NephroCAGE Federated Learning Infrastructure for Transatlantic Collaboration

Prof. Dr. Ali Sunyaev
Professor at KIT and Director at AIFB
Montréal, August 2nd, 2023

NEPHROCAGE
German-Canadian consortium on AI for improved kidney transplantation outcome
3nd International NephroCAGE Symposium, Aug 2, 2023

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag

Critical Information Infrastructures Research Group

We study internet technologies—their design, use, and symbiosis with our society.

Our research work accounts for the multifaceted use contexts of digital technologies with research on human behavior affecting information systems and vice versa.

To learn more, visit our website: https://cii.aifb.kit.edu/

NEPHROCAGE
German-Canadian consortium on AI for improved kidney transplantation outcome
3rd International NephroCAGE Symposium, Aug 2, 2023

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag

Our Project Team

M. Sc. Konstantin Pandl

- Research associate at KIT since 2019
- M. Sc. in electrical engineering and information technology
- Research interests: machine learning, distributed systems

M. Sc. Florian Leiser

- Research associate at the KIT since 2021
- M.Sc. in information systems
- Research interests: expert decision-making, informed ML, federated learning

Dr. Scott Thiebes

- Research associate with since 2014
- M.Sc. In information systems in 2014
- Research interests: health information systems, digital health, trustworthy AI and privacy

Prof. Dr. Ali Sunyaev

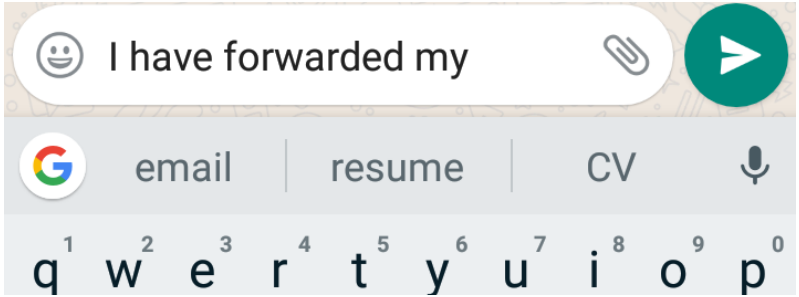
- Professor at KIT since 2018
- Research interests: trustworthy artificial intelligence, innovative health IT solutions

Supported by: Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

What is Federated Learning?

Introduced by Google in 2017

Initial use case: high-quality, Machine Learning (ML)-based word suggestions for the Android keyboard



Problem:

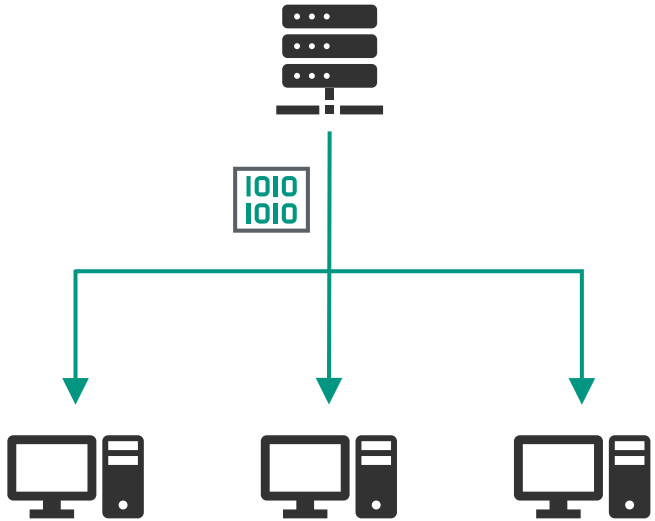
ML process typically runs on a large data set in the cloud

