Risk assessment and multicultural context analysis

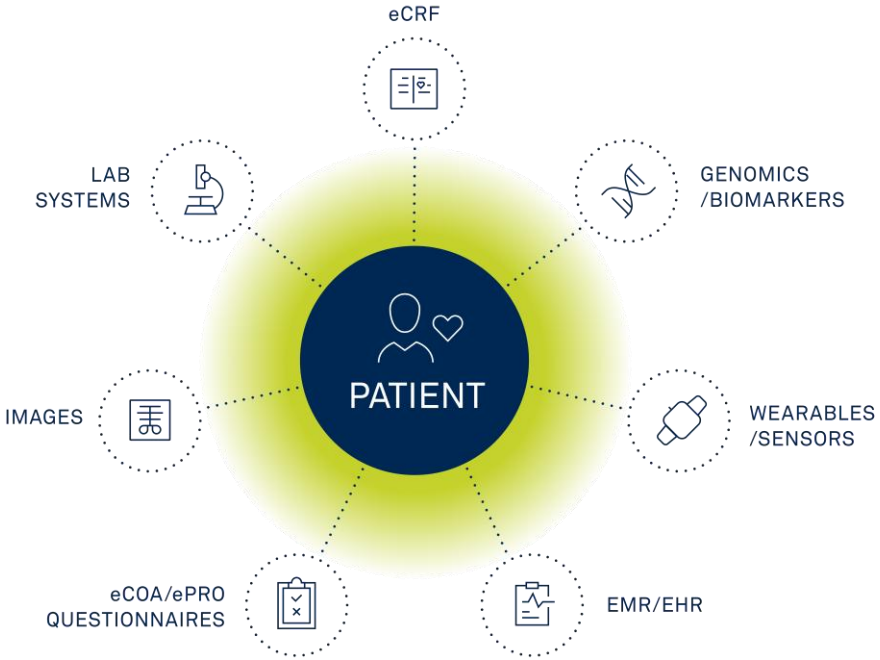
Gabriel Hanssen Kiss

Norwegian University of Science and Technology

Overview

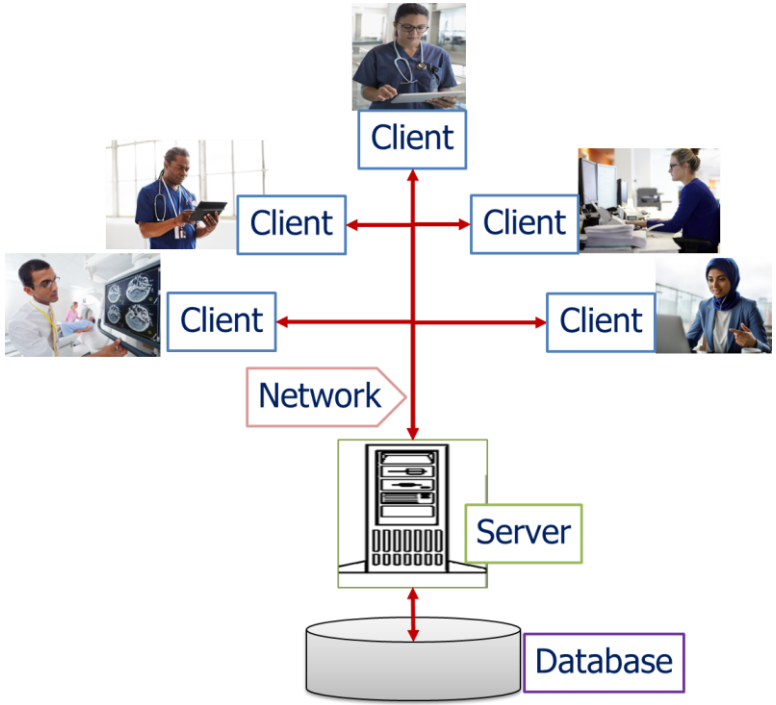
- High level risk assessment for projects working with clinical data: risks and mitigation
- Working with patient data: anonymity, distribution, collaboration, role of ethical committees, multi-cultural aspects
- 5 min break
- Questions / Reflection round the table on how this relevant to the candidate's projects

What is clinical data?

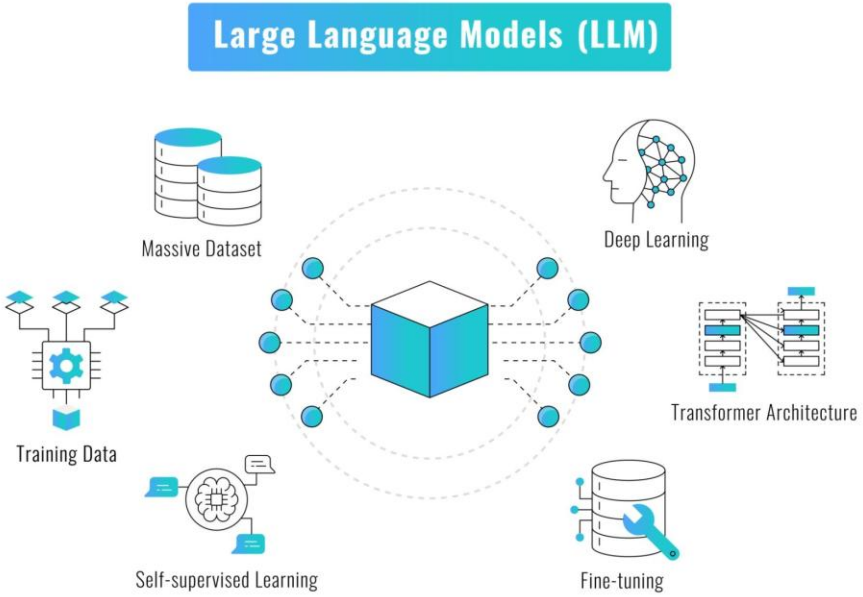


https://www.medidata.com/en/life-science-resources/medidata-blog/modernizing-clinical-data-management-and-capture/

Working with clinical data pre-AI



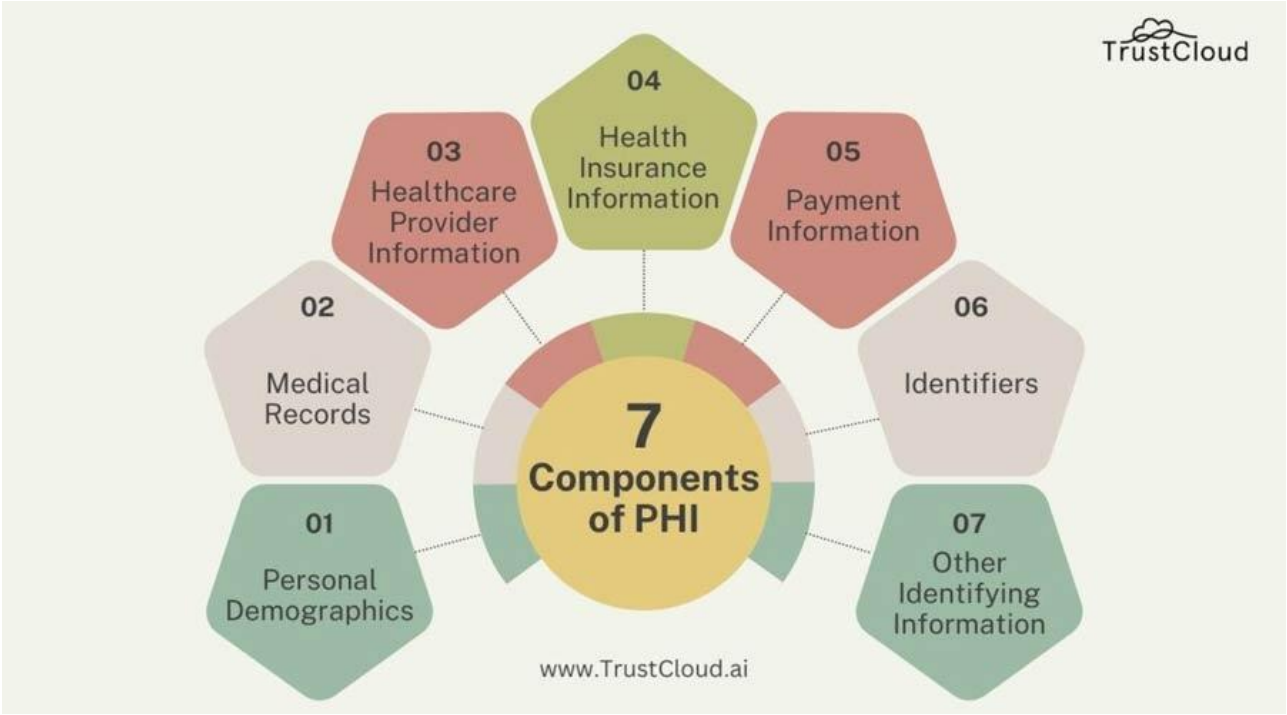
Working with Clinical Data post-AI



Clinical Data Risk Assessment

- Understanding the sensitivity of PHI (Protected Health Information)
- The 'Why': Safety, Legal Compliance (GDPR/HIPAA), and Trust
- The High Stakes: Identity theft and medical fraud

