Data Confidentiality vs Federated Learning

Dr. Balázs Pejó

www.crysys.hu

© 2022 CrySyS Lab, BME

Agenda

- MELLODDY
- Federated Learning
- Data Confidentiality
  - Membership Inference
  - Reconstruction Attack
  - Empirical Defenses
  - Secure Aggregation

TRANSACTIONS ON DATA PRIVACY 15 (2022) 87–107

Collaborative Drug Discovery: Inference-level Data Protection Perspective

Balazs Pejo*, Mina Remeli, Adam Arany***, Mathieu Galtier****, Gergely Acs*****

*CrySyS Lab, BME, Hungary, pejo@crysys.hu (Corresponding author)

** University of Cambridge, United Kingdom, mincsek@gmail.com (Work was done while at CrySyS Lab)

***Stadius, KUL, Belgium, adam.arany@kuleuven.be

****Owkin, France, mathieu.galtier@owkin.com

*****CrySyS Lab, BME, Hungary, acs@crysys.hu

Received 30 October 2021; received in revised form 27 April 2022; accepted 29 April 2022

Abstract. Pharmaceutical industry can better leverage its data assets to virtualize drug discovery through a collaborative machine learning platform. On the other hand, there are non-negligible risks stemming from the unintended leakage of participants' training data, hence, it is essential for such a platform to be secure and privacy-preserving. This paper describes a privacy risk assessment for collaborative modeling in the preclinical phase of drug discovery to accelerate the selection of promising drug candidates. After a short taxonomy of state-of-the-art inference attacks we adopt and customize several to the underlying scenario. Finally we describe and experiments with a handful of relevant privacy protection techniques to mitigate such attacks.

Keywords. Drug Discovery; Machine Learning; Privacy; Risk Analysis; Membership Inference

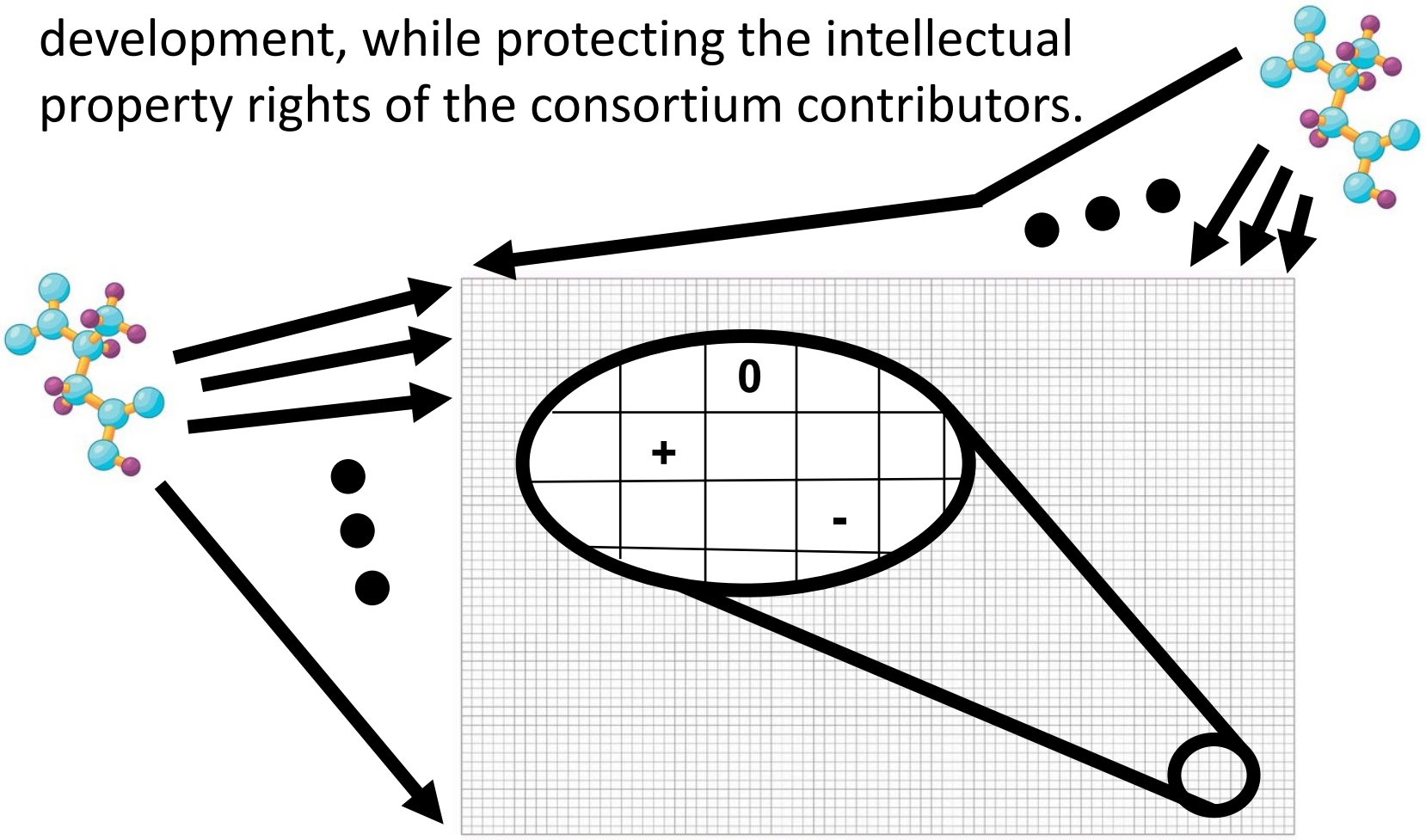
Balázs Pejó; Mina Remeli; Ádám Arany; Mathieu Galtier; Gergely Ács: "Collaborative Drug Discovery: Inference-level Privacy Perspective„ Transactions on Data Privacy (**TDP**), 2022.


