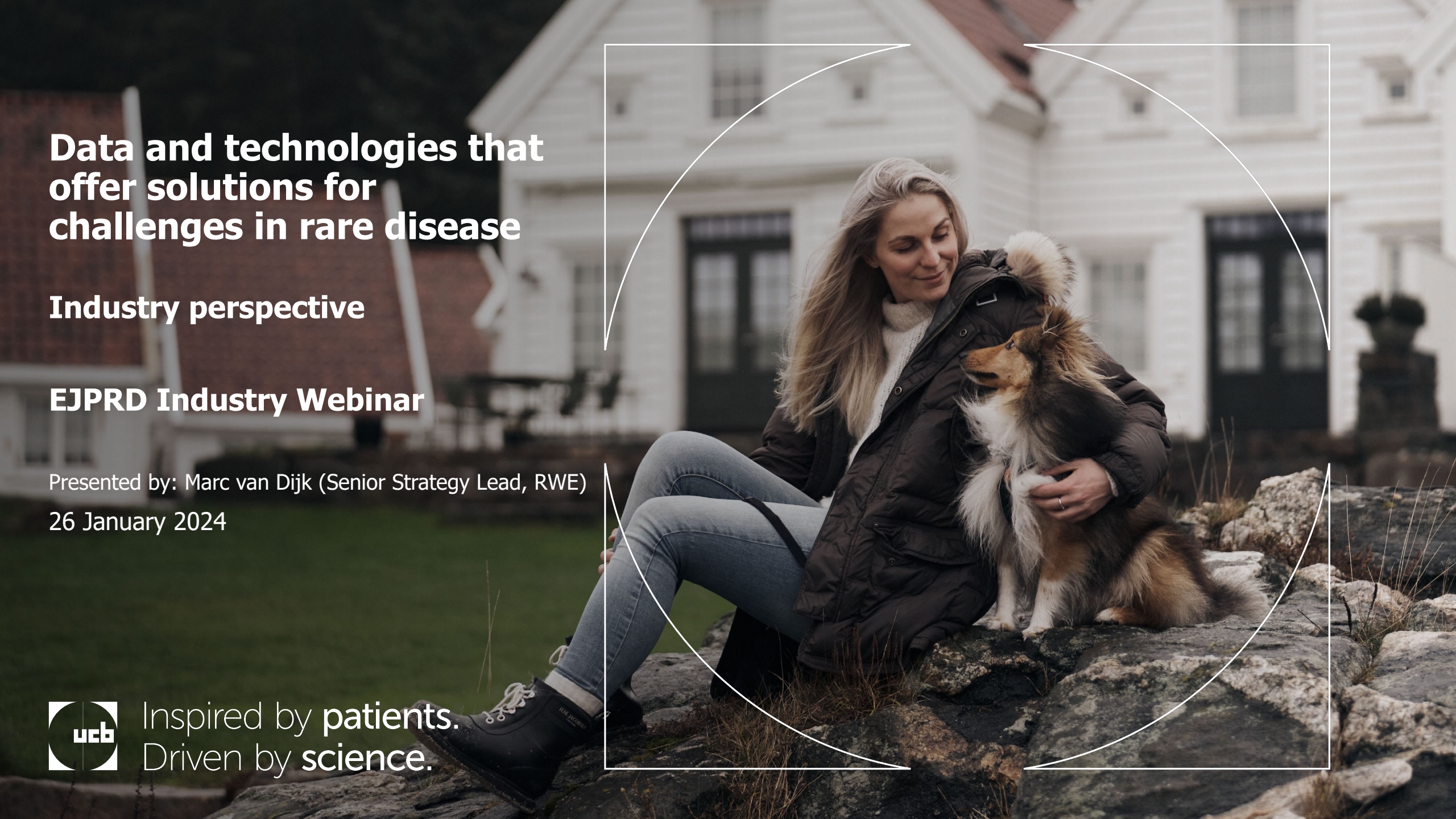
Data and technologies that offer solutions for challenges in rare disease

Industry perspective

EJPRD Industry Webinar

Presented by: Marc van Dijk (Senior Strategy Lead, RWE)

26 January 2024



Inspired by **patients.**
Driven by **science.**

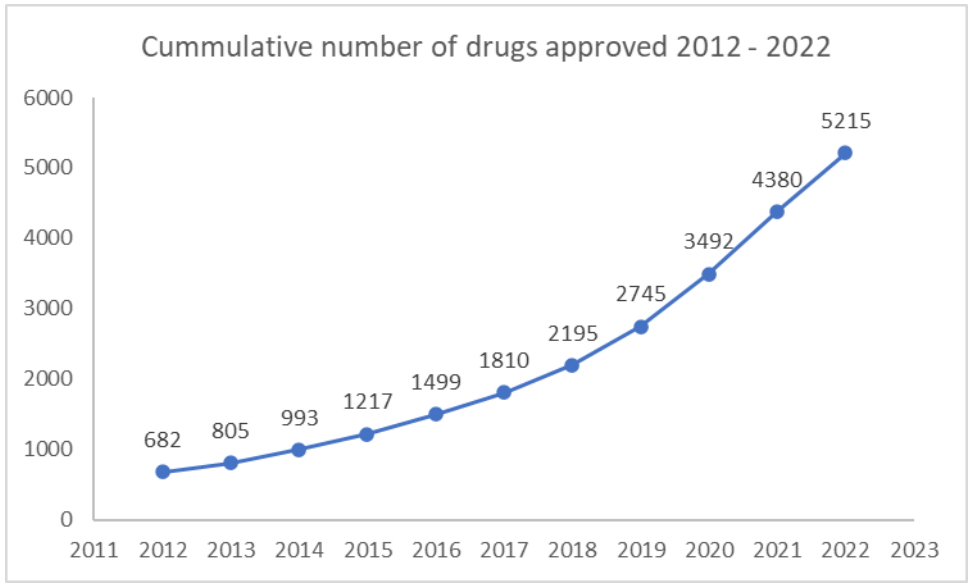
Disclaimer:

The opinions expressed in the presentation and on the following slides are solely those of the presenter and do not necessarily reflect those of UCB or any other organizations.