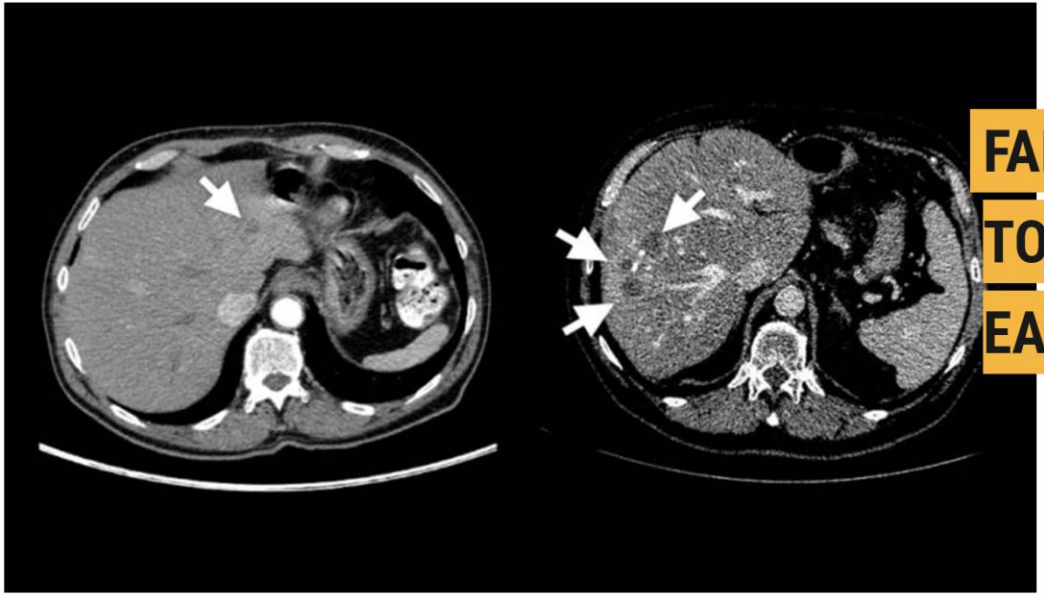
PHI (Protected Health Information)



GDPR = General Data Protection Regulation

https://www.usfhealthonline.com/resources/health-informatics/what-gdpr-means-for-healthcare-providers/

Fraud....



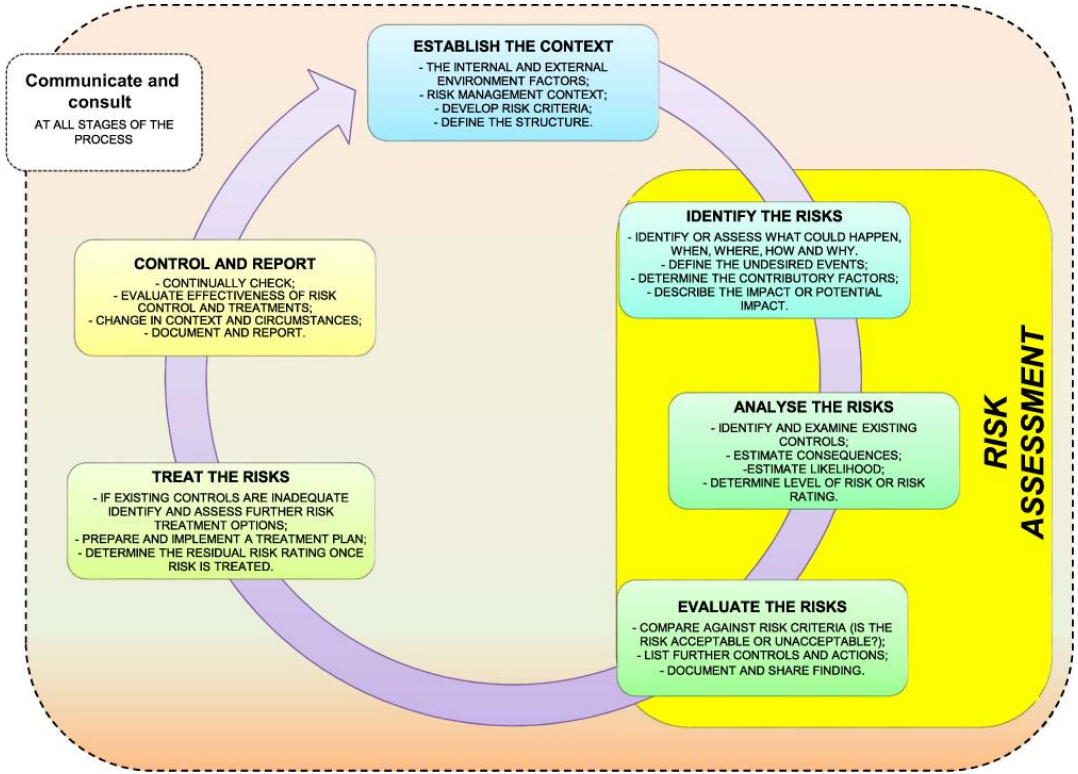
FAKE TUMORS TRAIN AI TO DETECT CANCER EARLY

JULY 1ST, 2024

POSTED BY JAIMIE PATTERSON-JOHNS HOPKINS

A real abdominal tumor, left, vs. synthetic tumors (Credit: Whiting School of Engineering/Johns Hopkins)

Risk assessment clinical study



- Reduce the risk to a quantifiable number

DOI https://doi.org/10.2147/RMHP.S309098

NTNU

Risk assessment

- Many ways of achieving this, can be as simple as this

Risk = L x I

Risk assessment for a new implant

PROBABILITY							
	Expected to occur in most circumstances	0,9	0,09	0,27	0,45	0,63	0,81
	Will probably occur in most circumstances	0,7	0,07	0,21	0,35	0,49	0,63
	Might occur occasionally	0,5	0,05	0,15	0,25	0,35	0,45
	Could happen some time	0,3	0,03	0,09	0,15	0,21	0,27
	May happen only in exceptional circumstances	0,1	0,01	0,03	0,05	0,07	0,09
			0,1	0,3	0,5	0,7	0,9
			Injuries requiring no treatment or first aid	Minor injury, first aid only required	Injury requiring medical treatment and some lost time	Serious injury, hospital treatment required	Death or permanent disability
			IMPACT				

RISK EXAMPLE	PROBABILITY	IMPACT	RISK LEVEL	RISK GRADING
PATIENT INJURY	0,3	0,5	0,15	MODERATE

RISK GRADING COLORS				
0,01-0,03 VERY LOW RISK	0,05-0,07 LOW RISK	0,09-0,27 MODERATE RISK	0,35-0,49 HIGH RISK	0,63-0,81 VERY HIGH RISK

ISO 31000 risk management framework

Working with Patient Data

- Anonymity vs. Pseudonymization
- The mechanics of secure distribution
- The role of ethical oversight (IRB/REC)