MELLODDY
Machine Learning Ledger Orchestration
For Drug Discovery

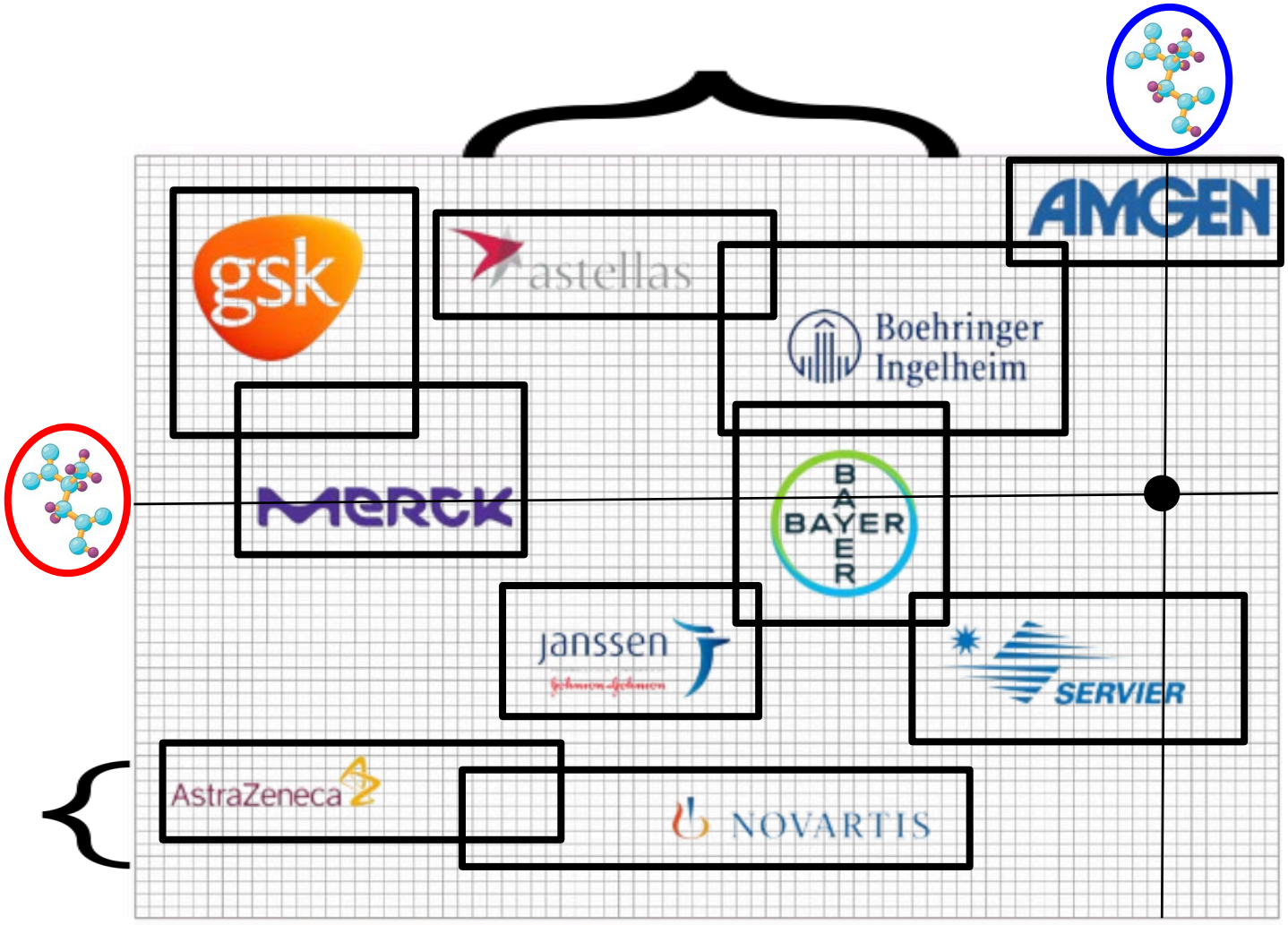
MELLODDY

MELLODDY

- To harness the collective knowledge of the consortium in identifying the most effective compounds for drug