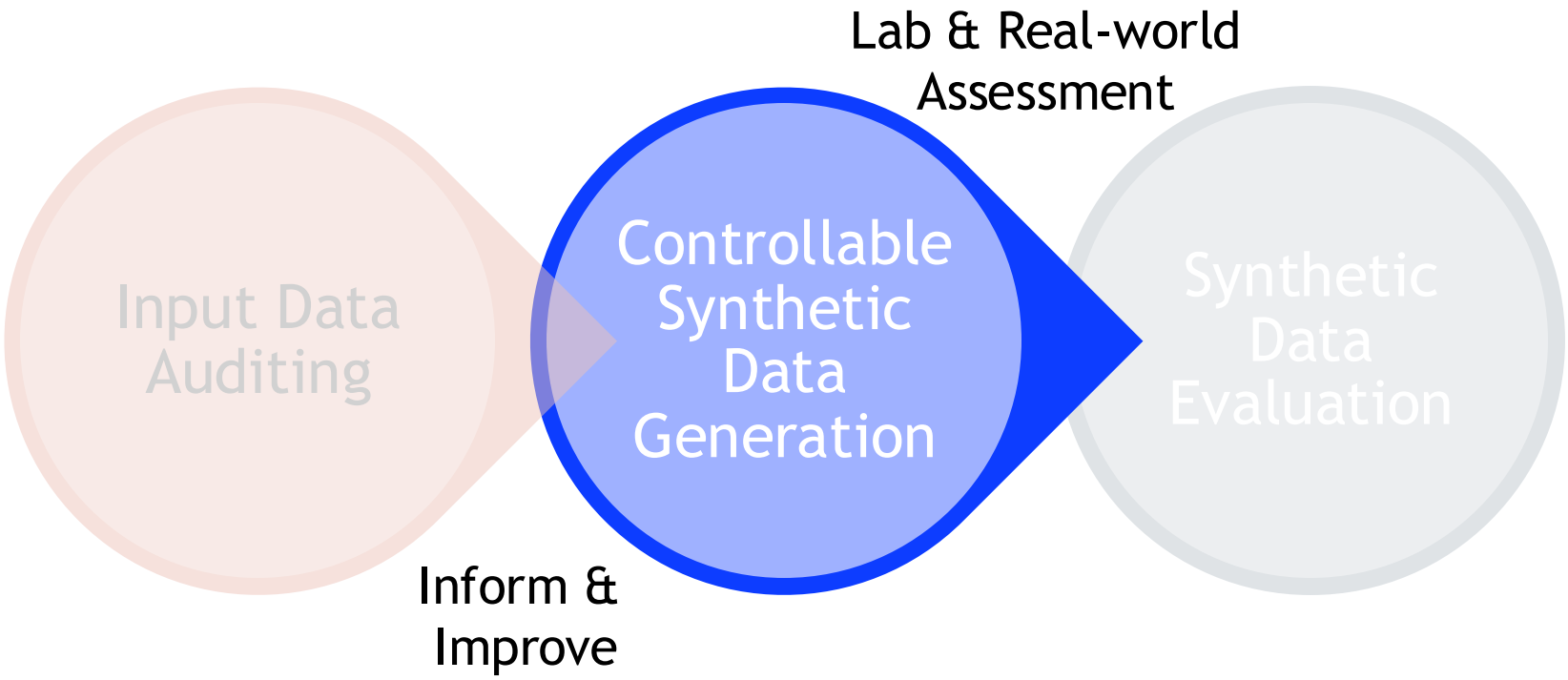
3 Metrics for **Image**, **Tabular** and **Time-Series** Data Modalities **categorized** on **evalution type** and **requirements**

Check out **pyMDMA Library** [1][2]

[1] https://github.com/fraunhoferportugal/pymdma.git

[2] Façoco, I., et al. (2025). pyMDMA: Multimodal data metrics for auditing real and synthetic datasets. SoftwareX, 31, 102256.

Synthetic Data Generation Pipeline



Controlable Synthetic Data

Synthetic Retinal Fundus Image Generation

Context

Retinal Fundus Images and OCT

A1 Retina photographs as the main tools for tracking diseases that affect vision

A2 **Large-Scale Public Datasets** for classification problems

- *EyePACS* - 88k images for DR
- *AIROGS* - 101k images for glaucoma
- *UCSD* - 109k OCT-B images for various diseases

A3 **Broad adoption** of small-scale datasets for classification benchmarking

- *IDRiD* - 516 images for DR grading
- *MESSIDOR* - 1.7k images for DR grading
- *OCT5k* - 1.2k OCT-B images for AMD Dry
- *OCT Data* - 600 OCT-B images for AMD Wet

Challenges and limitations

Datasets

B1 Poorly Annotated Instances (only target diseases)

B2 Low representativeness (demographic groups, acquisition criteria, comorbidities)

B3 Highly imbalanced datasets for target diseases

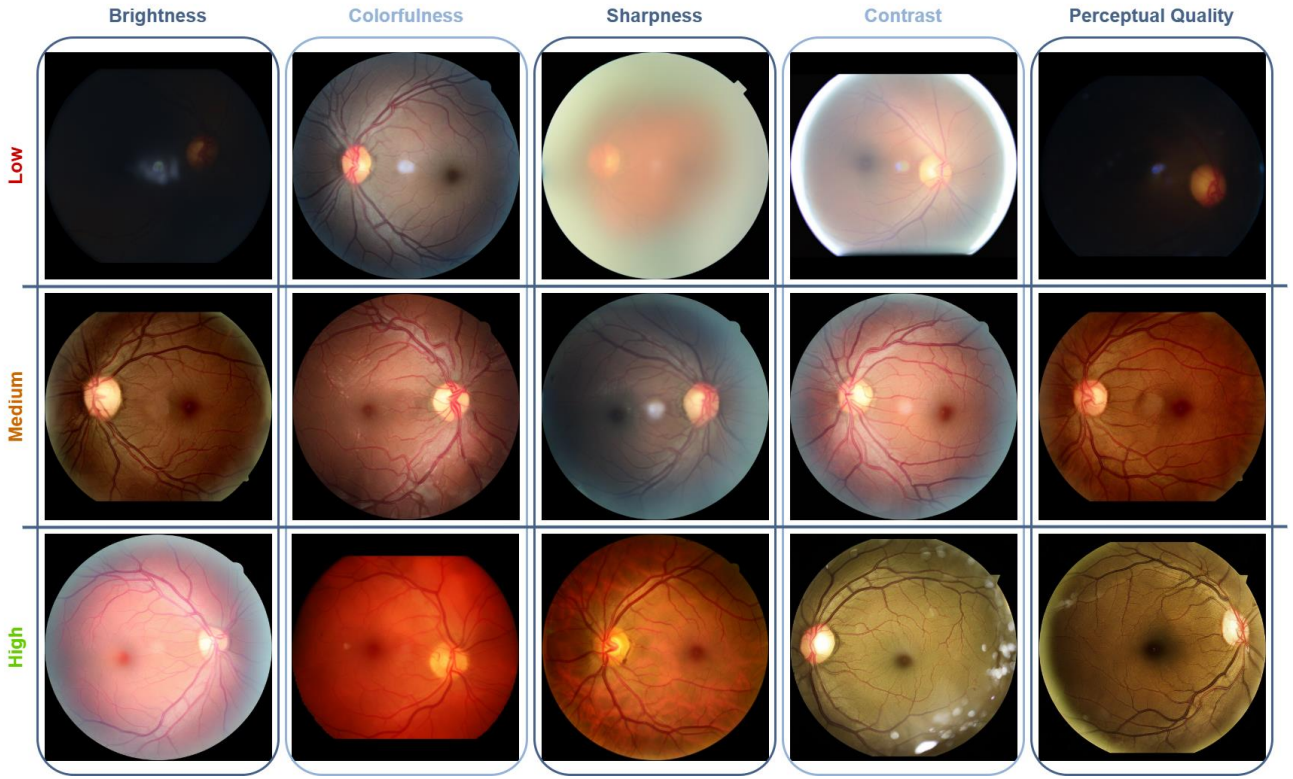
GenAI related

C1 Difficult to ensure realism of key structures (especially finegrained)