GDPR - For the patient

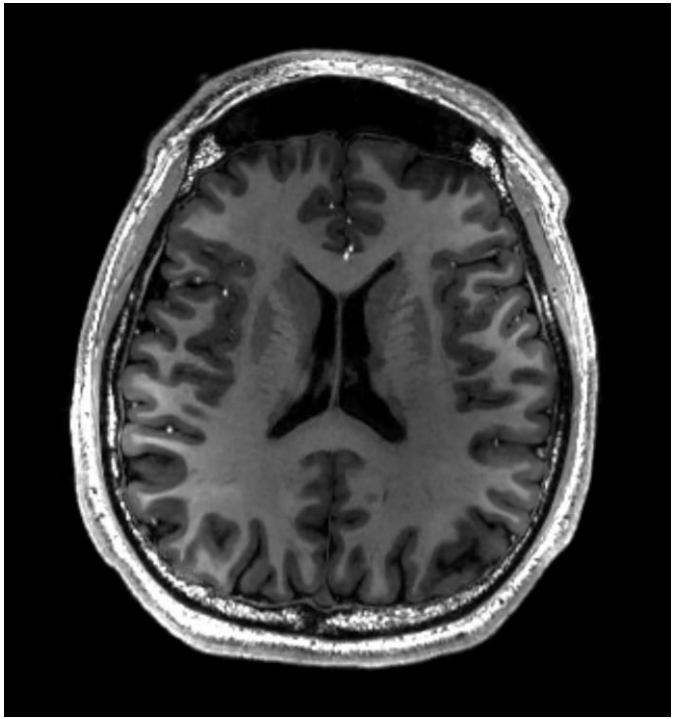
- **Right to be forgotten**. Consumers can withdraw consent from companies that collect their information and have the right to get that information deleted at any time.
- **Right to access.** European residents have access to their personal data stored by a business and can find out how their information is used – free of charge.
- **Right to be informed**. Citizens must give explicit permission to allow companies to gather data on them, and they must be informed about exactly what data is gathered and how it is used.

https://www.usfhealthonline.com/resources/health-informatics/what-gdpr-means-for-healthcare-providers/

Patient data

- Who owns it?
  - Patients
- How can you use it for your own research?
  - Informed consent

Is an MRI / US image anonymous?



Well…

What is anonymity?

- Clinical data anonymity = the process of modifying health information to remove or obscure direct and indirect identifiers, making it impossible for anyone to link the data back to a specific individual, thus protecting patient privacy while allowing for research and analysis.
- * In Norway it is even stricter, it should not be possible to identify a group of up to 5 persons including the one owning the data based on the provided information

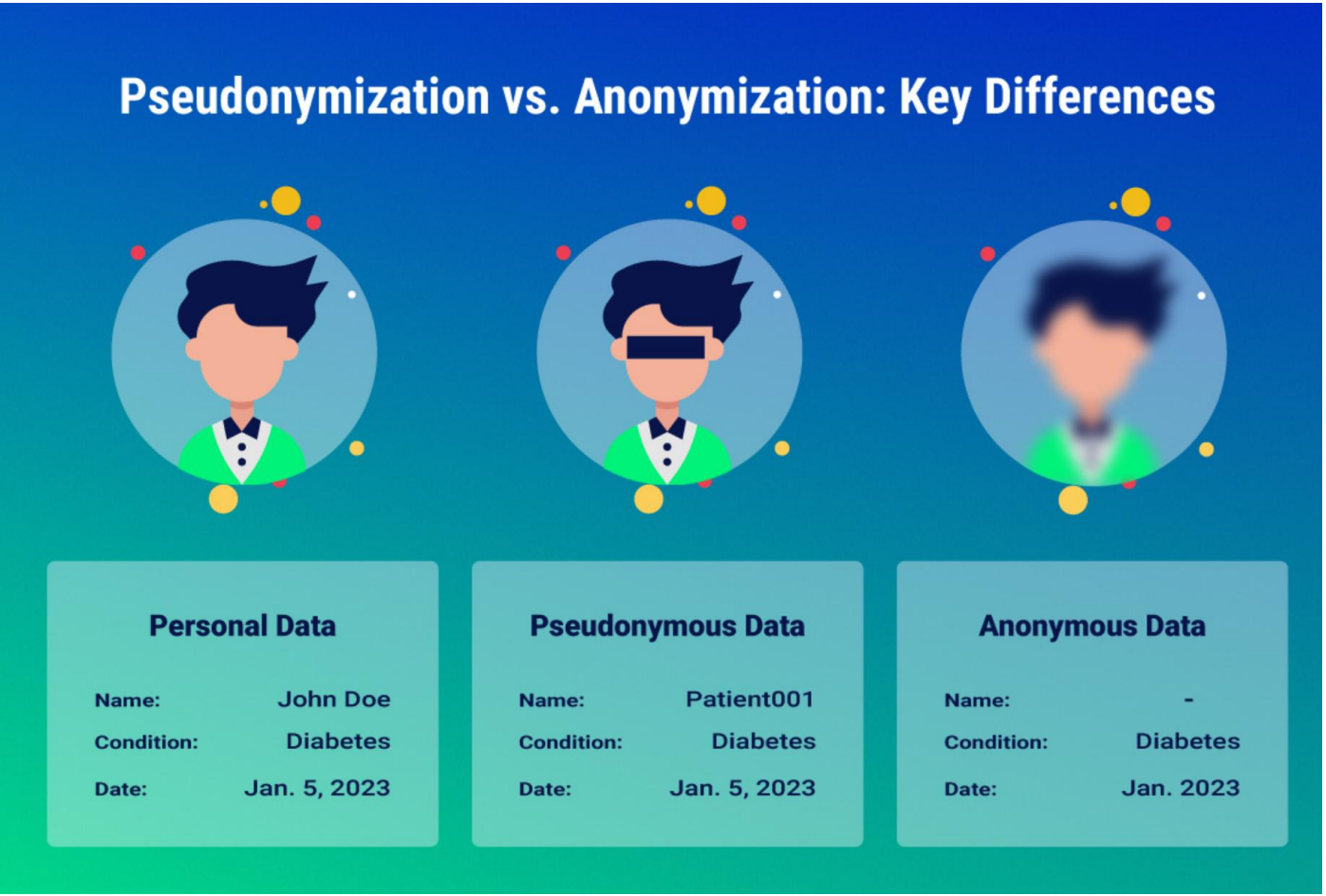
Anonymity vs. Pseudonymization

- Pseudonymization: Using keys (reversible)
- Anonymization: Irreversible removal of identifiers
- The Mosaic Effect: Risks of re-identification through data merging

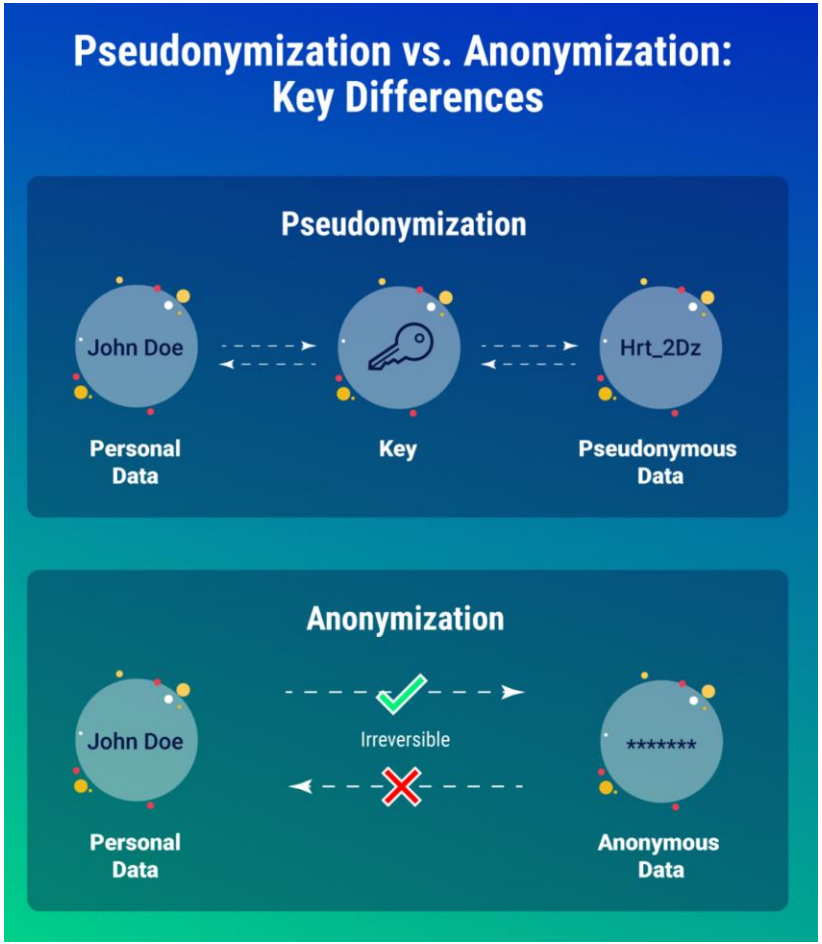
Pseudonymization

- Article 4(5) of the GDPR offers a definition for "pseudonymisation" as "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person."

