BDV BIG DATA VALUE ASSOCIATION

SYNTHETIC DATA IN HEALTHCARE

BENEFITS AND OPPORTUNITIES
TECHNOLOGICAL, CLINICAL, REGULATORY
GAPS & CHALLENGES
WAYS AHEAD FOR IMPACT MAXIMISATION

SEPTEMBER 2025

TABLE OF CONTENTS

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

List of authors(in alphabetical order):

Saurav Baidya, Philips, BDVA Healthcare Task Force Co-lead, ITEA4 IVVES project representative
Christian Buerger, Philips
Gorka Epelde Unanue, Vicomtech, H2020 VITALISE and Horizon Europe EOSC RAISE Projects representative
Helena Fernández López, Gradiant
Iñaki Fernández Pérez, Fundación CARTIF
Alberto Gutierrez-Torre, Barcelona Supercomputing Center, Horizon Europe SECURED project representative
Gianna Karanasiou, WINGS ICT Solutions, SNS TrialsNet project representative
Shane O'Seasnáin, Eindhoven University of Technology
Johan Plomp, VTT, ITEA4 IVVES project representative
Michal Rosen-Zvi, IBM Research, BDVA Healthcare Task Force Co-lead
Daniela Spajic, KU Leuven, Horizon Europe SECURED project representative
Sinem Tas, Philips
Sofia Tsekeridou, Netcompany SEE & EUI, Horizon Europe SYNTHEMA and H2020 NIGHTINGALE Projects representative
Yalin Yalic, TIGA Health, Horizon Europe AISym4Med Project representative

List of Abbreviations

Abbreviation	Definition
2D	Two dimensional
3D	Three dimensional
ADL	Activity of Daily Living
AI	Artificial Intelligence
AML	Acute Myeloid Leukemia
CARE	Collective Benefit, Authority to Control, Responsibility, Ethics
CT	Computer Tomography
DP	Differential Privacy
DPIA	Data Protection Impact Assessment
DT	Digital Twin
EHDS	European Health Data Space
EHR	Electronic Health Record
FAIR	Findable, Accessible, Interoperable, Reusable
FDA	Food and Drug Administration
FL	Federated Learning
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
HD	Hematological Disease
HIPAA	Health Insurance Portability and Accountability Act
ICT	Information and Communication Technologies

Abbreviation	Definition
ioECoG	Intraoperative Electrocorticography
IoT	Internet of Things
IP	Intellectual Property
MCI	Mass Casualty Incident
ML	Machine Learning
MRI	Magnetic Resonance Imaging
NIST	National Institute of Standards and Technology
PET	Privacy Enhancing Technologies
RAI	Research Analysis Identifier
R&D	Research & Development
SaMD	Software as a Medical Device
SCD	Sickle Cell Disease
SD	Synthetic Data
SDG	Synthetic Data Generation
SMPC	Secure Multi-party Computation
VAE	Variational Autoencoder

Executive Summary

Synthetic data is rapidly emerging as a transformative enabler in digital health innovation, offering new pathways to unlock data access, strengthen Artificial Intelligence (AI) development and advance inclusive research and care. Realising the significant potential of synthetic data, the Healthcare Task Force of BDVA has formulated a working group with relevant expertise to author this white paper that explores how synthetic data can strategically benefit and boost innovation in eight key domains: virtual patients, data scarcity and bias mitigation, data sharing and collaboration, clinical trial innovation, AI fairness & performance, privacy-respecting research addressing ethical, legal and business barriers, system readiness and reliability and health workforce training. Drawing from a wide range of use cases and state-of-the-art assessment, the paper demonstrates how synthetic data addresses long-standing challenges in healthcare, such as data scarcity, regulatory barriers and underrepresentation, while also enabling new forms of secure, scalable and ethically aligned collaboration. The paper maps current capabilities to real-world applications and identifies persistent research and policy gaps, from limited access to representative seed data to uncertainties around legal governance, benchmarking standards and sustainable regulated infrastructure.

To fully realise the potential of synthetic data, the paper proposes a coordinated set of recommendations spanning regulation, scientific and technical development, data governance and funding in the healthcare domain, many of which are further applicable in other sectors. These include establishing shared assessment benchmarks and sandboxes, developing multimodal and explainable AI models, clarifying and advancing legal frameworks and policies and investing in inclusive and privacy-enhancing technologies, as well as secure federated data and learning ecosystems. The European funding programs Horizon Europe, EU4Health and Digital Europe have a pivotal role to play in supporting synthetic data initiatives that align with the European health equity and innovation goals.

Together, these actions offer a strategic blueprint for embedding synthetic data into the European digital health ecosystem and infrastructure—building more inclusive, privacy-preserving and future-ready digital health systems.

Introduction: The Urgency of Ethical Health Data Research Collaboration & Innovation

Across global healthcare systems, data driven innovation remains constrained by data silos, systemic fragmentation, ethical and regulatory concerns and a lack of collaborative incentives. Despite the pressing need to tackle rare diseases, expedite clinical trials, respond to the needs of an ageing population and manage future public health emergencies, much of today's health data remains locked in institutional silos primarily due to the complexity of ethical and legal aspects regarding data sharing. Opportunities for innovation, equity and shared insight are lost before they begin.

Limits of Traditional Data Sharing

Initiatives like the European Health Data Space (EHDS) [1] and national data hubs mark important progress toward more connected and interoperable health systems. Yet, traditional data-sharing models continue to face significant hurdles in practice. Health data remains fragmented across organisations with differing legal frameworks, governance models and technical standards, leading to inconsistent access, unclear ownership and regulatory uncertainty, particularly around secondary use, i.e., for purposes other than the original intent for which data was collected. Even when collaboration is encouraged, liability concerns, divergent General Data Protection Regulation (GDPR) [2] interpretations and lengthy ethics reviews often cause delays or block progress. Technically, datasets are siloed, lack standardised metadata and suffer from semantic misalignment, making harmonisation difficult and costly. The result is not just inefficiency: research efforts frequently overlook rare disease groups, underserved regions and minority populations, reinforcing bias and slowing equitable innovation. Despite the ambition of recent initiatives, many promising projects stall because key data remains inaccessible when and where it is needed.

The Cost of Inaction: Innovation, Equity and Collaboration at Risk

Without a fundamental shift in how health data is governed and shared, healthcare systems risk falling behind in three critical areas. First, innovation slows down as promising AI tools and digital health solutions are constrained by limited access to high-quality, diverse datasets. Second, equity gaps widen as underrepresented groups, such as those with rare diseases or living in underserved regions, are not considered in research, AI model development and health tech innovation. Third, collaboration stalls, with legal, ethical and technical barriers continuing to prevent cross border, public private and multi-institutional efforts from reaching scale. Addressing these challenges is essential to unlocking the full potential of data-driven health innovation.

Persistent Challenges to Health Data Research Collaboration & Innovation

Challenge	Description
Siloed Ecosystems	Health data sits across fragmented systems, hospitals, registries, institutes, insurers and platforms, each with different governance structures and infrastructure and governed by different legal contexts. This fragmentation blocks the discovery of unified insights and limits cross border innovation.
Reluctance to Share	Institutions often avoid data sharing due to liability concerns, reputational risk and complex legal requirements (e.g. GDPR, Data Protection Impact Assessment (DPIA), informed consent for primary/secondary use of data). Ethical approval processes can delay or prevent access entirely.
Interoperability Gaps	Even when willing to collaborate, many stakeholders face incompatible data formats, missing compliance to widely accepted metadata standards and taxonomy/semantic mismatches. These issues hinder seamless heterogeneous data integration, slow down research and restrict scalability.
Lack of Equity & Representation	Many real-world datasets underrepresent minority, rural and rare disease populations, amplifying bias in AI tools and reinforcing inequalities in care and innovation.

The Synthetic Data Opportunity: From Concept to Ecosystem

To alleviate these barriers, synthetic data (SD) emerges as a high potential enabler of ethical, inclusive and innovation-friendly data collaboration, due to the mere fact that they can be more easily shared and used for different purposes without the need to re-access real-world seed data.

What is Synthetic Data and How is It Generated?

SD [3]-[6] refers to artificially generated datasets that replicate the statistical characteristics and structure of real-world data, across demographics, modalities and clinical pathways, without exposing identifiable information. It is not generated from individual patients but learned from underlying patterns within real data across patients, enabling anonymised simulation of clinical profiles, events and outcomes. It can be fully synthetic, entirely generated from trained validated AI models without including any real data used in training, or partially synthetic, when only sensitive variables are synthesised and thus anonymised. The generation of synthetic data is powered by machine learning methodologies, such as Generative Adversarial Networks (GANs) [7], Variational Autoencoders (VAEs) [3] and Transformers [4], which are appropriate for complex and heterogeneous data like EHRs. These approaches aim to balance high fidelity (i.e., statistical similarity to real data) with privacy protection.

Making use of SD is not about replacing real-world data, but about unlocking access, accelerating experimentation and building bridges between stakeholders.

Why is Synthetic Data a Strategic Enabler?

Unlike traditional data proxies or anonymised real-world datasets, SD offers a rare combination of technical utility and regulatory flexibility. This makes it strategically valuable to health systems aiming to unlock innovation while maintaining ethical and legal safeguards.

Strategic Benefit of SD	What it Enables
Privacy-by-Design	Enables ethical data sharing in sensitive domains (e.g. genomics, mental health) by ensuring data is effectively de-identified and GDPR-compliant.
Bias Mitigation & Inclusion	Fills gaps in real-world data by simulating rare events and underrepresented groups, reducing AI bias and improving fairness. As SD is measurement method agnostic, it does not introduce relevant AI biases.
On Demand Generation	Supports rapid prototyping and validation of digital tools in sandbox environments without waiting for real data access.
No need to share Real World Data	Enables organisations to only and more easily share SD while keeping real world data local, e.g. by adopting federated learning (FL) and Secure Multiparty Computation (SMPC) approaches for SD generation. This further enables seamless & faster cross-border clinical research and health innovation.

The Synthetic Data Generation & Use Lifecycle

Trustworthy SD is not a standalone output but part of a broader ethical and technology-driven ecosystem that starts with ethical data sourcing and curation, , which is a vital step given the significant current lack of data/semantic interoperability, continues with optimal synthetic data generation and validation and ends with structured governance and sharing, as shown in Figure 1 and detailed below. The lifecycle of SD generation & use is not just only technical, but also ethical and operational. Each phase must uphold privacy, utility and trust to deliver responsible health innovation at scale.

Seed data: Begins with ensuring curated real-world datasets (e.g., EHRs, imaging, medical sensor/device data) that are anonymised, harmonised and sampled for clinical and demographic relevance [9][11]. Institutions retain control, often via the use of secure federated data management environments [12].

