EDITH

Deep Thinkers Meeting

Breakout session 3

Data Privacy and Security regulatory landscape

Edwin Morley-Fletcher
President – Lynkeus

Lorenzo Cristofaro
EDITH *Senior Legal Counsel*

Rome 16th/17th May 2023

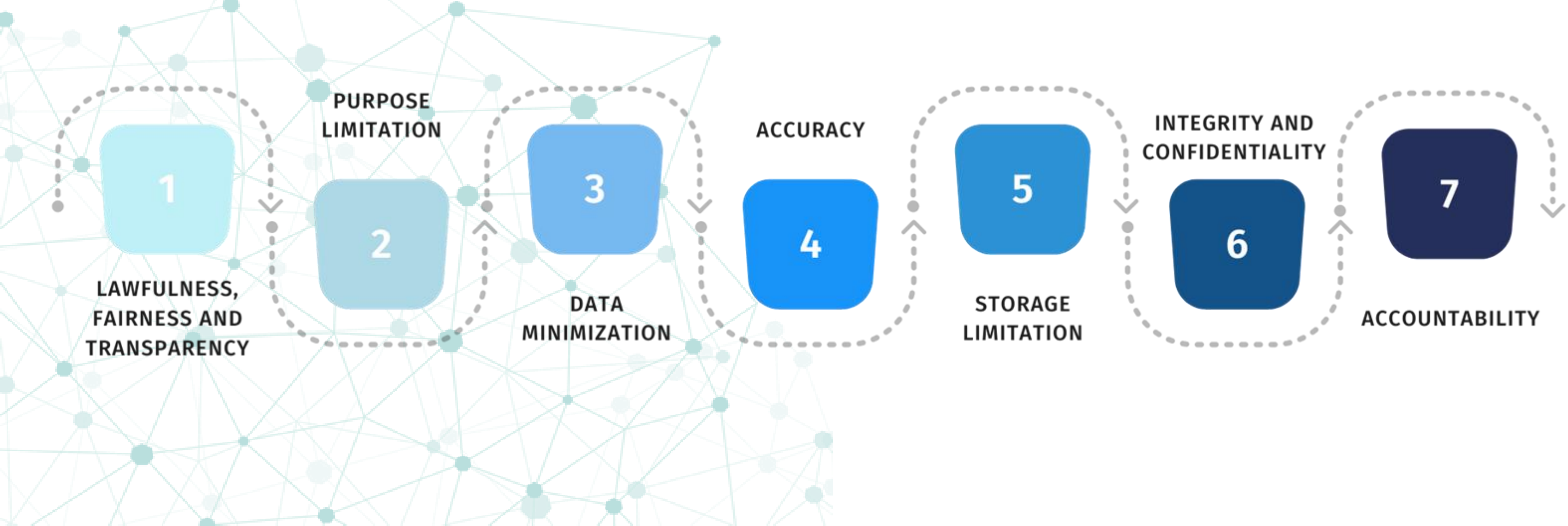
Requirements for artificial Intelligence applications

1. **Legal scenario and regulatory boundaries;**
2. **Health data reuse;**
3. **Specificities of AI-driven approaches;**
4. **Privacy-enhancing technologies to foster data sharing in virtual twins;**
5. **Ways and recommendations to engage with regulatory bodies towards adoption**

Data Privacy and Security regulatory landscape

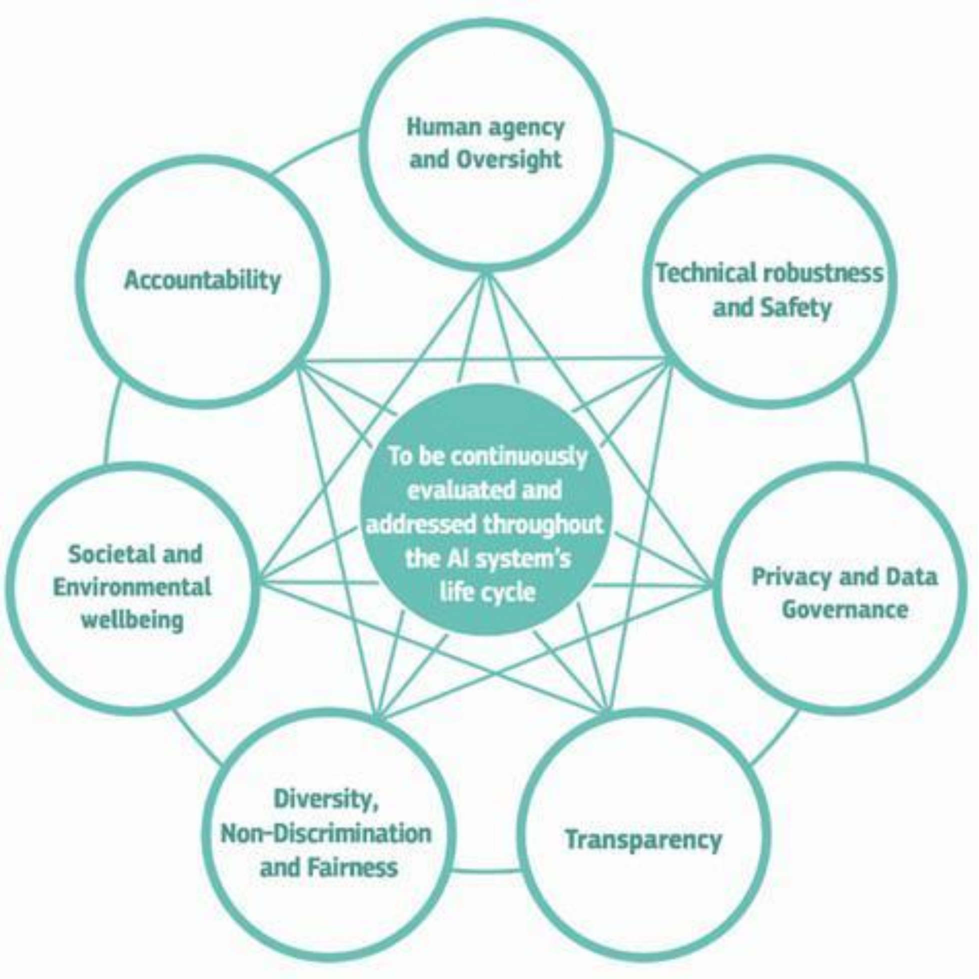
In the light of the currently applicable legislation, there are a number of gaps which need to be filled and issues that must be addressed. Some of the main and general answers arising in connection with Virtual Human Twins ('VHT') are:

1. ***which are the privacy conditions and safeguards that must be fulfilled so that personal data can be lawfully used in the context of VHT?***

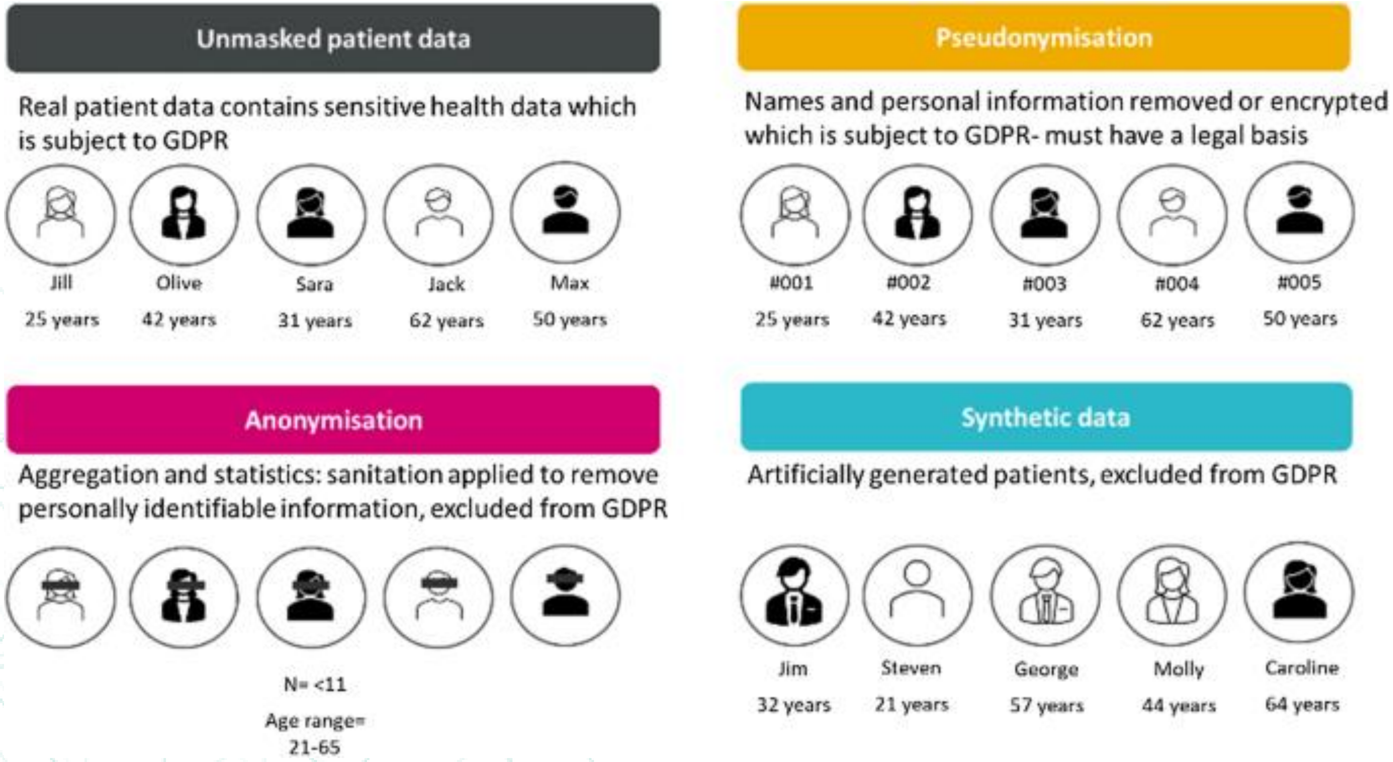


Data Privacy and Security regulatory landscape

2. *is the reuse of health data permitted in the EU for the purpose of delivering Artificial Intelligence-driven medical solutions?*
3. *which obligations apply to the developers and the users of AI-based models necessary to elaborate VHT and relevant ecosystems?*