Keyboard inputs are too sensitive to share them with a cloud server

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag

Procedure of (Centralized) Federated Learning

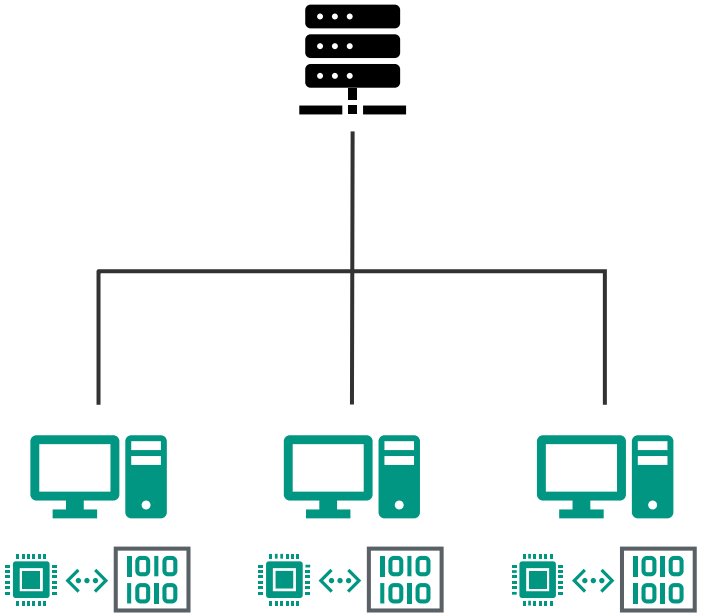
1. Central server broadcasts the model to all clients
2. Clients train the model with their local data
3. Clients send their updated model back to the server
4. Central server aggregates model updates and generates an updated model



Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

Procedure of (Centralized) Federated Learning

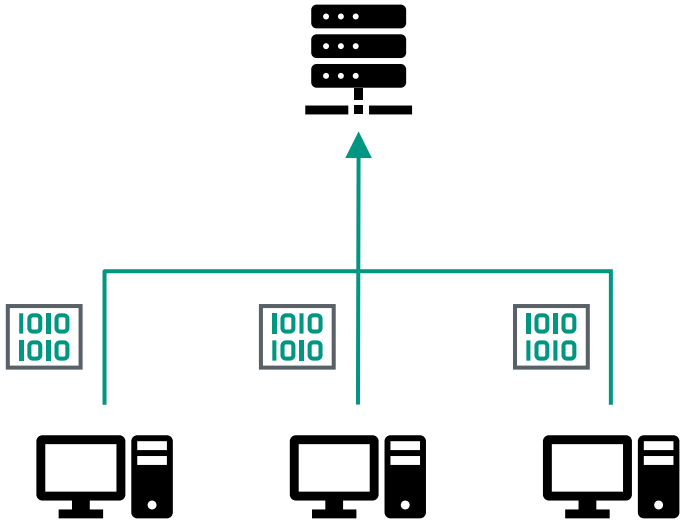
1. Central server broadcasts the model to all clients
2. Clients train the model with their local data
3. Clients send their updated model back to the server
4. Central server aggregates model updates and generates an updated model



Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

Procedure of (Centralized) Federated Learning

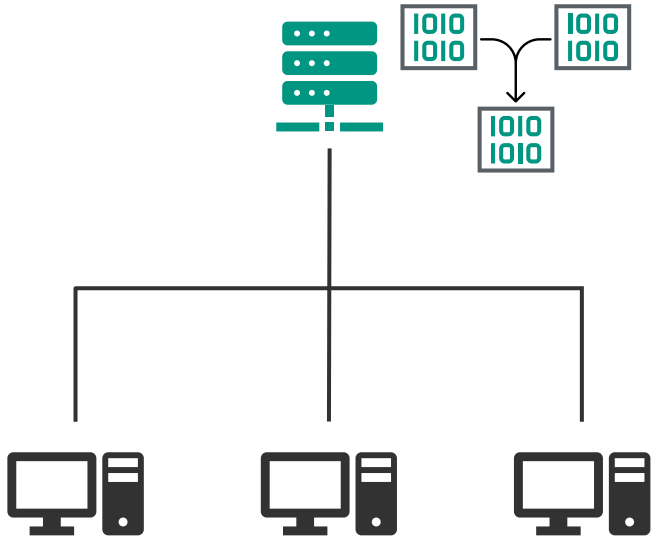
1. Central server broadcasts the model to all clients
2. Clients train the model with their local data
3. Clients send their updated model back to the server
4. Central server aggregates model updates and generates an updated model



Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

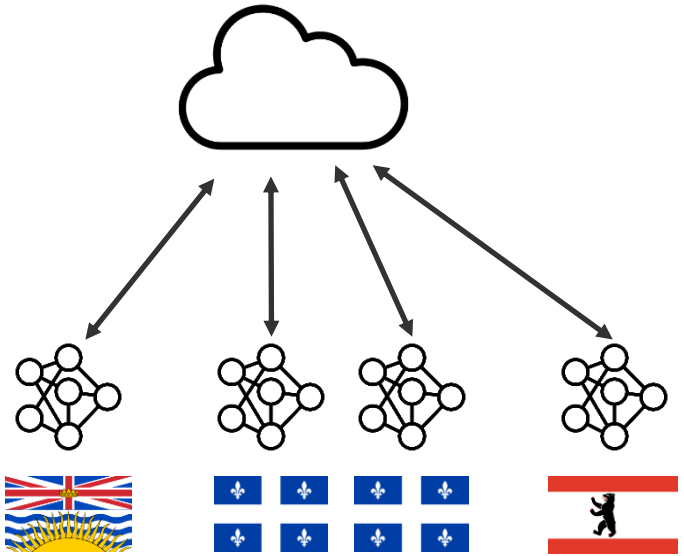
Procedure of (Centralized) Federated Learning

1. Central server broadcasts the model to all clients
2. Clients train the model with their local data
3. Clients send their updated model back to the server
4. Central server aggregates model updates and generates an updated model

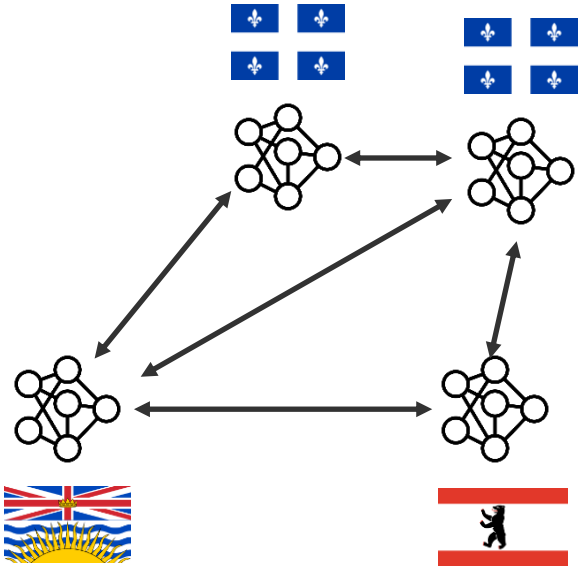


Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

Centralized and Decentralized Learning Infrastructure



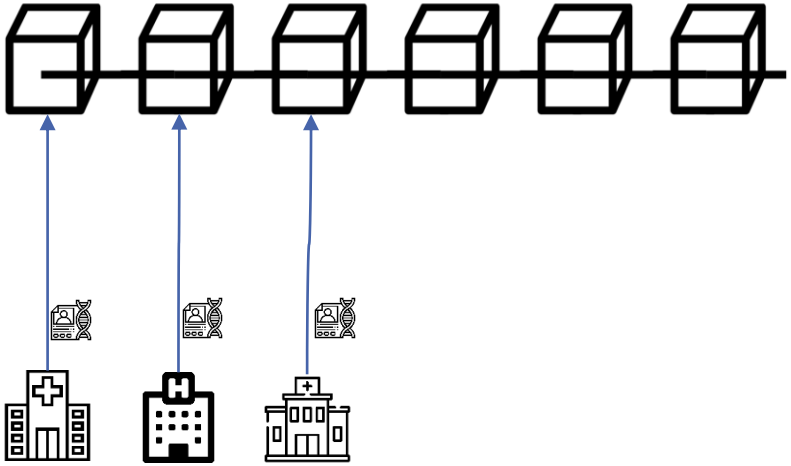
Low failure safety, low transparency, raises model ownership questions



Higher failure safety, high transparency and auditability, but higher network/storage cost