https://www.clym.io/blog/pseudonymization



https://www.clym.io/blog/pseudonymization



https://www.clym.io/blog/pseudonymization

Anonymization types

Original Dataset:	Pseudonymized Dataset:	Anonymized Dataset:
1. John Doe - Diabetes - Diagnosed on January 5, 2023	1. Patient001 - Diabetes - Diagnosed on January 5, 2023	1. Condition: Diabetes - Diagnosed in January 2023
2. Jane Smith - Hypertension - Diagnosed on February 10, 2023	2. Patient002 - Hypertension - Diagnosed on February 10, 2023	2. Condition: Hypertension - Diagnosed in February 2023
3. Alice Johnson - Asthma - Diagnosed on March 15, 2023	3. Patient003 - Asthma - Diagnosed on March 15, 2023	3. Condition: Asthma - Diagnosed in March 2023

https://www.clym.io/blog/pseudonymization

Anonymity

- Anonymization is when data is changed so much that there's no way to tell who it belongs to, not even by your organization. According to the GDPR, once the data is anonymized, it is no longer seen as personal data so the GDPR no longer covers it.

https://www.clym.io/blog/pseudonymization

Quick quiz!

- If I give you:
  - ZIP code
  - Date of Birth
  - Gender
  - for 100 people in the US (population more that 340 million)
  - How many persons you can identify uniquely?

  - **87**
  - "Mosaic Effect.“ = combining seemingly minor, separate data points (even public ones) to reveal sensitive or significant patterns

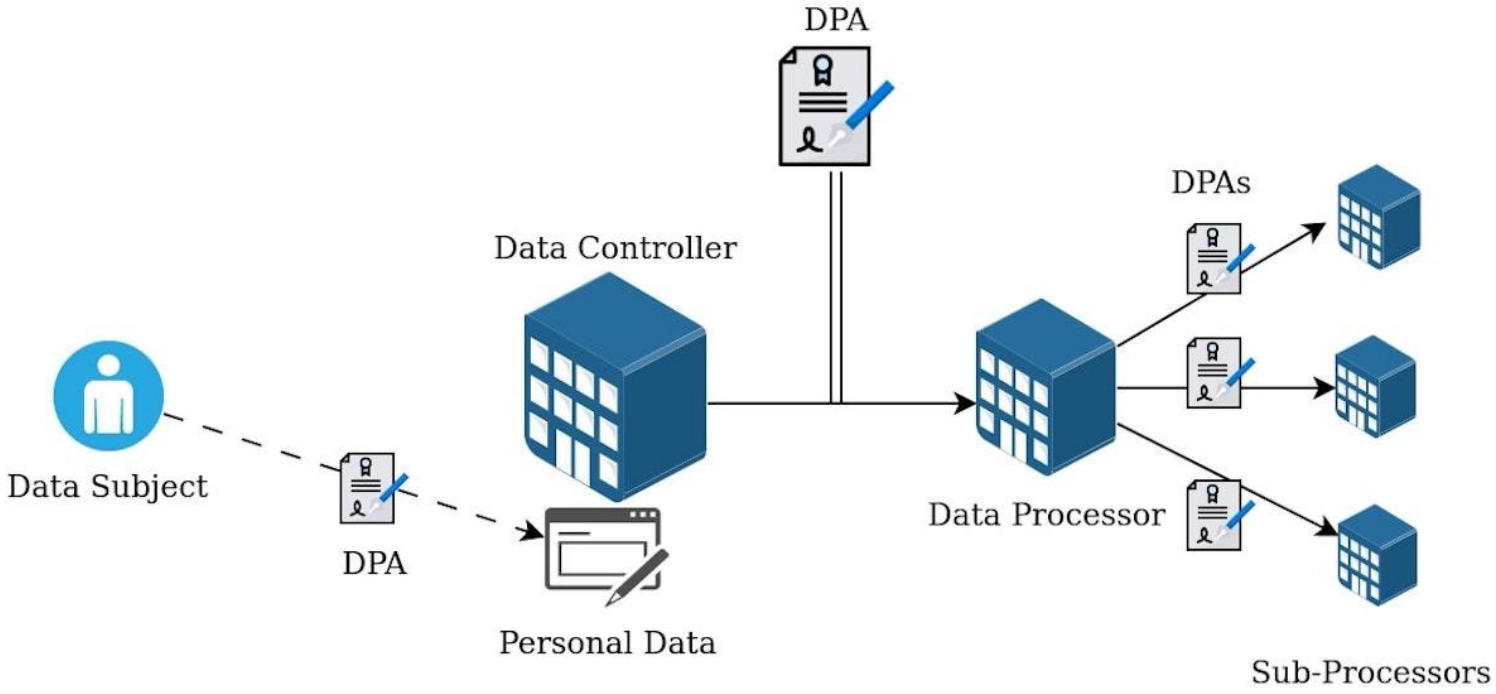
Secure Distribution of Patient Data

- Data Use Agreements (DUA): The legal framework
- Secure Transfer: SFTP and Trusted Research Environments (TRE)
- Federated Learning: Keeping data on-site while sharing insights

Data Processing Agreements (DPA)

- Data Processing Agreements (DPA) form a crucial legal framework for sharing sensitive data, acting as contracts that define permissible uses, security requirements, and restrictions (like prohibiting re-identification) to protect privacy, especially under regulations like HIPAA or GDPR, ensuring accountability and compliance by outlining terms for data handling, safeguarding, liability, and eventual data destruction or return between providers and recipients

Data Processing Agreements (DPA)



Ethical committees

- **The Belmont Report Principles:**
  - **Respect for Persons:** Patients must give Informed Consent and have the right to withdraw.
  - **Beneficence:** The research must intend to do good; the risks must be justified by the potential knowledge gained.
  - **Justice:** We must not experiment on vulnerable populations just because they are "available."

Ethical committee application

- Depends on your research:

Regional committees for medical and health research ethics

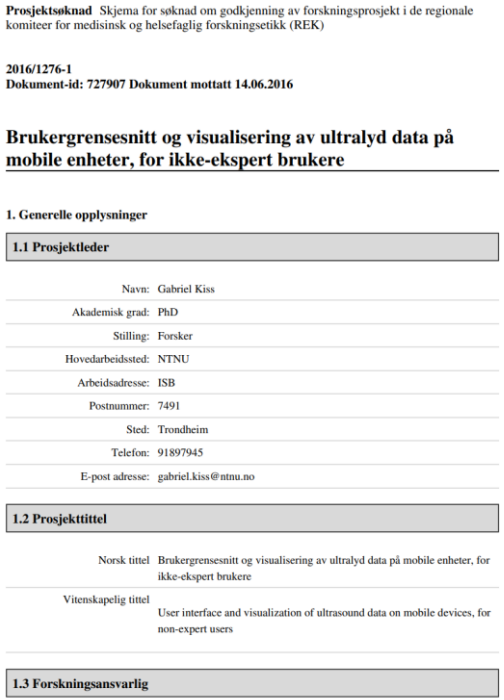
Go to the pages for those applying to REK

Committees for clinical trials of medicines and medical devices

Go to the pages for those applying to REK KULMU

https://rekportalen.no/en/

Ethical committee approval



Prosjektsøknad Skjema for søknad om godkjenning av forskningsprosjekt i de regionale komiteer for medisinsk og helsefaglig forskningsetikk (REK)

2016/1276-1
Dokument-id: 727907 Dokument mottatt 14.06.2016

Brukergrensesnitt og visualisering av ultralyd data på mobile enheter, for ikke-ekspert brukere