Human agency and Oversight
Technical robustness and Safety
Privacy and Data Governance
Transparency
Diversity, Non-Discrimination and Fairness
Societal and Environmental wellbeing
Accountability
To be continuously evaluated and addressed throughout the AI system's life cycle

Data Privacy and Security regulatory landscape



Unmasked patient data

Real patient data contains sensitive health data which is subject to GDPR

Jill 25 years · Olive 42 years · Sara 31 years · Jack 62 years · Max 50 years

Pseudonymisation

Names and personal information removed or encrypted which is subject to GDPR- must have a legal basis

#001 25 years · #002 42 years · #003 31 years · #004 62 years · #005 50 years

Anonymisation

Aggregation and statistics: sanitation applied to remove personally identifiable information, excluded from GDPR

N= <11

Age range= 21-65

Synthetic data

Artificially generated patients, excluded from GDPR

Jim 32 years · Steven 21 years · George 57 years · Molly 44 years · Caroline 64 years

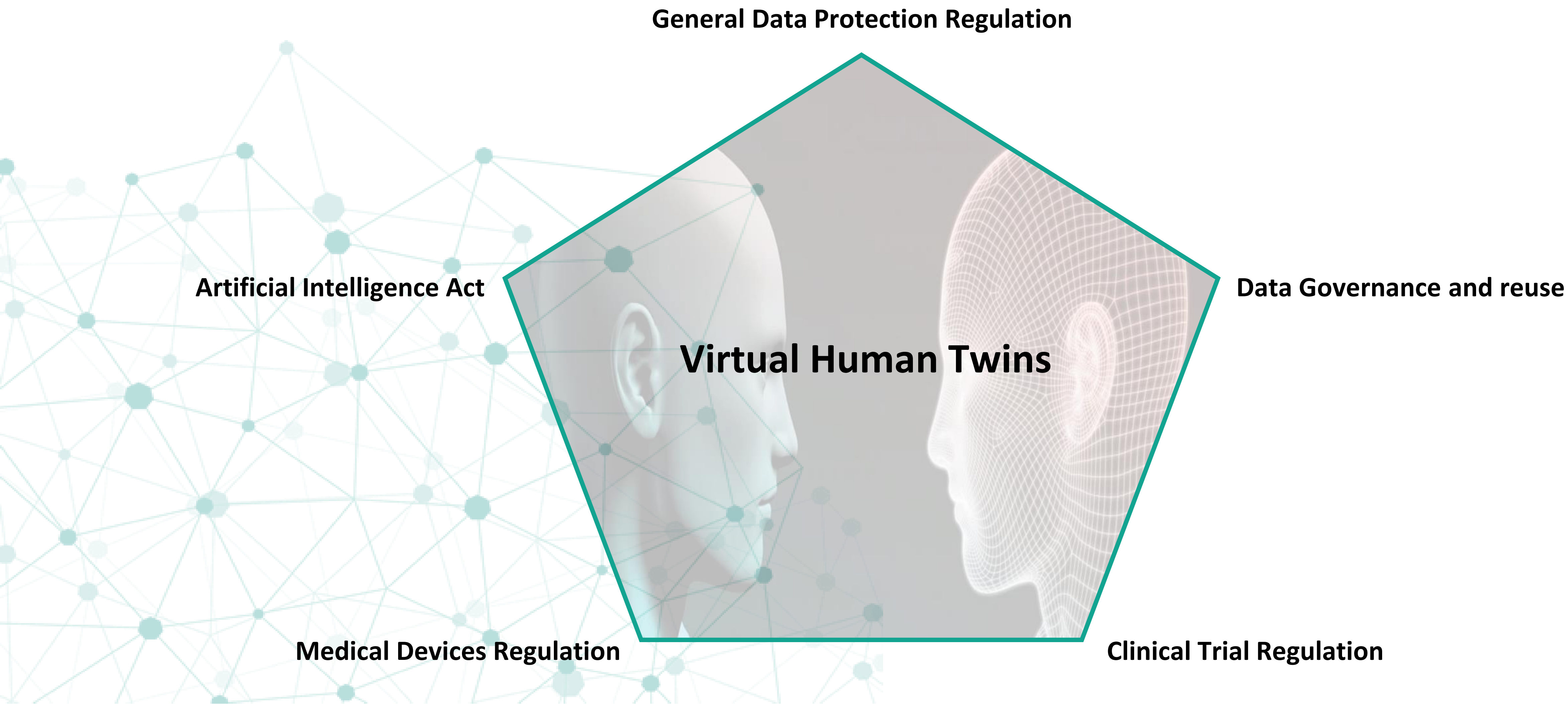
4. *Can specific Privacy-Enhancing Technologies help ensuring safe and compliant processing for the purpose of* in silico *medicine?*
5. *Can clinically reliable VHT be generated thanks to anonymous, pseudonymous or synthetic data?*

Data Privacy and Security regulatory landscape

6. ***Which regulatory recommendations can be made to policy-makers to ensure that the EU will soon be at the global forefront of the VHT sector?***

Legal scenario and regulatory boundaries

The main areas of legislation which are relevant to *In Silico* Medicine and so to Virtual Human Twins (‘VHT’) are to date:



General Data Protection Regulation

Artificial Intelligence Act

Virtual Human Twins

Data Governance and reuse

Medical Devices Regulation

Clinical Trial Regulation

Legal scenario and regulatory boundaries

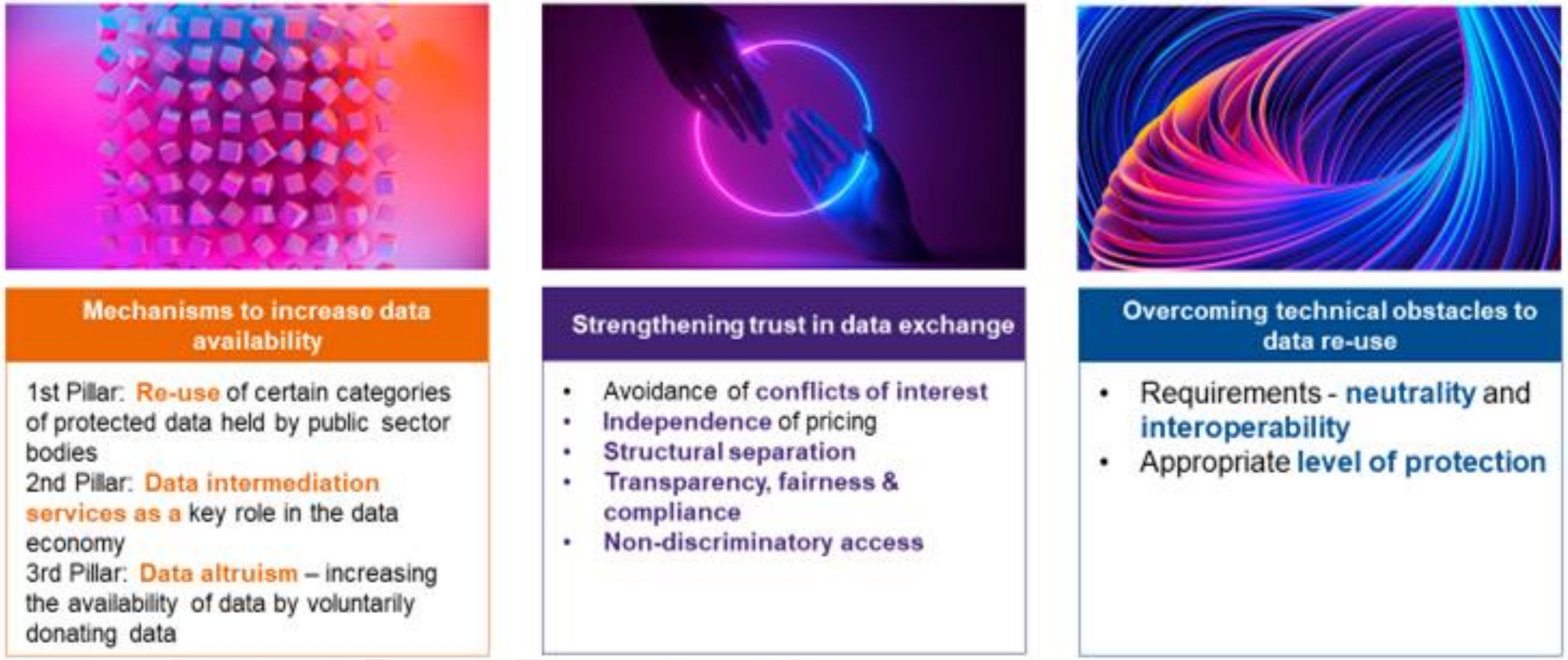
The main areas of legislation which are relevant to *In Silico* Medicine and so to Virtual Human Twins (‘VHT’) are to date:

1. **Privacy and Data Protection:** *in silico* medicine entails by definition the processing of personal data. As a consequence, privacy and data protection legislation (particularly **Regulation (EU) 2016/679, ‘GDPR’**) becomes relevant;
2. **Data Governance and reuse:** in the healthcare and especially *in silico* medicine fields, the reuse and sharing of data across healthcare facilities may be considered a key factor for the development, testing, and validation of reliable models. This is why the upcoming **Data Governance Act, Data Act and European Health Data Space** may become crucial;
3. **Clinical Trials:** the legislation concerning clinical trials, with particular reference to the the **Regulation no. 2014/536** (but including, more generally, also the Declaration of Helsinki, the Declaration of Taipei, Good Clinical Practice Directive) are of the utmost importance for *in silico* trials, even though there is a huge lack of clear regulatory pathways for this sector;
4. **Medical Devices and Medicinal Products:** from a product regulation perspective, *in silico* trials encompass medical devices and medicinal products, for which the **Medical Device Regulation no. 2017/745** and the ***In Vitro* Medical Device Regulation no. 2017/746** are of primary importance (jointly with other other set of rules such as the Community Code for Medicinal Products for Human Use Directive 2009/94, the Advanced Therapeutic Medicinal Product Regulation, and the Health Technology Assessment Regulation);
5. **Artificial Intelligence:** AI is the backbone and pre-condition for *in silico* medicine and trials. The EU was the first, at global level, to attempt to lay down a legal framework for this complex sector, thanks to the **Artificial Intelligence Act**.

Legal scenario and regulatory boundaries – DGA

The next three years will represent a turning point for EU policies on the use of data and related technologies. The Commission is intensively working to build the pillars of the European ‘Strategy for data’ and provide comprehensive and pioneering regulations on related technological areas:

- the **Data Governance Act** (applicable from 24 September 2023) intends to enhance trust in data sharing, strengthen mechanisms to increase data availability and overcome technical obstacles to the reuse of certain categories of data held by public sector bodies.



Mechanisms to increase data availability

1st Pillar: Re-use of certain categories of protected data held by public sector bodies
2nd Pillar: Data intermediation services as a key role in the data economy
3rd Pillar: Data altruism – increasing the availability of data by voluntarily donating data

Strengthening trust in data exchange

- Avoidance of conflicts of interest
- Independence of pricing
- Structural separation
- Transparency, fairness & compliance
- Non-discriminatory access

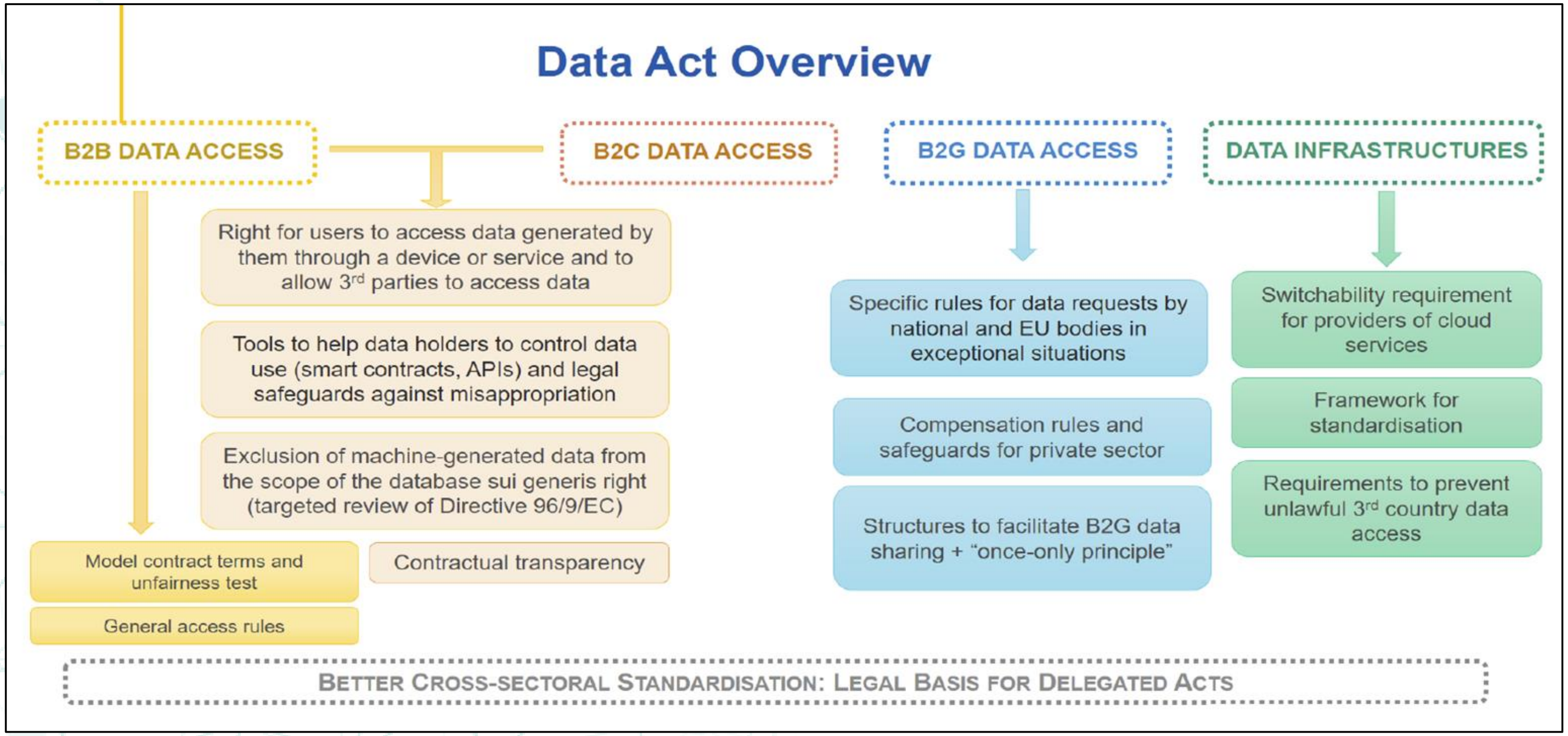
Overcoming technical obstacles to data re-use

- Requirements - neutrality and interoperability
- Appropriate level of protection

Credits: CMS Germany

Legal scenario and regulatory boundaries – Data Act

- The **Data Act** – still under negotiation (on the 23rd of March, the Council of the European Union, under the Swedish presidency, agreed on a negotiating mandate for the upcoming trilogue meetings) – complement the DGA. While the latter creates the processes and structures to facilitate data sharing, the Data Act clarifies who can access and use – and therefore control and benefit from – data, and under what conditions.



Data Act Overview

B2B DATA ACCESS

B2C DATA ACCESS

B2G DATA ACCESS

DATA INFRASTRUCTURES

Right for users to access data generated by them through a device or service and to allow 3rd parties to access data

Tools to help data holders to control data use (smart contracts, APIs) and legal safeguards against misappropriation

Exclusion of machine-generated data from the scope of the database sui generis right (targeted review of Directive 96/9/EC)

Model contract terms and unfairness test

Contractual transparency

General access rules

Specific rules for data requests by national and EU bodies in exceptional situations

Compensation rules and safeguards for private sector

Structures to facilitate B2G data sharing + "once-only principle"