Marc van Dijk is an employee of UCB and hold its stock and shares.

Approved drugs for rare diseases

- There are 7000 rare diseases affecting 300 million people
- There has been a steady increase in the availability of drugs approved for rare diseases globally*



- However, for more than 90% of rare diseases, there is still no approved treatment

* = Pharmaprojects database from Citeline

The current rate of drug approvals for rare diseases is sub-optimal

Why?

Challenges with drug development rare diseases	Issue	Major contributing factor
1) Scientific validation of therapeutic targets[1]	Limited disease insights	Real world data gap
2) Identifying clinically relevant outcomes[2]		
3) Estimating the prevalence[3] & identifying patients to participate in trials		
4) Creating external comparator arms* for clinical trials	Difficulty to match patients & outcomes	

Examples:
1)Alexion's drug development program for Wilson's disease: mechanistic trials were unable to show that copper was eliminated from the body.
2)Biogen's drug development program for X-linked retinitis pigmentosa: for 2 drugs, phase II/III trials failed to meet the primary endpoint.
3)Based on poor epidemiological data, Alexion projected sales of a drug for Lysosomal acid lipase deficiency at $2 billion per year. They subsequently had to revise this down to $200 million per year

The rate of rare disease drug approvals may decline

Why?

During the past years, there have been several examples of successful utilization of real world data for drug approval in rare diseases however, most of these were for "low hanging fruit" diseases:

- Diseases where it is easier to find patients:
  - Genetic diseases with mendelian or x-linked inheritance (e.g. dursulfase for Hunter Syndrome)
  - Rare as opposed to ultra-rare diseases
- Diseases where the relation between drug target and outcome has been clearly established:
  - With a clear therapeutic target in oncology / haematology / metabolic (e.g. Crizotinib kinase inhibitor for NSCLC; enzyme replacement therapies for Thrombotic Thrombocytopenic Purpura & Pompe's disease)
  - In the area of symptomatic treatments (e.g. ganaxolone a GABA-A receptor modulator for seizures in people with cyclin-dependent kinase-like 5 deficiency disorder)

The next wave of diseases will be more challenging.

For example : ultra-rare, monogenic, de-novo, pediatric, neuro-developmental disorders
(FOXG1 syndrome, Mowat-Wilson Syndrome, PURA syndrome, GNAO1 encephalopathy, SCN2A developmental and epileptic encephalopathy, SLC6A1, STXBP1, MED13L syndrome, GRIN2B-related neurodevelopmental disorder, Snijders Blok Campeau Syndrome, Baraitser-Winter Syndrome, Angelman syndrome, SYNPAG1-related encephalopathy, MDB5 Haploinsufficiency, Wiedemann-Steiner Syndrome, DLG4-related synaptopathy, CDK13-related disorder, Sotos Syndrome, Tuberous Sclerosis)

For these diseases, fewer real-world data, such as disease registries, are available and patients are more difficult to identify. This makes the drug development challenges…

- Scientific validation of therapeutic targets,
- Identifying clinically relevant outcomes
- Estimating the prevalence & identifying patients to participate in a trial

…even greater.

With a disease like Snijders Blok Campeau Syndrome, only 60 cases have been described thus far, world-wide. To collect data and learn about this disease experts will need to wait for a long time……

Rare diseases: at the core of current challenges

Real world data gap

Why?

- Many rare diseases have only been identified more recently. Because they are rare, and clinicians have little exposure, the learning process of these diseases is slow. So, no standardized data collection.
  - Many rare diseases do not have a disease code, so cases with rare diseases cannot be found in key data sources such as Electronic Health Records and Claims Databases
  - For many rare diseases the alternative data source - disease registries – are not available
  - Many of the curated data sets are not suitable for research

- The pathways to a confirmatory diagnosis is a long one, typically taking years
  - Process eliminating other reasons for the health complaints (differential diagnosis) takes time
  - Confirmation is often done by an ultra specialist (of which there may be only a handful in a country)

- Many individuals with a rare disease will never receive the diagnosis
  - Burden for patent/caregiver is high (e.g. travel to highly specialized centers – impact on underserved populations)
  - Presence of dis-incentives for confirmatory testing (e.g. implications positive test result for life insurance)

The data gap and therefore limited disease insights

Consequences for patients

- Not getting the right treatment (diagnosed late, receiving a mis-diagnosis)
- Even with diagnosis:
  - Uncertain future prospects