1. Generelle opplysninger

1.1 Prosjektleder

Navn:	Gabriel Kiss
Akademisk grad:	PhD
Stilling:	Forsker
Hovedarbeidssted:	NTNU
Arbeidsadresse:	ISB
Postnummer:	7491
Sted:	Trondheim
Telefon:	91897945
E-post adresse:	gabriel.kiss@ntnu.no

1.2 Prosjekttittel

Norsk tittel	Brukergrensesnitt og visualisering av ultralyd data på mobile enheter, for ikke-ekspert brukere
Vitenskapelig tittel	User interface and visualization of ultrasound data on mobile devices, for non-expert users

1.3 Forskningsansvarlig

- They require a project description but do not care about it from a research perspective
- What they evaluate is:
  - What data is acquired
  - How it is stored / is it anonymized
  - If a patient treatment changes from the current protocol what are the risks for the patients
  - Follow-up actions in case of emergency cases
  - Are companies involved if so do project participants own shares -> conflicts of interest
  - Planned publications
  - What happens to data after the project is completed

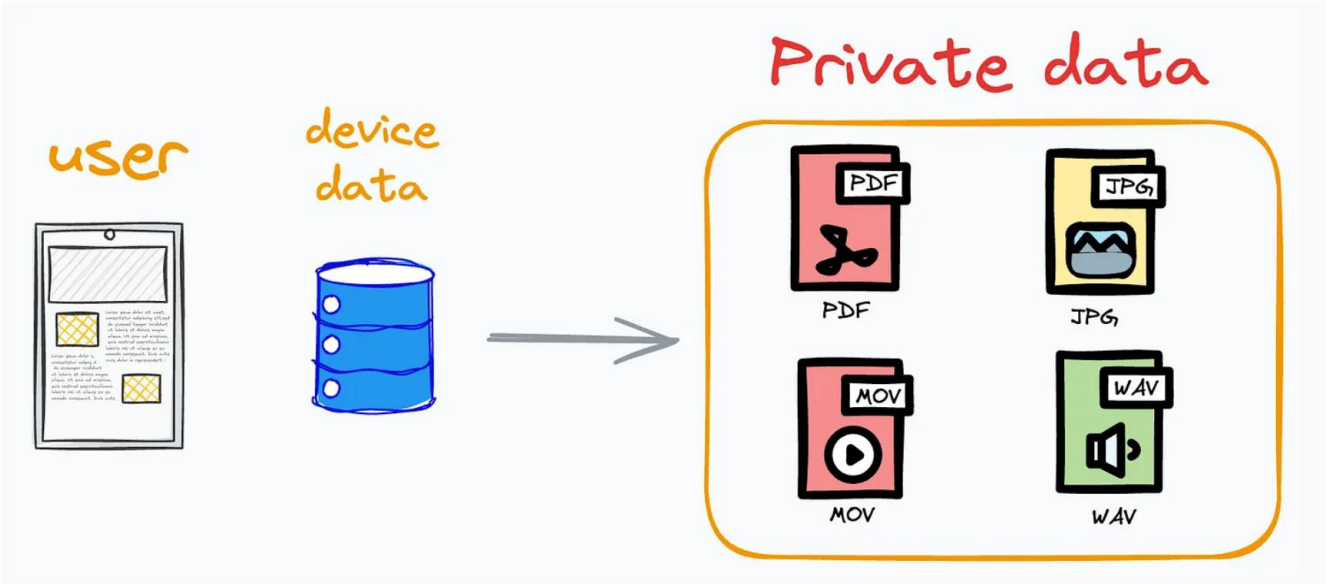
Data security

- **Encryption:** Protecting data in storage (at rest) and during transmission (in transit) using strong protocols.
- **Access Control:** Implementing Role-Based Access Control (RBAC) and Multi-Factor Authentication (MFA) to ensure only authorized personnel access necessary data.
- **Data Classification:** Identifying and labeling data sensitivity (e.g., PHI) to apply appropriate security measures, often using automated tools.
- **Regular Audits & Monitoring:** Continuously monitoring network traffic and auditing logs for suspicious activity.
- **Patch Management:** Keeping all software, operating systems, and medical devices updated to fix vulnerabilities.
- **Employee Training:** Educating staff on phishing, data handling, and security best practices.
- **Incident Response Plan:** Having a documented plan to quickly detect, respond to, and mitigate data breaches.
- **Vendor Management:** Ensuring third-party partners meet the same high security standards.
- **Compliance:** Adhering to standards like HIPAA (US), GDPR (EU), and HITRUST.

Standards / frameworks

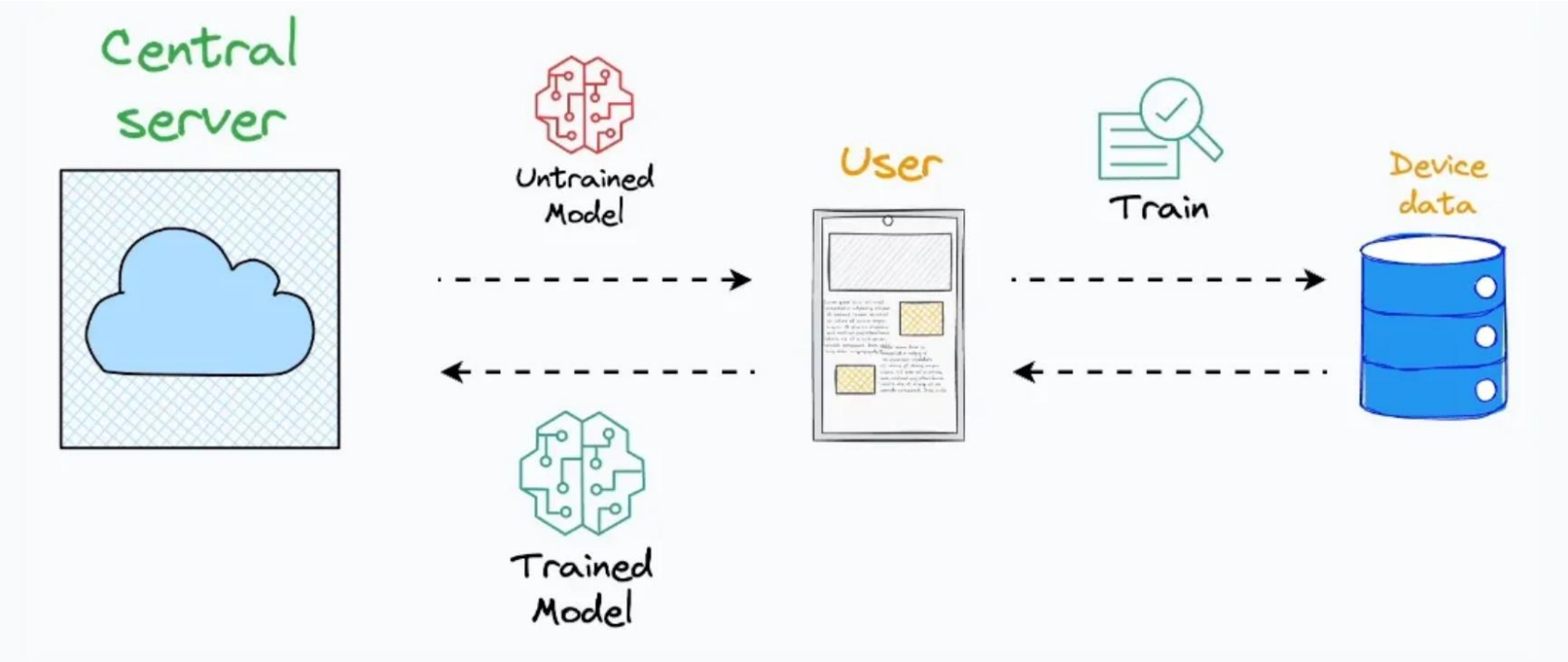
- **HIPAA Security Rule (US):** Mandates physical, technical, and administrative safeguards for PHI.
- **IEC 81001-5-1:** A specific cybersecurity standard for medical devices and health IT software.
- **HITRUST:** A framework for managing risk, security, and compliance in healthcare.