Switchability requirement for providers of cloud services

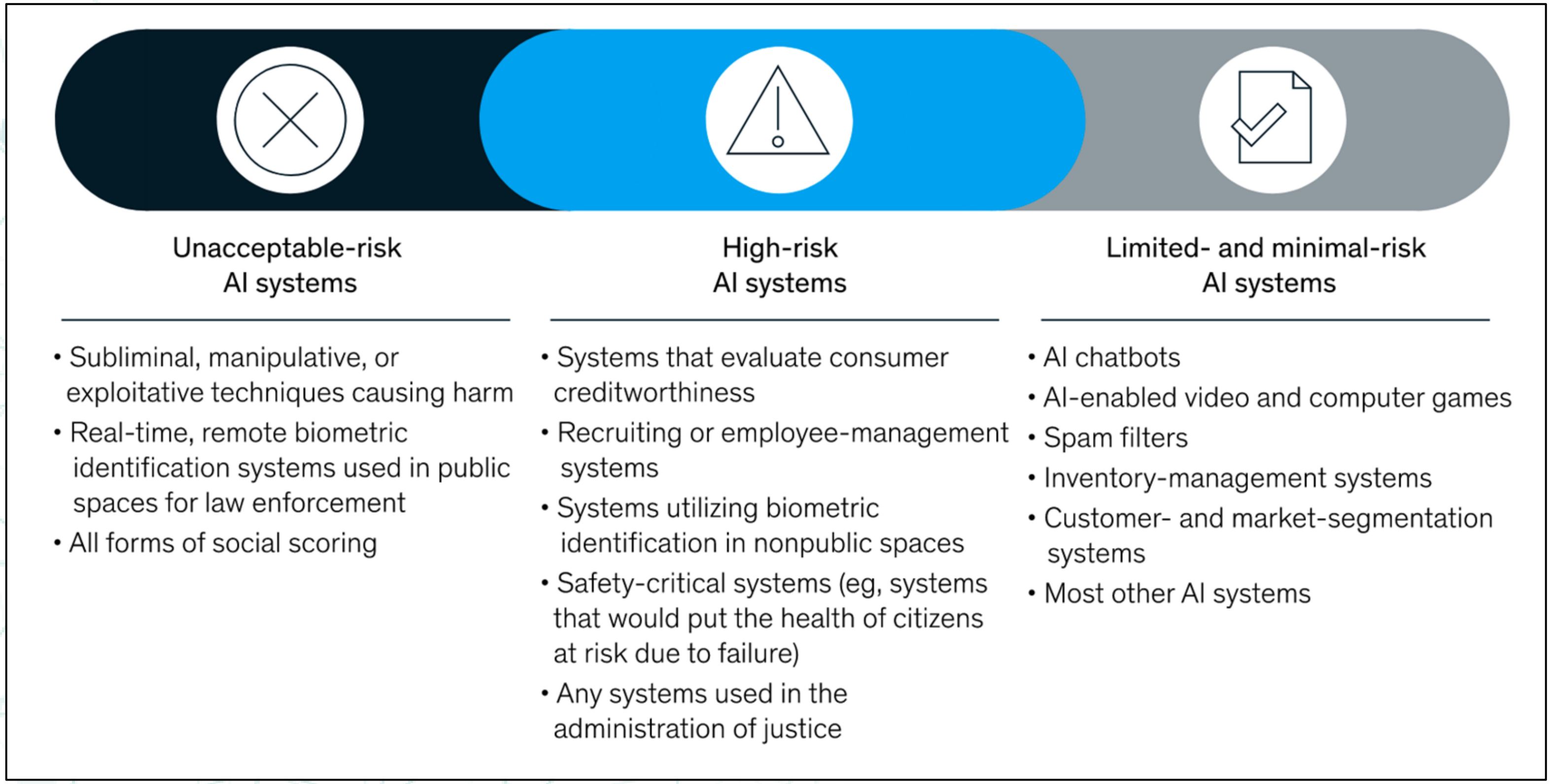
Framework for standardisation

Requirements to prevent unlawful 3rd country data access

BETTER CROSS-SECTORAL STANDARDISATION: LEGAL BASIS FOR DELEGATED ACTS

Legal scenario and regulatory boundaries – AI Act

- The **Artificial Intelligence Act** – still under negotiation (on the 27th of April, MEPs reached a provisional political agreement on an amended version of the EU Commission's draft) – establishes a uniform legal framework to regulate the development, commercialisation and use of any AI systems, in accordance with EU constitutional values and rights, by laying down clear requirements and obligations for relevant developers, deployers and users.



Unacceptable-risk AI systems

- Subliminal, manipulative, or exploitative techniques causing harm
- Real-time, remote biometric identification systems used in public spaces for law enforcement
- All forms of social scoring

High-risk AI systems

- Systems that evaluate consumer creditworthiness
- Recruiting or employee-management systems
- Systems utilizing biometric identification in nonpublic spaces
- Safety-critical systems (eg, systems that would put the health of citizens at risk due to failure)
- Any systems used in the administration of justice

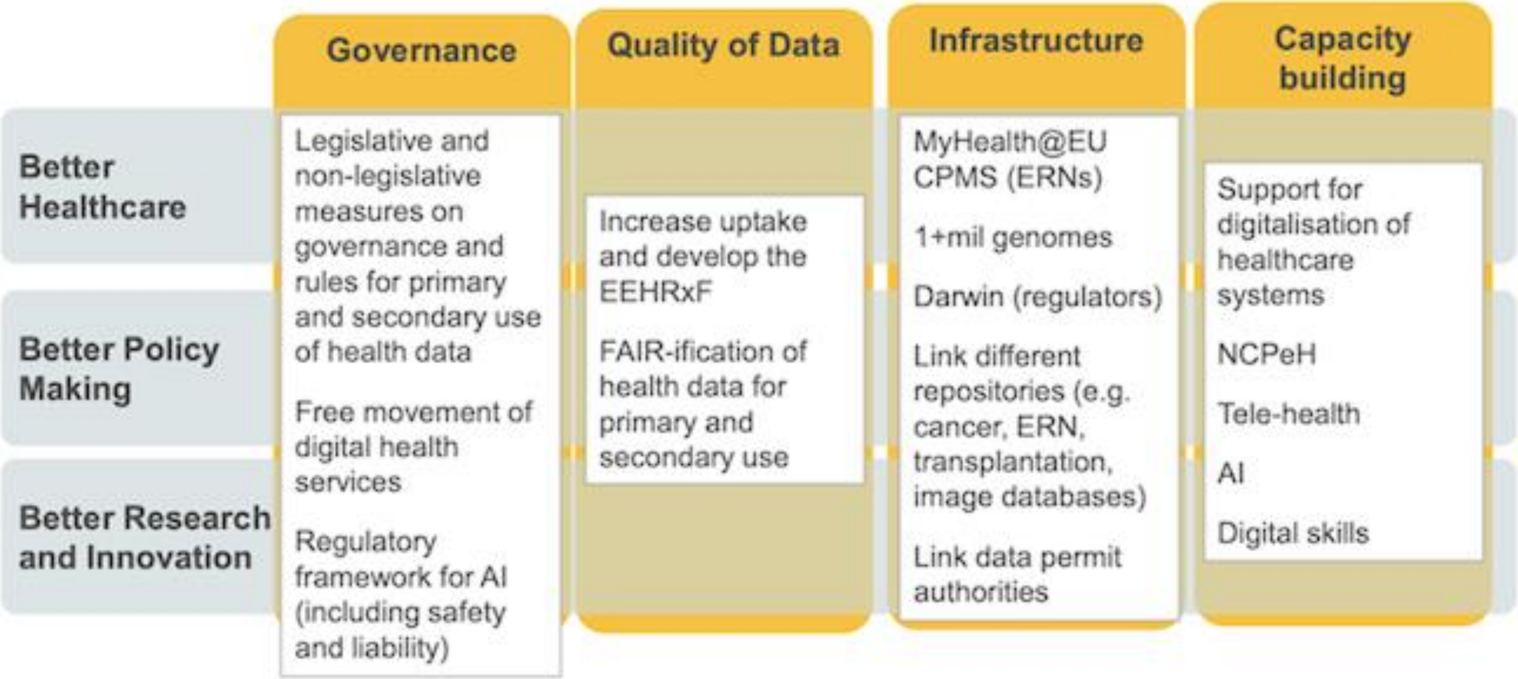
Limited- and minimal-risk AI systems

- AI chatbots
- AI-enabled video and computer games
- Spam filters
- Inventory-management systems
- Customer- and market-segmentation systems
- Most other AI systems

Credits: McKinsey & Company

Legal scenario and regulatory boundaries – EHDS

- The **European Health Data Space** – still under negotiation – establishes rules, common standards and practices, infrastructures and a governance framework with the aim to (i) empower individuals through increased digital access to and control of their electronic personal health data, both at national level and EU-wide, and foster a genuine single market for electronic health record systems, relevant medical devices and high risk AI systems (primary use); and (ii) provide a consistent, trustworthy and efficient set-up for the use of health data for research, innovation, policy-making and regulatory activities (secondary use).



	Governance	Quality of Data	Infrastructure	Capacity building
Better Healthcare Better Policy Making Better Research and Innovation	Legislative and non-legislative measures on governance and rules for primary and secondary use of health data Free movement of digital health services Regulatory framework for AI (including safety and liability)	Increase uptake and develop the EEHRxF FAIR-ification of health data for primary and secondary use	MyHealth@EU CPMS (ERNs) 1+mil genomes Darwin (regulators) Link different repositories (e.g. cancer, ERN, transplantation, image databases) Link data permit authorities	Support for digitalisation of healthcare systems NCPeH Tele-health AI Digital skills

Health data reuse – GDPR

EUROPEAN LEVEL

Article 6.4 GDPR reads that personal data can only be further processed for a purpose other than that identified and communicated to the individuals at the time when their data were collected, if the the secondary use is compatible with the initial purpose. However, when it comes to research (meeting methodological requirements, standards of research integrity and aiming to contribute to the common good), this must be read in conjunction with Article 5(1)(b) GDPR, which carves out a privileged position for this purpose, **stating that further processing for scientific research in accordance with Article 89(1) GDPR is by definition compatible with the original purpose**.