Consequences for industry R&D

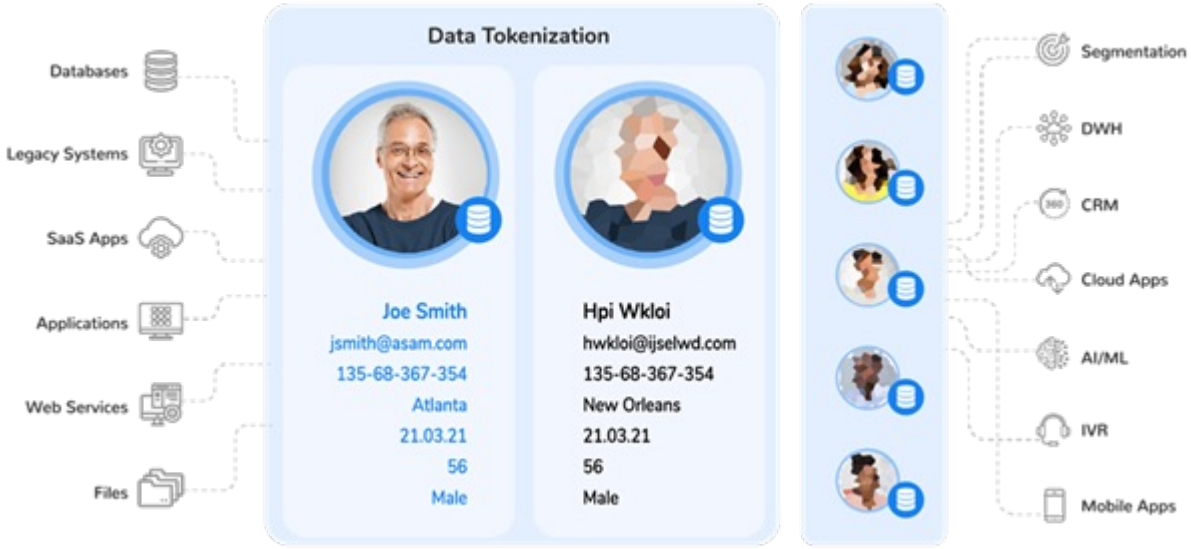
- Difficulty to reach out to patients benefiting most of treatment (e.g. early in disease)
- Difficulty in designing clinical trials
- Difficulty patient recruitment
- Difficulty to create comparative arms

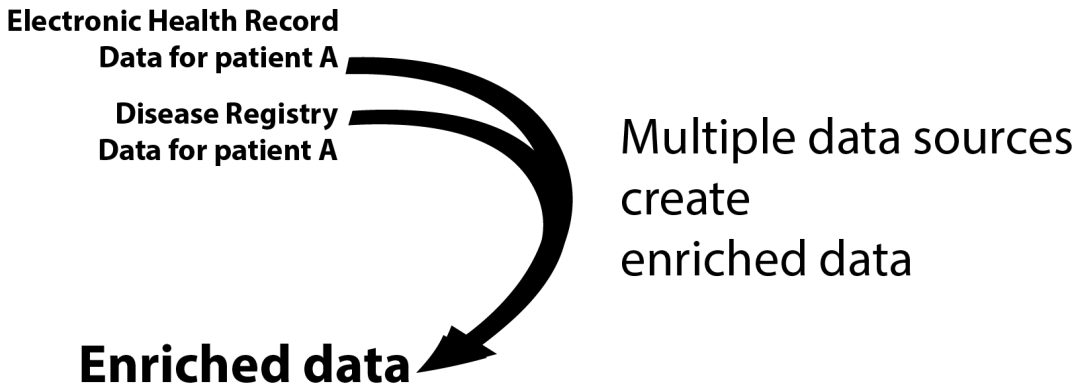
Data and technologies that offer solutions to challenges facing rare diseases

Technologies that offer solutions to the challenges for rare diseases

Tokenization

A process by which patient identifiers are anonymized through generation of a patient specific encrypted "token".





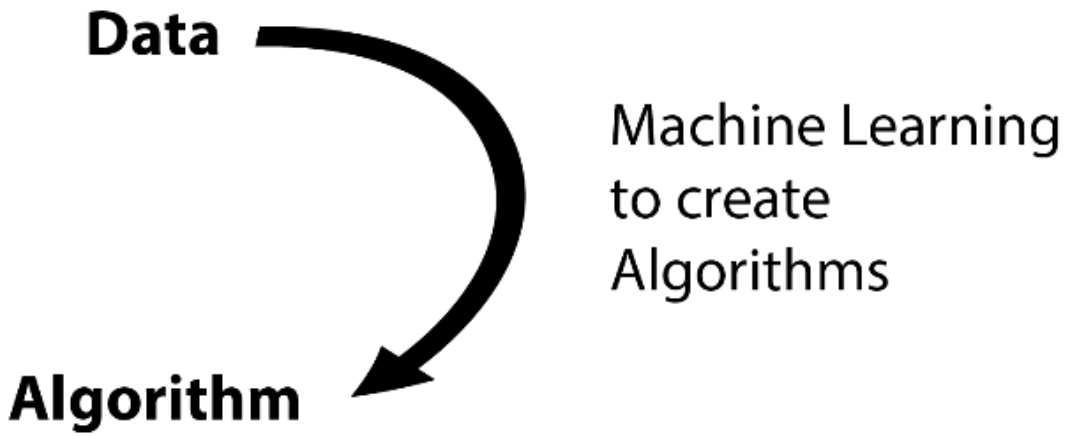
Example:

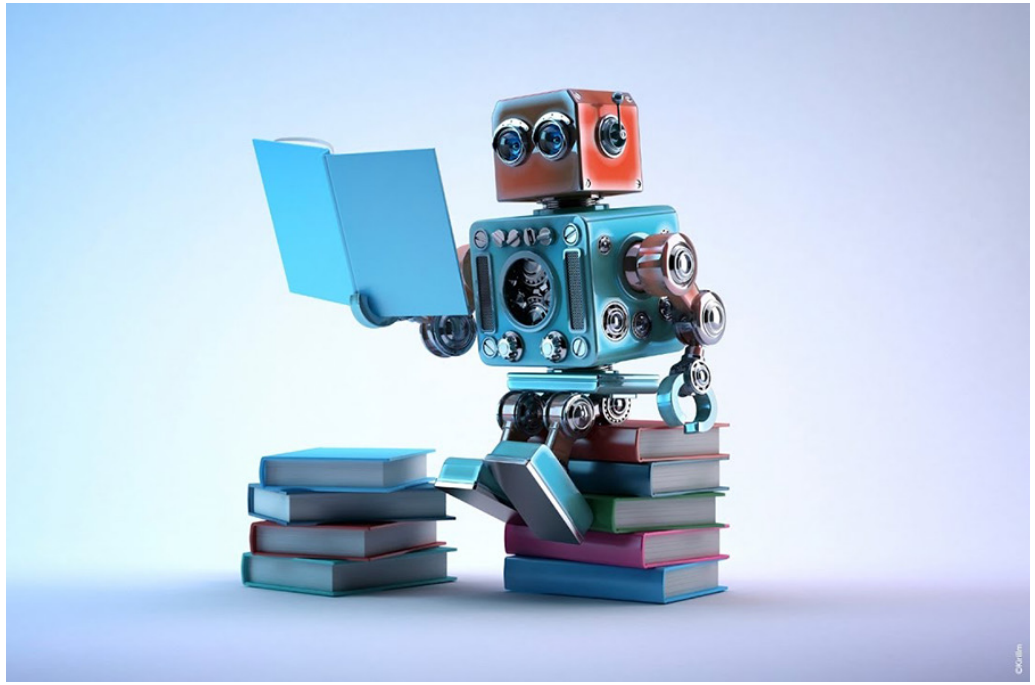
By tokenizing participants in a clinical trial, researchers gain the ability to link real world data from trial participant's medical data to their clinical trial data.