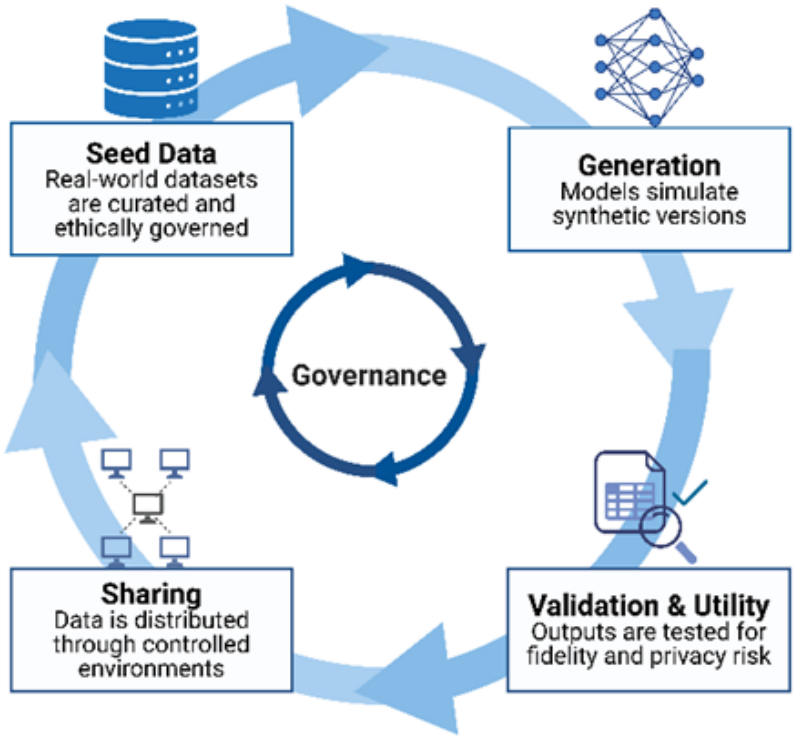

Figure 1. The Synthetic Data Lifecycle: From Ethical Sourcing to Responsible Reuse

Generation: Advanced models (e.g., GANs, Bayesian networks, Diffusion, Foundational Models, triggered with emerging Agentic AI [9]) learn statistical patterns from seed data without reproducing individual records in most cases. Their training and inference functions could employ secure SMPC based FL frameworks [13]-[15] when accessing distributed real-world data. Privacy safeguards such as differential privacy (DP) ensure outputs are de-identified [16][17].

Validation: SD is tested to confirm statistical and clinical fidelity, ensure privacy and verify utility in realistic scenarios (e.g. disease modeling, treatment response) [18]-[21].

Sharing: Synthetic datasets are shared through sandboxes, secure portals, or research commons with clear licensing and access controls, balancing openness with responsibility [22].

Governance: Continuous oversight ensures transparency, legal compliance (e.g. GDPR, EHDS) and public trust. Inclusive governance involves patients, regulators and data custodians in co-developing rules, audits and risk management [23].

Embedding Trust: Design and Governance Principles

To move from proof-of-concept to scalable use, SD ecosystems must be anchored in quality, standards and inclusive governance [24].

Priority Areas	Strategic Action
Seed Data Quality	SD is only as good as its source. Input datasets must be curated, denoised, imputed and validated to ensure clinical, demographic and contextual relevance.
Standards & Interoperability	Data generation must follow widely used domain taxonomies and standards (e.g., SNOMED CT [25], HL7 FHIR [26], OMOP CDM [27]) to ensure interoperability with relevant initiatives, ecosystems and infrastructures like EHDS.
Inclusive Governance	Governance should include patients, custodians and civil society to co-define norms for generation, access and reuse. Promising models include SD commons, ethical licensing and sandbox environments.
Outcome-Oriented Evaluation	Success must be measured beyond data fidelity, metrics should reflect public health impact, such as improved diagnostic access, bias reduction and system resilience.

High Innovation Potential in Healthcare: Unlocking Value with Synthetic Data

SD is rapidly transitioning from a privacy-preserving concept to a practical tool for solving a number of the most persistent challenges in healthcare. By replicating the statistical patterns of real-world data without compromising patient privacy, SD offers a unique combination of clinical utility, regulatory flexibility and ethical integrity. This section outlines eight strategic domains, as shown in Figure 2, where SD is already delivering measurable value, while expanding the frontier of what is possible in data-driven health systems.

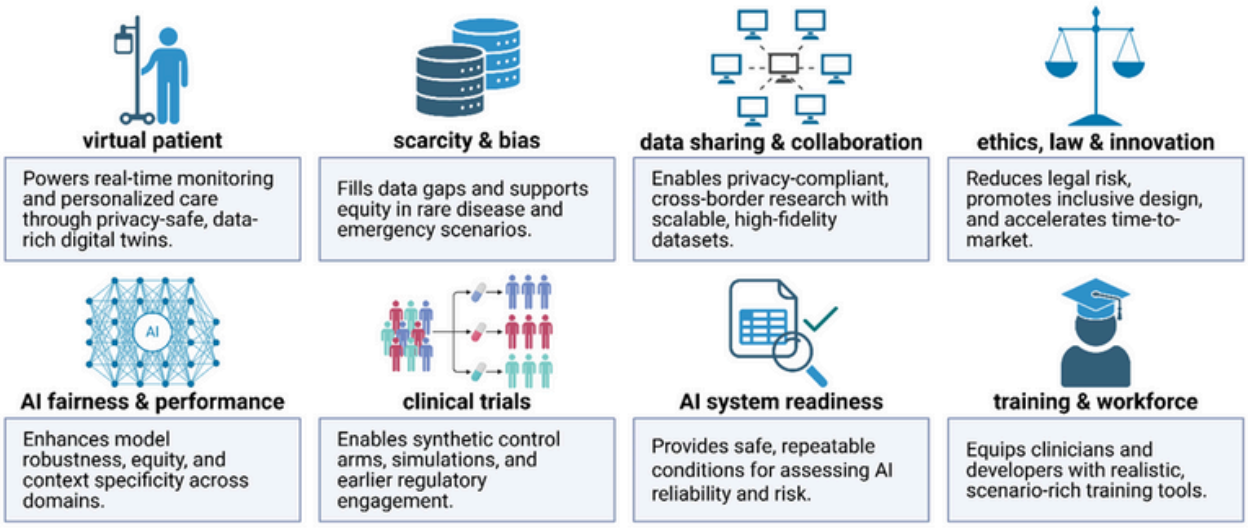

Figure 2. Eight strategic domains where synthetic data delivers scalable, inclusive and secure innovation in healthcare.

Virtual Patients: Enabling Personalised, Predictive Care

In the future, it is likely that patients will consult their clinicians alongside a digital representation or a digital twin (DT) of themselves. This virtual model offers an objective and understandable foundation to manage the patient's health, aimed both at specialised clinicians and at patients without any medical training: it enables the tracking of the patients' current health status, identify gradual changes and help to manage major events. Virtual patients, or DTs of patients, simulate individual health profiles by integrating data from EHRs, wearables, patient-reported outcomes and synthetic datasets.

These models enable real-time health monitoring, early alerts, prevention care and personalised treatment strategies, allowing clinicians to automate routine tasks and concentrate on complex decisions. SD plays a vital role in constructing these models, which is particularly instrumental in rare disease contexts where real data is scarce. By leveraging anatomical knowledge and large-scale population trends, SD fills critical gaps and enhances model accuracy without breaching patient privacy. These systems support proactive, cost-effective care and help avoid unnecessary interventions, improving both outcomes and patient engagement, as shown in Figure 3, left column. When governed ethically, such as through secure FL and patient-informed consent, virtual patients can also inform policy, advance research and promote equitable care. However, transparency in model design and safeguards against risks like data hallucination or model collapse are essential to ensure responsible deployment.

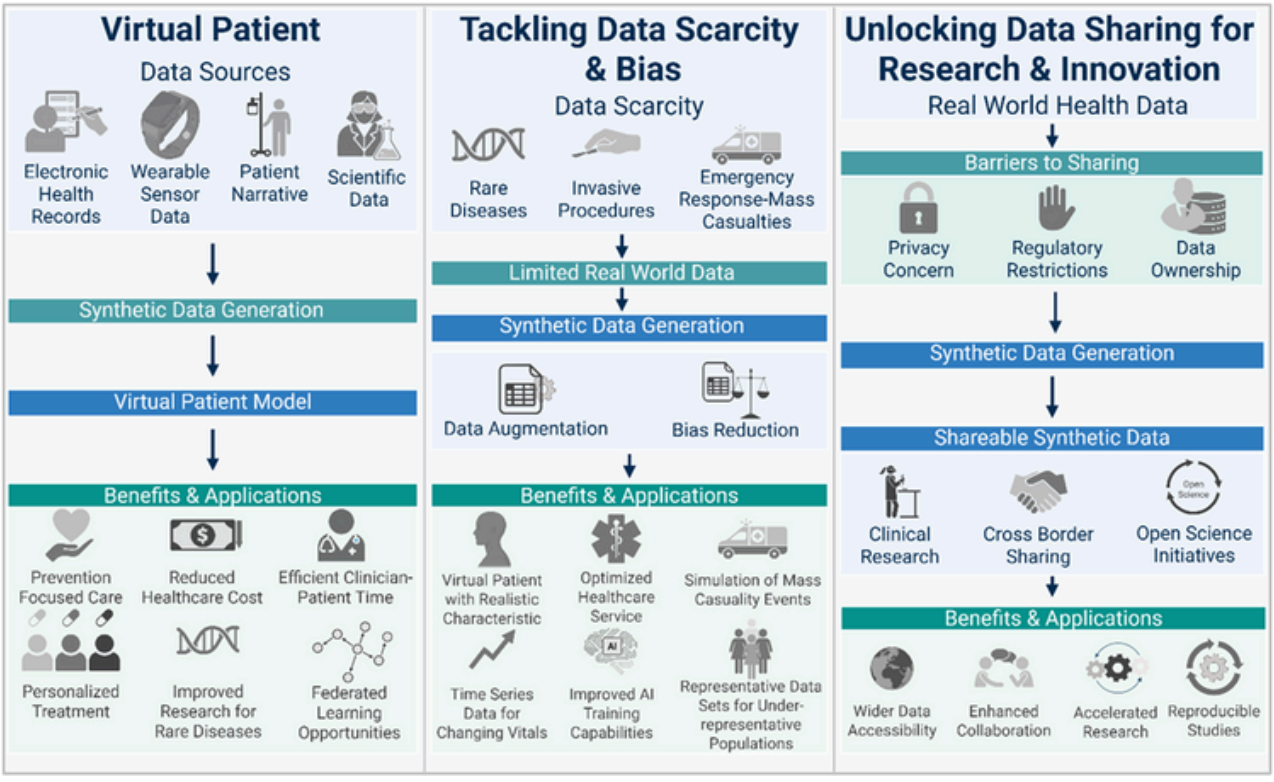

Figure 3. Context, Benefits & Applications illustrated for the SD strategic domains of Virtual Patients, Tackling Data Scarcity & Bias, Unlocking Data Sharing for Research & Innovation.

Tackling Data Scarcity and Bias in Critical Care Scenarios

SD proves especially valuable when real-world data sets are limited or hard to obtain such as in rare diseases and emergency situations, as shown in Figure 3, middle column. In rare conditions, SD enriches underrepresented cohorts and enables the creation of virtual patients with unique clinical signatures, with the aim in this case to augment data and improve diagnostic accuracy and treatment planning of such patients. In mass casualty or emergency scenarios, it facilitates the simulation of virtual victims, synthesising further trauma cases, complete with time-series vital signs and intervention pathways at prehospital contexts, differentiating in this way from virtual patients and disease cases and allowing AI systems to learn from complex, data-poor situations, while augmenting data.