NATIONAL LEVEL

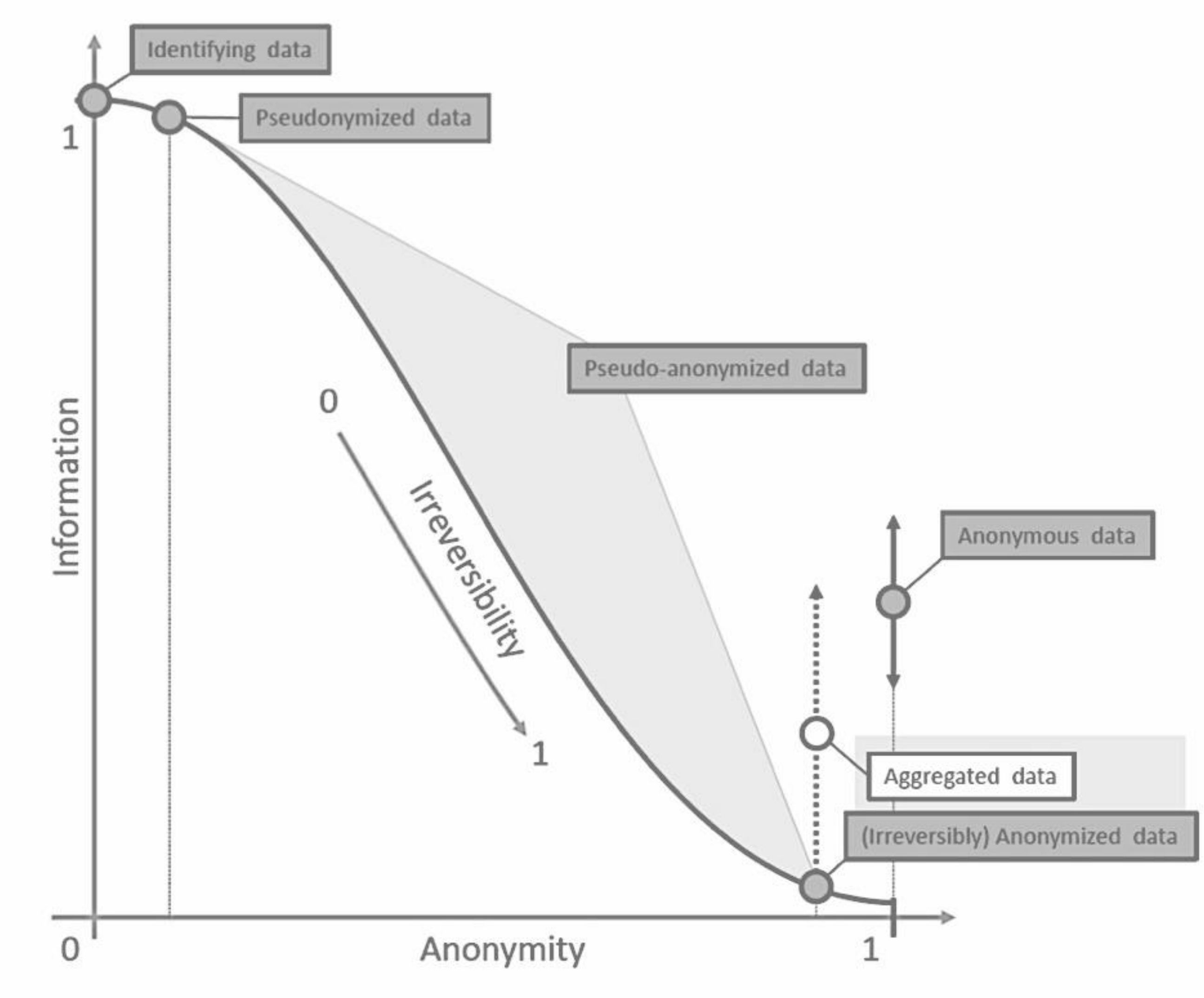
Notwithstanding the above, Art. 9.4 of GDPR provides that member States can "*maintain or introduce further conditions, including limitations, with regard to the processing of (…) data concerning health*". As it was likely, national legislators have not implemented this delegation in a homogenous way, «***resulting in a complex and fragmented landscape for researchers to navigate***. *Consequently, differences between member States in the way the GDPR is implemented and interpreted in the area of scientific research has made data exchange between Member State and EU bodies for research purposes difficult and in some cases highly technical*».

(EU Commission – DG Health and Food Safety, '*Assessment of the EU Member States' rules on health data in the light of GDPR*')

Health data reuse – National scenarios

Member State	Some examples of national regulations on the reuse of health data for research
🇪🇸	**Additional Provision 17, Par. 2, let. c) and d), of *Ley Orgánica* 3/2018**: **re-use of personal data for biomedical research purposes is considered lawful and compatible when, having obtained consent for a specific purpose, the data is used for purposes or areas of research related to the area in which the initial study was scientifically integrated**. In such cases, the person responsible for the specific research must publish a clear and comprehensive Privacy Policy in an easily accessible place on the corporate website of the centre where the research or clinical study is carried out and, where appropriate, on the website of the sponsor, and notify the data subjects of the existence of this information by electronic means. When the subjects do not have the means to access this information, they may request that it be sent in another format.
🇩🇪	***Bundesdatenschutzgesetz* (BDSG)- Section 22 § 2, in conjunction with Section 27 § 1**: by derogation from Article 9.1 GDPR, **special categories of personal data (such as health data) can be processed even in the absence of the data subject's consent for scientific research purposes, if such processing is necessary for these purposes and the interests of the controller in processing substantially outweigh those of the data subject (in not processing the data), so long as appropriate and specific measures are taken to safeguard the rights of the data subjects** (e.g. to constantly track the origin of the data; to train the staff involved and ensure confidentiality; to designate a DPO; to restrict the access to the data; to apply pseudonymisation or encryption; to ensure the ability, confidentiality, integrity, availability and resilience of processing systems and services; to regularly test, assess and evaluate the effectiveness of TOMs adopted).
🇮🇹	**Articles 110 and 110-*bis* of the Italian Privacy Code**: (i) **First-party research**: processing of health data without data subject consent is permitted for scientific research purposes in the medical, biomedical, and epidemiological fields, if (a) **the research is based on a national or EU law or regulation and the controller carries out and the makes publicly available a DPIA**, or (ii) informing the data subjects involves disproportionate effort or is likely to render impossible or seriously impair the achievement of the research purposes, on condition that the controller adopts suitable measures to protect the data subjects' rights and **obtains both the approval of competent Ethics Committee and the Garante's authorisation under Art. 36 GDPR**; (ii) **Third-party research**: the processing of health data for research purposes by third-party research organizations is **subject to the Garante's prior authorisation**, insofar appropriate minimization measures are put in place (e.g. pseudonymisation), when informing the data subjects involves disproportionate effort or is likely to render impossible or seriously impair the achievement of the research purposes.

Health data reuse - Anonymisation



"***There is a significant conceptual gap between legal and mathematical thinking around data privacy***" (*Towards formalizing the GDPR's notion of singling out*, A. Cohen & K. Nissim. 2019).

National Data Protection Authorities «*work on the basis that full anonymity can never be achieved for health-related data while still keeping the data useful for research, others believe anonymity within the meaning of GDPR can be achieved. In the literature, it was noted that anonymous data, to the highest standards without any residual risk for reidentification, may lose their value for nuanced research (…). Similar differences in interpretation also exist with respect to pseudonymisation (Article 4(5) GDPR). The legal definition of pseudonymisation under the GDPR is quite far-ranging (…). As a result, a number of misconceptions have arisen as to its meaning*» (EU Commission – DG Health and Food Safety, '**Assessment of the EU Member States' rules on health data in the light of GDPR**').

Health data reuse – Anonymisation

EDPS/AEPD: 10 Misunderstandings related to anonymisation:

1. “*Pseudonymisation is the same as anonymisation*” - **pseudonymisation is not the same as anonymisation**;
2. “*Encryption is anonymisation*” - **encryption is not anonymisation, but it can be a powerful pseudonymisation tool**;
3. “*Anonymisation of data is always possible*” - **it is not always possible to lower the re-identification risk below a previously defined threshold, whilst retaining a useful dataset for a specific processing**;
4. “*Anonymisation is forever*” - **risks exist that some anonymisation processes could be reverted in the future. Circumstances might change over time and new technical developments and the availability of additional information might compromise previous anonymisation processes**;
5. “*Anonymisation always reduces the probability of re-identification of a dataset to zero*” - **anonymisation process and the way it is implemented have a direct influence on the likelihood of re-identification risks**;

Health data reuse – Anonymisation

6. “*Anonymisation is a binary concept that cannot be measured*” - **it is possible to analyse and measure the degree of anonymisation (e.g. through Differential Privacy)**;
7. “*Anonymisation can be fully automated*” - **automated tools can be used during the anonymisation process, however, given the importance of the context in the overall process assessment, human intervention is needed on a case-by-case basis**;
8. “*Anonymisation makes the data useless*” – **a proper anonymisation process keeps the data functional for a given purpose**;
9. “*Following an anonymisation process that others used successfully will lead our organisation to equivalent results*” **- anonymisation processes need to be tailored to the nature, scope, context and purposes of processing, as well as the risks of varying likelihood and severity for the rights and freedoms of the data subjects**;
10. “*There is no risk and no interest in finding out to whom this data refers to*” **- personal data has a value in itself.Re-identification of an individual could have a serious impact for his rights and freedoms**.

Health data reuse – Anonymisation

TEHDAS

There is no common European interpretation of what constitutes ‘sufficient anonymisation’ to transform personal data to non-personal data. Researchers, stakeholders and policymakers reported a lack of guidance on anonymisation at national and international level as a key barrier to data reuse and sharing. In particular, a lack of clarity between "absolute" and "relative" anonymisation has been identified as a key issue, as well as a lack of guidance in relation to anonymization of medical images, genomic data, longitudinal data and rare diseases.