Federated learning - Need

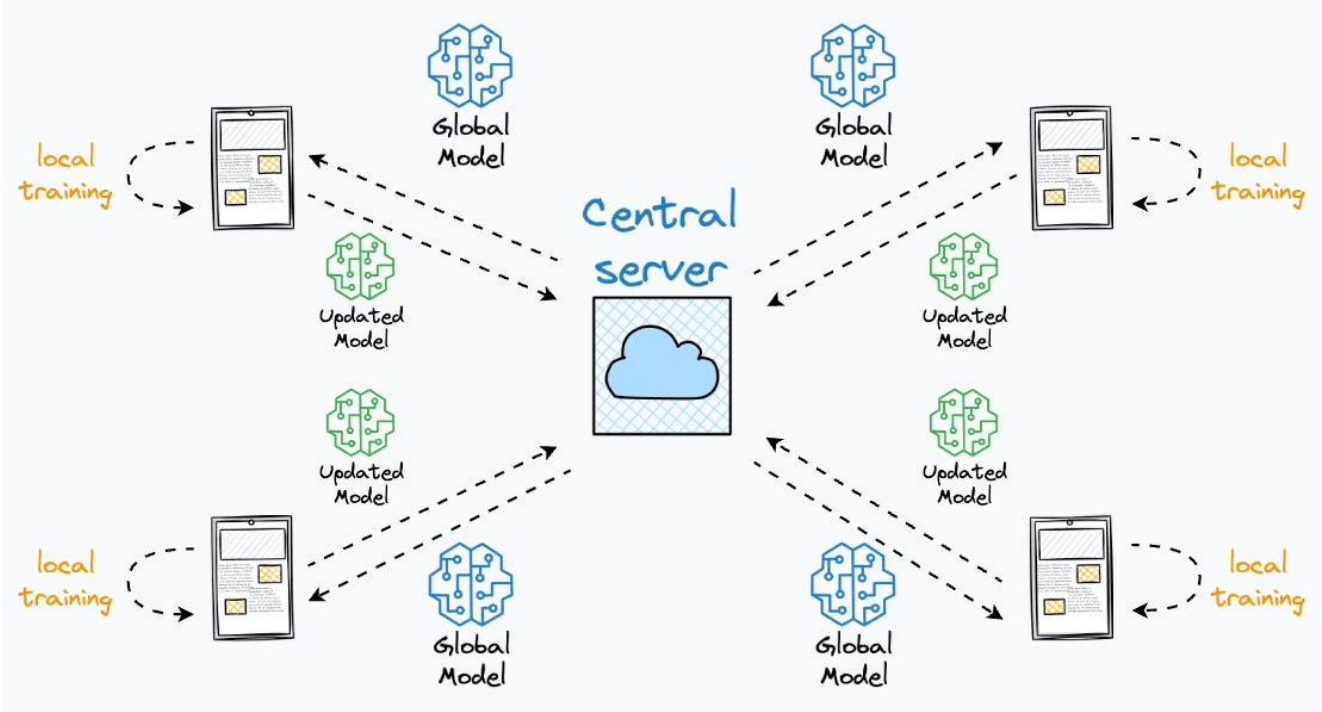


- What happens when you want to include multiple sites, different hospitals in different countries?

Federated learning - Principle



Federated learning - Overview



https://blog.dailydoseofds.com/p/introduction-to-federated-learning

Federate learning – Side note

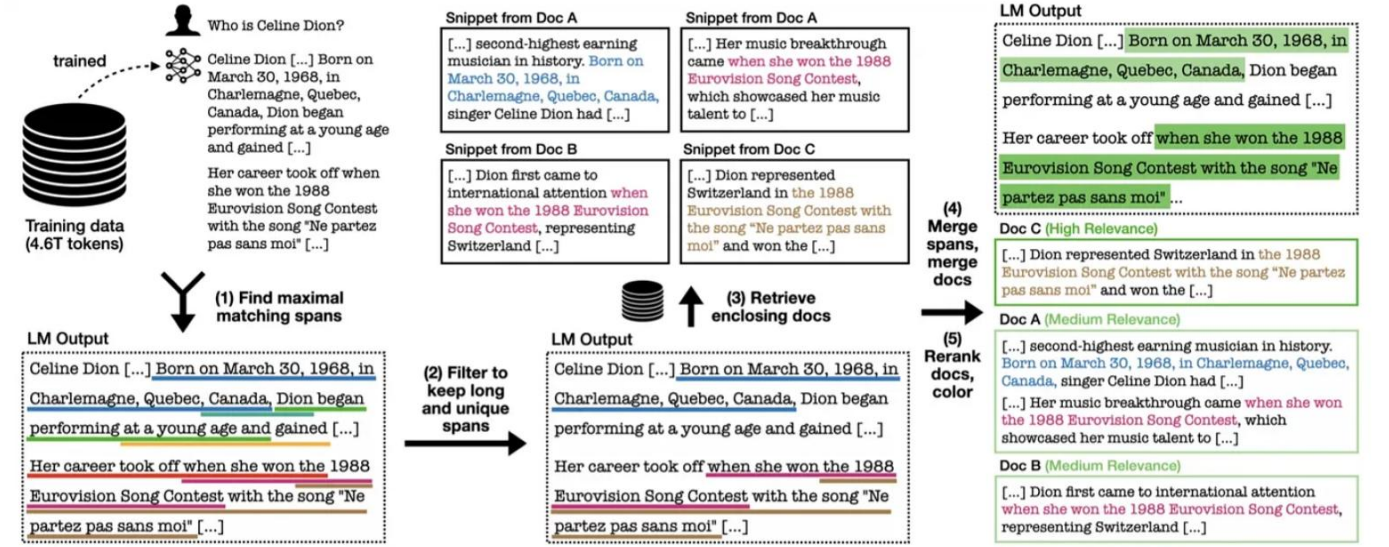


Figure 2: The OLMoTRACE inference pipeline, as described in §3. For better illustration, we slightly adjusted the highlighted spans and document relevance from the actual example.

"OLMOTRACE: Tracing Language Model Outputs Back to Trillions of Training Tokens" by Liu et al. (2025),

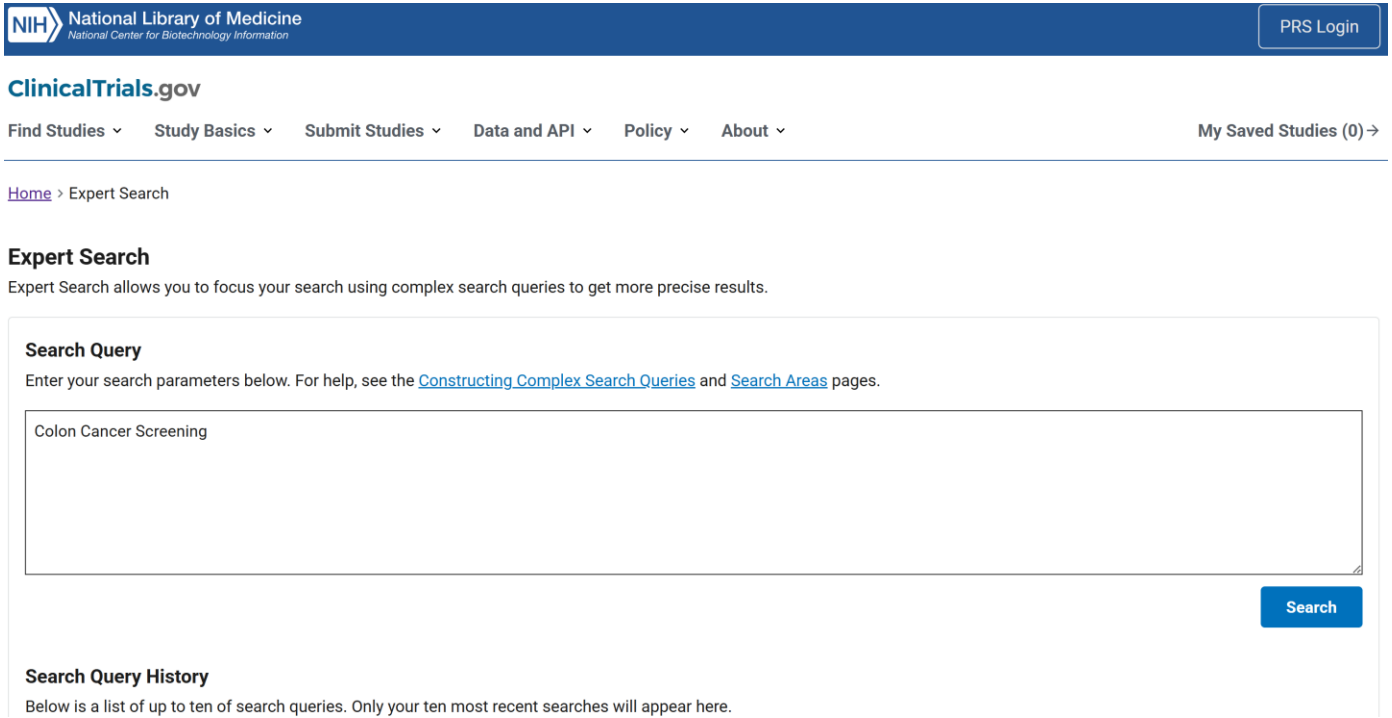
It is still possible to do research?