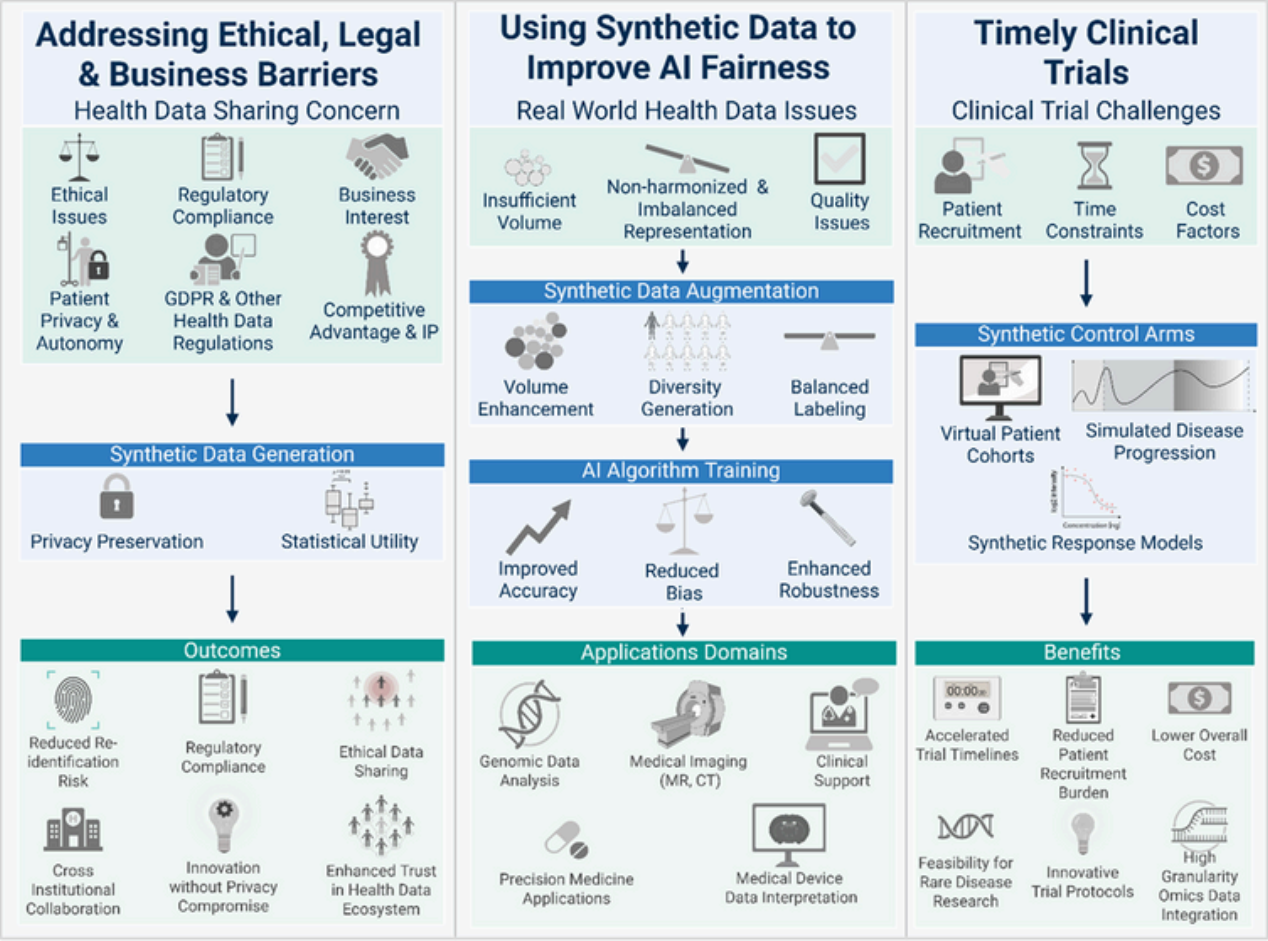

Figure 4. Context, Benefits & Applications illustrated for the SD strategic domains of Addressing Ethical, Legal & Business Barriers, Using Synthetic Data to improve AI Fairness, Timely Clinical Trials.

Unlocking Data Sharing for Research and Innovation

Because SD is anonymised and privacy-safe, it bypasses many regulatory current hurdles that hinder real world data sharing. This makes it an ideal asset for open science and cross-border research. Researchers can access scalable, high-fidelity synthetic datasets without triggering GDPR or Health Insurance Portability and Accountability Act (HIPAA) restrictions, accelerating collaboration, reproducibility and innovation, as illustrated in Figure 3, right column.

Addressing Ethical, Legal and Business Barriers

SD eliminates direct identifiers and reduces the re-identification risk, at the extent possible according to privacy versus utility trade-offs and the used SDG methods, thus enabling broader data utility while maintaining compliance with data protection laws. For businesses, it supports rapid, low risk prototyping and early validation of digital health solutions, accelerating time to market. It can also be combined with real-world data using privacy-preserving techniques like DP or secure FL to retain model utility without sacrificing security. Ethically, SD facilitates inclusive research by simulating underrepresented populations, aligning with frameworks such as Findable, Accessible, Interoperable, Reusable (FAIR) and Collective Benefit, Authority to Control, Responsibility, Ethics (CARE). Regulatory developments, including the AI Act [25] and Food and Drug Administration (FDA) AI/ML Software as a Medical Device (SaMD) [29] regulations, have already begun to provide validation guidelines, underscoring the need for strong governance, embedded fairness assessments and strategic use of SD alongside real data. These multiple outcomes of the use of synthetic data are illustrated in Figure 4, left column.

From Gaps to Gains: Using Synthetic Data to Improve AI Fairness

High-performing healthcare AI requires large, diverse and representative datasets, something that real-world clinical data often fails to deliver. SD closes these gaps by augmenting datasets with rare conditions, catering for balanced class distributions and greater demographic variability.

In fields like medical imaging, synthetic Magnetic Resonance Imaging (MRI) or Computer Tomography (CT) scans expand sample diversity and reduce underrepresentation. In distributed or Internet of Things (IoT)-enabled settings, it generates user-specific data for training activity recognition systems, facilitating in-home care while preserving privacy. Crucially, the targeted creation of edge cases and minority subgroups strengthens fairness, reduces bias and improves performance across all populations. The result is more accurate, equitable and trustworthy AI systems, as shown in Figure 4, middle column.

Reimagining Clinical Trials with Synthetic Data

SD is reshaping clinical trials, making them more timely, efficient, inclusive and patient-centric, as explained in Figure 4, right column. Its use is especially critical in studies involving rare diseases, pediatric populations, or invasive procedures where real data is limited. One of its key applications is the generation of synthetic control arms, virtual cohorts that statistically mimic real populations, reducing the need for placebo groups and minimising patient exposure to ineffective treatments. Synthetic datasets also support trial simulations and in-silico trials driven by DTs and generative AI, enabling researchers to refine designs, stratify patients more effectively and project long-term outcomes before live implementation. These capabilities enhance trial power, reduce costs and shorten the path to regulatory engagement. Pre-labelled synthetic datasets further improve labelling precision and trial efficiency, while modeling of rare and complex scenarios supports algorithm validation and clinical decision-making. Collectively, these innovations position SD as a foundation for next-generation clinical research.

Advancing Reliability and Readiness of AI Systems with Synthetic Data

Before clinical deployment, AI-driven health technologies must be validated under diverse and demanding scenarios, many of which are underrepresented in real datasets. SD provides a safe, controllable environment to simulate rare conditions, extreme cases and evolving disease trajectories.

This enables transparent evaluation of AI behaviour across varied conditions and platforms, improving explainability, auditability and regulatory compliance. Synthetic datasets can be replicated and modified for performance benchmarking across different software and hardware environments, fostering scalable, resilient and accountable AI systems ready for clinical use, as illustrated in Figure 5.

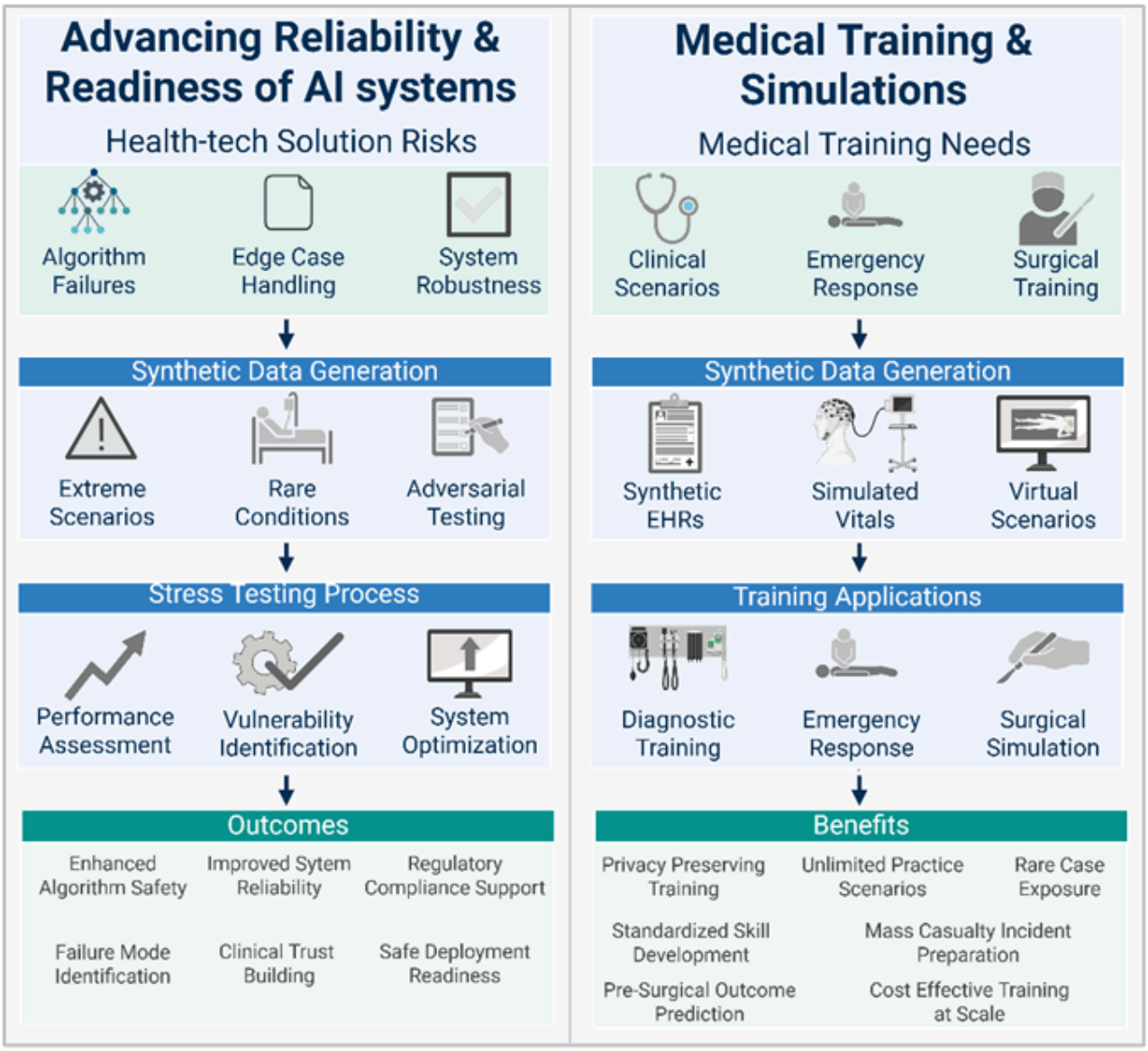

Figure 5. Context, Benefits & Applications illustrated for the SD strategic domains of Advancing Reliability & Readiness of AI Systems and Medical Training & Simulations.