Negative impacts:

✓ interpretation of applicable methods for anonymisation varies significantly among regional, national and European authorities, causing huge interoperability issues;

✓ risk-averse behaviours due to lack of clarity. Some stakeholders reported that they treat all data as personal data due to this lack of clarity;

✓ stakeholders stated that some countries apply a stricter definition of ‘sufficient anonymisation’ which further limits the sharing of data for research on the basis that the individual could potentially be traced and re-identified, e.g. due to the rarity of their disease;

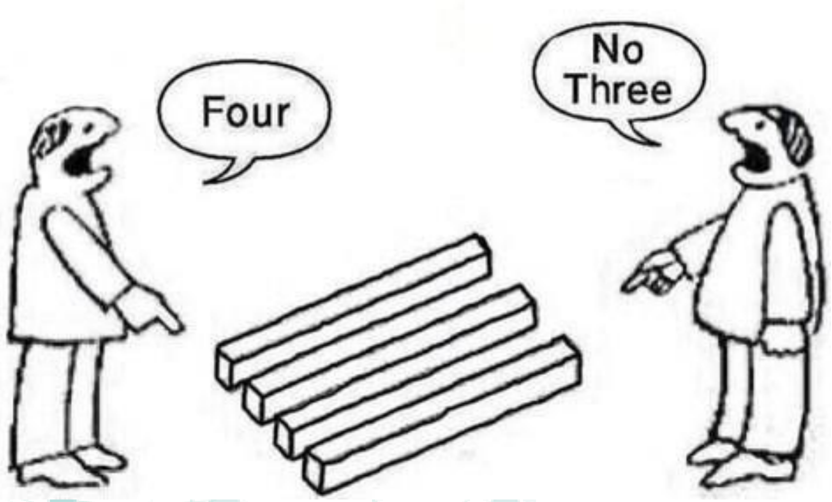
Health data reuse – Anonymisation

- ✓ speed of innovation is reduced or impeded;
- ✓ unclear public communication around health data use;
- ✓ uncertainties about how and if certain types of personal health data can be accessed. Secondary impacts include delays and financial costs;
- ✓ difficulties in following the patient through the health care system when more than one care provider, each with their own interpretation of ‘sufficient anonymisation’, is involved in their care;
- ✓ ‘over anonymisation’ can reduce data quality, usability and reliability to the point that the data could potentially be inaccurate. Over-anonymisation reduces data usability in research as it is often important to do correlation studies where individual data linkage is essential.

Health data reuse – Anonymisation

Is the scenario about to change due to a super recent landmark ruling by Court of Justice of the European Union (dated 26 April 2023, in *Case T-557/20*)?

The Court highlighted that in order to determine whether pseudonymised information transmitted to a data recipient constitutes personal data, it is necessary to consider the data recipient's perspective: **if the data recipient does not have any additional information enabling it to re-identify the data subjects and has no legal means available to access such information, the transmitted data can be considered anonymized** and therefore not personal data. **The fact that the data transmitter has the means to re-identify data subjects is irrelevant** and does not mean that the transmitted data is automatically also personal data for the recipient.



It's always a matter of perspective…

VHT – Requirements for AI applications

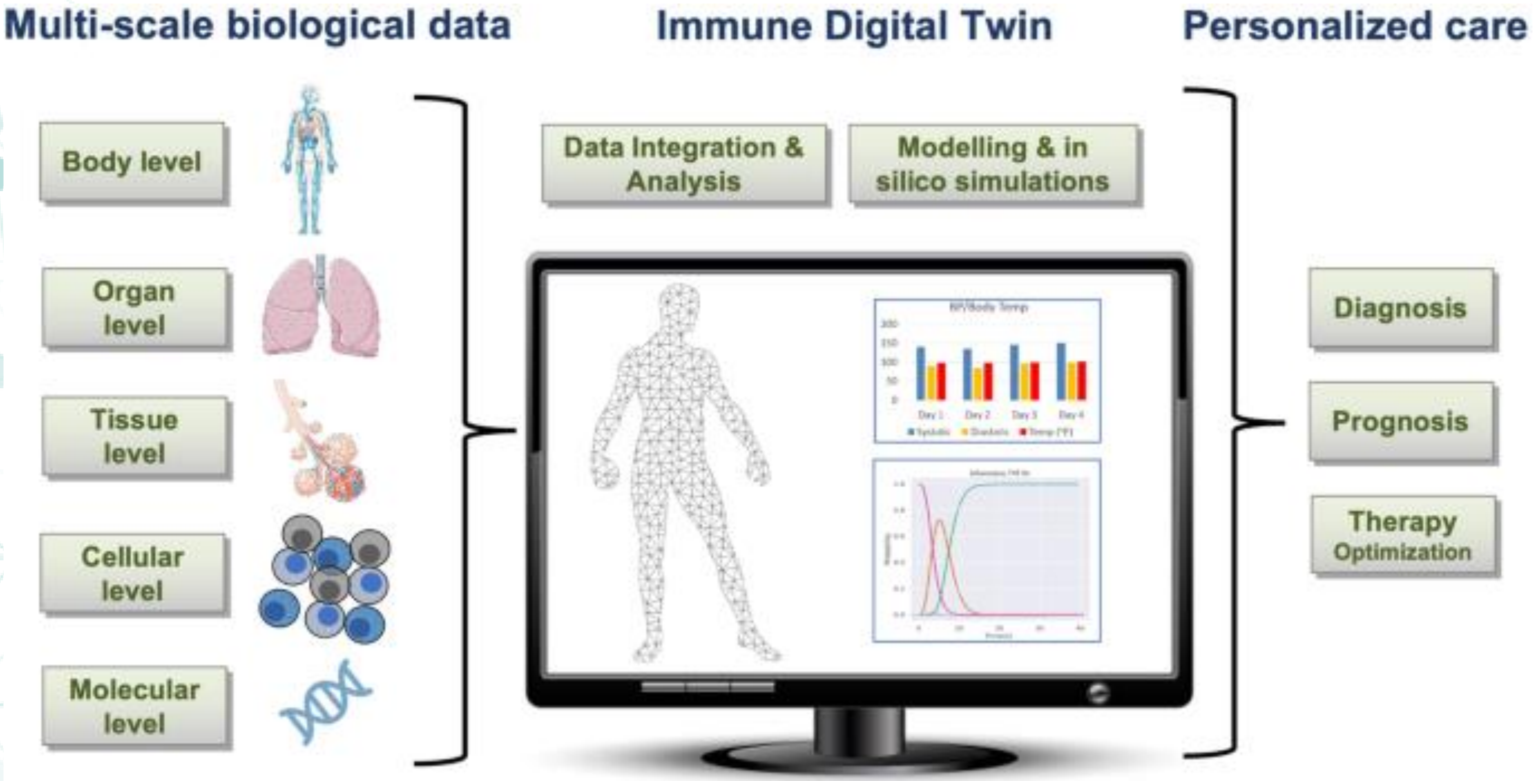
The ‘subject’ of the Artificial Intelligence Act is set out as follows:

> “*The purpose of this Regulation is to* ***promote the uptake of human centric and trustworthy artificial intelligence*** *and to* ***ensure a high level of protection of health, safety, fundamental rights, democracy and rule of law, and the environment from harmful effects of artificial intelligence systems*** *in the Union while supporting innovation.*
>
> *This Regulation lays down:*
>
> *a) harmonised rules for the placing on the market, the putting into service and the use of artificial intelligence systems in the Union;*
>
> ***b) prohibitions of certain artificial intelligence practices****;*
>
> ***c) specific requirements for high-risk AI systems*** *and obligations for operators of such systems;*
>
> *d) harmonised transparency rules for certain AI systems;*
>
> *e) rules on market monitoring, market surveillance, governance and enforcement;*
>
> *f) measures to support innovation, with a particular focus on SMEs and start-ups, including on setting up regulatory sandboxes and targeted measures to reduce the regulatory burden on SMEs and start-ups;*
>
> *g) rule for the establishment and functioning of the European Union artificial intelligence Office*”.

VHT – Requirements for AI applications

VHT will always configure ‘high-risk systems’ under the AI Act, whenever they will qualify as *Medical Devices* (or SaMD, *Software as Medical Devices*).

In other cases, a case-by-case assessment must be carried out to ascertain whether the VHT amount or not to an AI high-risk system.



VHT – Requirements for AI applications

Medical Device Regulation - Rule 11

Software intended to provide information which is used to take decisions with diagnosis or therapeutic purposes is classified as class IIa, except if such decisions have an impact that may cause: (i) death or an irreversible deterioration of a person's state of health, in which case it is in class III; or (ii) a serious deterioration of a person's state of health or a surgical intervention, in which case it is classified as class IIb.

Software intended to monitor physiological processes is classified as class IIa, except if it is intended for monitoring of vital physiological parameters, where the nature of variations of those parameters is such that it could result in immediate danger to the patient, in which case it is classified as class IIb. All other software is classified as class I.