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

Introducing DLT

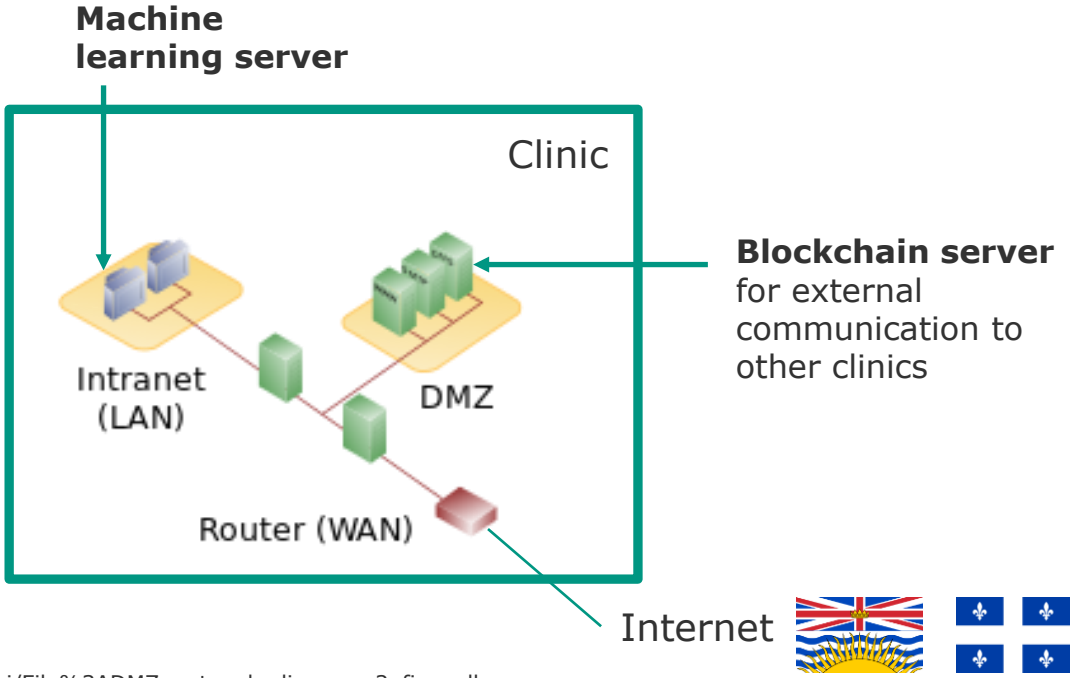


Contents of each block:

- Sending Institution (as hash)
- Message content
- Transmission fee

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

Two-Server Setup



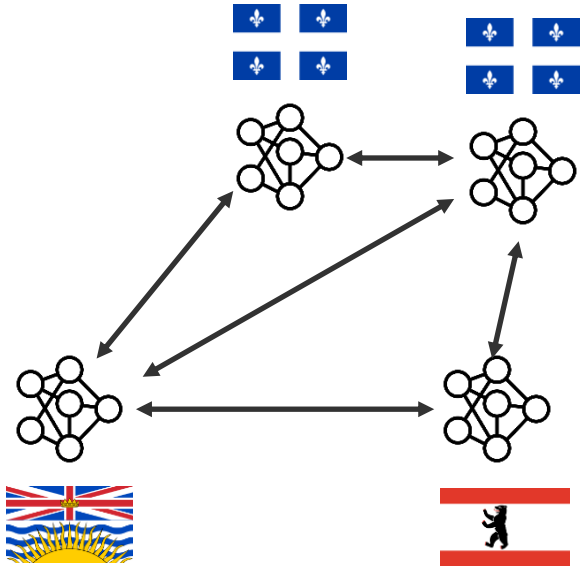
Blockchain server
for external
communication to
other clinics

Image source:
https://commons.wikimedia.org/wiki/File%3ADMZ_network_diagram_2_firewall.svg

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

Decentralized Federated Learning Infrastructure

- Build Ethereum-based DLT
- Decentralized aggregation requires high synchronicity and therefore an organizing clients
- Without synchronicity multiple aggregation models could exist simultaneously



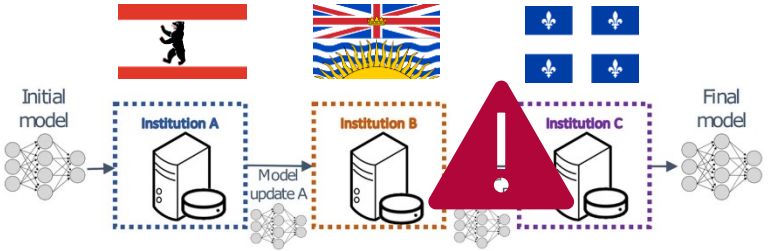
NEPHROCAGE
German-Canadian consortium on AI for improved kidney transplantation outcome
3nd International NephroCAGE Symposium, Aug 2, 2023

Hasso Plattner Institut · KIT Karlsruher Institut für Technologie · UBC The University of British Columbia · Université de Montréal · Polytechnique Montréal · GenomeQuébec · PIRCHE · Charité Universitätsmedizin Berlin · Centre universitaire de santé McGill McGill University Health Centre · CHUM · GenomeCanada · Genome British Columbia

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag

Decentralized Federated Learning Infrastructure

- Build Ethereum-based DLT
- Decentralized aggregation requires high synchronicity and therefore an organizing clients
- Without synchronicity multiple aggregation models could exist simultaneously
- Institutional incremental learning avoids organization and allows institutions to train whenever new data is available