C2 Difficult to map location of specific biomarkers given the target disease

Controlable Synthetic Data

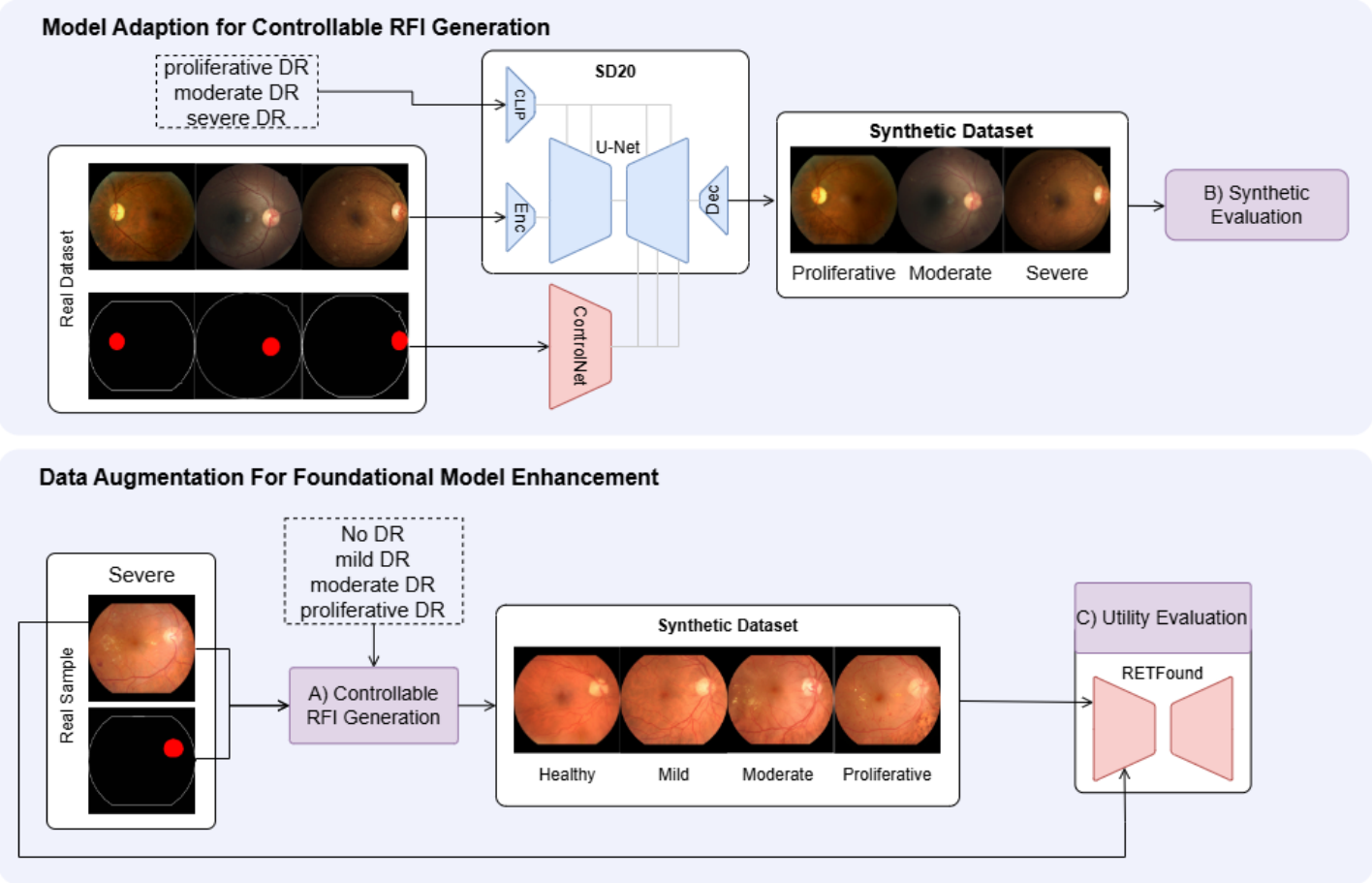
Leveraging Input Data Auditing

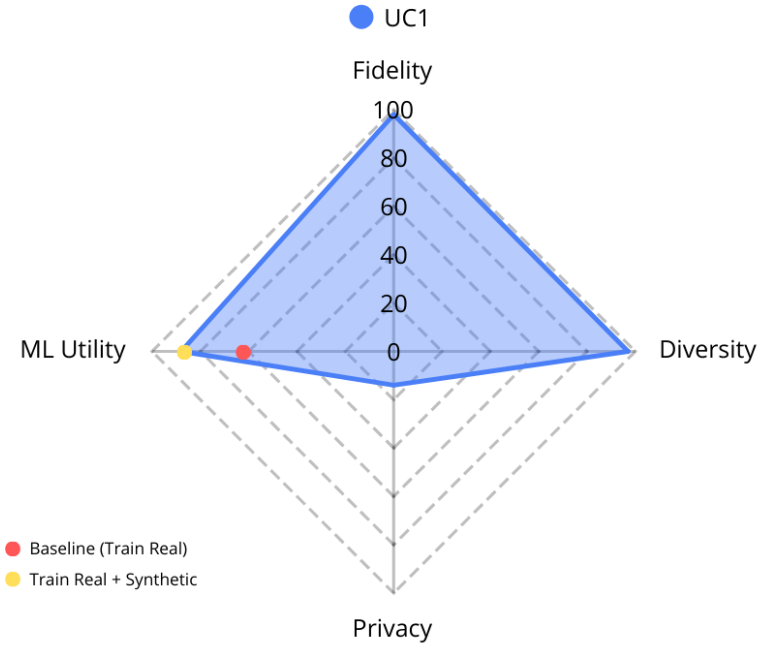


- Conditioning on the disease is not sufficient to retain original data variability.
- Leveraged pyMDMA metrics to sample from discretized distributions to create a specialized prompt.

[1] Façoco, I., et al., "Adapting Stable Diffusion Models for Domain-Specific Medical Imaging: A Case Study in Synthetic Retinal Fundus Image Generation", ECML-PKDD, 2025.