development, while protecting the intellectual property rights of the consortium contributors.



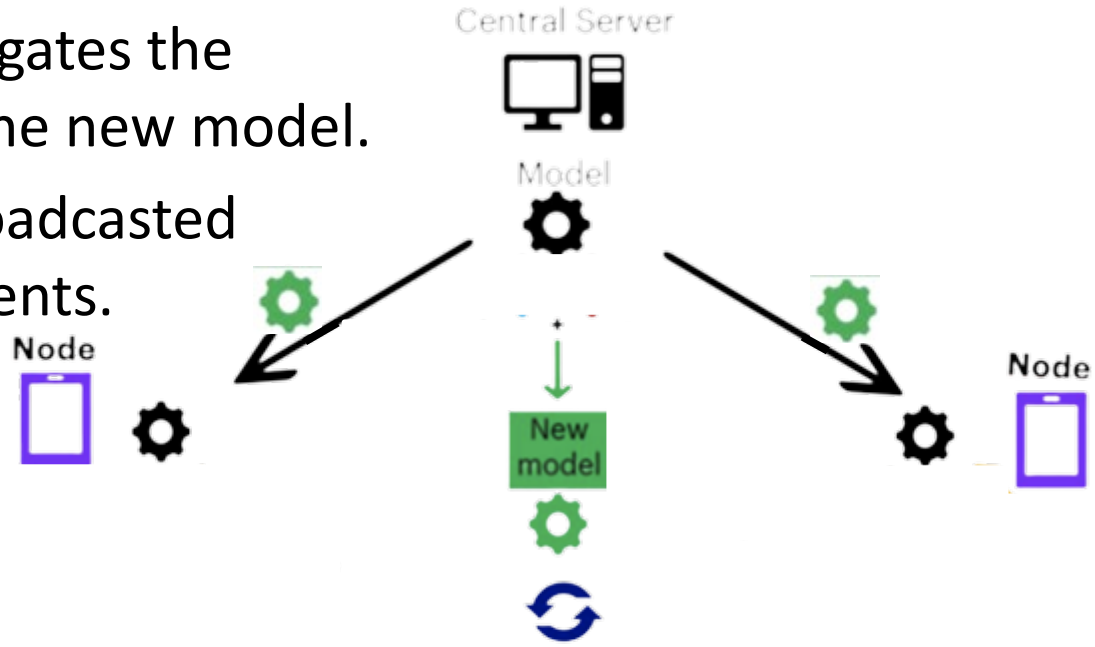
Consortium


gsk
astellas
AMGEN
Boehringer Ingelheim
MERCK
BAYER
janssen
SERVIER
AstraZeneca
NOVARTIS

