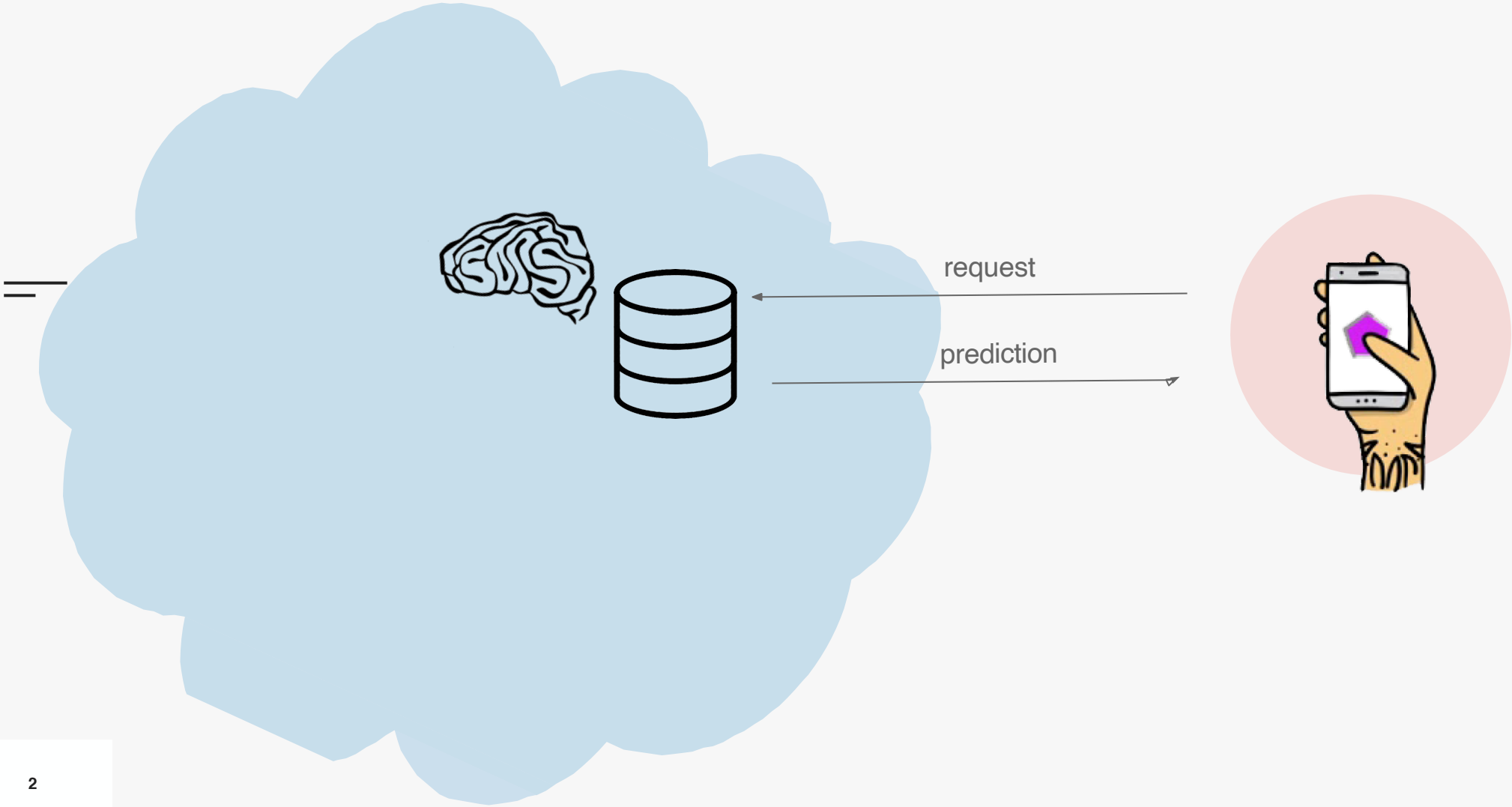
CMPE-295

Federated Learning

For Medical Institutions

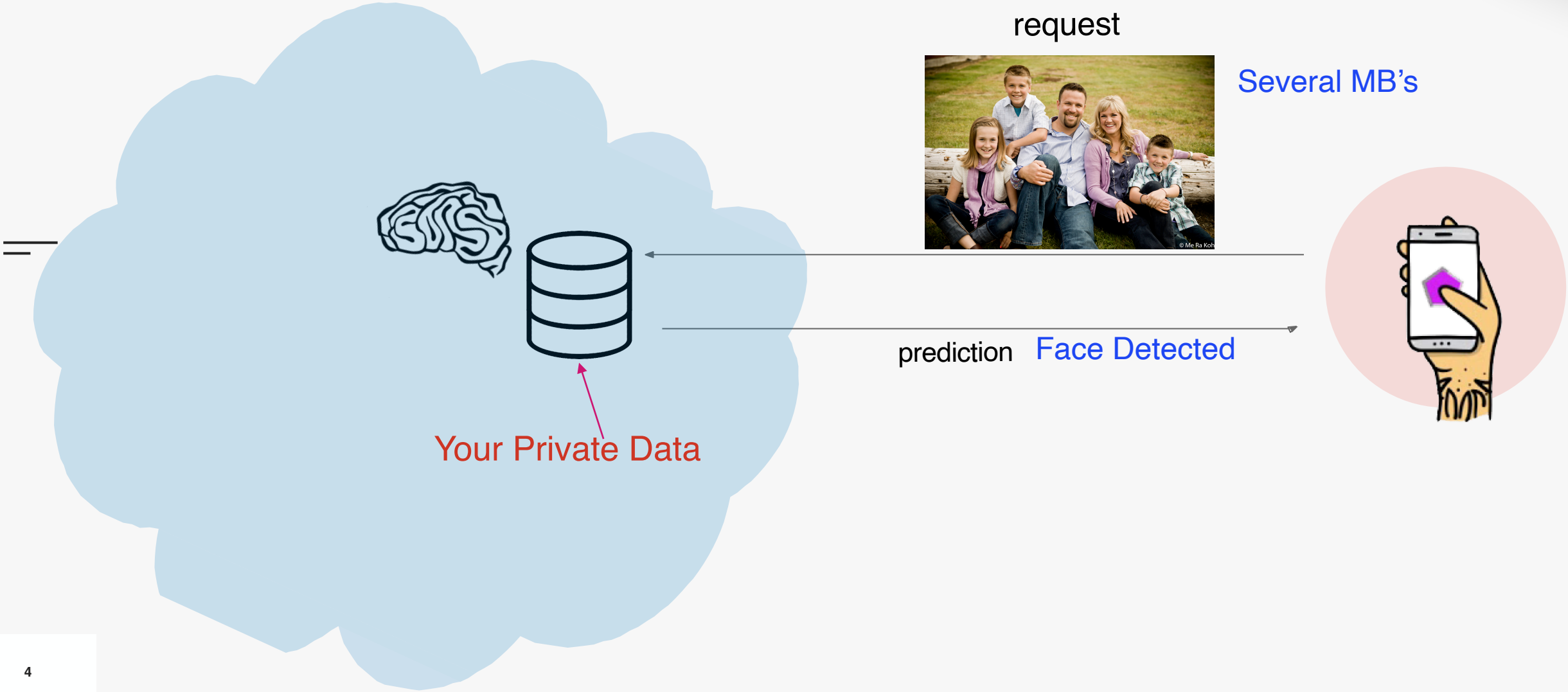
Magdalini Eirinaki