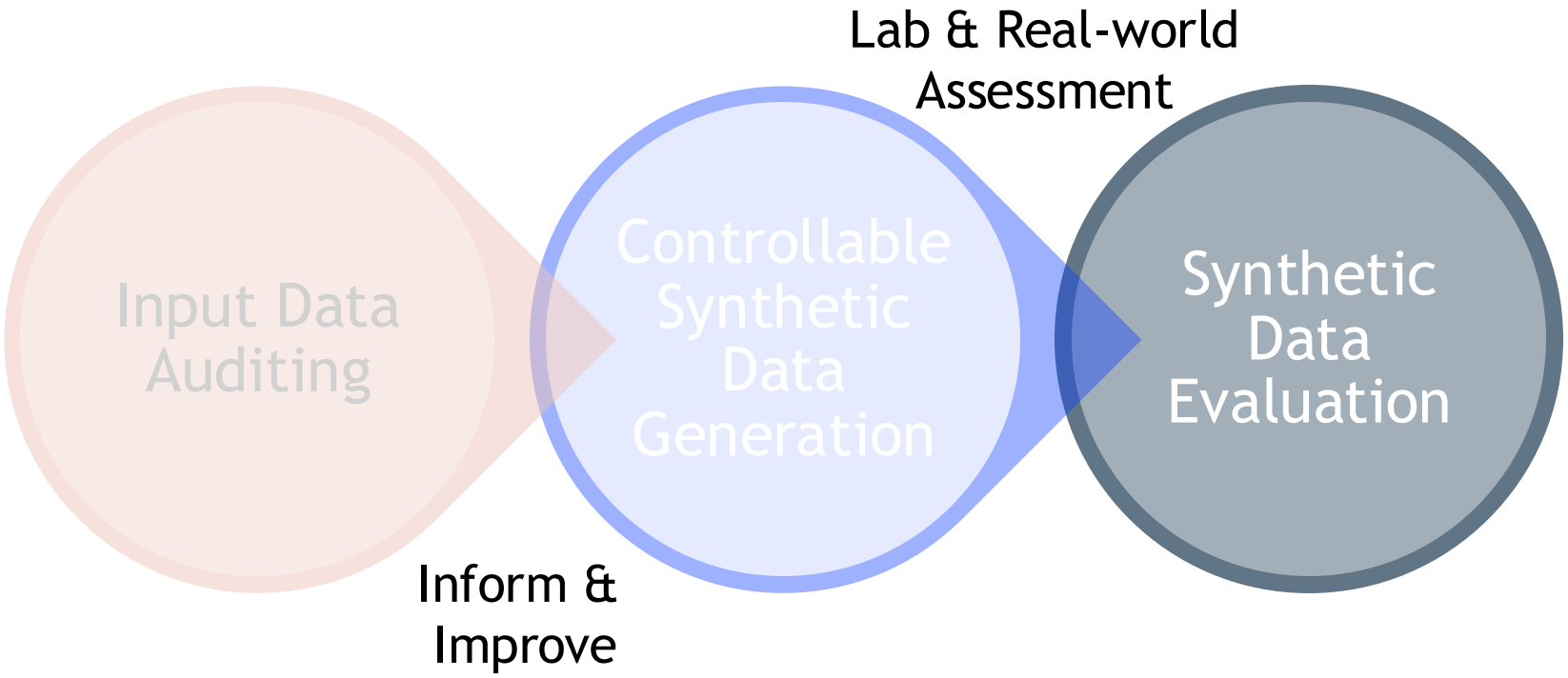
Controlable Synthetic Data





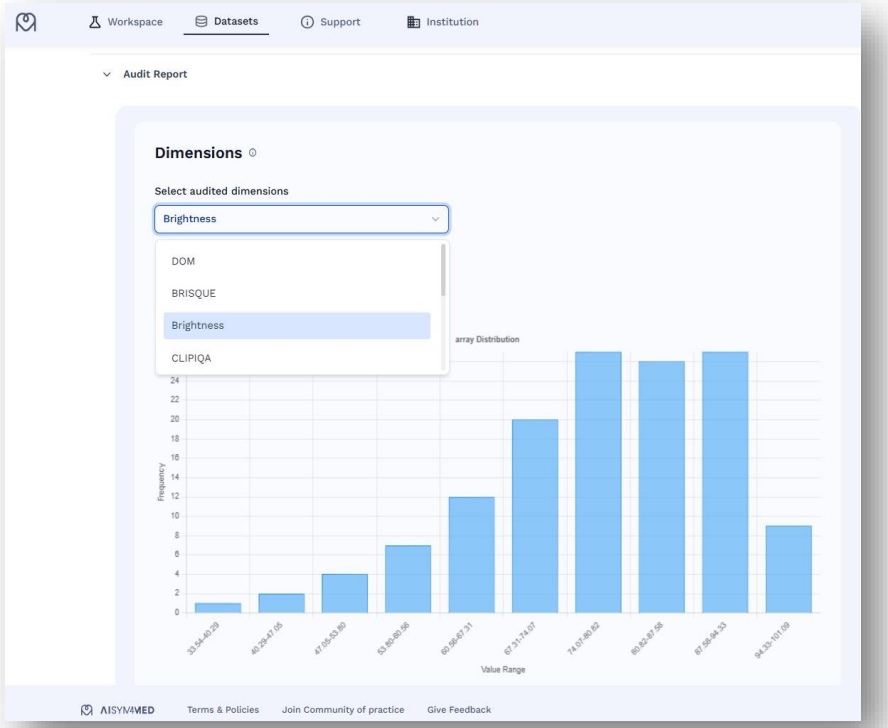
[1] F. Lúcio, I. Façoco and L. Rosado, "Controllable Diffusion Models for Synthetic Fundus Image Generation: Leveraging Text and Image Guidance to Enhance DR Detection", 2026.

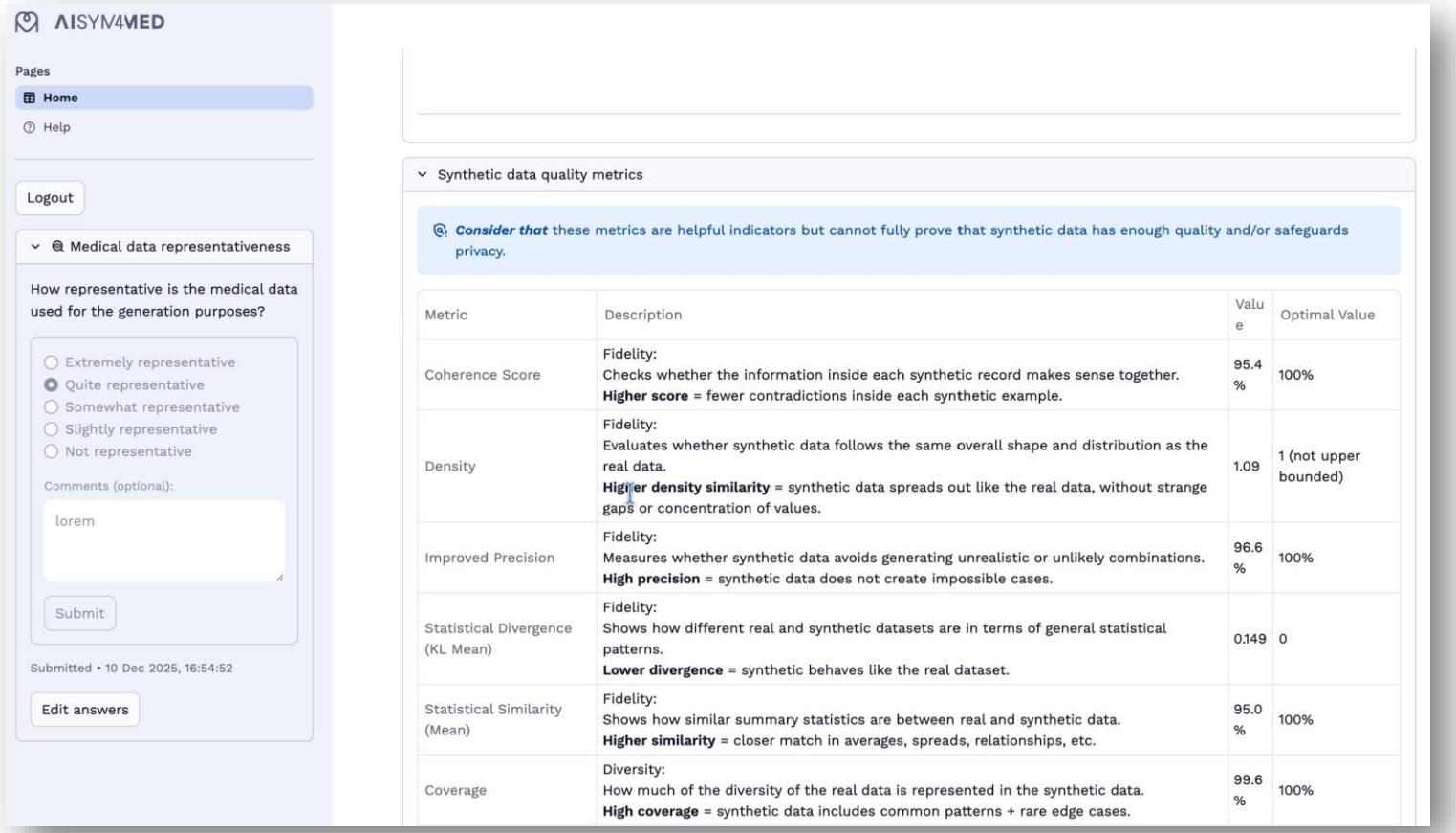
Synthetic Data Generation Pipeline



Auditing Report

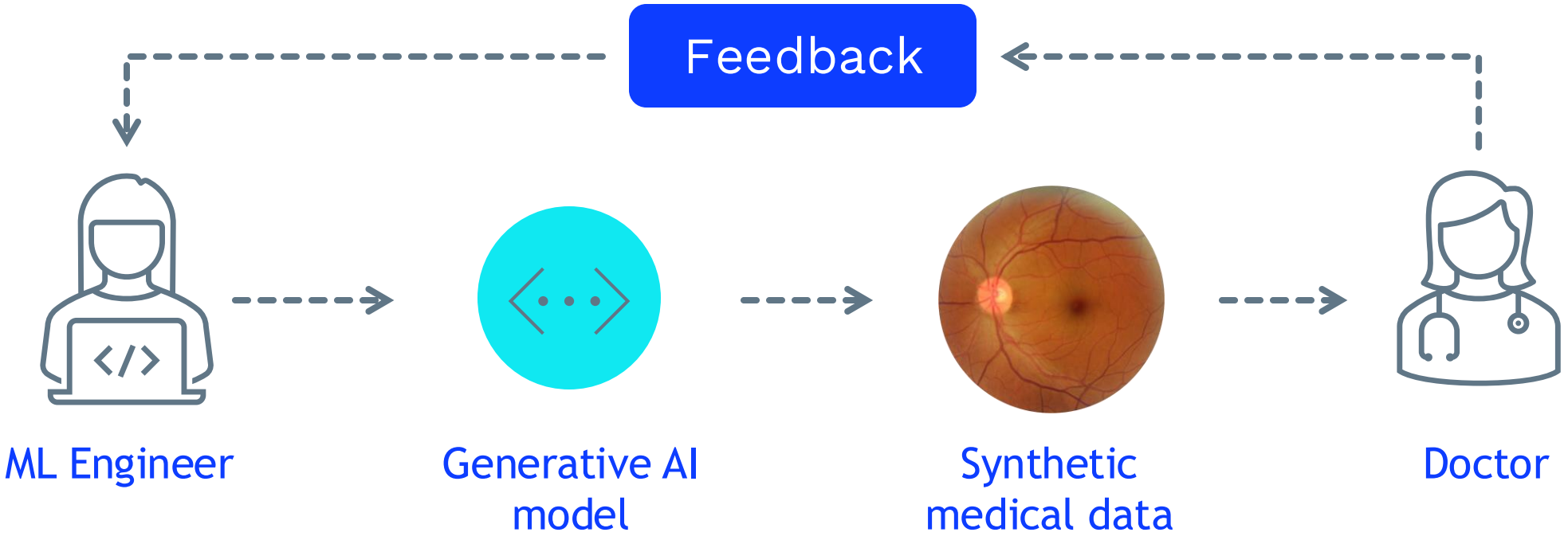
Quantitative Evaluation





Doctor-in-the-Loop

Qualitative Evaluation of Synthetic Data



Thank you!