Machine Learning

Federated Learning

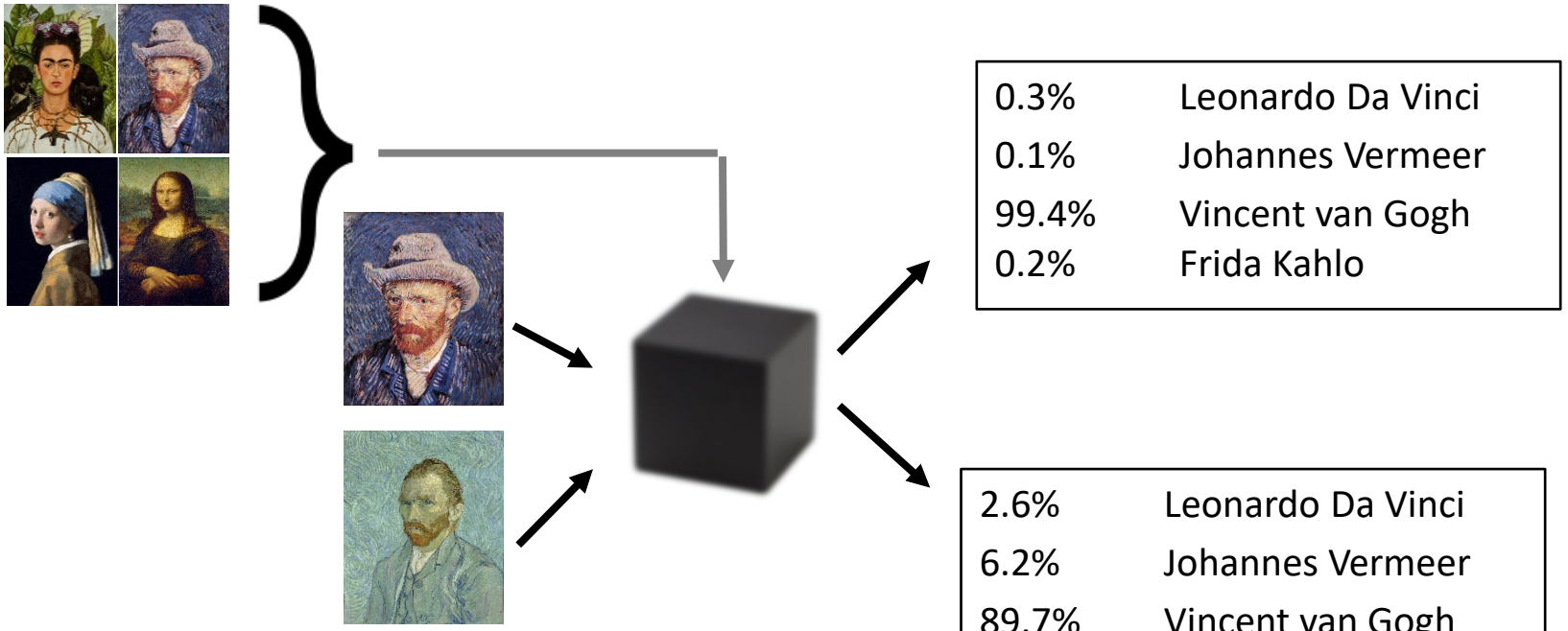
0) The aggregator server initializes the model, i.e., determine the hyperparameters.
1) The aggregator broadcast the model to some clients.
2) Those clients train that model on their local dataset and send the update to the aggregator.
3) The aggregator aggregates the model updates into the new model.
4) The final model is broadcasted to all participating clients.
5) Repeat from 1) until convergence is reached.