Technologies that offer solutions to the challenges for rare diseases

Machine learning

A field of study in Artificial Intelligence concerned with the development and study of statistical algorithms
that can learn from data,
and then generalize to unseen data



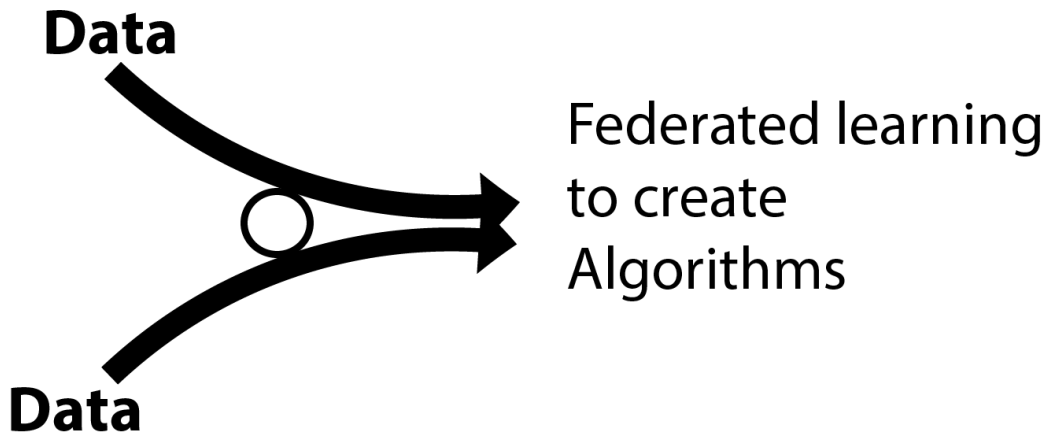


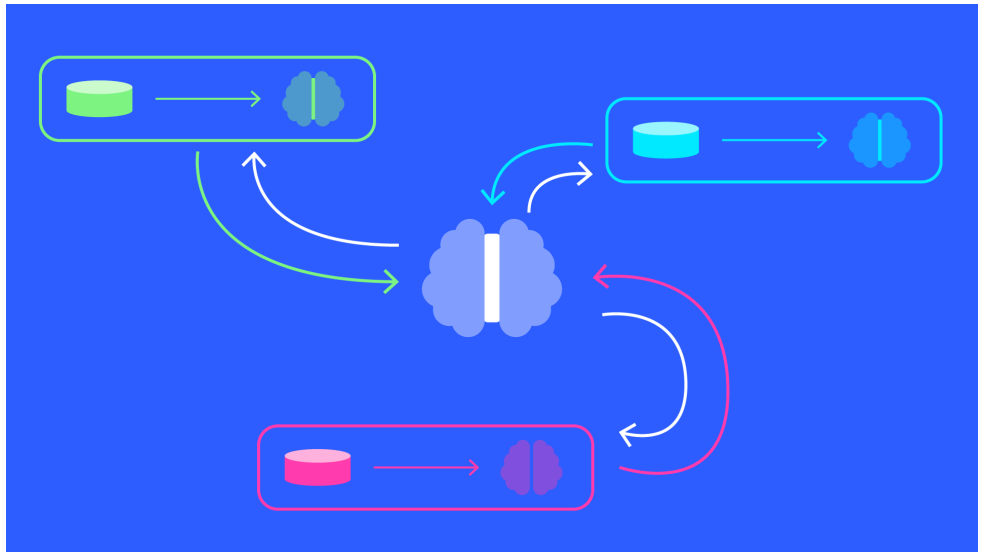
Example:
Machine learning on mammography images has made it possible to detect early cases of breast cancer that doctors would have missed.

Technologies that offer solutions to the challenges for rare diseases

Federated Learning

A machine learning technique that trains an algorithm via multiple independent sessions





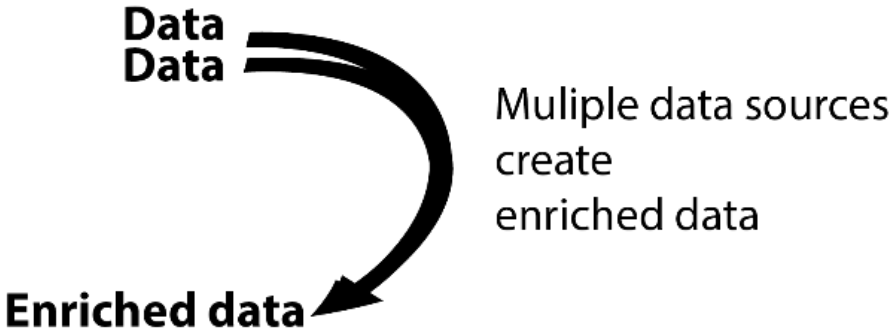
Example:

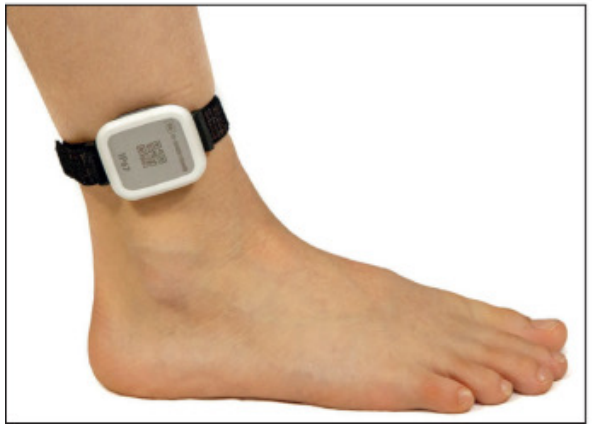
Federated learning across 20 institutes around the globe was used to create a model for oxygen requirements for symptomatic COVID-19 patients. The model was created at speed and was an improvement compared to models created at individual institutes.

Creating data that offer solutions to the challenges for rare diseases

Remote digital health technologies

- Wearables
- Apps
- Home monitoring devices





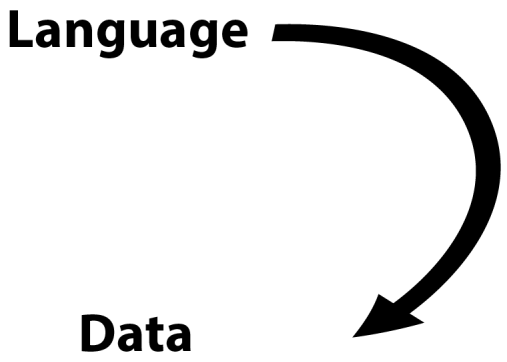
Example:

In 2019, stride velocity 95th centile (SV95C) became the first wearable-derived digital clinical outcome assessment (COA) qualified by the European Medicines Agency (EMA) for use as a secondary endpoint in trials for Duchenne muscular dystrophy (a rare disease).

Creating data that offer solutions to the challenges for rare diseases

Natural Language Processing (NLP)

A branch of artificial intelligence that enables computers to comprehend, generate and manipulate human language.



Natural Language Processing

*"Morphological signs of chronic **hepatopathy**. There is no signs of **portal hypertension**. Focal **liver lesion** in segment 4, which is hyperintense on T2-weighted sequence, shows shine-through effect on diffusion-weighted series and is so intense to liver parenchyma on delayed post contrast acquisitions. It most probably represent a **hemangioma**.*

***Prostate** mildly **enlargement** with dimensions of 47 x 34 x 41 mm. Expansion of central zone secondary to benign adenomatous **prostatic hyperplasia**. In left apical peripheral zone, there is a **nodule** of 12 x 7 x 15 mm, hypointense on T2-weighted sequence and functional data of **malignancy** (reduced ADC value and hyper perfusion in the dynamic series). **Biopsy** is recommended."*

Concept Unique Identifier	Term find at text	Preferred term in UMLS
C0023895	hepatopathy	Liver diseases
C0020541	portal hypertension	Hypertension, Portal
C0577053	liver lesion	Lesion of liver
C0023884	liver	Liver
C0018916	hemangioma	Hemangioma
C0033572	Prostate	Prostate
C1293134	enlargement	Enlargement procedure
C0033572	prostatic	Prostate
C2937421	prostatic hyperplasia	Prostatic Hyperplasia
C0028259	nodule	Nodule
C0006826	malignancy	Malignant Neoplasms
C0005558	Biopsy	Biopsy

Example:

The medical reports of a children with febrile seizures at a pediatric hospital department were used for NLP to identify concepts that help with early diagnosis of Dravet Syndrome.

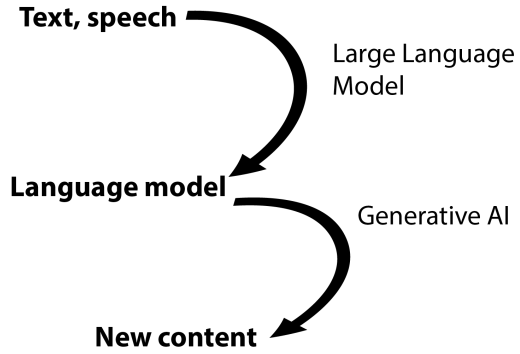
Creating data that offer solutions to the challenges for rare diseases

Large Language Models (including Generative AI)

Large language models (LLMs) are a type of AI system that works with language. A LLM aims to model language, i.e., to create a simplified—but useful—digital representation. The "large" part of the term describes the trend towards training language models with more parameters.

Generative AI is a broad term that can be used for any AI system whose primary function is to generate content. This is in contrast to AI systems that perform other functions, such as classifying data (e.g., assigning labels to images), grouping data (e.g., identifying customer segments with similar purchasing behavior), or choosing actions (e.g., steering an autonomous vehicle).

ChatGPT and Bard are based on a large language model and have generative AI capabilities.