Ricardo Santos

ricardo.santos@aicos.fraunhofer.pt

ricardobbsantos

Fraunhofer PORTUGAL

AISYM4MED

AISym4MED

aisym4med.eu

@AISym4MED_Project

DATAWEEK JOIN.LEARN.SHARE.GET VALUE

ORGANISED BY BDV BIG DATA VALUE ASSOCIATION

SUPPORTED BY The Research Council of Norway

IN COLLABORATION WITH IFE NTNU SINTEF VESTLANDSFORSKING

DATAWEEK
JOIN.LEARN.SHARE.GET VALUE

BDVA Task Forces
Healthcare

Validating Synthetic Data for Clinical Use: Lessons Learned and Open Challenges

Eleonora Iascone, IRCCS Humanitas Research Hospital

ORGANISED BY BDV BIG DATA VALUE ASSOCIATION
SUPPORTED BY The Research Council of Norway
IN COLLABORATION WITH IFE NTNU SINTEF WESTERN NORWAY RESEARCH INSTITUTE VESTLANDSFORSKING

Synthetic Data Ecosystem

SYNTHEMA GEMINI synthia GENOMED4ALL

European consortia under the HORIZON EUROPE program, with ambitious collaborations between public and private institutions to facilitate the responsible use of Generative AI in healthcare applications such as Synthetic Data (SD) generation and advanced technology for precision medicine.

Multimodal data

Clinical expertise

IRCCS HUMANITAS RESEARCH HOSPITAL

The **AI Center** (born in 2019) designs, implements, optimizes and applies ML/DL/AI predictive models for solving complex data science tasks in health.

SAFE Synthetic vAlidation FramEwork

train

Train is a spin-out from Humanitas Research Hospital founded in 2023, developing Generative AI solutions in healthcare.

Multimodal Generation

Synthetic Data Validation and Human-in-the-loop

Real-world Applications and Clinical Trials

Synthetic Data Validation

Why validating synthetic data matters?

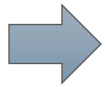
*If synthetic data **looks** like real patient data, how do we know it actually **behaves** like real patient data?*

For **clinicians**: *«Does it preserve clinical meaning?»*

For **data scientist**: *«Does it generalize beyond statistics?»*

For a **regulatory body**: *«Does it protect patients?»*

Validation is not a quality check, but the prerequisite for any downstream use of synthetic data



Generate → **Validate** → Trust → Use

Synthetic Data Validation

The Validation Landscape in Healthcare

Validation Dimensions

Statistical Fidelity

*Do synthetic data **replicate real distributions** and statistical properties?*

Clinical Utility

*Are synthetic data capable of reproducing the same **clinical outcomes** and conclusions?*

*Are synthetic data **useful** for downstream **clinical applications**?*

Privacy Preservation

*Do synthetic data **expose real patients**?*

*Are **sensitive attributes** sufficiently protected?*

Challenges

No shared standard

Every study uses different metrics, thresholds and definitions of «good enough»

Siloed evaluation

Statistical, clinical and privacy dimensions are often assessed separately

Lack of clinical utility

Clinical utility is the most neglected dimension, but necessary to assess real-world applicability and build trust

Synthetic Data Validation

The Validation Landscape in Healthcare

Validation Dimensions

Statistical Fidelity

*Do synthetic data **replicate real distributions** and statistical properties?*

Clinical Utility

*Are synthetic data capable of reproducing the same **clinical outcomes** and conclusions?*

*Are synthetic data **useful** for downstream **clinical applications**?*

Privacy Preservation

*Do synthetic data **expose real patients**?*

*Are **sensitive attributes** sufficiently protected?*

Challenges

No shared standard

Every study uses different metrics, thresholds and definitions of «good enough»

Siloed evaluation

Statistical, clinical and privacy dimensions are often assessed separately

Lack of clinical utility

Clinical utility is the most neglected dimension, but necessary to assess real-world applicability and build trust

Synthetic Data Validation

Clinical Utility

*Are synthetic data capable of reproducing the same **clinical outcomes** and conclusions?*

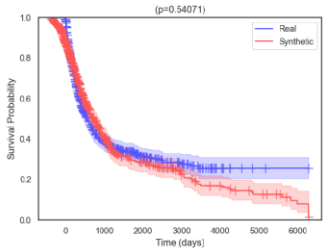
*Are synthetic data **useful** for downstream **clinical applications**?*

Replication of Clinical Studies

Re-run published analyses on synthetic data. Verify **consistency of clinical endpoints**, effect directions and statistical significance.

Survival Analysis

Compare Kaplan-Meier curves (real vs. synthetic) via log-rank test.



Train on Synthetic/ Test on Real (TSTR)

Train **predictive models** on synthetic only and evaluate on real held-out data. Performance parity demonstrates downstream clinical utility.

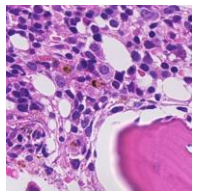
Human-in-the-loop

Clinicians and **experts** are directly involved in the validation process, defining validation criteria upfront and review results downstream.

Applications

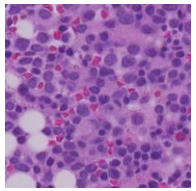
SYNTHEMA

Histopathological Images Validation



REAL

Generative Model

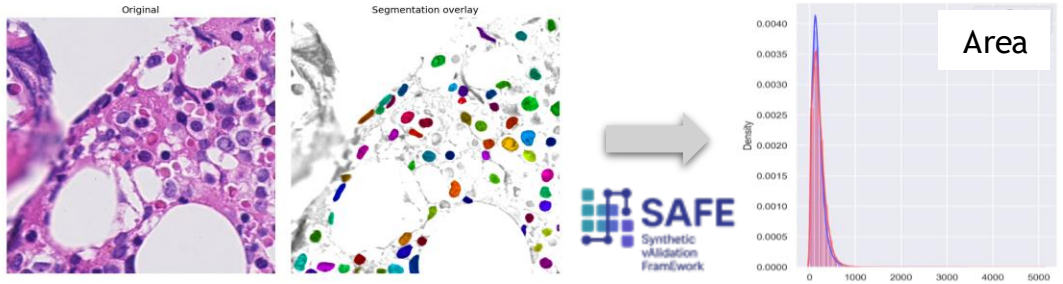


SYNTH

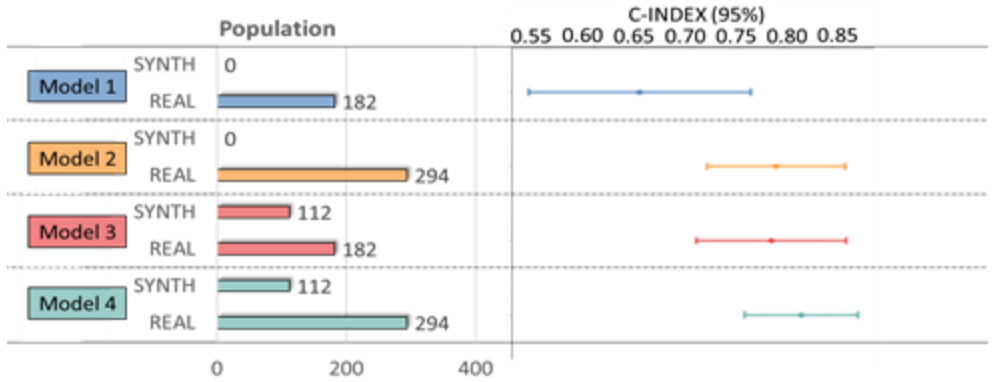
Validation Pipeline

1. **Statistical Metrics for images** (MS-SSIM, FID)
2. **Morphological features extraction and SAFE comparison**
3. **Prognostic Models (Cox Proportional Hazard)**
   Clinical data + Imaging-extracted features
4. **Blinded classification** by experts

2



3



Hybrid approach combining **automated metrics** with **expert judgement** (**Human-in-the-loop**)

Applications

Synthetic Data for Regulatory Use (EMA) - Paediatric Neuroblastoma

Target population: Paediatric patients with relapsed/progressing neuroblastoma

Barrier: ethical complexity of randomizing vulnerable children into a non-treatment/placebo arm

Study design: interventional, single-arm with **Synthetic Control Arm (SCA)** used as External Control Arm (ECA) to benchmark treatment efficacy

Clinical Goal: comparison of expected 2-year Event-Free-Survival (EFS) of 50% in the treatment group vs 30% in the SCA

Bambino Gesù
OSPEDALE PEDIATRICO · IRCCS

Phase II study of anti-GD2 Chimeric Antigen Receptor-Expressing T cells in pediatric patients affected by relapsed/progressing Neuroblastoma.