Project Guide

Current System

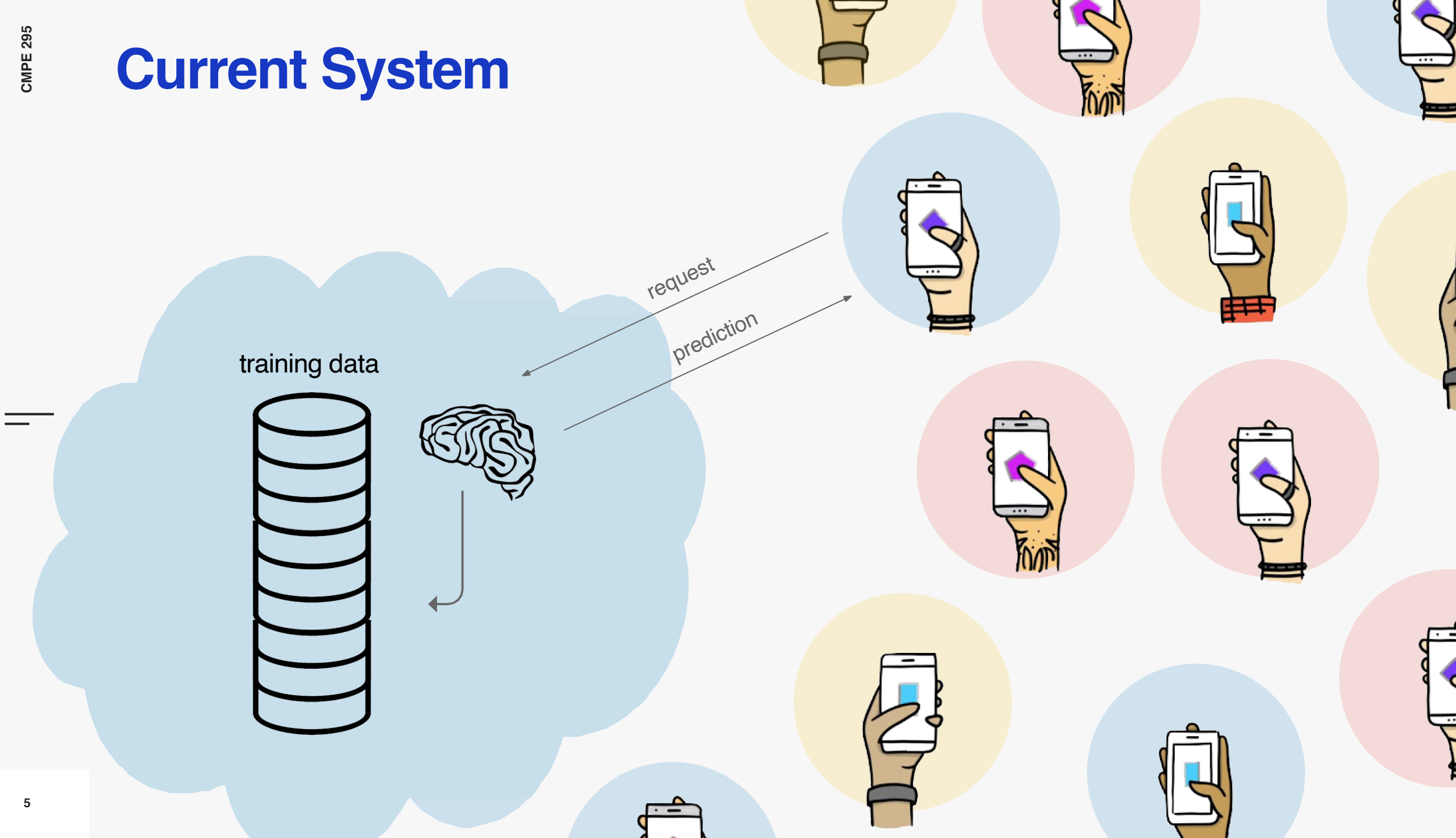


Problems

Current System



Current System



The Solution

Distribute the model, make predictions locally

Advantage of this System

Privacy

Data Privacy is preserved. No personal or sensitive data need to be sent to server.