Training Tomorrow's Health Workforce

Training with real patient data is often restricted by privacy, legal and logistical barriers. SD offers a privacy-safe, high-fidelity alternative for education and skill-building across clinical, data science and informatics domains. Synthetic datasets, as shown in Figure 5, right column, simulate real patient records and clinical scenarios, enabling learners to engage in hands on training for diagnostic interpretation, workflow testing and AI tool usage. Educators can use SD to recreate rare or emergency conditions for simulation-based learning that mirrors real-world complexity. Its adaptability makes it suitable for replicable, targeted training across specialties and geographies, ensuring that the next generation of health professionals is well-equipped for the demands of a digital, data-driven healthcare system.

These use domains demonstrate that synthetic data is not merely a workaround, it is a foundational asset for building ethical, inclusive and efficient healthcare systems. By enabling safer access, targeted experimentation and fair, equitable participation, synthetic data helps healthcare organisations move from constraint to capability and from data protection to data activation.

Real-World Use Cases: Mapping the Value of Synthetic Data in Healthcare

SD is already addressing real-world challenges in clinical and research settings. This section presents a set of indicative use cases that illustrate how it is generated, why it is used and the barriers it helps overcome, highlighting shared benefits, recurring challenges and its growing strategic role in healthcare innovation.

Use Case 1: Modelling Rare Hematological Diseases for Research and Diagnosis

Synthetic data enables more equitable, secure and collaborative research on rare blood disorders, bringing innovation within reach of patients and healthcare systems alike.

Summary	Hematological diseases, though individually rare, affect a substantial global population and place significant pressure on healthcare systems. Despite ongoing national and EU research efforts, data scarcity and fragmentation, as well as existing data silos, continue to limit effective diagnosis and treatment, as in cases of sickle cell disease (SCD) and acute myeloid leukemia (AML). To address these challenges, cross-border data hubs are being established to enable GDPR-compliant research and AI development. By using FL, SMPC and DP, synthetic datasets can be created from multimodal clinical data while preserving patient privacy. This privacy-centric framework connects clinical and computing centers to produce SD that supports robust AI model training, global aggregation and real-time prediction. These datasets must be validated for clinical relevance/fidelity, statistical utility and residual privacy risk. Ethical and legal safeguards are essential, including privacy-by-design principles and co-creation of trustworthy AI systems.		
SD in Action	Overall survival prediction (AML) Treatment response prediction from clinical-genomic data (AML) Genomic variants/disease phenotypes association (for SCD) MRI feature-based prediction of brain vascular events (for SCD)		
Key Benefits	Expands privacy-safe, cross-border access to scarce rare disease data Enhances data diversity and reduces bias in AI training Builds transparency and performance confidence through rigorous design Enables AI-driven predictions for survival, treatment and complications Enables transparent, benchmarked validation frameworks Simulates rare cases to enhance precision and fairness		
	Scientific/Technical	**Ethical/Clinical**	**Regulatory**
Challenges & Gaps	Need for FL and privacy-enhancing technologies (PETs) Trade-offs between data fidelity, utility and privacy protection Absence of standardised frameworks for SD quality validation and benchmarking	Lengthy processes for ethical approvals, seed data access and informed consent Variability in seed data formats, missing values and context-specific seed data definitions Concerns about re-identification, consent scope and trust in synthetic data	Complexities of GDPR compliance and cross-border data governance Emerging obligations under the EU AI Act and EHDS Lack of clear policies on synthetic data ownership, liability and the use of SD in clinical decision-making
Target Users	Clinicians, patients, data scientists, AI developers		
Involved Stakeholders	Clinicians, AI developers, data scientists, data privacy experts, policymakers and research community		
Relevant Projects	Horizon Europe SYNTHEMA project		

Use Case 2: Synthetic Data to Advance Epilepsy Surgery

Synthetic intraoperative electrocorticography (ioECoG) data enables safer, more accurate localisation of epileptic zones, enhancing AI model training, clinical decision-making and privacy-compliant innovation in epilepsy care.

Summary	Epilepsy affects over 50 million people globally, with nearly one-third of patients remaining resistant to medication. In severe cases, surgical intervention guided by ioECoG helps localise and remove seizure-generating brain tissue. However, real ioECoG datasets are limited due to the invasiveness and complexity of recording procedures. This limits the performance and generalisability of AI models used to map epileptic zones. Synthetic ioECoG data offers a powerful solution; enabling scalable, privacy-safe simulation of complex neural signals that improve diagnostic accuracy, support training and reduce reliance on invasive procedures. It also opens the door for better predictive modelling of surgical outcomes and non-invasive diagnostics.		
SD in Action	Training AI models to distinguish between healthy and epileptic brain signals Simulating surgical scenarios and predicting patient outcomes Supporting research in non-invasive, privacy compliant diagnostic tools Enhancing algorithm robustness for rare or complex epilepsy cases		
Key Benefits	Expands access to high-quality datasets without patient risk Enhances AI accuracy through balanced, representative simulations Enables transparent, benchmarked validation frameworks Simulates rare cases to enhance precision and fairness Builds trust via explainable, context-aware data generation		
	Scientific/Technical	**Ethical/Clinical**	**Regulatory**
Challenges & Gaps	Difficult to replicate the physiological complexity of real ioECoG signals Risk of bias or overfitting if synthetic signals reflect flaws in source data SD validation methods are not yet standardised High computational cost for generating large, realistic datasets	SD may miss clinical nuances in rare or atypical epilepsy cases Low trust in AI tools trained on non- real data Consent and representation issues when real patient data used to generate synthetic variants	Lack of regulatory clarity on SD use in diagnostics and surgery. Unclear accountability if AI models trained on SD misguide clinical action Variability in standards across jurisdictions (GDPR, HIPAA, etc.)
Target Users	Neurologists, neurosurgeons, clinical researchers, AI developers, medical device companies		
Involved Stakeholders	Hospitals, AI developers, academic institutions, technology firms, ethics boards		
Relevant Projects	Horizon Europe AISym4MED project		

Use Case 3: Enhancing Neurological Disorder Assessment with Synthetic MRI Data

Synthetic MRI data enhances AI-driven diagnostics for neurological disorders, improving accuracy and innovation in clinical assessment.

Summary	Modern AI tools hold promises for detecting and assessing neurological disorders caused by tumors, traumatic injuries, or neurodegenerative conditions such as Alzheimer's disease. Yet these models require large, diverse datasets of both healthy and affected brain MRIs, resources that are typically scarce, restricted to clinical use, or insufficiently labelled. SD offers a privacy-safe way to generate realistic, high-resolution 3D brain images, helping fill data gaps and improve the reliability of AI-powered diagnostics. However, ensuring the clinical utility, diversity and privacy compliance of these synthetic datasets remains a key challenge.		
SD in Action	Indicative use cases of SD for advancing diagnostic accuracy and assessment of neurological disorders include tumor classification and impact analysis, traumatic injury severity analysis and Alzheimer's disease detection and progression.		
Key Benefits	Enables privacy-safe, high-resolution AI training Expands access to clinically realistic brain images Supports simulation of edge cases and uncommon clinical presentations Enhances diversity and generalisability in model development		
	Scientific/Technical	**Ethical/Clinical**	**Regulatory**
Challenges & Gaps	Most generation methods are 2D-based; 3D MRI synthesis risks discontinuities. High-quality 3D SD is hard to generate from limited real samples High-resolution image generation demands intensive computational power No standard exists for evaluating image quality, realism, or dataset coverage Synthetic images too similar to real ones may compromise patient privacy The field evolves rapidly, lacking stable benchmarks and methods	Access to well-labeled, high-quality MRI data for clinical use is limited Risk of privacy leakage from synthetic images if poorly generated. Clinical validity of synthetic data for intended diagnostic use must be proven.	Ownership and IP rights for synthetic datasets remain unclear.
Target Users	AI developers (medical image analysis), clinicians (radiologist, neurologist, neurosurgeon)		
Involved Stakeholders	Clinicians (radiologist, neurologist, neurosurgeon), AI developers (medical image analysis), data privacy experts, Information and Communication Technology (ICT) providers and research community		
Relevant Projects	Eureka Cluster Program ITEA4-IVVES project		

Use Case 4: Enabling Emergency Preparedness with Synthetic Data for Mass Casualty Incidents

Synthetic data enables AI-driven decision support for mass casualty response, enhancing emergency preparedness and clinical coordination in data-scarce crisis settings.

Summary	Mass casualty incidents (MCIs) are rare but highly disruptive, requiring rapid, coordinated medical responses. Real-world data on such events is limited due to their infrequency, complexity and sensitivity, hindering the development of AI tools that can guide timely triage, treatment and evacuation decisions. SD fills this gap by simulating realistic victim scenarios, enabling model training and team preparedness without compromising privacy. This improves prehospital care, strengthens decision-making under pressure and supports cross-organisational collaboration. While SD helps address critical data gaps, ensuring its realism, clinical relevance and regulatory alignment remains a key challenge for deployment at scale.		
SD in Action	Training AI-powered tools for continuous triage, treatment prioritisation and hospital dispatch Simulating diverse emergency scenarios to support multi-agency collaboration and training Modeling victim deterioration for early intervention guidance		
Key Benefits	Enables privacy-safe training with realistic, high-fidelity synthetic datasets Simulates diverse scenarios and personalised clinical responses Mitigates legal exposure through anonymised synthetic data Powers real-time decision support for critical prehospital interventions Enables cross-agency preparedness through common training datasets		
	Scientific/Technical	**Ethical/Clinical**	**Regulatory**
Challenges & Gaps	Scarcity of prehospital seed data due to event rarity and variability. Complex simulation of victim deterioration and diverse injury types. Computational and modeling complexity for creating scalable and context-rich datasets. Difficulty ensuring continuous accuracy and realism as real-world data and scenarios evolve.	Requires contextual medical expertise for clinically meaningful synthetic data generation Concerns around data sharing, trust in synthetic data and cultural readiness in emergency agencies. Necessity of prior patient consent and public trust in data collection during MCIs. Varying incident types and use cases increase the model design complexity	Lack of regulatory guidance on SD use in life-saving interventions. Alignment with GDPR, AI Act and health data regulations remains unclear.
Target Users	Emergency medical teams, civil protection units, policy makers, AI developers (emergency response systems)		
Involved Stakeholders	First responder organisations, health authorities, AI developers, ICT and simulation providers, civil society organisations		
Relevant Projects	H2020 NIGHTINGALE Project SNS TrialsNet Project		

Use Case 5: Creating Synthetic Control Arms for Smarter Clinical Trials

Synthetic control arms offer a faster, more ethical and data-driven alternative to traditional control groups, accelerating trial timelines and improving patient inclusion, especially in rare and complex diseases.