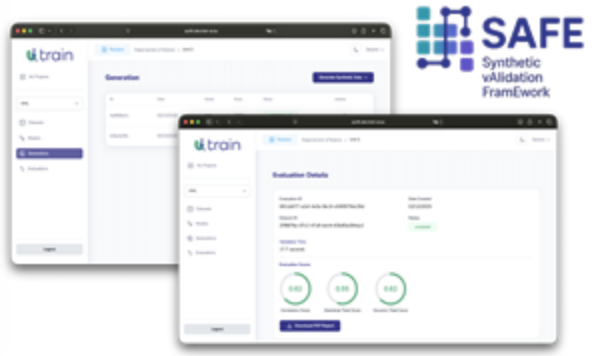
PI:
Prof. F. Locatelli

Ref:
ClinicalTrials.gov ID: NCT03373097

Synthetic Data:
The design of the study is based on the construction of a **Synthetic Control Arm (SCA),** as an External Control Arm (ECA) alternative, to benchmark the efficacy of the pivotal GD2-CART01 treatment.

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Synthetic Data Generation and Validation Platform



Broader implication: clinical utility validation is not just academic, but it is the prerequisite for **regulatory acceptance** of SCA in clinical trials

Iascone et al, Blood 2025 (submitted)

Lessons Learned and Open Challenges

What we learned

Clinical validation ≠ statistics: replicating clinical endpoints and involving clinicians is what builds trust. Statistical fidelity alone is not sufficient.

Context-dependent clinical utility: alignment is highly use case-dependent, there is no universal «good enough» threshold.

Human-in-the-loop to build trust: combining automated metrics with expert judgement increases rigor and catches implausible patterns.

What still needs work

No shared validation standard: every study defines its own metrics and thresholds, a community-agreed framework is needed.

Regulatory frameworks needed: transparency and regulatory endorsement are needed to scale adoption.

Regulatory sandboxes are missing: controlled environments to test and validate SD before final submission.

Privacy-utility trade-off unresolved: no quantitative thresholds available and very dependent from the final application.

Thank you!

Eleonora Iascone eleonora.iascone@humanitas.it

DATAWEEK
JOIN.LEARN.SHARE.GET VALUE
ORGANISED BY BDV BIG DATA VALUE ASSOCIATION
SUPPORTED BY The Research Council of Norway
IN COLLABORATION WITH IFE NTNU SINTEF VESTLANDSFORSKING

DATAWEEK
JOIN.LEARN.SHARE.GET VALUE

Synthetic Data
Limitations & Opportunities

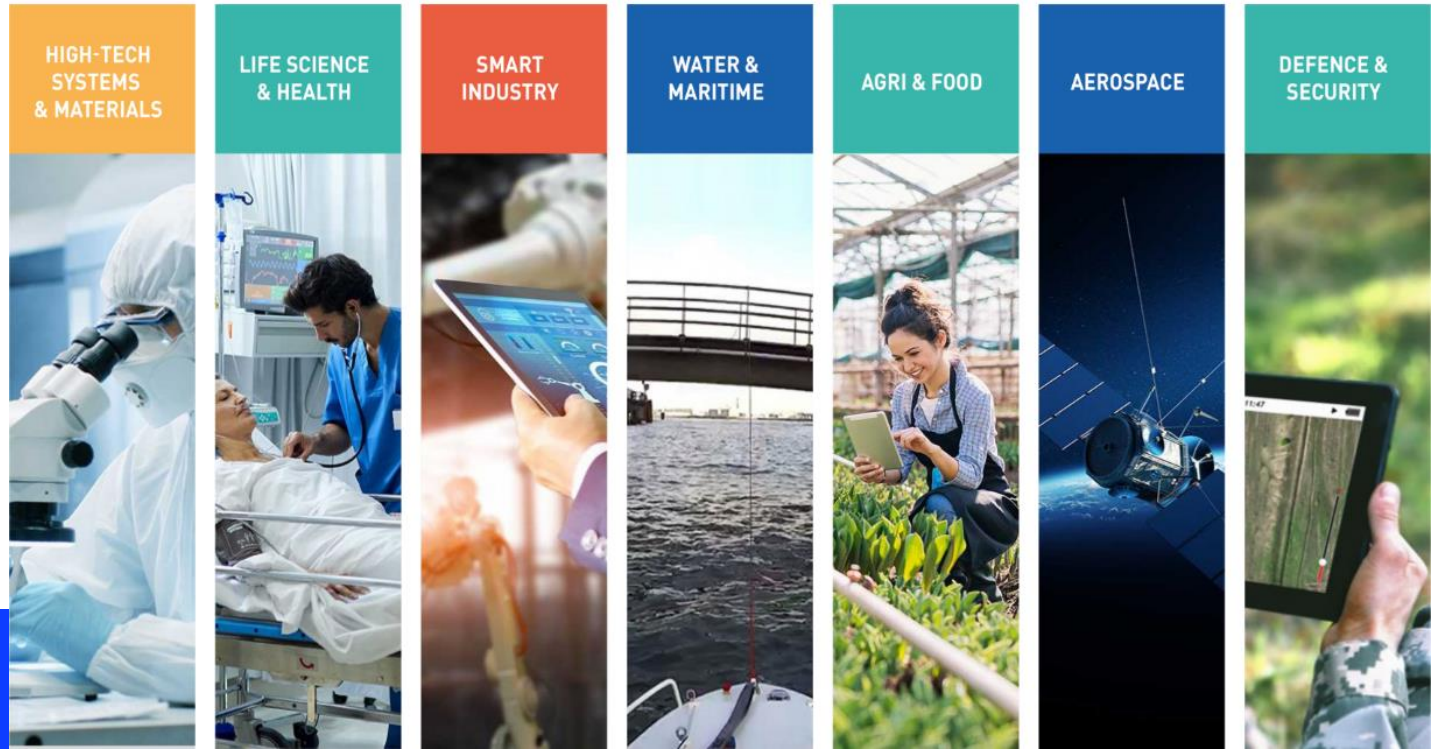
Vincent Bos, DEMCON

ORGANISED BY BDV BIG DATA VALUE ASSOCIATION
SUPPORTED BY The Research Council of Norway
IN COLLABORATION WITH IFE NTNU SINTEF WESTERN NORWAY RESEARCH INSTITUTE VESTLANDSFORSKING

DEMCON

Full-Service Design House

- Design, development & manufacturing
  - Energy
  - Defence & Security
  - High-Tech Mechatronic Systems
  - Life Sciences & Health
  - Optical Systems
  - Data Science & Algorithms
- Contract R&D
- Founded 1993
- Based in the Netherlands
- 1200+ Employees
- ISO 13485, 9001, 45001