Research: Openly available datasets

- Challenges

Challenge name	Acronym	DOI
Advancing Generalizability and Fairness in Breast MRI Tumour Segmentation and Treatment Response Prediction	MAMA-MIA	10.5281/zenodo.15052677
AIMS-TBI - Automated Identification of Moderate-Severe Traumatic Brain Injury Lesions	AIMS-TBI	10.5281/zenodo.15084119
Automated Lesion Segmentation in Whole-Body PET/CT and Longitudinal	autoPET/CT IV	10.5281/zenodo.15045095
Benchmarking of Artificial Intelligence and Radiologists for Lung Cancer Screening in CT: The LUNA25 Challenge	LUNA25	10.5281/zenodo.15094630
Calibration and Uncertainty for multiRater Volume Assessment in multiorgan Segmentation 2nd Edition	CURVAS - PDACVI	10.5281/zenodo.15045199
CARE 2025: Comprehensive Analysis & computing of REal-world medical images	CARE2025	10.5281/zenodo.15045249
Challenge for Vision-Language Modeling in 3D Medical Imaging	VLM3D	https://doi.org/10.5281/zenodo.15052707
Deep-learning Evaluation for Enhanced Prognostics - Prostate Specific Membrane Antigen	DEEP-PSMA	10.5281/zenodo.15094694
Dehazing Echocardiography Challenge 2025	DehazingEcho 2025	10.5281/zenodo.15083973
Combining HIstology, Medical imaging and molEcular data for medical pRognosis and diAgnosis	CHIMERA	10.5281/zenodo.15045552

Clinical trials



Nice overview of available data

Unknown status

Verified **2016-10** by John F. Gray MD FACG AGAF, Renown Regional Medical Center

Last known status was: Active, not recruiting

FIT Mailing Protocol-For Cancer Screening Navigation

ClinicalTrials.gov ID NCT02934958

Sponsor Renown Regional Medical Center

Information provided by John F. Gray MD FACG AGAF, Renown Regional Medical Center (Responsible Party)

Last Update Posted 2016-10-17

Download | Save | + Expand all content | — Collapse all content

Study Details | Researcher View | No Results Posted | Record History

On this page

- Study Overview
- Contacts and Locations
- Participation Criteria
- Study Plan

Study Overview

Brief Summary

We plan to study whether the impact of offering the choice of a pre-colonoscopy physician visit or direct referral to colonoscopy will increase adherence to colonoscopy relative to usual care in a large fecal immunochemical test (FIT) mailing campaign.

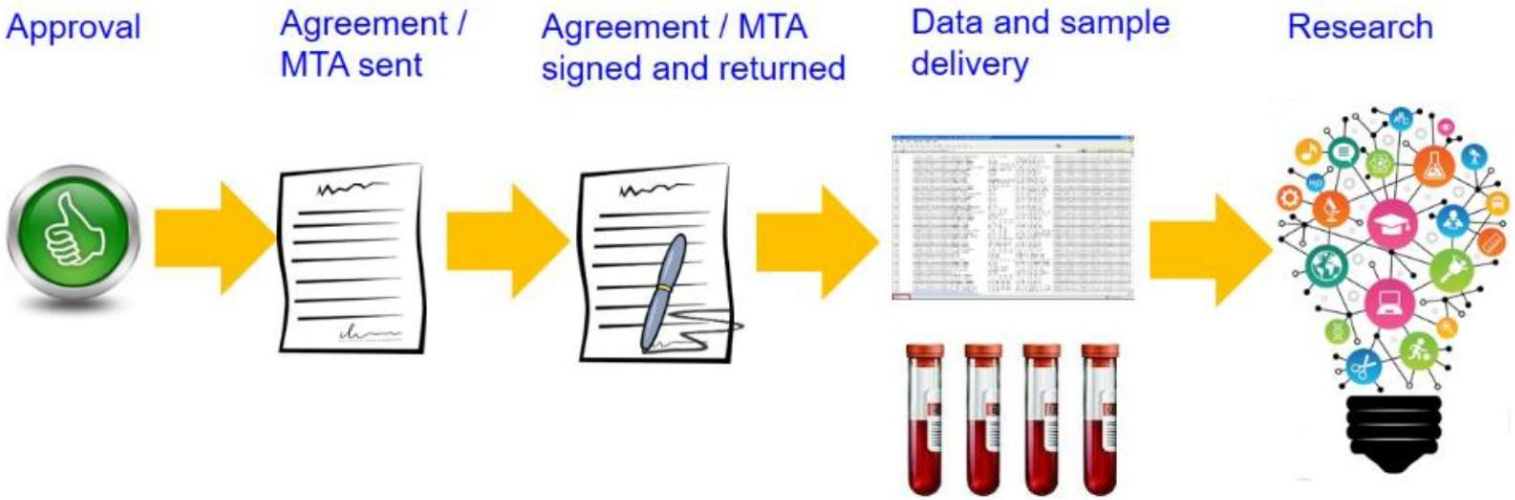
Study Start

2015-10

Primary Completion (Estimated)

What are other options?

- Biobanks



Bring compute to the data

For researchers

Digital labs for sensitive data science

Focus on your research with confidence, assured that your data is secure and compliant with all regulations.

Our labs provide you with the creative freedom you need, secured within our high-trust framework.

CONTACT US ›

HUNT cloud: https://about.hdc.ntnu.no/en/researchers/

Why bother….

- Sometimes it goes terribly wrong even if the intentions are good

Sweden: surgeon convicted of bodily harm over synthetic trachea transplant

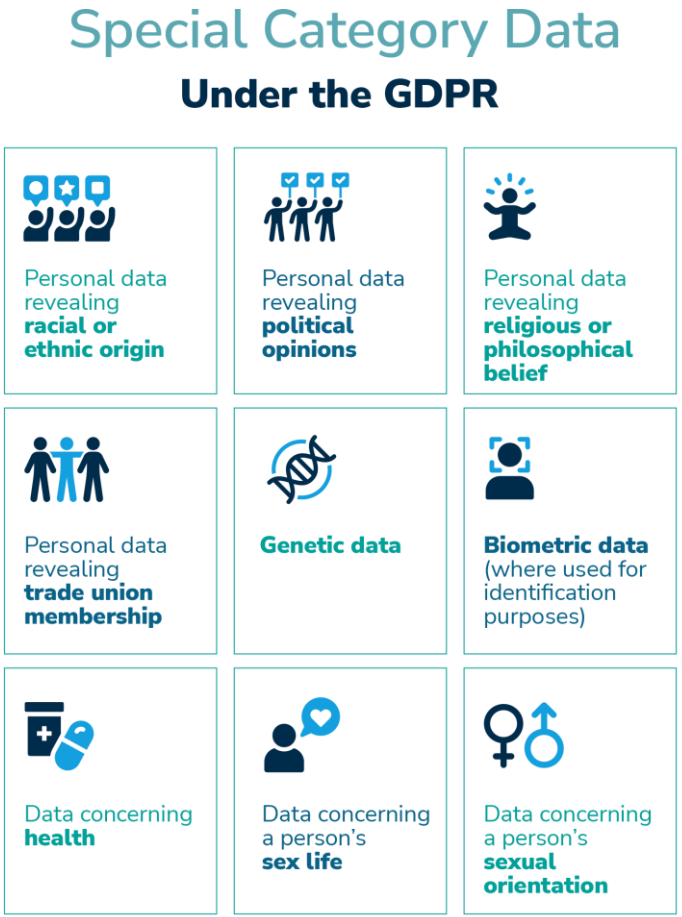
Court finds that Paolo Macchiarini carried out experimental procedure on patient who was not critically ill

A photograph of Paolo Macchiarini released in 2011 by Stockholm's Karolinska University hospital, where he carried out the transplants. Photograph: Karolinska University Hospital/AFP/Getty Images

A Swedish court has found an Italian surgeon, once hailed for pioneering windpipe surgery, guilty of causing bodily harm to a patient, but cleared him of assault charges.