Summary	Traditional randomised clinical trials require control groups that often expose patients to placebos or standard treatments with limited benefit. Synthetic control arms use retrospective patient data to simulate these groups, reducing recruitment barriers and ethical concerns. With AI, multimodal data such as omics (i.e., datasets generated from various fields of biological study that end in -omics, like genomics), imaging and clinical records can now be used to generate high-quality synthetic cohorts that reflect standard-of-care outcomes. This approach is especially valuable in rare disease trials and oncology, where patient heterogeneity and trial complexity are high. Regulatory interest in this approach is growing, though challenges around data quality, model robustness and privacy remain.		
SD in Action	Rare disease trials with limited eligible participants Oncology trials using omics-driven eligibility criteria Clinical studies require multimodal data integration (omics, pathology, imaging, clinical)		
Key Benefits	Offers ethical, synthetic alternatives to traditional control groups Supports robust, AI-driven patient cohort modeling Facilitates integration of omics, imaging, pathology and clinical records Reduces patient burden and supports trial designs with greater inclusion and transparency Accelerates timelines and lowers cost via data reuse and automation		
	Scientific/Technical	**Ethical/Clinical**	**Regulatory**
Challenges & Gaps	Real-world seed data is often incomplete, noisy, or inconsistently formatted AI methods for multimodal data generation remain experimental and require further validation.	Use of sensitive seed data (e.g. omics) requires strict privacy controls Ensuring clinical credibility and reproducibility of synthetic cohorts is essential	Recognition of synthetic control arms is increasing, but regulatory frameworks for validation and reliability are still evolving
Target Users	Pharmaceutical and biotech companies, clinical trial designers		
Involved Stakeholders	Regulators, AI developers, researchers, digital health companies (e.g., Exploristics, InsilicoTrials)		
Relevant Projects	N/A		

Use Case 6: Synthetic Data for Ethical and Scalable Health & Wellbeing Research

Synthetic data enables privacy-safe, high-quality research in health and well-being by replicating real-world insights without compromising sensitive personal information.

Summary	Privacy concerns and strict regulations often limit access to real-world health and well-being data, hindering research collaboration and innovation. Living Labs offer rich behavioral and physiological insights, yet sharing this data externally remains a challenge. SD provides a promising solution by replicating the statistical properties of real datasets without compromising individual privacy. This use case demonstrates how privacy-preserving SD can be used for analysis, model development and remote validation across distributed Living Lab nodes. Supported by centralised metadata generation, data access management and persistent identifiers, the approach enables ethical, scalable research. Ongoing pilots across sectors—from health to mobility—aim to further validate and adapt SD technologies for broader, cross-disciplinary impact.		
SD in Action	Remote execution of research AI models on synthetic proxies of real datasets Cross-sectoral studies (e.g., health, mobility, environment) Metadata automation and centralised access management via Research Analysis Identifiers (RAI)		
Key Benefits	Allows secure, shareable datasets that preserve anonymity Enables controlled, ethical model testing and validation Drives technology maturation and cross-domain application Supports federated research via Living Lab-hosted nodes		
	Scientific/Technical	**Ethical/Clinical**	**Regulatory**
Challenges & Gaps	Limited synthetic generation capabilities for non-tabular seed data Difficulty in automating model selection and retraining Lack of benchmarks for data quality and privacy metrics. Integration with existing data collection pipelines remains complex.	Risk of bias propagation in synthetic proxies. Uncontrolled use of synthetic datasets for non-ethical purposes	Absence of unified standards for privacy guarantees, AI security and responsible sharing of SD
Target Users	Academic researchers, data scientists and AI developers		
Involved Stakeholders	Living Labs, technology providers, data privacy officers, policy makers, EU regulators		
Relevant Projects	H2020 VITALISE (Health & Wellbeing living labs) HE EOSC RAISE (Pilot in Health)		

Use Case 7: Monitoring Elderly Wellbeing at Home Using Synthetic Data

Synthetic Activities of Daily Living (ADLs) enables personalised, privacy-safe monitoring to support elderly wellbeing in smart home environments.

Summary	The growing need to support ageing populations calls for intelligent, privacy-preserving home monitoring. SD enables the training of decentralised AI models for recognising ADLs in elderly individuals, without compromising personal privacy. By simulating realistic, user-specific behavior patterns from IoT and smart devices, these synthetic datasets support AI systems tailored to diverse living environments. Leveraging edge computing and FL, this approach ensures sensitive data remains local while enabling accurate activity detection, early intervention and better quality of life for elderly individuals at home.		
SD in Action	SD allows in-home AI systems to detect and respond to elderly behavior patterns while respecting personal privacy		
Key Benefits	Addressing data scarcity and bias by enabling availability of data of appropriate quality and volume Enables privacy-preserving and ethical training and testing of AI models for ADL		
	Scientific/Technical	**Ethical/Clinical**	**Regulatory**
Challenges & Gaps	Massive decentralisation and diversity of real-world data Interoperability with diverse IoT environments for real world data collection Real world data volume and quality issues for training AI models Real world data drift in non-stationary problems	Use of sensitive personal seed data in homes requires strict privacy controls	Compliance to GDPR and national data protection regulations for seed data access Compliance to AI Act for SD generation for decision support Need for Regulated Infrastructure
Target Users	Elderly individuals and family members, care providers, occupational therapists and home-based healthcare workers		
Involved Stakeholders	AI developers, smart home solution providers, health and social service administrators, researchers (robotics, human-computer interaction, gerontology)		
Relevant Projects	EIAROB HOMETEC		

Use Case 8: Transforming Medical Training via Synthetic, Privacy-Safe Data

Synthetic data enables hands-on, privacy-safe medical training by expanding access to realistic clinical case simulations.

Summary	Medical education depends on access to diverse clinical cases to support experiential learning for healthcare professionals. However, privacy concerns and limited real world data availability often restrict data sharing across institutions and borders. SD offers a scalable, GDPR-compliant solution by generating anonymised, high-fidelity patient cases that replicate real-world scenarios. These datasets allow doctors and trainees to explore a wide range of clinical patterns, interventions and outcomes without risking patient privacy. This fosters better decision-making and prepares healthcare professionals to manage complex cases with confidence, while promoting cross-border collaboration in medical training.		
SD in Action	SD generated across clinical sites enables secure cross-border collaboration, expanding access to realistic training resources for medical students, doctors and healthcare professionals.		
Key Benefits	Enables access to large volumes of realistic, privacy-compliant data Facilitates experiential education without risking patient privacy Enhances learning through exposure to millions of diverse, anonymised medical data and contexts		
	Scientific/Technical	**Ethical/Clinical**	**Regulatory**
Challenges & Gaps	Current SD models are not tailored to clinical training. Insufficient methods to assess model privacy or data leakage risks. Limited benchmarks to evaluate clinical relevance of synthetic datasets.	Barriers to real world data sharing (e.g., ethical approvals, data sharing agreements, DPIAs) Institutional reluctance to share real world clinical data for education and AI training. Bias and inconsistency across datasets of different institutions.	Cross-border data use must comply with multiple legal frameworks No specific regulation on SD standards for anonymisation or quality assessment Data minimisation and sensitivity (e.g. genetics) complicate training data definition
Target Users	Medical student, healthcare professionals, medical researchers		
Involved Stakeholders	Hospitals, universities, AI developers, data privacy experts, training platform provides.		
Relevant Projects	Horizon Europe SECURED project		

Unlocking Impact: Insights on Synthetic Data in Cross Domain Healthcare Use Cases

This section distills insights on shared benefits, common gaps and challenges across the healthcare domains addressed by the use cases presented above, highlighting how SD advances clinical research and innovation. It analyses recurring technological, ethical, clinical and regulatory aspects in the process of both generating but also using SD, while emphasising the strategic value SD provides across domains, from rare diseases (Use Case 1) to elderly care (Use Case 7) and medical education (Use Case 8). Across the eight presented use cases, SD demonstrates a remarkable ability to address critical needs, from empowering DTs of patients and enhancing clinical trials to addressing data scarcity and enabling bias mitigation, secure data sharing and next-generation medical education. The elaborated use cases do not exist in isolation; rather, they form a web of application areas that align with the already identified eight core domains in which SD delivers strategic value, as shown in Table 1.

Strategic Domain	Use Cases							
	1	2	3	4	5	6	7	8
Virtual Patients: Enabling Personalised, Predictive Care	✓			✓				✓
Tackling Data Scarcity and Bias in Critical Care Scenarios	✓	✓	✓	✓	✓		✓	
Unlocking Data Sharing for Research and Innovation	✓		✓	✓		✓		✓
Addressing Ethical, Legal and Business Barriers	✓		✓			✓	✓	✓
From Gaps to Gains: Using Synthetic Data to Improve AI Fairness	✓	✓	✓	✓	✓		✓	
Reimagining Clinical Trials with Synthetic Data					✓			
Advancing Reliability and Readiness of AI Systems with Synthetic Data	✓	✓		✓				
Training Tomorrow's Health Workforce		✓		✓				✓

Table 1. Mapping uses cases across the Synthetic Data Strategic Domains

The mapping of use cases reveals that certain strategic domains, particularly Tackling Data Scarcity and Bias and Addressing Ethical, Legal and Business Barriers, cut across nearly all use cases. This underscores the synthetic data foundational role in expanding data access, addressing under-representation and accelerating safe and inclusive innovation. More targeted domains, such as Virtual Patients and Clinical Trials Innovation or Training Tomorrow's Health Workforce, play equally vital roles. For instance, applications like rare hematological disease modeling (Use Case 1) and synthetic control arms in clinical trials (Use Case 5) demonstrate how synthetic data enables both predictive care and experimental validation without compromising patient privacy. Meanwhile, six use cases, from rare diseases to emergency triage, clinical trials and medical training (Use Cases 1-5 and 8), reflect the strategic importance of fairness-focused data generation to fill gaps in clinical coverage and demographic representation. Use cases like rare diseases (Use Case 1), epilepsy surgery (Use Case 2) and wellbeing monitoring in Living Labs (Use Case 6) highlight the SD growing role in Privacy-Respecting Research and Innovation, lowering ethical, legal and business barriers to data access, fostering cross-institutional R&D collaboration and aligning with evolving legal and ethical standards. Similarly, use cases such as mass casualty simulations (Use Case 4) and AI-powered epilepsy diagnostics (Use Case 2) illustrate the importance of Advancing Reliability and Readiness of AI Systems, where synthetic datasets enable robust testing under edge-case or high-risk conditions. Finally, Training Tomorrow's Health Workforce surfaces in education-focused use cases (Use Case 8), where SD supports hands-on, privacy-safe learning, from medical schools to smart homes, democratising access to clinical knowledge in increasing digital care settings. Overall, the matrix reveals the SD emerging role as a foundational tool for scalable, secure and inclusive healthcare innovation.

Critical Gaps and Challenges in Synthetic Health Data Research

Despite growing adoption and experimentation, the field of synthetic health data continues to face a number of critical research, policy and infrastructure gaps. These gaps hinder the safe, equitable and scalable integration of SD across the healthcare value chain. The following points outline the most pressing gaps and challenges identified across literature and the presented use cases:

Clinical Data and AI Model Development for Synthetic Data Generation

Limited Access to Real-World Quality Seed Data: High-quality, representative real-world data of adequate volume is essential to SD generators. Yet in many settings such data remains fragmented, imbalanced, noisy/incomplete or inaccessible due to ethical, legal, business or technical barriers.