DATAWEEK
JOIN.LEARN.SHARE.GET VALUE

Synthetic Sensor Data at Demcon

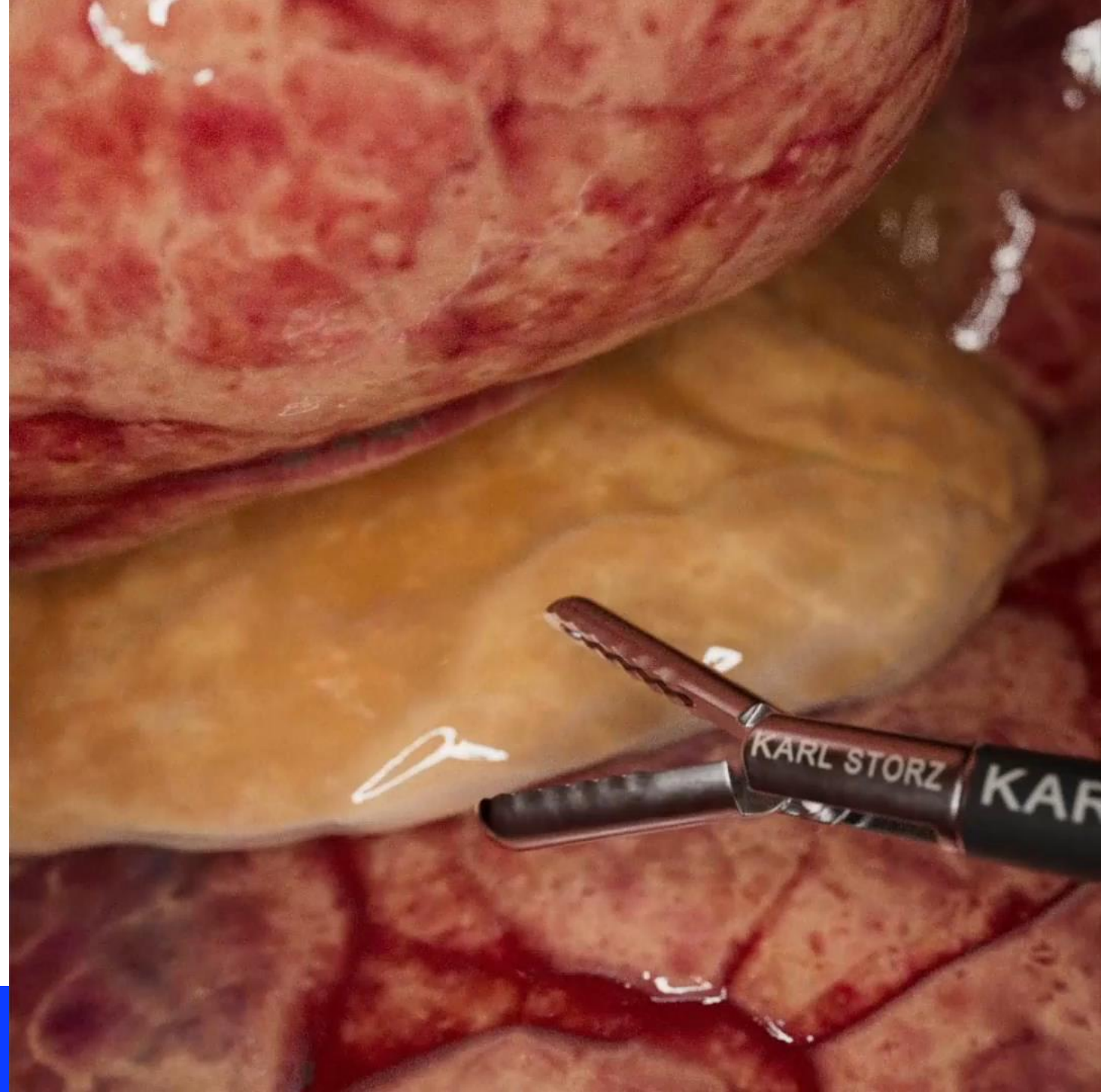
Geometry

↓

Simulation

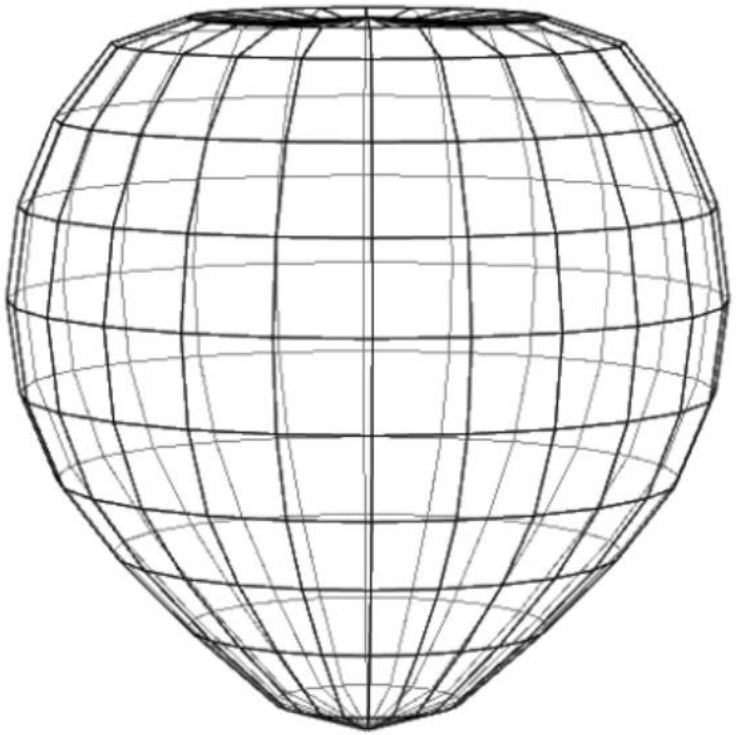
↓

Data Management

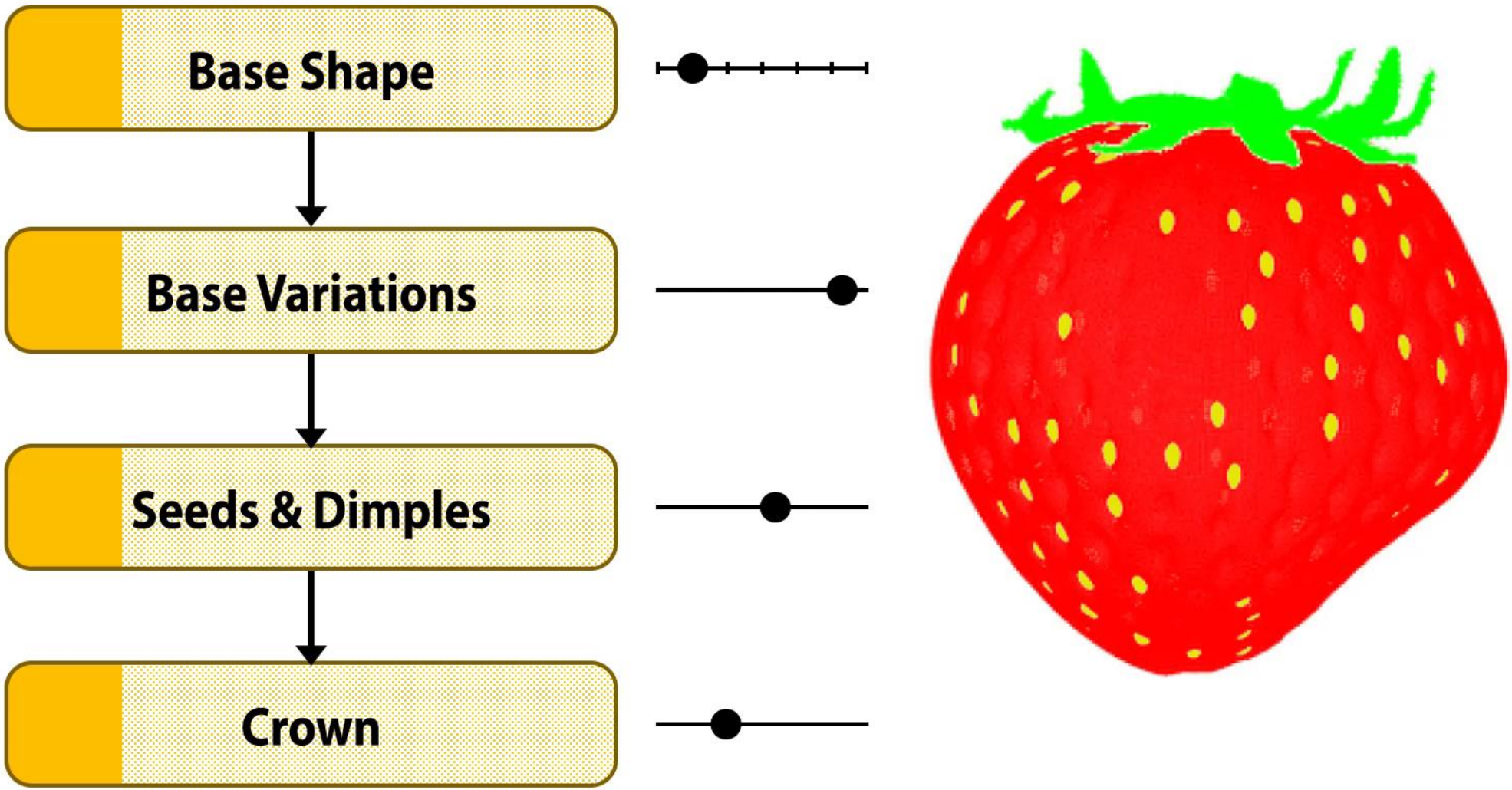


Strawberry Generator

Base Shape



Strawberry Generator



Pulmonary Nodules

Synthetic Data & New ML Outlooks

DATAWEEK
JOIN.LEARN.SHARE.GET VALUE

ORGANISED BY
BDV BIG DATA VALUE ASSOCIATION

SUPPORTED BY
The Research Council of Norway

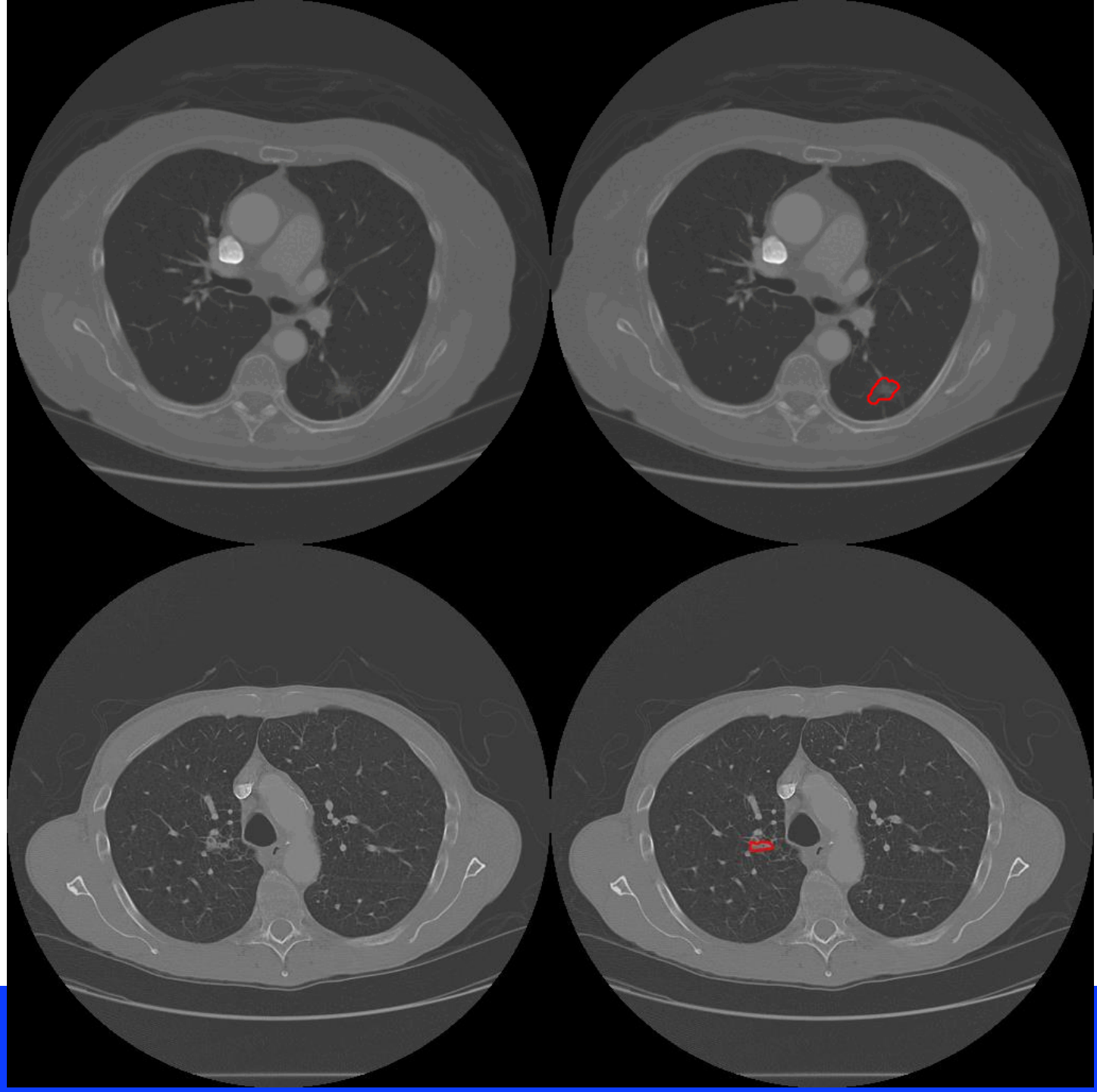
IN COLLABORATION WITH
IFE
NTNU
SINTEF
WESTERN NORWAY RESEARCH INSTITUTE
VESTLANDSFORSKING

Pulmonary Nodules

- Lung cancer is the top cause of cancer death at over 1,8M deaths annually
- Early detection is crucial but challenging, even for experienced radiologists