Attacks

Membership Inference

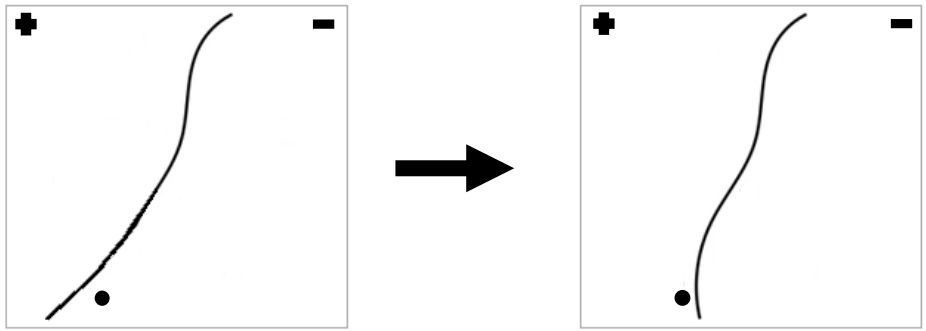
- Leaks the usage of data samples about the overall training data.
- Machine learning models often behave differently on training sample versus a sample that they 'see' for the first time.

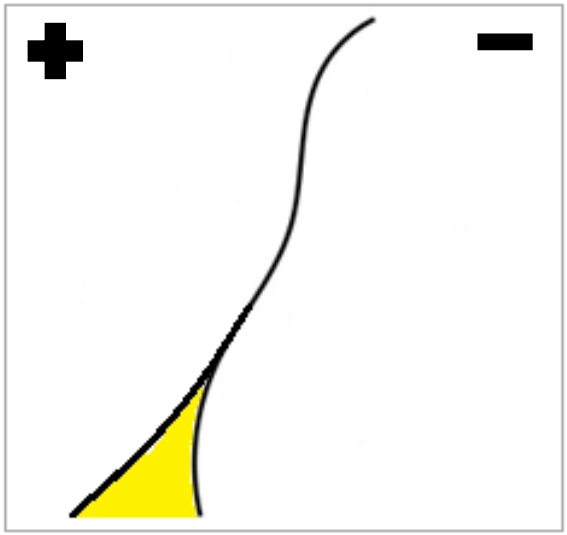


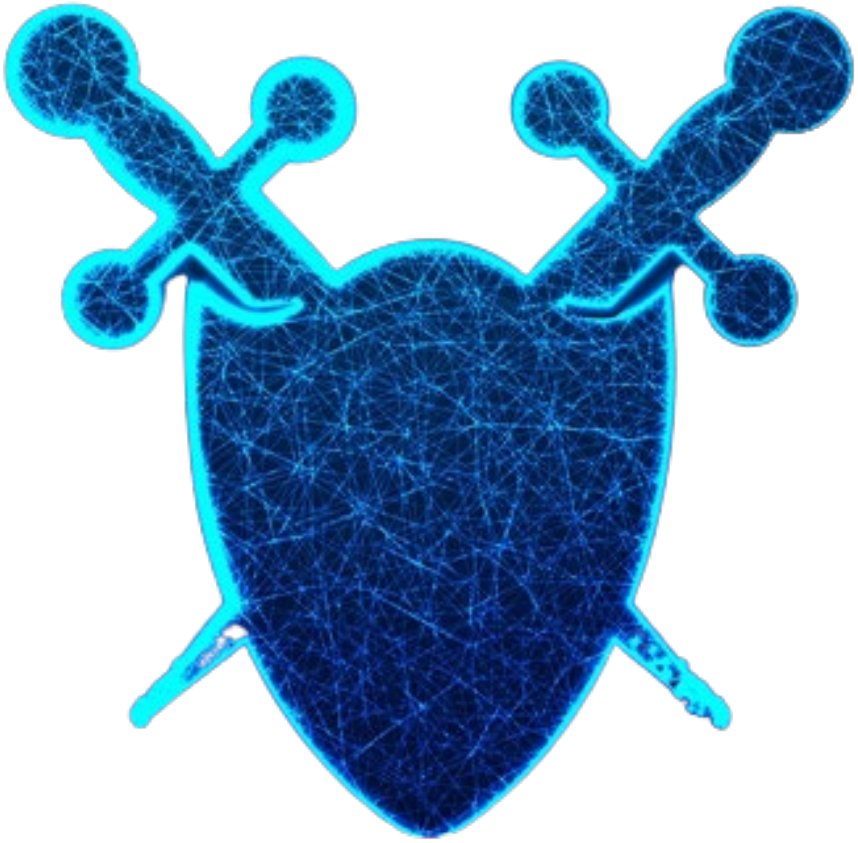
- Used to audit ML models.