Lack of Common Assessment and Benchmarking Frameworks: There is no standardised or commonly agreed set of metrics or formal evaluation frameworks to assess the quality in terms of fidelity, privacy and clinical utility of synthetic datasets for their intended use, as such metrics cannot be objectively measured without context-specific benchmarks. This gap impairs trust, reliability and regulatory readiness.

Underdevelopment of Multimodal and Cross-Domain Synthesis: Current AI approaches are optimised for specific data types (e.g., tabular EHR or static images), but lack capacity for generating longitudinal, multimodal datasets (e.g., integrating imaging, sensor, genomic and clinical data). The inability to simulate complex patient trajectories limits real-world impact.

Inadequate Model Accuracy, Explainability and Interpretability: Many SD generators remain opaque in how data distributions are learned and replicated. Existing SD generators are not sufficiently tailored for medical contexts, which limits their ability to produce data that is representative and useful for healthcare applications. This constrains transparency, validation and auditability, particularly in regulated or clinical applications.

Lack of acceptance and trust in AI by clinical stakeholders: Reluctance often stems from the stringent provisions of the AI Act, which emphasise compliance and ethical considerations. Additionally, cultural resistance among end users, as to how supportive the technology is, can limit innovation. Clinical stakeholders are often skeptical regarding the clinical reliability of synthetic data and may question the accuracy of insights derived from AI models trained on synthetic data.

Where Synthetic Data Struggles Most: Shared Challenges Across the Domains

Despite growing momentum, synthetic data solutions consistently face friction at the intersection of technology, ethics and regulation.

From the high computational burden of multimodal synthetic data generation to trust gaps in clinical deployment, these issues signal the urgent need for systemic coordination.

Strategic guidance, harmonised standards and ethical governance are no longer optional, they are prerequisites for scale and impact.

Governance and Legal Ambiguities

Unclear Data Ownership and Exploitation Rights: There is no legal consensus on the ownership of SD, especially when derived from seed data pooled across multiple clinical institutions. Questions of attribution, reuse and monetisation remain unresolved.

Absence of Fit-for-Purpose Data Governance Structures: Trusted intermediaries and federated data infrastructures for sharing, cataloguing and validating SD remain underdeveloped. Without such systems, large-scale collaboration and transparency are difficult to achieve.

Gaps in Data Sharing Agreements: Existing legal instruments and informed consent mechanisms are often incompatible with the scale and flexibility synthetic data demands. There is the need for standardised clauses and agreements that address secondary use, cross-border access and risk-sharing for SD generation purposes.

Infrastructure, Regulatory and Policy Gaps

Lack of Regulatory Sandboxes and Testbeds: Few jurisdictions offer structured, compliant environments for testing synthetic data in clinical or cross-border contexts, catering for compliance assessment to applicable regulations of the developed models. This slows down experimentation, validation and responsible adoption, particularly for early-stage innovation.

Limited Support for Cross-Border Collaboration: Despite the progress towards the EHDS, regulatory and operational barriers continue to prevent seamless data access across national systems. This stifles the development of SD models trained on diverse, representative populations.

Insufficient Funding for Scalable Innovation: Large-scale, multi-stakeholder efforts to develop SD generation engines tuned to local health system needs remain underfunded. Targeted investment is required to build trustworthy engines that address structural bias and regional diversity.

Scientific, Technical and Security Gaps

Lack of Research on Privacy Guarantee Composition: There is limited understanding of how privacy guarantees (e.g., DP parameters) accumulate across SD pipelines. This makes it difficult to assess cumulative privacy risks in complex systems and augments the fear of re-identification and the privacy concerns of clinical stakeholders who become reluctant to collaborate.

Gaps in Post-Quantum Security Readiness: Emerging SD architectures are not yet designed to withstand forthcoming quantum computing threats. Research into quantum-resilient Privacy Enhancing Technologies (PETs) for synthetic data generation is currently minimal.

Usability Constraints: Most privacy-enhancing technologies remain difficult for non-specialists to implement. Thus, usability constraints limit broader adoption of SD in clinical, industry/SME and academic settings.

A complete comparative analysis of the identified scientific/technical, clinical/ethical & regulatory gaps and challenges across the eight studied use cases, is summarised in Table 2, Table 3 and Table 4 respectively. The definitions of the presented challenges/gaps in these Tables are provided in Annex I – Definitions of Identified Gaps & Challenges.

Scientific/Technical Gaps & Challenges	Use Cases							
	1	2	3	4	5	6	7	8
Seed Data Volume and Quality	✓			✓	✓		✓	
Need for Distributed Data Access and Federated AI For Scarce Seed Data	✓						✓	
Need for Infrastructure and Compute Power	✓	✓	✓			✓		
Lack of Data and Systems Interoperability	✓						✓	
Data Governance Gaps Among Multiple Stakeholders	✓							
Privacy, Security and Data Protection	✓		✓	✓			✓	✓
Lack Of Metrics and Benchmarking Framework for Target Use	✓	✓	✓	✓	✓	✓		✓
Testing and Experimentation Provisions	✓	✓	✓	✓			✓	
Synthetic Data Quality, Realism and Accuracy Limitations	✓	✓	✓	✓				
AI maturity for Synthetic Data Generation	✓		✓	✓	✓	✓		✓
Operationalisation of AI for Synthetic Data Generation		✓				✓		
Reliability of Clinical Decision Support Systems		✓		✓	✓			

Table 2. Scientific/Technical gaps and challenges across use cases

Ethical & Clinical Gaps & Challenges	Use Cases							
	1	2	3	4	5	6	7	8
Ethical Process for Seed Data Access	✓				✓		✓	✓
Anonymisation vs. Fidelity/Utility Trade-offs	✓							
Patient Consent need	✓	✓		✓	✓		✓	
Bias and Fairness	✓	✓				✓		✓
Reluctance to Share, Fear of Re-Identification	✓							✓
Privacy Concerns	✓		✓	✓	✓		✓	
Misuse of Synthetic Data						✓		
Extent of Clinical Relevance	✓	✓	✓	✓				
Cross-Site Clinical Data Collection/ Representation	✓							✓
Lack Of Acceptance and Trust	✓	✓		✓				✓

Table 3. Ethical and clinical gaps and challenges across use cases

Regulatory & Standardisation Gaps & Challenges	Use Cases							
	1	2	3	4	5	6	7	8
Compliance with Legal Frameworks	✓	✓		✓	✓	✓	✓	✓
Cross-Border Data Sharing Complexity	✓					✓		✓
Open Data Policy for Synthetic Data						✓		✓
Regulated Infrastructure Requirement	✓						✓	
Synthetic Data Ownership and Business Governance Gaps	✓		✓			✓		
Gaps in Synthetic Data Reliability and Evaluation standards	✓				✓	✓		✓
Gaps in Accountability for Use of Synthetic Data in Decision Support Systems	✓	✓		✓				

Table 4. Regulatory and standardisation gaps and challenges across use cases

Recommendations for Scaling Synthetic Data in Health Innovation & Beyond

Realising the transformative potential of synthetic data in healthcare will require coordinated action across regulation, technology, ethics and ecosystem development. The following recommendations present a consolidated roadmap for responsible, scalable and high-impact synthetic data adoption.

The foundational need is to **strengthen access to high-quality, real-world seed data**. Synthetic models are only as robust as the data they learn from. Yet, in many domains, particularly those involving rare diseases or vulnerable populations, real-world datasets remain fragmented or inaccessible. Investments in **secure federated, privacy-preserving data management infrastructures** across multiple data providers, along with adopting **secure FL**, are essential to enable **secure data access and ethical and significantly bias-free SD generation** across borders. Simultaneously, efforts must support the creation of **"synthetic-ready" real-world datasets,** as well as **"AI-ready" synthetic datasets**, that are accessible, complete, noise and bias-free and semantically interoperable, aligning with existing widely adopted data standards (e.g., SNOMED CT, HL7 FHIR, OMOP CDM) to ensure cross-system consistency.

Building trust also depends on the **definition and adoption of shared assessment frameworks**, **benchmarks and standards**. The field currently lacks universal evaluation frameworks for assessing fidelity, privacy and utility of SD and of DTs of patients for the same or even across use contexts. Co-developed benchmarks, particularly in high-impact areas such as rare diseases, oncology or neurology, need to be developed in standardised form to enable regulatory readiness. Open benchmarking platforms can foster transparency, comparability and clinical relevance and trust.

Building clinical stakeholders confidence requires transparent validation processes, robust evidence of efficacy and continuous engagement with clinicians and end users to address their concerns and demonstrate the tangible benefits of synthetic data in healthcare.

Advancing scientific and technical capabilities of AI-driven SD generation is equally vital. Current AI-driven synthetic data generation methods often focus on single data types, limiting their usefulness. Future models should integrate multimodal inputs, such as time-series vitals, genomics, imaging and clinical notes, while also improving explainability. Hybrid approaches that combine deep generative models with privacy-preserving methods like FL, DP or SMPC can enhance both realism and privacy in sensitive applications.

Clarifying data ownership and governance is another priority. Legal ambiguity persists around rights to SD, especially when it is generated from pooled, multi-institutional sources. Policymakers must define attribution and reuse rights across research, validation and commercial contexts. Trusted intermediaries at national or EU levels could support governance by overseeing consent, cataloguing data assets and ensuring equitable reuse. Effective governance is essential to foster trust and cooperation among parties, facilitating the seamless integration and utilisation of SD.

Advent of European Data Union Strategy: setting the ground and widening the scope of Synthetic Data

Triggered by the newly introduced **European Data Union Strategy** [30], it becomes evident that **synthetic data** brings **significant value and impact** to **a multitude of other domains** spanning from manufacturing to mobility and transport, energy or public services.

This is because synthetic data addresses a series of the Strategy **key objectives:**

- more open, easily accessible, privacy-safe, bias-free and wider synthetic data availability, accessibilit and sharing across systems and stakeholders to boost collaboration, research and innovation.
- ensuring desired quality and volume to train robust, fair and reliable AI applications
- better addressing EU's data regulatory landscape to facilitate cross border data flows.

BDVA [31], with the support of many of its Task Forces, including the Healthcare Task Force [32], has already responded with recommendations, including relevant ones pertaining to synthetic data, to the public consultation on the European Data Union Strategy, to maximse their impact.

Legal instruments should also evolve. Standard informed consent forms and data sharing agreements are rarely equipped to cover the cross-border and multi-use nature use purposes of SD. SD-specific clauses are needed to define liability, enable collaborative reuse and protect patient rights. This will improve legal certainty and foster ethically grounded partnerships.

Meanwhile, **regulatory sandboxes and real-world testbeds** are critical for safe experimentation. Co-designed with regulators, clinicians and researchers, these environments would enable early validation of clinical trials, digital twins, or diagnostic tools. Federated testbeds and Living Lab infrastructures can further support responsible cross-border experimentation. Privacy and security aspects remain foundational. Ensuring compliance involves establishing robust data governance frameworks, conducting regular audits and fostering transparency to build trust among stakeholders and effectively leverage SD while safeguarding patient privacy and ensuring adherence to legal standards.

