AI and PETs: mitigate privacy risks of patient's data

Dr. Kyriakos Stefanidis
Principal Researcher
Industrial Systems Institute / ATHENA Research Center

Why Privacy Matters in Healthcare/Oncology

- Sensitivity of oncology-related data
  - Genomic profiles
  - Treatment outcomes
  - Rare diagnoses
- Privacy Risks
  - Re-identification
  - Insurance discrimination
  - Secondary use without consent
- Legal responsibilities (GDPR)
- Impact: Patients unwilling to share data or make them available for training due to privacy concerns

AI and Privacy Risks

AI is already used in most domains

- Diagnostics
- Prognosis
- Treatment planning and decision support
- Clinical trials optimization

AI models and applications can expose data (new problems)

Training data leakage

Membership inference

Model inversion

Improper model sharing

Privacy Attack Surface

Membership Inference Attacks

- An attacker queries a model and determines whether a specific patient’s data was part of the training set.
- Impact (example): knowing a person was part of a clinical trial on aggressive cancer could reveal sensitive health status

Model Inversion Attacks

- The attacker uses access to an AI model to reconstruct input data (e.g., a gene expression profile)
- Impact (example): Given a model trained on CT scans; inversion could reconstruct the patient’s anatomical image

Data Reconstruction / Extraction Attacks

- AI models, especially large ones, can memorize and leak portions of training data when prompted cleverly

Linkage Attacks

- Combining anonymized data with external datasets (e.g., voter registries or social media) to re-identify patients
- Impact (example): Even if names are removed, rare cancer types or zip codes may uniquely identify someone.

Privacy-Enhancing Technologies (PETs)

Goal

- Use and analyze patient data safely—without exposing personal information
- Gain insights from sensitive data (like medical records or scans) while protecting patient privacy

PET Categories

Federated and distributed analytics (e.g., federated learning)

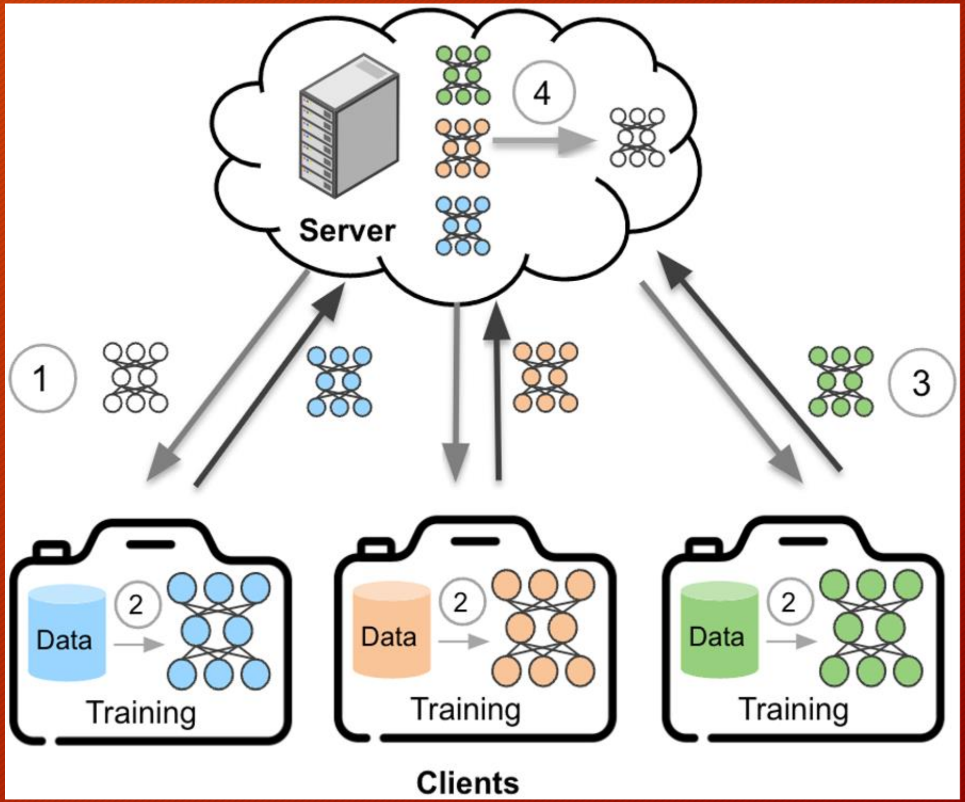
Data obfuscation (e.g., differential privacy, anonymization)