Example:

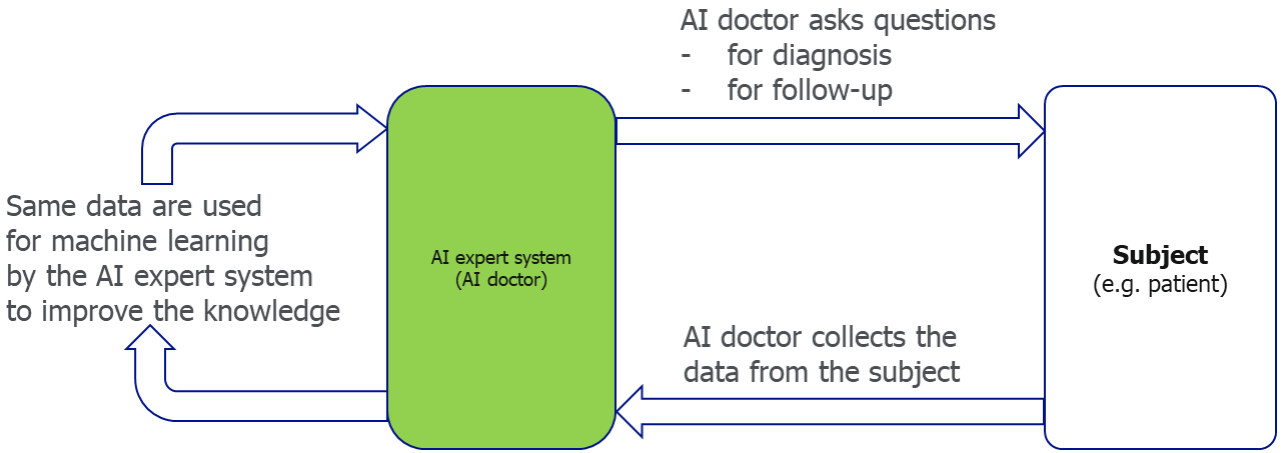
- Google's (AMIE). A chatbot to conduct medical interviews based on LLM. AMIE was more accurate than board-certified primary-care physicians in diagnosing respiratory and cardiovascular conditions

Potential uses

Creating data that offer solutions to the challenges for rare diseases

AI expert systems (AI doctors)

- An AI expert system is a computer system emulating the decision-making ability of a human expert. They are designed to solve complex problems by reasoning through bodies of knowledge and experience.



Example:

AdInfer, an AI doctor, was used to interview a representative sample of citizens in the USA. The data collected and knowledge gained, helped to :

- determine the prevalence of diseases (also identifying cases not yet diagnosed by a doctor)
- revise diagnostic criteria for sleep disorders as described by the International Classification of Sleep Disorders and mood disorders as described by the Diagnostic and Statistical Manual of Mental Disorders.

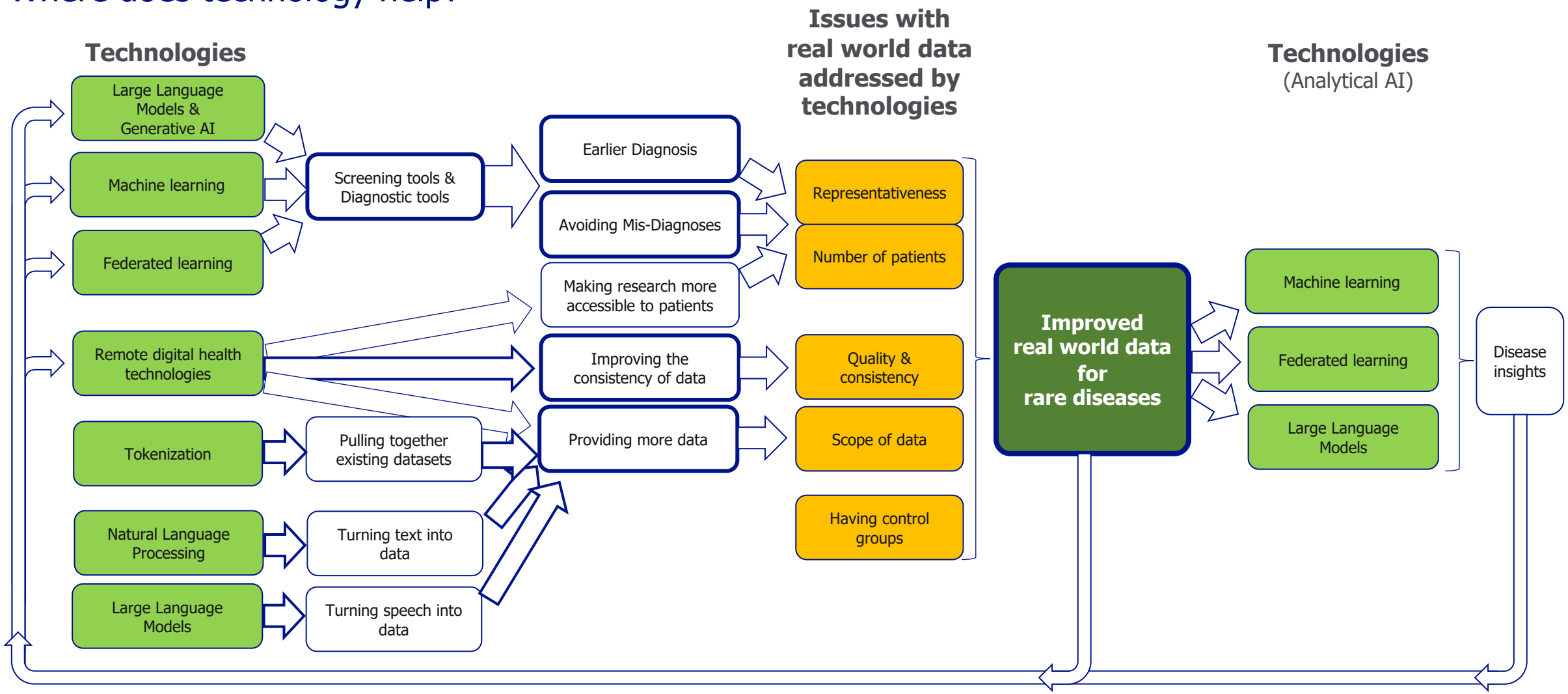
Lack of real world data for rare diseases

What are the issues?