Privacy-enhancing technologies must be further developed for usability, scalability and post-quantum resilience. Composable privacy frameworks are needed to assess cumulative risks across SD generation pipelines, while accessible PET tools will help broaden adoption in clinical and SME settings.

Finally, success depends on **sustainable funding and viable business models**. Although momentum is growing, the ecosystem still lacks adequate investment, particularly for initiatives addressing structural bias, equity and regional diversity. EU-level programs such as Horizon Europe, EU4Health and Digital Europe could play a strategic role by supporting targeted SD missions aligned with public health priorities. Emerging models like **"Data Generation as a Service"** along with open-source toolkits and modular platforms, offer scalable pathways for broader access. Moving from pilot to practice will require coordinated public-private partnerships and mission-driven funding to ensure impact and long-term sustainability.

Together, these actions provide **a strategic foundation for embedding synthetic data into Europe's digital health ecosystem,** supporting systems that are **more equitable, fair, privacy-preserving and innovation-ready**.

Credits & Acknowledgement

The white paper has been produced by the **Big Data Value Association** [31], and specifically by an expert working group of its **Healthcare Task Force** [32] working in different research fields pertaining to synthetic data, AI, data and privacy-enhancing technologies. The **Big Data Value Association (BDVA)** [31] is an industry-driven research and innovation association with a mission to develop the innovation ecosystem that enables and accelerates the Data and AI economy with European values and focus. Its **Healthcare Task Force** [32] brings together experts from the health and care sectors with the goal to create a platform and a network of collaboration within the BDVA community to identify key technologies and research trends, such as the use of synthetic data in the healthcare domain. The Task Force has also links to the pharma and drug development fields.

Last but not least, the authors would like to express their gratitude to all participating EU-funded and national research projects and their representatives for the crucial inputs and insights openly shared to maximise impact and further research on open still issues.

Annex I – Definitions of Identified Gaps & Challenges

The following Tables outline the definition of each one of the identified gaps & challenges in generating and using SD in healthcare.

Scientific/ Technical Gaps & Challenges	Definition
Seed Data Volume and Quality	The volume and quality of seed data are crucial in order to achieve accurate and SD generation. Access to data is often challenged by scarcity, e.g. in rare diseases, leading to difficulties in obtaining sufficient training data. In addition, seed data is frequently noisy and incomplete, negatively impacting their quality, necessitating processes like denoising, imputation, harmonisation and standardised description to enhance its utility.
Need for Distributed Data Access and Federated AI For Scarce Seed Data	Scarcity of seed data is a significant problem for AI. To guarantee appropriate volume and quality of seed data and to reduce population biases, distributed data access and management from multiple data providers is required, along with adopting secure FL approaches for SD generation model training, since sensitive health data must not migrate to third party platforms, while fully anonymised data loses significant information for analysis.
Need for Infrastructure and Compute Power	The need for infrastructure and compute power in clinical data management is critical to address the computational demands of generating, for example, high-resolution synthetic 3D images. Furthermore, scalability challenges arise when producing large-scale, high-quality SD, as this process demands significant computational resources and time. Clinical data holders are called to address such infrastructure and compute power needs to effectively manage the demanding computational requirements for local SD generation (as seed data reside at the clinical data holder site).
Lack of Data and Systems Interoperability	In scenarios of care provision, a lack of data and systems interoperability significantly impacts the use of SD, primarily due to the massively distributed architecture often involved for access to seed data. Achieving effective interoperability is crucial for leveraging SD to enhance healthcare delivery.
Data Governance Gaps Among Multiple Stakeholders	The generation of SD in healthcare necessitates robust data governance structures, especially among multiple stakeholders such as clinical and ICT solution providers, since clear such governance frameworks defining roles and responsibilities among multiple stakeholders for the generation and use of SD are not yet fully established.

Privacy, Security and Data Protection	In the realm of AI model training, privacy and data protection are paramount due to the sensitive nature of health data. Privacy-enhancing technologies and robust security measures are critical to safeguarding access and use of seed data, with DP and SMPC protocols serving as effective methods to mitigate such risks. Furthermore, AI models are not always adequately assessed for vulnerabilities, such as inference attacks, which could lead to unintended leakage of personal information. Ensuring comprehensive data protection also necessitates rigorous anonymisation processes and secure communication channels to prevent unauthorised access and maintain confidentiality of data.
Lack Of Metrics and Benchmarking Framework for Target Use	The lack of metrics and a robust evaluation or benchmarking framework for the target use of SD poses significant challenges in assessing synthetic data quality, particularly in healthcare, along with fidelity, utility and privacy trade-offs, as these elements cannot be objectively measured without context-specific benchmarks. Methods to validate SD against real clinical data are not always well-established, making it hard to gauge how useful or reliable the synthetic data is.
Testing and Experimentation Provisions	As with the infrastructure and compute power needs to manage data and train SD generation algorithms, relevant testing and experimentation provisions are necessary for further evaluating the trained models and the resulting SD objectively against a commonly agreed set of metrics for the intended use. Although TEFs have been established for similar purposes in target domains, including Health, these did not consider the specific needs on evaluation methods and testing tools for synthetic data.
Synthetic Data Quality, Realism and Accuracy Limitations	The quality of SD is highly dependent on the quality of real-world seed data used in the SD generation process; any flaws or incompleteness in the latter can propagate into the synthetic data, potentially compromising their usefulness and reliability in practical applications. Additionally, the generation of 3D data faces significant challenges due to the predominance of methods initially designed for 2D data, which can result in through-plane discontinuities. SD often faces realism and accuracy limitations due to its inability to fully capture the complexity and nuances present in real-world data, which can lead to AI models that perform inadequately in practical applications.
AI maturity for Synthetic Data Generation	The maturity of AI methods for SD generation is still evolving as new approaches continue to emerge. While multimodal AI approaches are gaining traction in academic research, designing algorithms capable of accurately capturing the complex biology of diseases remains an unresolved challenge. SD generation models lack maturity, especially for non-tabular data, while the data diversity and variability inherent in medical emergency settings are not effectively captured by current AI or simulation techniques. Existing SD generation approaches are not sufficiently tailored for medical contexts, which limits their ability to produce data that is representative and useful for healthcare applications.

Operationalisation of AI for Synthetic Data Generation	Operationalising AI for SD generation involves strategic decision-making and integration challenges, particularly in determining when to retrain models in dynamic environments where data collection and availability is incremental. Automating the choice of which SD generation model to employ for each data source demands intelligent systems capable of evaluating the specific characteristics and requirements of the data and intended purpose. Predicting the necessary computing resources for model training and inference adds another layer of complexity, impacting efficiency and cost-effectiveness. Successfully integrating SD generation pipelines into existing data-producing workflows requires seamless coordination to ensure that synthetic datasets are consistently updated and accurately reflect real-world conditions. As real-world data evolves, maintaining the accuracy, fidelity and reliability of SD necessitates regular assessments and updates, ensuring its ongoing relevance and effectiveness in supporting diverse applications.
Reliability of Clinical Decision Support Systems	The reliability of clinical decision support systems (DSS) is significantly impacted by the quality and integrity of the synthetic data used in their development. Generative AI models often rely on existing datasets that may embed inherent biases or inaccuracies, which can be perpetuated in the synthetic data they produce. If these biases are not identified and mitigated, they can lead to skewed outcomes and unfair decision-making in clinical applications. Additionally, poorly generated SD can result in models that lack accuracy and are not generalisable, undermining their effectiveness in real-world clinical settings.

Ethical & Clinical Gaps & Challenges	Definition
Ethical Process for Seed Data Access	Accessing seed data from one or more clinical sites to feed SD generation processes involves navigating complex and often tedious ethical processes. These include undergoing time-consuming Ethical Approval processes by Ethics Committees at each clinical site, securing Joint Controllers Agreement (JCA) and conducting DPIAs to ensure compliance with regulations and standards. Secondary use of retrospective seed data necessitates additional processes to be considered, including efficient data anonymisation procedures to protect patient privacy, which may lead to information loss.
Anonymisation vs. Fidelity/Utility Trade-offs	Data anonymisation, a critical process for ensuring privacy and GDPR/EHDS compliance when sharing seed health data to third parties, often presents a trade-off between maintaining data fidelity and utility. By altering or removing identifiable information to protect individual privacy, the process can inadvertently reduce the granularity and richness of the real-world health data, thereby reducing its usefulness and accuracy for AI-driven SD generation processes, ultimately affecting the quality of the generated SD. Therefore, balancing the need for robust anonymisation while maintaining desired levels of data utility and fidelity is crucial in SD generation.

Patient Consent need	Patient consent is a fundamental step in the ethical use of health seed data, particularly when collecting real world prospective data for primary use. Informed consents ensure that patients are fully aware of how their data will be utilised, fostering trust and transparency. However, obtaining consent prior to e.g. health emergencies can be challenging, emphasising the need for pre-established relevant frameworks. Furthermore, while SD offers a pathway to protect patient privacy, ethical concerns persist regarding whether patient consent is necessary when real world data serves as the basis for generating synthetic datasets.
Bias and Fairness	Addressing bias and fairness in SD generation is critical to ensuring that AI models trained on such data do not perpetuate or amplify existing inequalities. Ensuring that synthetic data accurately and equitably represents diverse patient populations is essential to avoid skewed outcomes that may disadvantage certain groups. When SD serves as a private proxy for real data, there is a risk of reproducing or even amplifying inherent biases present in the original data. To mitigate these biases, it is crucial to implement rigorous checks both in the input seed data used for generating synthetic data and in the output data produced. This involves employing advanced techniques and validation processes to ensure fairness and equity in data representation.
Reluctance to Share, Fear of Re-Identification	The reluctance to share data, whether real or synthetic, is significantly influenced by the fear of re-identification and the potential privacy breaches that may occur and hinders stakeholder collaboration in clinical research. This reluctance further extends to sharing data for AI training with other stakeholders, as the perceived risks associated with personal data exposure and privacy infringement often outweigh the potential benefits of collaborative health tech innovation. Furthermore, the need for data sharing agreements adds another layer of complexity.
Privacy Concerns	Privacy concerns are paramount when generating and using SD, especially in contexts where highly sensitive personal information is involved, such as in omics data and disease progression. It is further crucial to ensure that synthetic images derived from original datasets do not inadvertently contain privacy-sensitive details that could compromise privacy. This is particularly challenging in distributed environments, such as monitored homes, where the risk of privacy breaches is amplified due to the diverse and decentralised nature of data collection.
Misuse of Synthetic Data	The misuse of SD is of concern, particularly when control mechanisms are insufficient to prevent its utilisation in training non-ethical algorithms. Once SD is shared, it can be challenging to monitor and regulate its application, potentially leading to scenarios where it is used inappropriately or maliciously, such as developing biased AI models.
Extent of Clinical Relevance	The clinical relevance of SD is often limited by its potential inability to capture the full spectrum of clinical variability, including rare or atypical cases, which are crucial for developing comprehensive and effective AI models. This limitation poses a significant risk, as models trained on such data may inadequately address the diverse needs of patients in real-world clinical settings, potentially leading to suboptimal decision-making and care.