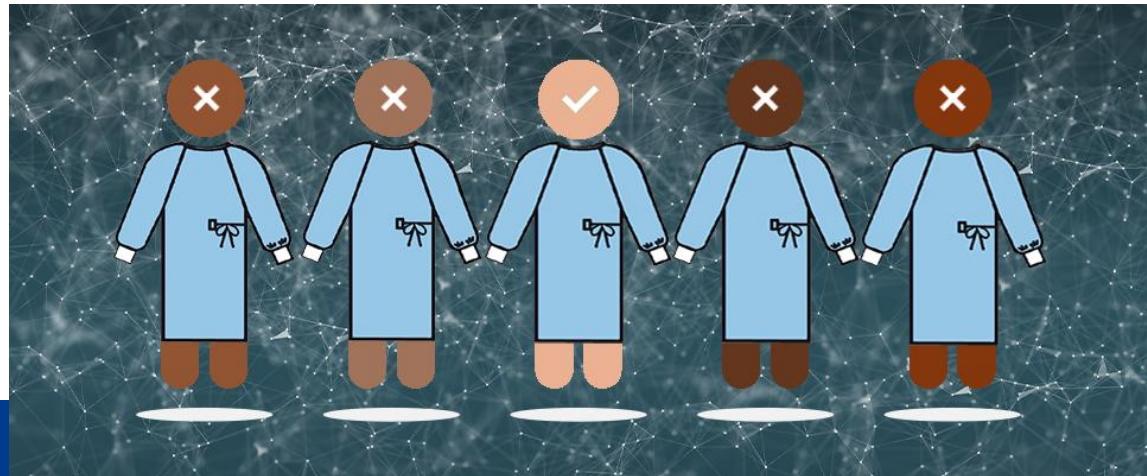
Patient benefits

Avoids discrimination



AI discrimination

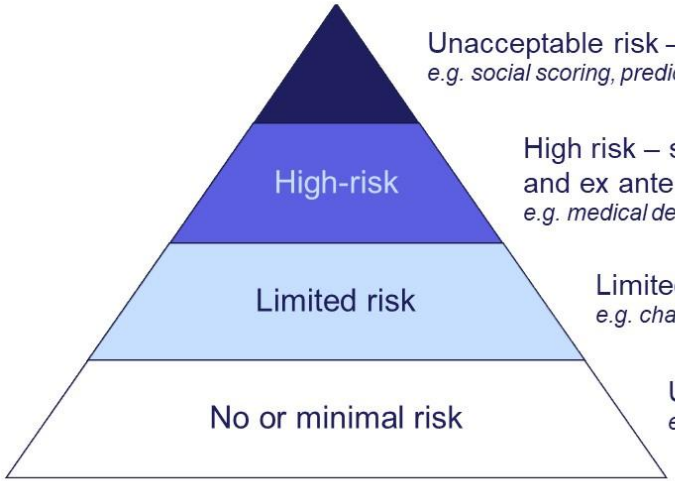
- Many reasons
  - Biased historical data
  - Unequal healthcare availability
  - Research funding bias towards high developed countries
- Biased training data
- Biased models



Openness – algorithmic discrimination

- **Generative AI systems**: certain generative AI systems, such as chatbots, do not seem dangerous at first glance, but can still pose threats to fundamental rights and freedoms, for example when they generate hate speech. Generative AI is not classified as high-risk as such. However, some generative AI systems that harm fundamental rights might be included in the high-risk system list that regulators can update periodically, as laid down in the AI Act.
- **Autonomous cars**: discrimination could also arise in insurance calculation or AI machine vision systems – for instance, if autonomous cars are developed in a way that will detect more accurately pedestrians with lighter than those with darker skin.
- **Job recruitment and employment**: algorithms selecting job applicants might contain bias, for example with respect to gender or health.
- **Credit scoring and banking**: AI systems are increasingly used in the banking and credit sectors as a support tool to assess the granting of loans or mortgages. These systems may follow decisional processes that hide discrimination or mask bias based on factors such as clients' residence or ethnicity.

More to come: AI act



Unacceptable risk – prohibited
e.g. social scoring, predictive policing behavioural manipulation

High risk – subject to mandatory requirements and ex ante conformity assessment
e.g. medical devices, vehicles, law enforcement, education, immigration

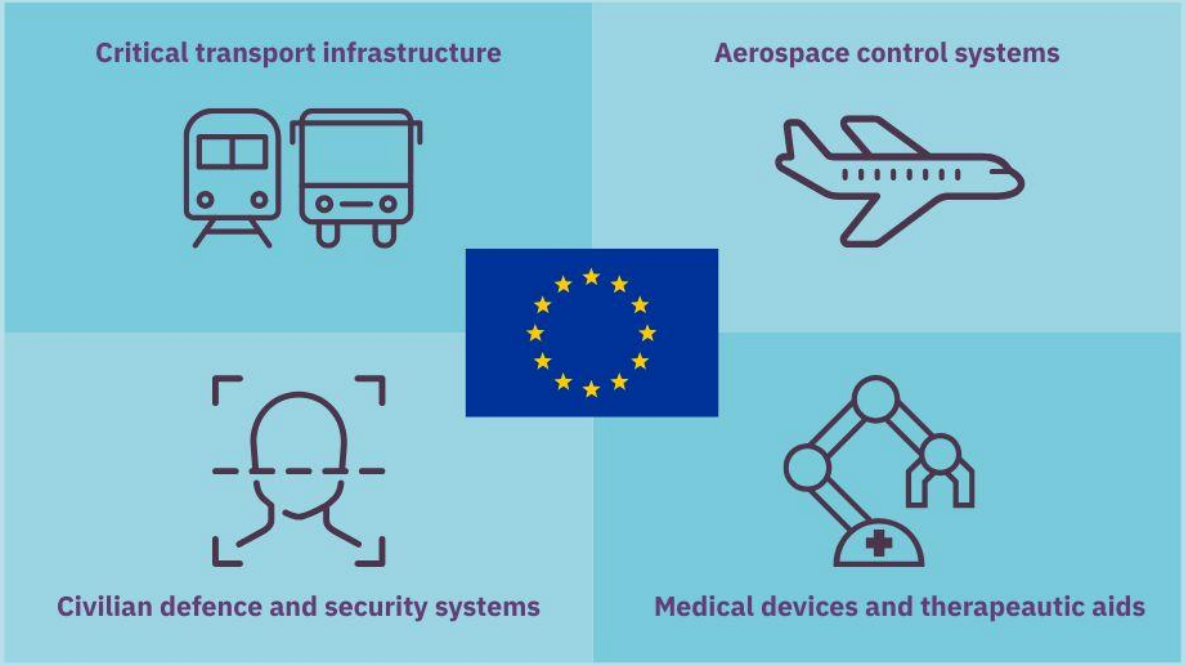
Limited risk – subject to transparency obligations
e.g. chatbots, deep fakes

Unregulated – voluntary codes of conducts
e.g. spam filters, games

+ general-purpose AI is subject to separate specific regulations
e.g. GPT-4

https://www.sorainen.com/publications/the-ai-act-is-here-the-time-to-act-is-now/

High-risk AI systems



Source: EU AI Act Annex III

Bluefruit Software

NTNU

Discussion Questions

- Is 100% security possible or even desirable in research?
- Does a 1% risk of re-identification justify a potential cure?
- Should patients have a 'right to be forgotten' in medical datasets? i.e. they ask to be removed after you trained your fancy AI models
- What data are you using in your projects?
- What paperwork do you need to fill out?