Reconstruction Attack

- Recovers “exact” training samples using the model updates.
- During training, the model is updated with the change corresponds to a data sample.
  - Data is given (•), model is given (⁄).
  - Change is computed (⁄).
- An attacker can swap what is given and what is computed.
  - Model is given (⁄), change is given (⁄).
  - Data is computed (•).



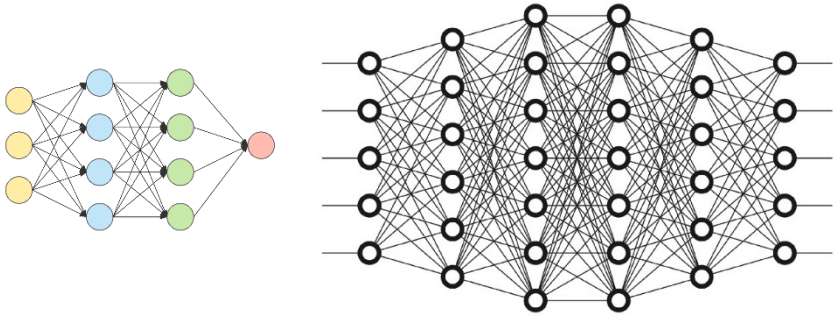




Defenses

Empirical Defenses

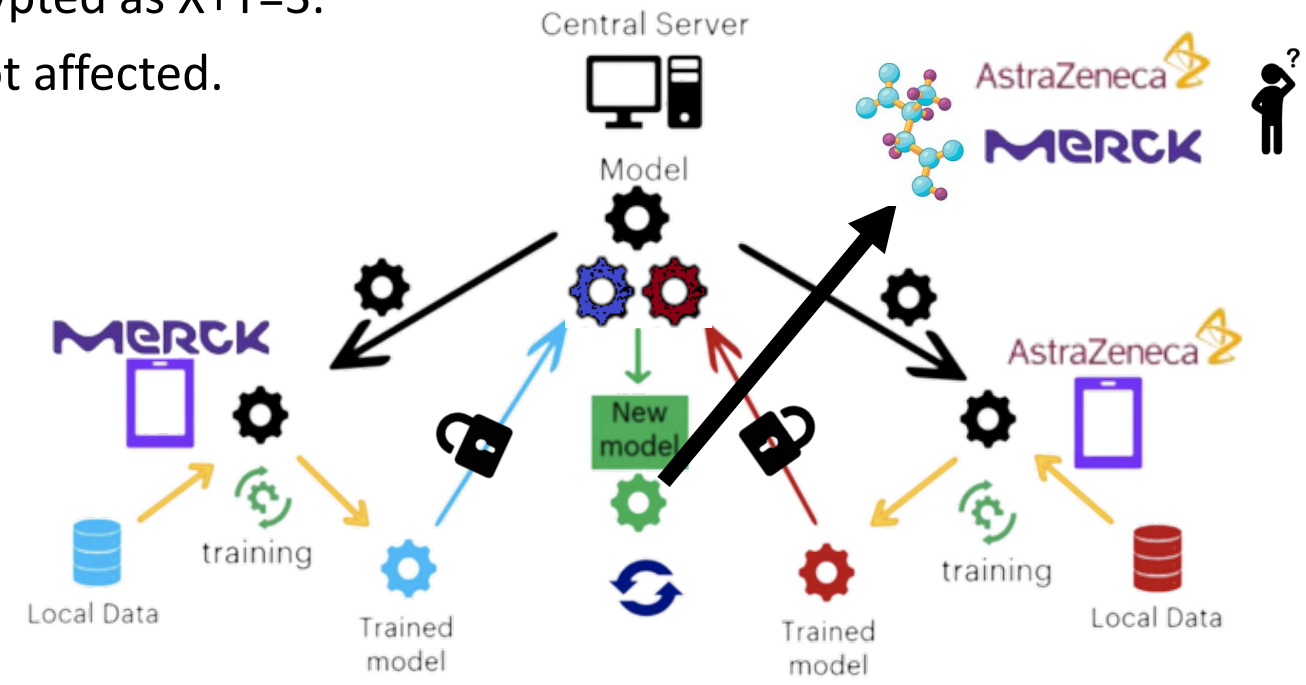
- Pre-Processing
  - Remove Sensitive Data
  - Include Fake Data
- Post-Processing
  - Round probabilities
  - Only return predictions
- Before Training
  - Chose Adequate Model
- During Training
  - Use Randomization
  - Use Compression