Cross-Site Clinical Data Collection/ Representation	Cross-site clinical data collection and representation are often complicated by the disparate data formats employed by different institutions. Each healthcare provider may utilise its own systems and standards for recording and storing patient information, leading to significant variability in data structure and content. These differences pose challenges in harmonising and integrating data from multiple sites, which is essential for generating comprehensive synthetic datasets that are representative of diverse patient populations. Inconsistent data formats can hinder the seamless exchange of information, complicating efforts to create unified and interoperable SD for clinical research and AI model training.
Lack Of Acceptance and Trust	The lack of acceptance and trust in AI, especially in the context of SD generation, is a multifaceted issue that hinders the adoption and integration of these technologies in healthcare. Reluctance often stems from the stringent provisions of the AI Act, which emphasise compliance and ethical considerations. Additionally, cultural resistance among end users, as to how supportive the technology is, can limit innovation. Clinical stakeholders are often skeptical regarding the clinical reliability of SD and may question the accuracy of insights derived from AI models trained on SD.

Regulatory & Standardisation Gaps & Challenges	Definition
Compliance with Legal Frameworks	Ensuring compliance with legal frameworks such as GDPR, HIPAA, EHDS regulation and the AI Act presents a complex challenge for synthetic data generation in healthcare. Since SD is derived from real-world patient information, it carries an inherent risk of privacy breaches and data misuse. Therefore, organisations must navigate a multifaceted regulatory landscape to ensure that the generation and use of SD align with national and international data protection and privacy regulations. This includes implementing stringent safeguards, especially when handling sensitive information such as genetic data, which demands even higher levels of protection. Moreover, organisations must also consider the AI Act's requirements, which emphasise ethical use and data protection.
Cross-Border Data Sharing Complexity	Cross-border data sharing requires complying with regulations from multiple countries, which presents an even more complex regulatory landscape to manage in a time-wise optimal manner.

Open Data Policy for Synthetic Data	SD can be much more easily shared across stakeholders than real world health data, since it poses less privacy and data protection concerns brought by relevant policies and regulations. As such, SD could be considered as a form of open shareable data for research and innovation, however respective policies for sharing SD as open data do not currently exist.
Regulated Infrastructure Requirement	The regulated infrastructure requirement underscores the necessity for computing environments that meet stringent regulatory and data protection standards, when training AI models with real world health data to produce SD in healthcare. Such infrastructure must be designed to prevent unauthorised access and ensure the confidentiality, integrity and availability of sensitive data. This often involves using controlled, sandbox-like environments, which provide isolated spaces for data processing and model training, minimising the risk of data breaches or leaks.
Synthetic Data Ownership and Business Governance Gaps	Gaps in SD ownership and business governance present significant challenges, particularly concerning the establishment of clear guidelines on ownership and intellectual property rights of generated datasets. As SD is derived from real-world data, often involving multiple stakeholders such as data providers, AI developers and end users, determining ownership and co-ownership becomes complex. These complexities are further exacerbated by the lack of standardised business governance frameworks that define who holds rights to the synthetic data and how it can be used, shared, or commercialised. Without clear delineation of ownership, there may be disputes over control and profit-sharing, potentially hindering collaboration and innovation.
Gaps in Synthetic Data Reliability and Evaluation standards	Current SD reliability assessment methods are often insufficient, lacking standardised frameworks and metrics to evaluate the fidelity, privacy and utility of synthetic data in clinical contexts. This lack of standardised approaches can lead to inconsistencies and uncertainties in how SD are utilised and interpreted, potentially undermining their effectiveness and credibility. In addition, particularly for synthetic control arms, despite consensus among regulators in the US and Europe on their potential value in clinical trials and other research settings, significant gaps persist in establishing standards and regulations for reliability assessment of such data.
Gaps in Accountability for Use of Synthetic Data in Decision Support Systems	There is currently no relevant regulatory framework or policy to regulate accountability related to the use of SD for clinical decision support and clarify who would be responsible if a model trained on SD fails in a clinical setting. Establishing clear guidelines for accountability when using SD for AI model development is a major regulatory hurdle.

BDV BIG DATA VALUE ASSOCIATION

SYNTHETIC DATA IN HEALTHCARE

BENEFITS AND OPPORTUNITIES

TECHNOLOGICAL, CLINICAL, REGULATORY

GAPS & CHALLENGES

WAYS AHEAD FOR IMPACT MAXIMISATION

Bibliographic References

[1] European Health Data Space regulation. Available from: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en

[2] General Data Protection Regulation. Available from: https://www.edps.europa.eu/general-data-protection-regulation_en

[3] Giuffrè, M., Shung, D.L. Harnessing the power of synthetic data in healthcare: innovation, application and privacy. npj Digit. Med. 6, 186 (2023). https://doi.org/10.1038/s41746-023-00927-3gfgs

[4] S. Islam, H. Elmekki, A. Elsebai, J. Bentahar, N. Drawel, G. Rjoub, W. Pedrycz. A comprehensive survey on applications of transformers for deep learning tasks. Expert Systems with Applications, 241 (2024). https://doi.org/10.1016/j.eswa.2023.122666

[5]Gonzales A, Guruswamy G, Smith SR. Synthetic data in health care: A narrative review. PLOS Digit Health 2(1) (2023) https://doi.org/10.1371/journal.pdig.0000082

[6] The NIGHTINGALE Consortium (incl. Netcompany SEE & EUI, Sofia Tsekeridou, Filopoimin Lykokanellos, Evangelos Agorogiannis), Marta Caviglia (2025). Bridging Data Gaps in Emergency Care: The NIGHTINGALE Project and the Future of AI in Mass Casualty Management. Journal of Medical Internet Research, vol. 7, (2025). https://doi.org/10.2196/67318

[7]Hernandez, M., Epelde, G., Alberdi, A., Cilla, R., & Rankin, D. Synthetic data generation for tabular health records: A systematic review. Neurocomputing, 493, 28–45 (2022). https://doi.org/10.1016/j.neucom.2022.04.053

[8] Isasa, I., Hernandez, M., Epelde, G., Londoño, F., Beristain, A., Larrea, X., Alberdi, A., Bamidis, P., & Konstantinidis, E. Comparative assessment of synthetic time series generation approaches in healthcare: Leveraging patient metadata for accurate data synthesis. BMC Medical Informatics and Decision Making, 24(1), 27 (2024). https://doi.org/10.1186/s12911-024-02427-0

[9] Jin, R., Zhang, Z., Wang, M., & Cong, L. (2025). STELLA: Self-Evolving LLM Agent for Biomedical Research. arXiv preprint arXiv:2507.02004. https://arxiv.org/abs/2507.02004

[10] Cheng, C., Messerschmidt, L., Bravo, I. et al. A General Primer for Data Harmonisation. Sci Data 11, 152 (2024). https://doi.org/10.1038/s41597-024-02956-3

[11] Lehne, Moritz, et al. Why digital medicine depends on interoperability. NPJ digital medicine 2.1: 79 (2019).
[12] Raab, René et al. Federated electronic health records for the European Health Data Space. The Lancet Digital Health, Volume 5, Issue 11, e840 - e847 (2023).
[13] Rieke, Nicola, et al. The future of digital health with federated learning. NPJ digital medicine 3, 119: 1-7 (2020). https://doi.org/10.1038/s41746-020-00323-1
[14] Xu, J., Glicksberg, B.S., Su, C. et al. Federated Learning for Healthcare Informatics. J Healthc Inform Res 5, 1–19 (2021). https://doi.org/10.1007/s41666-020-00082-4
[15] Ulf Mattsson. Privacy-Preserving Analytics and Secure Multiparty Computation. ISACA Journal, Vol. 2 (2021). Available from: https://www.isaca.org/resources/isaca-journal/issues/2021/volume-2/privacy-preserving-analytics-and-secure-multiparty-computation
[16] The United Nations Guide on Privacy-Enhancing Technologies for Official Statistics. Available from: https://unstats.un.org/bigdata/task-teams/privacy/guide/2023_UN%20PET%20Guide.pdf
[17]OECD, Emerging privacy-enhancing technologies: Current regulatory and policy approaches, OECD Digital Economy Papers, No. 351, OECD Publishing, Paris (2023). https://doi.org/10.1787/bf121be4-en
[18] Hernandez, M., Epelde, G., Alberdi, A., Cilla, R., & Rankin, D. Synthetic Tabular Data Evaluation in the Health Domain Covering Resemblance, Utility and Privacy Dimensions. Methods of Information in Medicine. (2023). https://doi.org/10.1055/s-0042-1760247
[19] Ozonze, O., Scott, P. J., & Hopgood, A. A. Automating electronic health record data quality assessment. Journal of Medical Systems, 47(1), 23 (2023). https://doi.org/10.1007/s10916-022-01892-2
[20] Chen, H., Chen, J., & Ding, J. Data evaluation and enhancement for quality improvement of machine learning. IEEE Transactions on Reliability, 70(2), 831-847 (2021). https://doi.org/10.1109/TR.2021.3070863
[21] P. A. Osorio-Marulanda, G. Epelde, M. Hernandez, I. Isasa, N. M. Reyes and A. B. Iraola. Privacy Mechanisms and Evaluation Metrics for Synthetic Data Generation: A Systematic Review. IEEE Access, vol. 12, pp. 88048-88074 (2024). https://doi.org/10.1109/ACCESS.2024.3417608

[22] Hurtado Ramírez, D., Porras Díaz, L., Rahimian, S., Auñón García, J.M., Irigoyen Peña, B., Al-Khazraji, Y., Gavín Alarcón, Á., González Fuente, P., Soler Garrido, J. and Kotsev, A. Technological Enablers for Privacy Preserving Data Sharing and Analysis, Publications Office of the European Union, Luxembourg (2023). https://doi.org/10.2760/427718 , JRC134350.
[23]DD. Adam. Synthetic data can aid the analysis of clinical outcomes: How much can it be trusted? Proc. Natl. Acad. Sci. U.S.A. 121 (32) (2024). https://doi.org/10.1073/pnas.2414310121
[24]Earning Trust for AI in Health: A Collaborative Path Forward. World Economic Forum White Paper, published 11 June 2025. Available from: https://www.weforum.org/publications/earning-trust-for-ai-in-health-a-collaborative-path-forward/
[25] What is SNOMED CT. Available from: https://www.snomed.org/what-is-snomed-ct
[26] HL7 FHIR, Release 5. Available from: https://www.hl7.org/fhir/
[27] The OMOP Common Data Model (CDM). Available from: https://www.ohdsi.org/data-standardisation/
[28]Artificial Intelligence (AI) Act. Available from: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
[29] Artificial Intelligence and Machine Learning in Software as a Medical Device (AI/ML SaMD) Regulation, FDA. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device
[30] Forthcoming European Data Union Strategy. News item available from here: https://data.europa.eu/en/news-events/news/commission-unveils-new-strategy-how-productivity-will-be-boost-digital-tech
[31] Big Data Value Association. Available from: https://bdva.eu/
[32] Healthcare Task Force of BDVA. Available from: https://bdva.eu/task-forces/healthcare/

BDVA
Data, AI and Robotics (DAIRO) aisbl
Avenue des Arts, 56
1000 Bruxelles
Belgium

BDVA.eu
info@bdva.eu