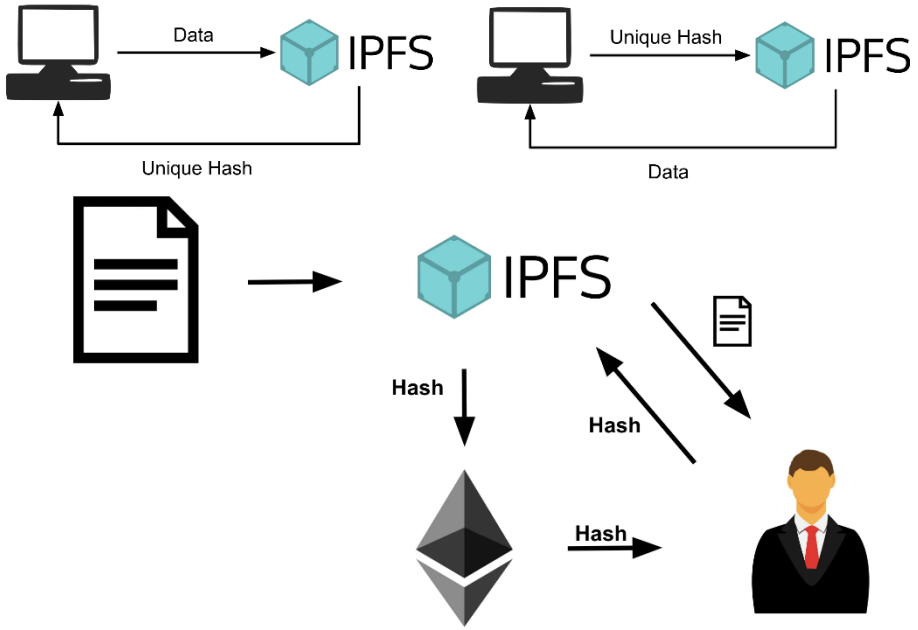
(c) Data-private Collaborative Learning using Institutional Incremental Learning

Image Source: Sheller, M.J., Edwards, B., Reina, G.A. *et al.* Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep* **10,** 12598 (2020). https://doi.org/10.1038/s41598-020-69250-1

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag

IPNS as a peer-to-peer file sharing system

- First efforts were just changing transaction information in blockchain
- For further scalability, and to keep the blockchain lightweight, we implemented the InterPlanetary File System (IPFS)
- Runs on the blockchain server to retrieve files by hashes
- Extending with InterPlanetary Name System (IPNS) for updating (otherwise immutable) files
- We developed additional hashes indicating current training of model (reduced potential overlap to 15 seconds)

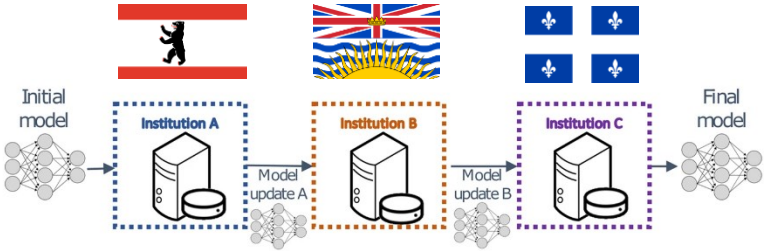


Images from: https://medium.com/pinata/ethereum-and-ipfs-e816e12a3c59

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag

Decentralized Federated Learning Infrastructure

- Build Ethereum-based DLT
- Decentralized aggregation requires high synchronicity and therefore an organizing clients
- Without synchronicity multiple aggregation models could exist simultaneously
- Institutional incremental learning avoids organization and allows institutions to train whenever new data is available



(c) Data-private Collaborative Learning using Institutional Incremental Learning

Image Source: Sheller, M.J., Edwards, B., Reina, G.A. *et al.* Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep* **10,** 12598 (2020). https://doi.org/10.1038/s41598-020-69250-1

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag

Ongoing Research Efforts

Icons from https://www.flaticon.com

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

NephroCAGE Federated Learning Infrastructure for Transatlantic Collaboration

Prof. Dr. Ali Sunyaev
Professor at KIT and Director at AIFB
Montréal, August 2nd, 2023

NEPHROCAGE
German-Canadian consortium on AI for improved kidney transplantation outcome
3nd International NephroCAGE Symposium, Aug 2, 2023

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag