NephroCAGE Work Package 1 – Federated Learning Infrastructure

Konstantin Pandl

Our project team

M. Sc. Konstantin Pandl

- Research associate at KIT since 2019
- M. Sc. in electrical engineering and information technology in 2018
- Research interests: machine learning, digital health, distributed systems

M. Sc. Scott Thiebes

- Research associate with Prof. Sunyaev since 2014
- M. Sc. in information systems in 2014
- Research interests: digital health, patient-centric health care, gamification

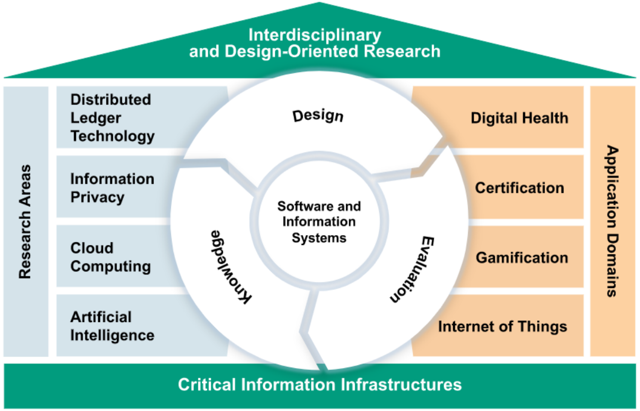
Prof. Dr. Ali Sunyaev

- Professor at KIT and director of the AIFB since 2018
- Previous professorships at the University of Cologne and Kassel
- Research interests: trustworthy artificial intelligence, innovative health IT solutions

Our research group and KIT

Our research group

- We study internet technologies – their design, their usage, and their symbiosis with our society.



Karlsruhe Institute of Technology

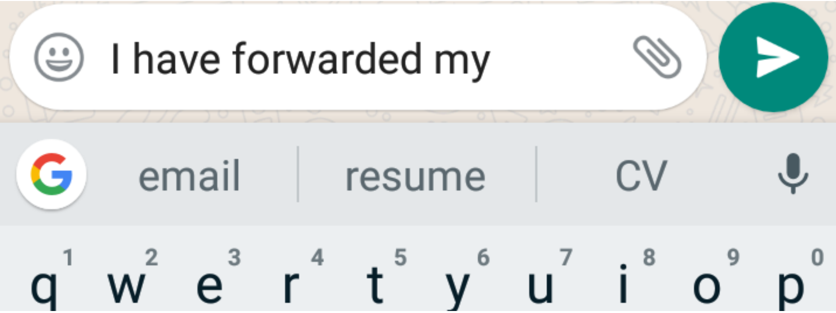
- Located in Karlsruhe, Germany
- One of the largest research and educational institutions in Germany, ca. 25k students and 10k employees
- Originated from the University of Karlsruhe founded in 1825

Goal of the work package

- Design a **federated learning infrastructure**, that
  - preserves the **confidentiality** of the training data
  - is based on blockchain / distributed ledger technology, and thus, **robust** and **auditable**
- **Develop** this infrastructure
- **Deploy** this infrastructure in the clinics
- **Evaluate** its utility

History of federated learning

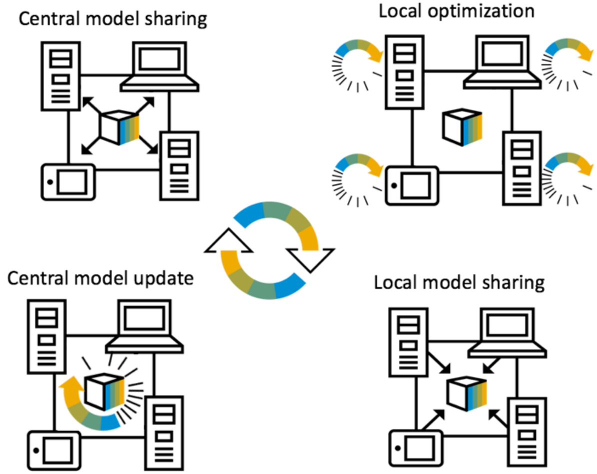
- Introduced by Google in 2017
- Initial use case: high-quality, machine learning (ML)-based word suggestions for the Android keyboard



- Problem: ML process typically runs on a large data set in the cloud, but keyboard inputs are too sensitive to share them with a cloud server
- Solution: federated learning

Functional principle of federated learning

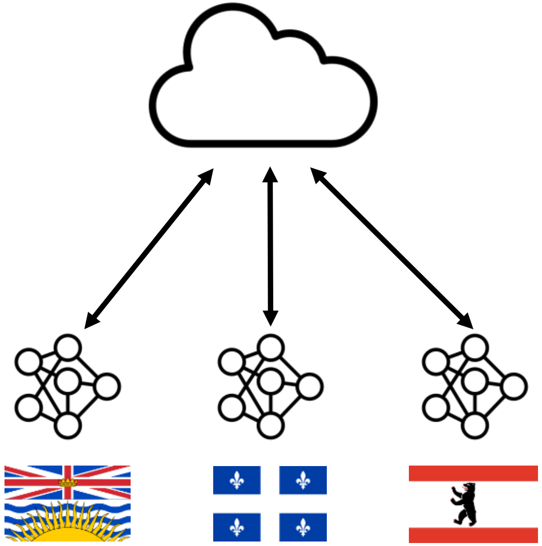
- Goal: ML across private data silos
- Key idea: train ML models on local data and only exchange ML models
- Process consists of repeated rounds, each round comprises 4 steps
- Central model update through averaging the local model parameters



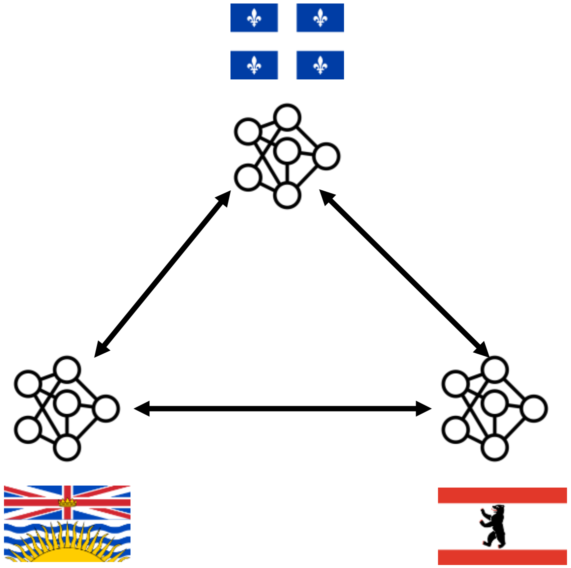
Procedure of a federated learning round

Image: https://medium.com/sap-machine-learning-research/client-sided-differential-privacy-preserving-federated-learning-1fab5242d31b

Centralized and decentralized federated learning



Low failure safety, low transparency, raises model ownership questions



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

Blockchain

- Originated with the emergence of Bitcoin, today a variety of different blockchain networks exist
- In the scope of our project: **private peer-to-peer network** of **institutions**
- Characteristics:
  - **Replicated** ledger (i.e., each institution stores a copy of the blockchain locally)
  - **Immutable** (i.e., data can be written but cannot be removed)
  - Each institution has **equal rights**
  - **Transparent** and, thus, **auditable**
- Goals in the project:
  - **Communication** between the institutions **through the blockchain**
  - **Store ML models** (or a representation such as hashes) on the blockchain ledger

Expected results of federated learning

- Highly anticipated
  - The federated ML model is trained on more data and, thus, performs better than locally trained models on a general test data set
- Potentially
  - The federated ML model may still perform worse on locally generated test data sets
  - The federated ML model is trained on more diverse data and, thus, performs better especially on minorities

Example model evaluation results (dice coefficient) —ProstateX challenge dataset

		ProstateX (n = 343)
Private models	NCI	0.872 ± 0.062*
	SUNY	0.838 ± 0.043*
	UCLA	0.812 ± 0.136*
FL Model		**0.889** $\pm$ 0.036

*Significantly lower than FL model ($P < .001$).

Sarma, K. V., Harmon, S., Sanford, T., Roth, H. R., Xu, Z., Tetreault, J., ... & Arnold, C. W. (2021). Federated learning improves site performance in multicenter deep learning without data sharing. *Journal of the American Medical Informatics Association*.

NephroCAGE
German-Canadian consortium on AI for improved kidney transplantation outcome
1st NephroCAGE Symposium, May 20, 2021

Thank you very much! Any questions?

Konstantin Pandl

Institute of Applied Informatics and Formal Description Methods
Karlsruhe Institute of Technology
konstantin.pandl@kit.edu