Secure Aggregation

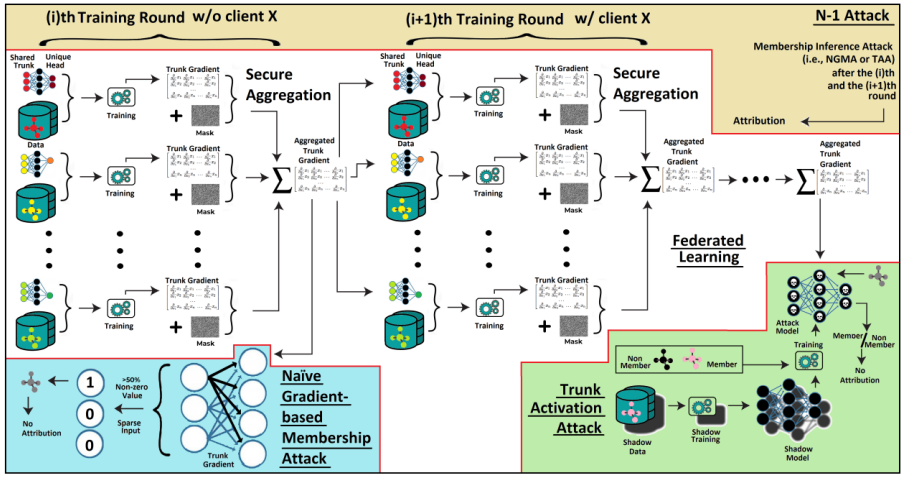
- In FL the server learns the individual changes; hence, it can get information about underlying sensitive data.
- Secure Aggregation hides individual gradients with masks which cancel out after aggregation.
  - Relies on cryptography.
    E.g., 1+2=3 encrypted as X+Y=3.
  - Final model is not affected.

- Attribution is not possible without background knowledge.

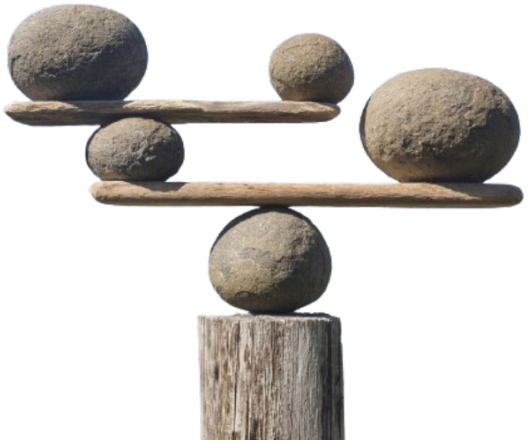


Conclusion

- Machine Learning models potentially leak sensitive information about the underlying training data.
  - Membership Inference reveals the usage of a data sample.
  - Reconstruction Attack reverse engineers the training data itself.



- Defense techniques exists, but they come with a compromise.
  - The stronger the defense, the more it effects the performance of the model.



- https://www.tdp.cat/issues21/abs.a449a21.php