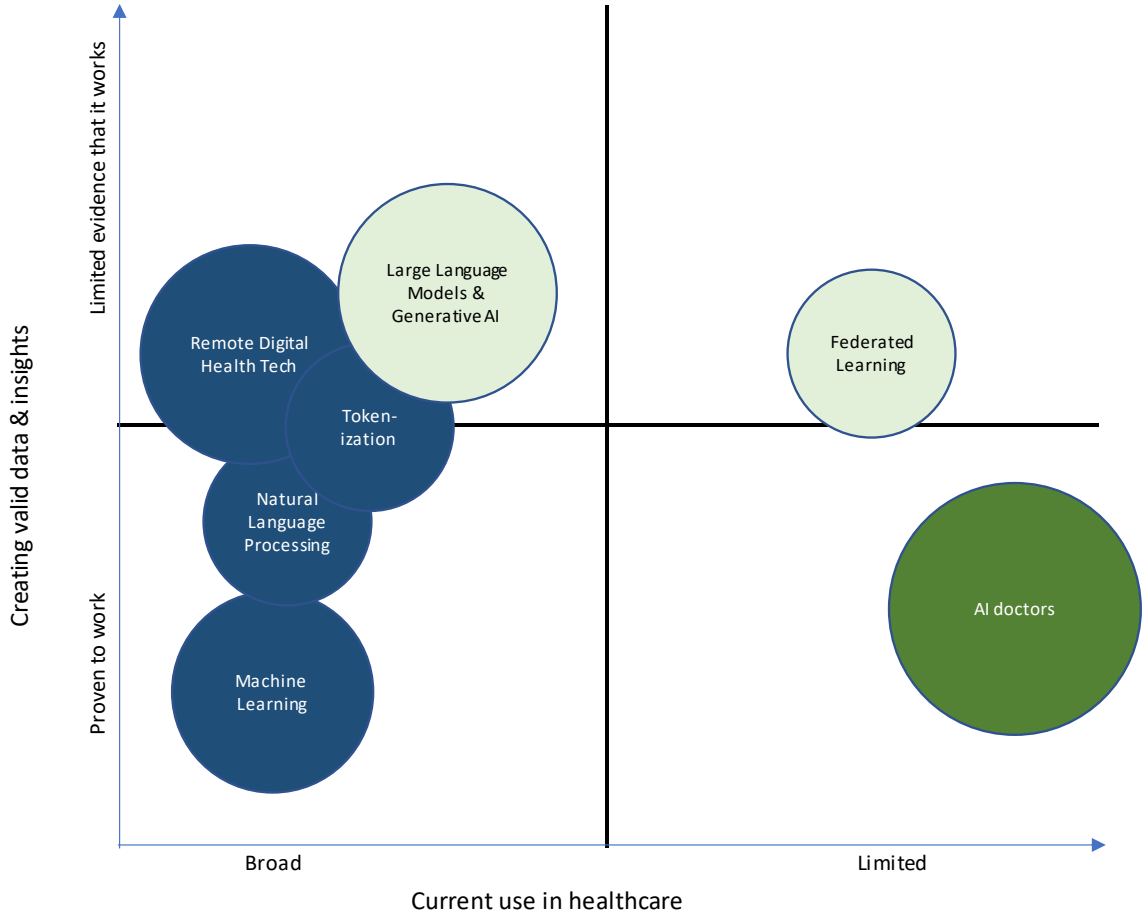
- Representativeness:
  - The set of patients we have data for: do they represent all people with the disease?
  - The shortcomings in rare diseases: patients are identified by highly specialized clinics, so an over-representation of more severe, late, more co-morbid patients
- Number of patients
  - Our disease insights: from how many patients did we get the insights?
  - The shortcomings in rare diseases: the number may be too small for ultra rare disease to even start defining a disease / syndrome
- Scope
  - Did we collect all information that is relevant for the disease?
  - The shortcomings in rare diseases: we do not know al lot about rare diseases, so we do not know what is relevant and miss a lot of information.
- Quality & consistency
  - Are the data collected accurate? Are they complete?
  - The shortcomings in rare diseases: a lot of rare disease data come from individual doctors / clinics, but there is no standardization, no common data model. So no consistency in what is collected and how data are collected.
- Having control groups
  - Do we know what the unique characteristics of the disease and outcome measures that are most relevant to the disease?
  - The shortcomings in rare diseases: a lot of rare disease data come from individual doctors / clinics but they do not collect data from individuals without the disease, so we cannot compare the population with without the disease for identify the characteristics that define the disease and outcome measures that are most relevant.

Improving real world data for rare diseases

Where does technology help?



Technologies: those that have been around for years to those that are only just appearing on our doorsteps



= Technologies which have been around for yeai

= Technologies making inroads

= Technologies that are on our doorsteps

= Impact on rare diseases likely to be small

= Impact on rare diseases likely to be large

How we use these technologies in rare diseases?
Where do they work?
What are the limitations?
What need to be done to improve?

New technologies and data for rare diseases

Overall, these technologies help in 2 areas:

1. Enhancing the scope of data
2. Getting data from a larger, more representative population (earlier diagnosis is key)

Enhancing the scope of data

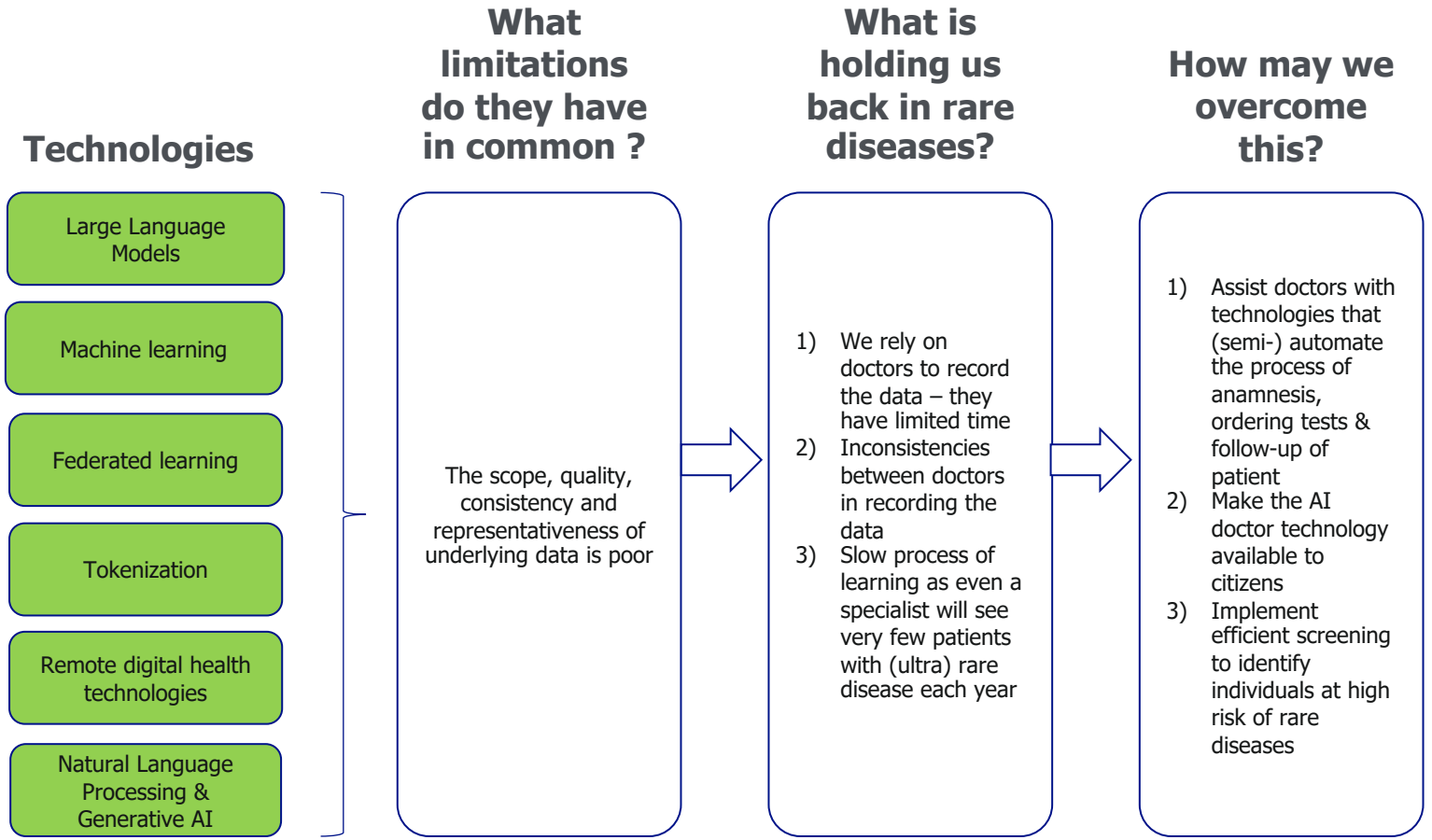
Technology	Where could it be used in rare diseases	Success in rare disease drug development	Limitations in rare diseases
Tokenization	Pulling data together for instance: disease registries / clinical trial data <-> Electronic Health Records	No known success stories	Overlap (tokenized patients in 2 databases) is often small. Quality and completeness and interoperability of underlying data is poor.
Wearables, Apps, home monitoring devices	Clinical trials Devices for orphan diseases	AI-Powered Wearable Tech using motion capture to monitor Duchenne Muscular Dystrophy and Friedreich's Ataxia. (Faisal AA 2023)	Over-reliance on the instrument data as contextual data around results are lacking.
NLP	Qualitative research for hypothesis generation	No known success stories	Limited availability of texts for patients that are early in the disease evolution.
	Enhancing EHR	No known success stories	Over-reliance on inconsistent text recordings.
Large Language Models & Generative AI	Helping doctors to ask the right questions, order the right test and record data.	No known success stories - yet	Limitation is in the underlying information to create the large language models.

Getting data from a larger, more representative population (earlier diagnosis is key)

Technology	Where could it be used in rare diseases	Success in rare diseases	Limitations in rare diseases
Machine learning / Deep Learning to create algorithms	Early diagnosis	Analysis of blood cells to detect rare subsets of cells (Arvaniti E 2017) Radiology report interpretation (Jia X 2020) Limited success using claims databases to detect: - a rare blood disorder (Li W 2018) - a rare pancreatic disorder (Yu K 2019)	Does not help much with earlier diagnosis as these type of test are done by specialist centers. The route to get to the right specialist tends to be the longest. 1) Rare disease chosen are exceptions as they have a diagnostic code to identify cases in claims databases. Many rare diseases do not have a diagnostic code. 2) Low level of specificity of the resulting algorithm A broader use of machine learning in rare diseases is held back by limitations of the data that are available (quality, completeness, interoperability, representativeness and lack of control data).
Federated Learning	Rare disease detection	Tumor boundary detection for glioblastoma using X-ray images. (Pati S 2022)	Does not help much with earlier diagnosis as these patients are already seen by the right specialist. The route to get to the right specialist tends to be the longest. Limitation is in the underlying data for federated learning (quality, completeness, interoperability etc.)
NLP	Screening for rare diseases	In a highly specialized children's hospital: use electronic health records to identify patients who may have Dravet syndrome (Lo Barco 2021)	Does not help much with earlier diagnosis as these patients are already seen by the right specialist. The route to get to the right specialist tends to be the longest. Limitation is in the text and data to create screening tools for citizens or primary care (quality, completeness, interoperability etc.).
Large Language Models	Screening for rare diseases	No known success stories - yet	Limitation is in the underlying text and data to create the large language models (quality, completeness, interoperability etc.)

The limitations of the technologies

What needs to be done to improve?



And always keep in mind….

… data privacy

Especially with (ultra) rare diseases: without taking the right measures, it is too easy to identify individuals with a rare disease.

Thank you!