MedTech Europe
from diagnosis
to cure

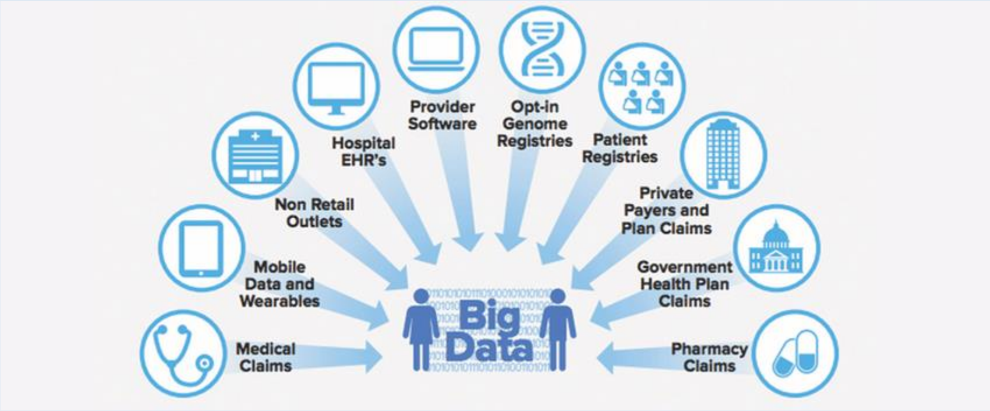
«Medical technologies can be assigned to a range of risk classes under the MDR/IVDR. While the AIA is not meant to change sectoral classification, it would put most cases of AI in/as a medical technology in the highest risk class under the Act. This is a clear deviation from how medical technologies are regulated in Europe and around the world. It would drive confusion among regulators and manufacturers and would create additional, unnecessary complexity in the regulatory approval process that could hinder highly innovative technology to reach citizens in a timely manner».

VHT – Requirements for AI applications

Art. 10.3 AI Act:

Training datasets, and where they are used, validation and testing datasets must be:

- **relevant;**
- **sufficiently representative;**
- **appropriately vetted for errors;**
- **as complete as possible in view of the intended purpose**.

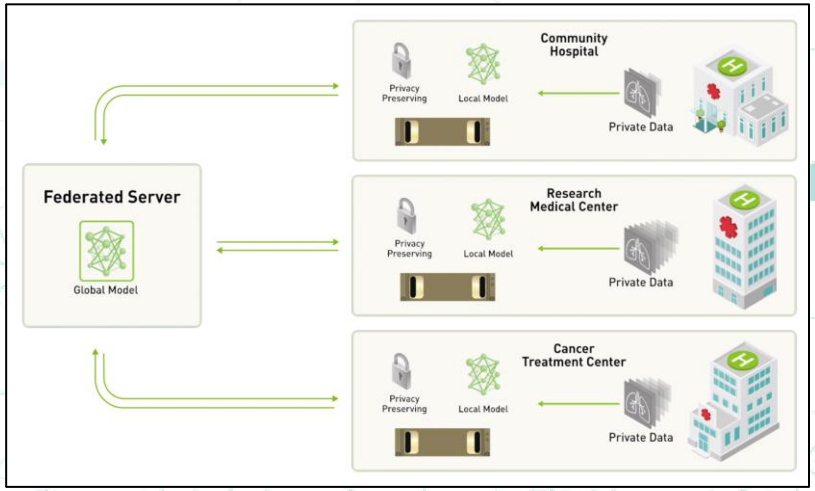


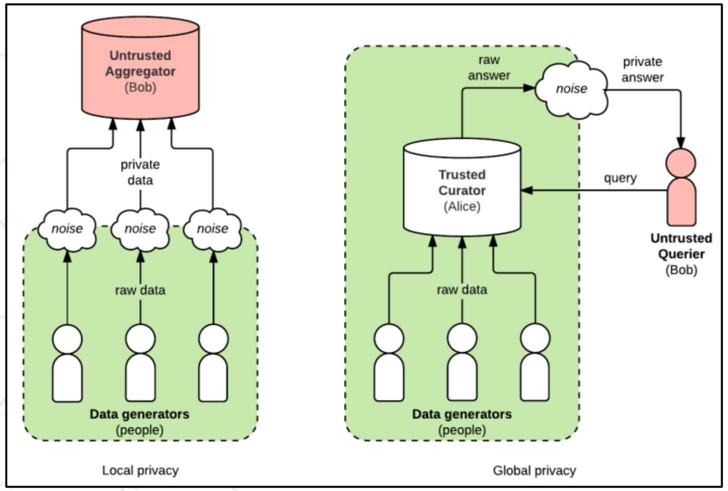
Such datasets must have the **appropriate statistical properties**, including, where applicable, as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used. These characteristics of the datasets shall be met at the level of individual datasets or a combination thereof.

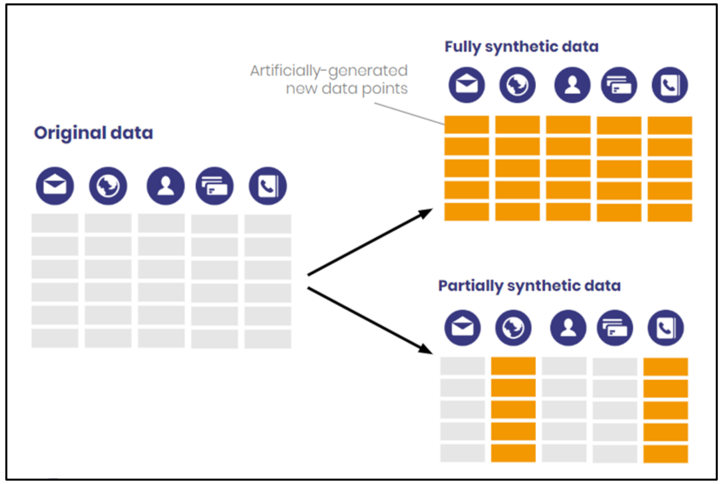
It may be vey complex for stakeholders to find a suitable legal basis (under Art. 6 GDPR) and, when special categories of data are involved, also a valid condition (according to Art. 9 GDPR) to collect and process big data from different appropriate sources, to ensure that the above characteristics are satisfied.

Privacy-Enhancing Technologies fostering VHT

All the hurdles and negative effects – also in terms of competitive disadvantages between member States – deriving from the lack of uniform approach in the EU regarding reuse of personal data in the healthcare sector and application of anonymization must be addressed by policy-makers, by examining more in depth the **benefits and opportunities of Privacy Enhancing Technologies, with particular reference to *Federated Learning*, *Differential Privacy* and *Data Synthesis*, also combined together**.







FEDERATED LEARNING

To implement data minimization and security

DIFFERENTIAL PRIVACY

To prevent re-identification and reduce the risk of data breach

DATA SYNTHESIS

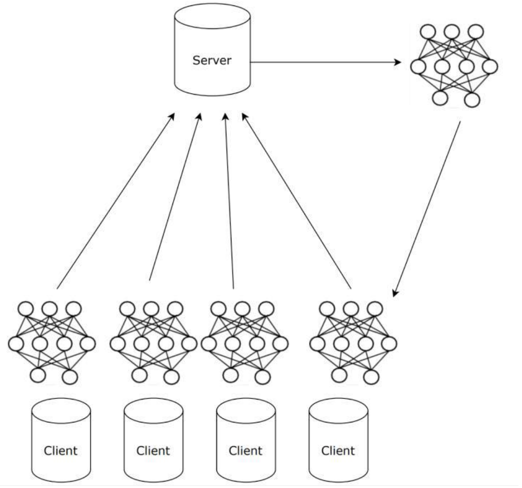
To enhance minimization and safely train AI models and reduce bias without need for big data

Privacy-Enhancing Technologies fostering VHT

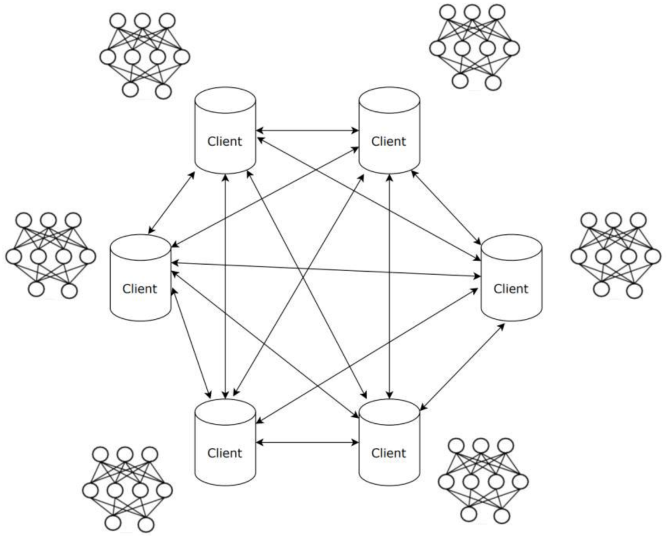
The European Data Protection Supervisor regularly publishes the so-called ‘*TechSonar*’ aiming to anticipate and analyse, from a data protection perspective, the most promising emerging technologies. Among the 5 ‘*Forseen Trends*’ explored by the EDPS (together with Metaverse, Fake News Detection and Central Bank digital currency), there are **Synthetic Data** and **Federated Learning**. For each of this PET, pros and cons are identified.

Federated learning is a technique which allows multiple different parties to train AI models on their own data (‘local’ models). They then combine some of the patterns that those models have identified (known as ‘gradients’) into a single, more accurate ‘global’ model, without having to share any training data with each other.

In **centralised FL**, a co-ordination server creates a model or algorithm, and duplicate versions of that model are sent out to each distributed data source. The duplicate model trains itself on each local data source and sends back the analysis it generates. That analysis is synthesised with the analysis from other data sources and integrated into the centralised model by the co-ordination server. This process repeats itself to constantly refine and improve the model.