- A lot of data is available for training but is typically lacking in certain areas:
  - No true ground truth labelling available, only best judgement by domain experts
  - Especially problematic for tricky cases where no data points with a hard ground truth are available

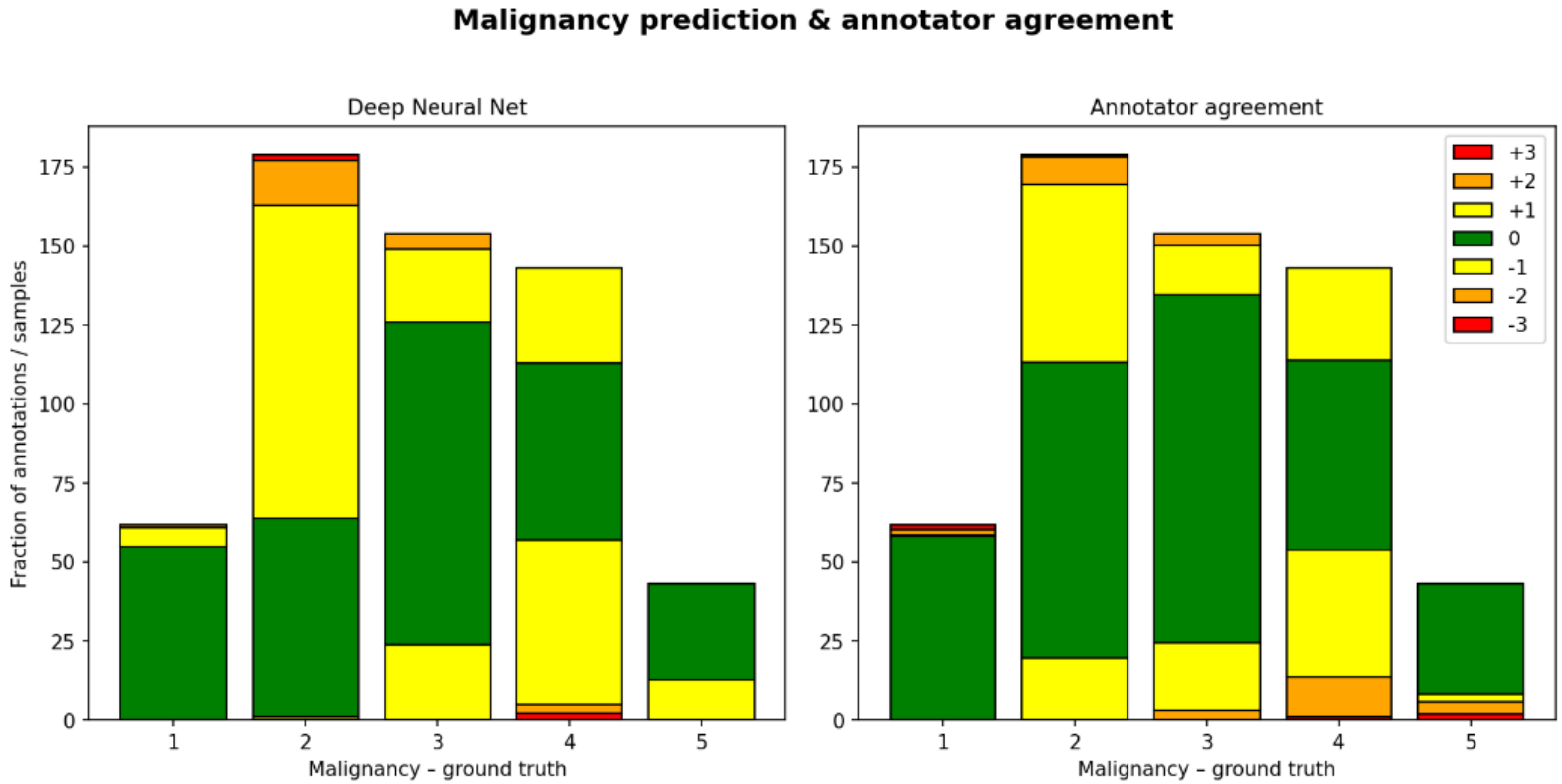
LIDC IDRI Dataset

- Established by the National Cancer Institute (NCI)
- Each nodule annotated by 3-4 experienced radiologists
- Rich annotation
  - Segmentation
  - Likelihood of malignancy
  - Ordinal morphological features:
    - Spiculation
    - Lobulation
    - Texture
    - Sphericity
    - Margin
  - Categorical morphological features:
    - Internal structure
    - Calcification