Encrypted data processing (e.g., homomorphic encryption, multi-party computation)

Federated Learning

- How it works: Train AI models across many hospitals or clinics, without sharing patient data.
  - Patient data stays local (on each hospital's system)
  - Only model updates (not raw data) are sent to a central server.
  - The central server combines these updates to improve a global AI model.
- Example
  - “The Federated Tumor Segmentation (FeTS) Challenge”
    - Train a model to detect and outline gliomas in MRI scans.
    - Enabled training on diverse patient populations across institutions, without moving any MRI data.

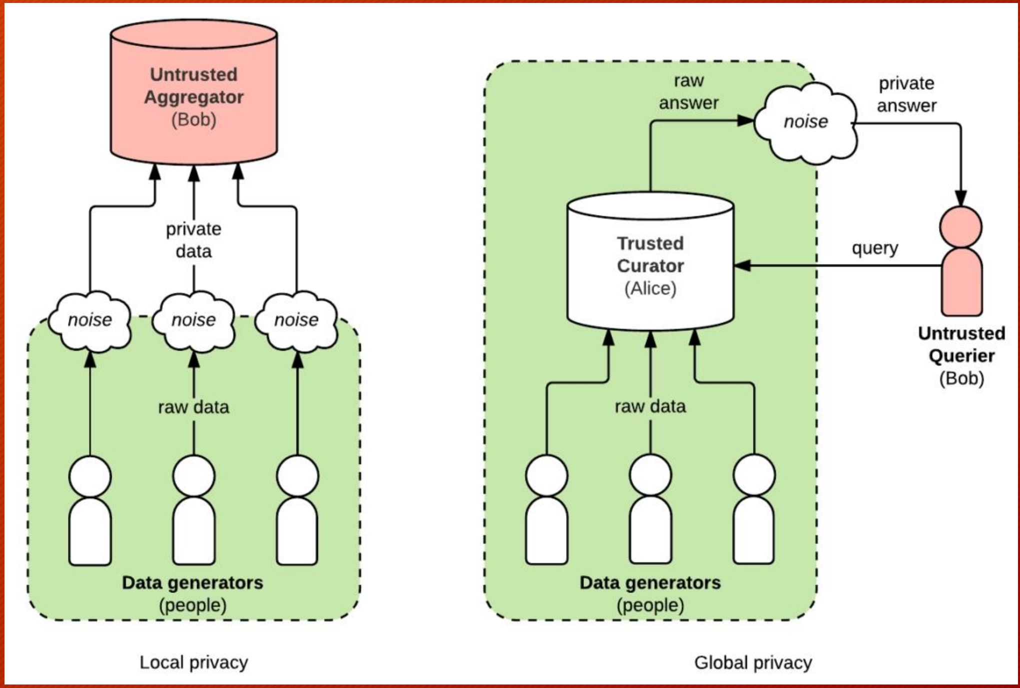
https://www.synapse.org/Synapse:syn54079892/wiki/626481



https://ai.sony/blog/Recent-Breakthroughs-Tackle-Challenges-in-Federated-Learning/