Privacy-Enhancing Technologies fostering VHT



In **decentralised FL**, there is no central co-ordination server involved. Each participating entity communicates with each other, and they can all update the global model directly. The decentralised design has some advantages since processing on one server may bring potential security risks or unfairness and there is no single point of failure.

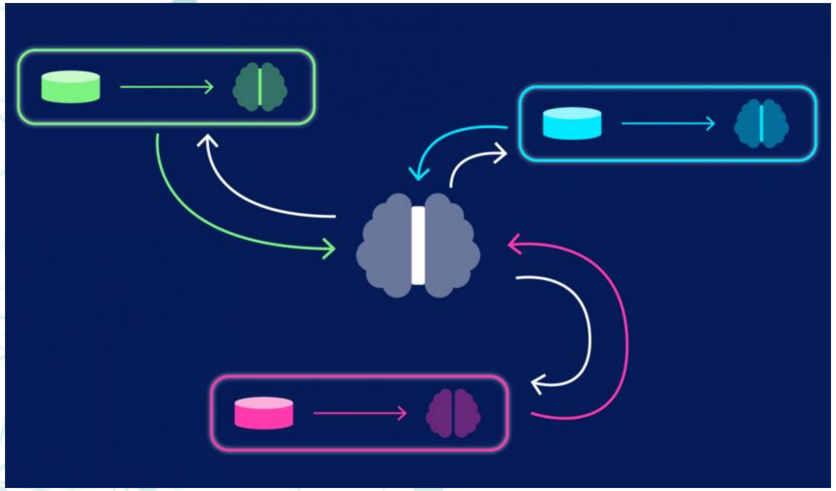
POSITIVE IMPACTS (EDPS)

- **Decentralisation**: by leveraging on distributed datasets, federated learning avoids data centralisation and allows the parties to have better control over the processing of their personal data;
- **Data minimisation**: federated learning reduces the amount of personal data transferred and processed by third parties for machine-learning model training;
- **International cooperation**: when the shared parameters are anonymous, federated learning facilitates the training of models with data coming from different jurisdictions.

Privacy-Enhancing Technologies fostering VHT

NEGATIVE IMPACTS (EDPS)

- **Interpretability**: machine-learning developers often rely on the analysis of the training dataset to interpret the model behaviour. The developers using federated learning do not have access to the full training dataset, which can reduce the models' interpretability.
- **Fairness**: some federated learning settings may facilitate bias toward some parties, for example towards devices hosting the most common model types;
- **Security issues**: the distributed nature of federated learning facilitates some types of attacks (e.g. model poisoning). Classic defence mechanisms do not currently provide sufficient protection in a federated learning setup. Ad hoc defence methods still have to be developed and tested.



Privacy-Enhancing Technologies fostering VHT

Synthetic data are data generated by Artificial Intelligence with the aim of reproducing the statistical properties of an original dataset. This is made by learning relevant distributions of real data using a generative model, for then mimicking and sampling them to produce realistic, but totally fake dataset having the very same statistical properties of the original one, so enhancing the protection of personal data and individuals' privacy, while maintaining the intrinsic utility of the novel dataset for statistical analysis and medical research.

There are two main types of synthetic data:

- ✓ 'partially' synthetic data, for which only some variables of the original data are synthesised;
- ✓ 'fully' synthetic data, where all initial variables are synthesised.

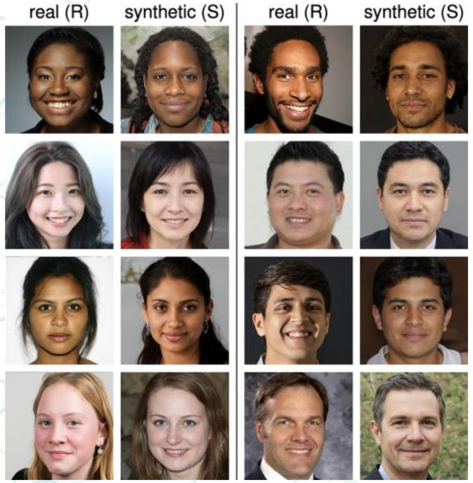
POSITIVE IMPACTS (EDPS)

- **Enhancing privacy in technologies:** from a data protection by design approach, this technology could provide, upon a privacy assurance assessment, an added value for the privacy of individuals, whose personal data does not have to be disclosed;
- **Improved fairness:** synthetic data might contribute to mitigate bias by using fair synthetic datasets to train artificial intelligence models. These datasets are manipulated to have a better representativeness of the world (to be less as it is, and more as society would like it to be), For instance, without gender-based or racial discrimination.

Privacy-Enhancing Technologies fostering VHT

NEGATIVE IMPACTS (EDPS)

- **Output control could be complex:** especially in complex datasets, the best way to ensure the output is accurate and consistent is by comparing synthetic data with original data, or human-annotated data. However, for this comparison again access to the original data is required;
- **Difficulty to map outliers**: synthetic data can only mimic real-world data; it is not a replica. Therefore, synthetic data may not cover some outliers that original data has. However, outliers in the data can be more important than regular data points for some applications;
- **Quality of the model depends on the data source**: the quality of synthetic data is highly correlated with the quality of the original data and the data generation model. Synthetic data may reflect the biases in original data. Also, the manipulation of datasets to create fair synthetic datasets might result in inaccurate data.



Prospective recommendations to policy-makers

The purposes for which data reuse is admitted according to the EHDS are:

- activities for reasons of public interest in the area of public and occupational health, such as protection against serious cross-border threats to health, **public health surveillance** or **ensuring high levels of quality and safety of healthcare and of medicinal products or medical devices**;
- to support public sector bodies or Union institutions, agencies and bodies including regulatory authorities in the health or care sector, to carry out their tasks defined in their mandates;
- to produce national, multi-national and Union level official statistics related to health or care sectors;
- education or teaching activities in health or care sectors;
- **scientific research related to health or care sectors**;
- development and **innovation** activities for products or services **contributing to public health or social security, or ensuring high levels of quality and safety of health care, of medicinal products or of medical devices**;
- **training, testing and evaluating of algorithms, including in medical devices, AI systems and digital health applications, contributing to the public health or social security, or ensuring high levels of quality and safety of health care, of medicinal products or of medical devices**;
- providing **personalised healthcare** consisting in assessing, maintaining or restoring the state of health of natural persons, based on the health data of other natural persons.

Prospective recommendations to policy-makers

✓ **Data sharing by Data holder**: (Recital 37) the EHDS establishes an obligation for data holders to make the electronic health data they have collected available to data access bodies (Art. 6.1, c) GDPR) and provides for the conditions to derogate the prohibition to process health data in accordance with Articles 9(2) (h),(i),(j) GDPR;

✓ **Data permit issued by Data access bodies**: (Art. 33.5 EHDS) «*Where the consent of the natural person is required by national law, health data access bodies shall rely on the obligations laid down in this Chapter to provide access to electronic health data*»;

✓ **Data access by Data user**: (Art. 45) data applicants must demonstrate that both an appropriate legal ground (Art. 6 GDPR) and a valid condition exist (Art. 9.2 GDPR) which permit them to lawfully access and process the data for one or more of the admitted purposes.

The EDPS and EDPB question how the above «*may be reconciled with Article 9(4) GDPR and the possibility for Member State law to introduce further conditions, including limitations with regard to the processing of genetic data, biometric data or data concerning health*» (Par. 89 of the EDPB-EDPS 'Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space' adopted on 12 July 2022).

Prospective recommendations to policy-makers

A possible solution?

In the attempt to foster clinical research and *in silico* medicine, the whole triangulation (*i.e.* all data flows) between (i) Data Holders, (ii) Data Access Bodies and (iii) Data Users (in their role as Data applicant) should rely on the legal framework established by the EHDS – also in connection with Data Act – **by derogation of Art. 9.4 GDPR**, meaning that any eventual limitation or stricter condition for accessing and reusing health data laid down by member States at national level must not apply, being overridden by the obligations posed by the EHDS (and the Data Act).

Prospective recommendations to policy-makers

«*The EDPB and the EDPS note a lack of proper delineation of the purposes*» for which the electronic health data can be reused (as listed under Article 34 EHDS) «*and in particular express concern with regards to Articles 34(1)(f) and (g) of the Proposal, which possibly encompass any form of 'development and innovation activities for products or services contributing to public health or social security' or 'training, testing and evaluation of algorithms, including in medical devices, AI systems and digital health applications, contributing to public health or social security'. The EDPB and the EDPS strongly recommend for the Proposal to further delineate these purposes and circumscribe when there is a sufficient connection with public health and/or social security*» (Par. 85 of the EDPB-EDPS 'Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space' adopted on 12 July 2022).

The purposes for which data reuse is admitted under Art. 34 EHDS are:

e) *scientific research related to health or care sectors, including (…) in silico research, medicine and trials;*

f) *development and innovation activities for products or services contributing to public health or social security, or ensuring high levels of quality and safety of health care, of medicinal products or of medical devices, including by means of in silico medicine, research and trials;*

g) *training, testing and evaluating of algorithms, including in medical devices and for in silico research and medicine, AI systems and digital health applications, contributing to the public health or social security, or ensuring high levels of quality and safety of health care, of medicinal products or of medical devices.*

EDITH

Thanks for your attention!