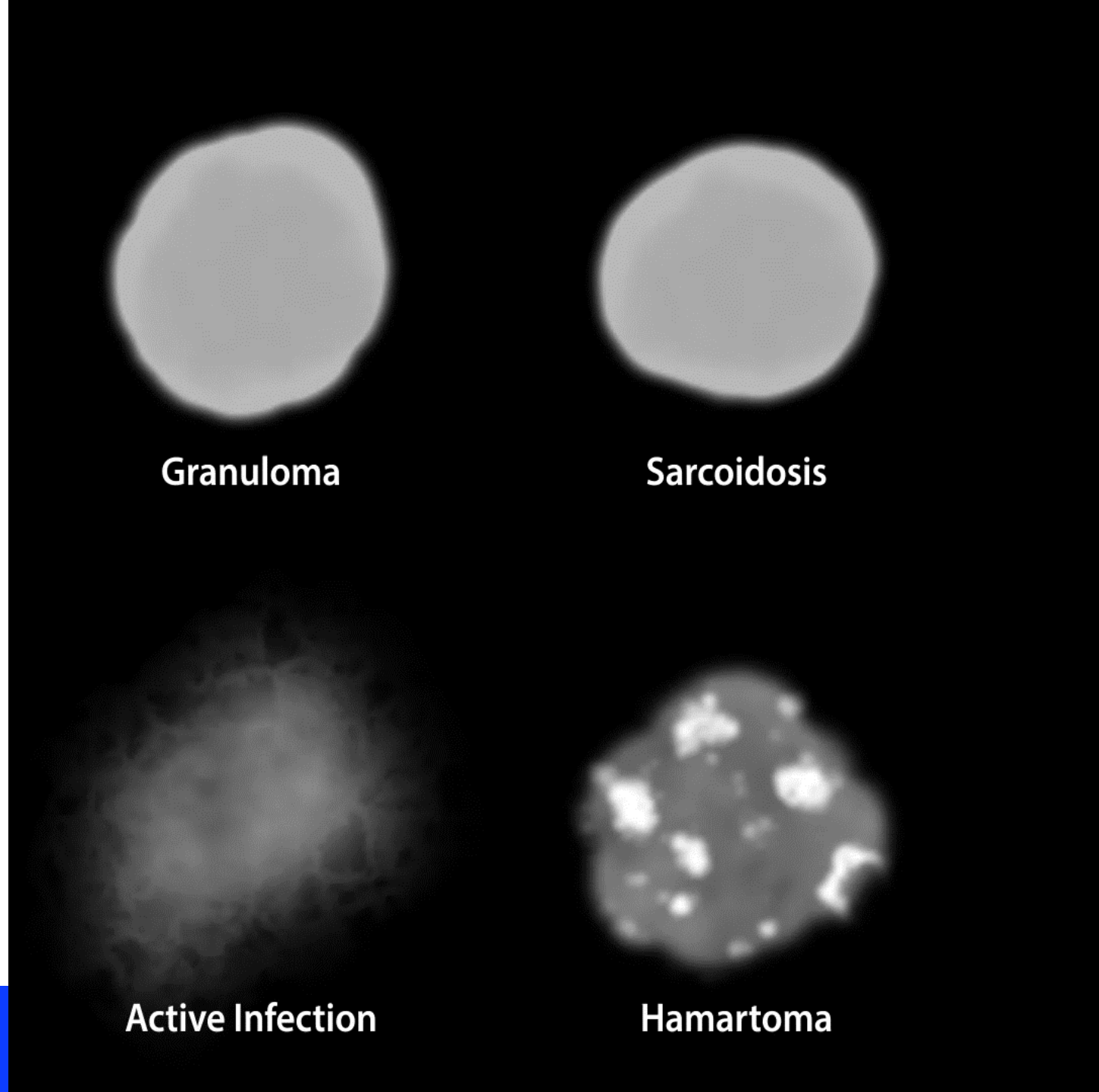
Synthetic Data
Limitations & Opportunities

- How to annotate for likelihood of malignancy in synthetic data?
- Fundamental problem
- Can we use morphological features for classification?



Synthetic Data Opportunities

- Simulate a large range of scanner types: hardware informed ML
- Morphological features: employ the perfect ground truth available
- Morphological, context and patient features in an explainable model which reflects accepted guidelines
- Not just likelihood of malignancy but a grounded probability distribution of nodule sub types



Thank you!

Vincent Bos vincent.bos@demcon.com

DATAWEEK
JOIN.LEARN.SHARE.GET VALUE
ORGANISED BY BDV BIG DATA VALUE ASSOCIATION
SUPPORTED BY The Research Council of Norway
IN COLLABORATION WITH IFE NTNU SINTEF VESTLANDSFORSKING

DATAWEEK
JOIN.LEARN.SHARE.GET VALUE

From Factory Floor to Hospital Floor

What Healthcare Can Learn from Synthetic Data in Industry

Industry

Healthcare

Dr. Sotiris Koussouris
Suite5 Data Intelligence Solutions Ltd
sotiris@suite5.eu

ORGANISED BY BDV BIG DATA VALUE ASSOCIATION
SUPPORTED BY The Research Council of Norway
IN COLLABORATION WITH IFE NTNU SINTEF WESTERN NORWAY RESEARCH INSTITUTE VESTLANDSFORSKING

7-minute keynote · a view from outside the clinic

Different domains, same structure

Both sectors suffer from data that is fragmented, biased, rare and hard to share

	HEALTHCARE	MANUFACTURING
Siloed ecosystems	hospitals, registries, insurers	plants, OEMs, Tier-1 suppliers, service partners
Reluctance to share	GDPR, consent, liability, ethics review	IP protection, trade secrets, contractual risk
Interoperability gaps	FHIR, OMOP, SNOMED still uneven across sites	OPC-UA, AAS, ISA-95 adopted unevenly across assets
Under-represented cases	rare diseases, minorities, edge clinical events	rare faults, new product ramp-ups, disruption scenarios

The four healthcare challenges are industrial challenges too. The problem class is shared; the vocabulary is different.

Synthetic data on the factory floor

AI REDGIO 5.0 — Horizon Europe

AI-REDGIO 5.0

Edge AI for Industry 5.0

Synthetic production data for SME-scale AI, edge deployment and bias-free training when plant data is scarce or locally sensitive.

XMANAI — MAKING AI UNDERSTANDABLE — H2020

XMANAI

eXplainable Manufacturing AI

Synthetic data to augment defect and process datasets so explainable models can be validated on rare quality events without exposing shop-floor IP.

AI-DAPT — Horizon Europe

AI-DAPT

AI Data & Process Pipelines

Treating synthetic data generation as a first-class step in the data lifecycle and for getting hands on more representative sets — reusable pipelines across manufacturing, health and energy.

Four recurring industrial use-patterns

- **Fill historical blind spots**
- **Repair noisy live streams**
- **Simulate rare critical events**
- **Safer cross-org collaboration**

Four lessons industry learned the hard way

01 Don’t ask "is it synthetic?" — ask ”is it fit for purpose?"

A dataset good enough for predictive maintenance is not good enough for safety certification. Fidelity has no meaning without a decision context.

Whitepaper’s call for outcome-oriented evaluation & shared benchmarks.

02 Validate in context, not in abstraction

Industry burned cycles on generic fidelity scores. What worked: co-designed evaluation with line engineers on the actual downstream task.

Clinical validity: validate against the clinician’s workflow, not generic statistics.

03 Augment reality — don’t replace it

Synthetic + real hybrids consistently outperform either alone. The win is enrichment of edge cases and rare failures, not wholesale substitution.

Bias mitigation for rare diseases, minorities, emergency scenarios.

04 The strategic value goes beyond AI training

Workflow simulation, stress-testing, regulatory sandboxes, third-party onboarding — synthetic data earns its keep as operational infrastructure.

Hospital workflow optimisation and AI system readiness.

Where healthcare has the harder problem

Industry lessons don’t transfer one-for-one.
Three dimensions make healthcare structurally harder!

Cost of the "wrong" answer

INDUSTRY

A mispredicted defect costs €€€ — recoverable.

HEALTHCARE

A missed diagnosis costs a life — irreversible. Validation bar is categorically higher.

Consent & seed data

INDUSTRY

Machine data; contractual negotiation with owners.

HEALTHCARE

Patient consent, DPIAs, ethics committees, secondary-use rules. No industrial analogue.

Accountability chain

INDUSTRY

CE marking, ISO process audits — mature.

HEALTHCARE

AI Act + MDR/SaMD + EHDS still being wired together. Liability on SD-trained DSS is open.

Copy the engineering patterns, not the risk appetite.

An industry-2-healthcare crossover playbook

	FROM INDUSTRY	TO HEALTHCARE
1	Manufacturing benchmarking consortia (e.g. defect-detection challenges)	**Shared clinical-utility benchmarks for SD — co-owned by clinicians, patients, regulators.**
2	Digital twins of production lines	**Workflow twins of hospital departments — test bed reconfigurations before touching reality.**
3	Federated MLOps across plants (AI-DAPT pattern)	**Federated SD generation across hospitals — seed data never leaves the ward.**
4	Regulatory sandboxes for CE-marked AI	**EHDS + AI Act sandboxes that treat SD as first-class evidence, not a curiosity.**
5	"Data-Generation-as-a-Service" emerging in Industry 5.0	**Same service model for EU hospital networks — aligned with the European Data Union Strategy.**

Thank You!

Dr. Sotiris Koussouris
Suite5 Data Intelligence Solutions Ltd
sotiris@suite5.eu

DATAWEEK
JOIN.LEARN.SHARE.GET VALUE

ORGANISED BY
BDV BIG DATA VALUE ASSOCIATION

SUPPORTED BY
The Research Council of Norway

IN COLLABORATION WITH
IFE
NTNU
SINTEF
WESTERN NORWAY RESEARCH INSTITUTE
VESTLANDSFORSKING

BDV BIG DATA VALUE ASSOCIATION

Home / Task Forces / Healthcare

BDVA Task Forces

Healthcare

Thank you!

Main editors

Dr. Sofia Tsekeridou (Netcompany SEE & EUI)
e-mail: sofia.tsekeridou@netcompany.com

Dr. Iñaki Fernández Pérez (CARTIF)
e-mail: inafer@cartif.es

Dr. Sinem Tas (Philips)
e-mail: sinem.tas@philips.com

Saurav Baidya (Philips)
e-mail: saurav.baidya@philips.com