Differential Privacy

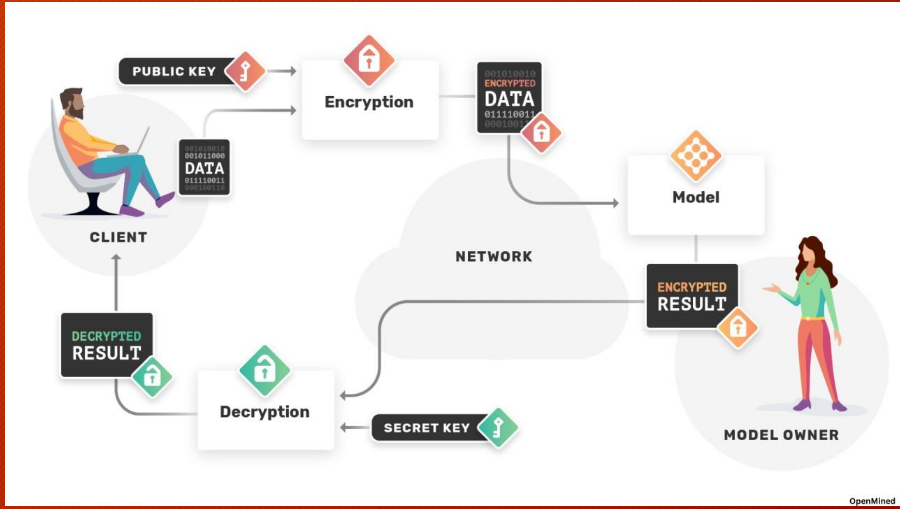
- How it works: Allows AI models to learn from patient data while mathematically guaranteeing that no individual's data can be identified or singled out
  - Before data is used for training, a small amount of noise (randomness) is added
  - Noise hides the influence of any single patient in the dataset.
- Example
  - Predicting 5-year survival for cancer patients based on treatment history.
  - DP makes it safe to publish or share model outputs (e.g., feature importance), since no patient-specific treatment path can be reconstructed.



https://research.aimultiple.com/differential-privacy/

Homomorphic Encryption

- How it works: Allows computations to be performed directly on encrypted data, without ever decrypting it.
  - Hospital encrypts sensitive patient data (e.g., tumor size, genetic variants).
  - A (cloud/local) AI model processes the encrypted data without decrypting it.
  - The output is also encrypted and only the hospital can decrypt the result.
- Example
  - Personalized treatment: AI recommends the best therapy combination for a patient with metastatic tumor based on tumor markers and clinical history.
  - With HE all patient inputs (labs, pathology, genomics) stay encrypted while the AI makes a decision.
- Enables safe(r) cloud AI use



https://www.linkedin.com/pulse/future-proofing-privacy-securing-ai-llms-data-mark-kovarski-hw31c/

Combined and Emerging Approaches

- Use of synthetic data in AI model training
- AI model training on anonymized/pseydonimized data
- Hybrid PETs
  - Combination of PETs based on use case requirements

Combined and Emerging Approaches

FL + DP in Oncology

- FL allows hospitals to collaboratively train an AI model (e.g., for tumor detection) without centralizing data
- Federated updates can leak patterns if intercepted.
- DP Injects noise into model updates before they are shared
- DP Ensures that individual patient contributions are mathematically untraceable

SMPC + HE for Cross-Border Trials

- In international clinical trials (e.g., rare cancers), data may be fragmented and subject to different jurisdictional controls
- SMPC(Secure Multi Party Computation) splits sensitive computations across different parties so that no single entity sees the full dataset.
- HE allows AI models to be applied directly to encrypted datasets
- Enables complex analyses like survival modeling or biomarker prediction while the raw patient data never leaves its host country or appears in plaintext

Final Takeaways and Challenges

- The use of AI exposes new attack surface layers
- PETs should balance data usage in AI operations vs patient privacy
- Each use case calls for a suitable combination of PETs
- Trust and validation of PET-powered AI tools
- Need for clinical interpretability and regulatory approval

- UNCAN-Connect (HORIZON-MISS-2024-CANCER-01) – "Decentralized Collaborative Network for Advancing Cancer Research and Innovation"