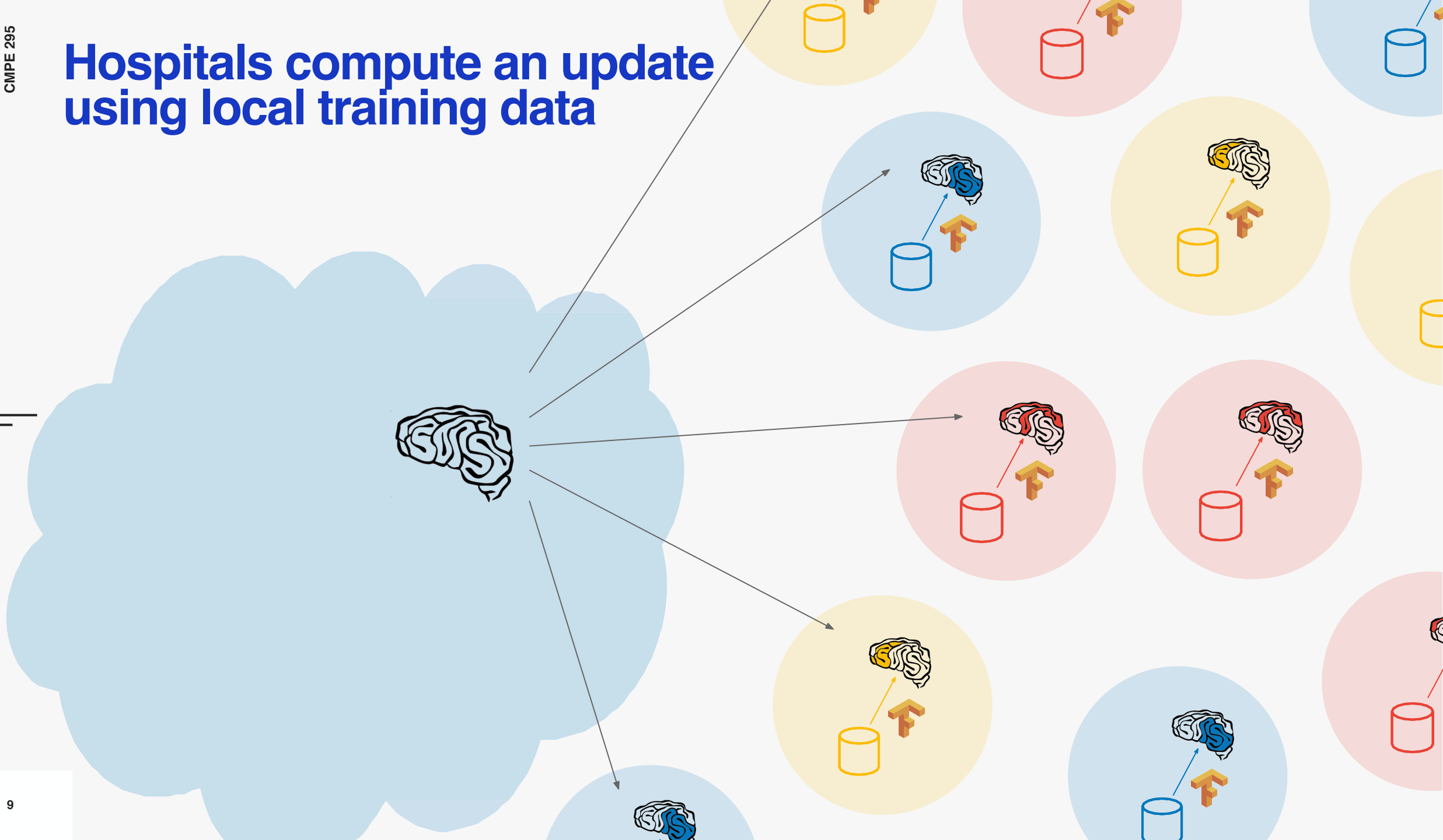
Data Caps

Works best when data is capped when Roaming. **No extra bandwidth required** to transfer full data.

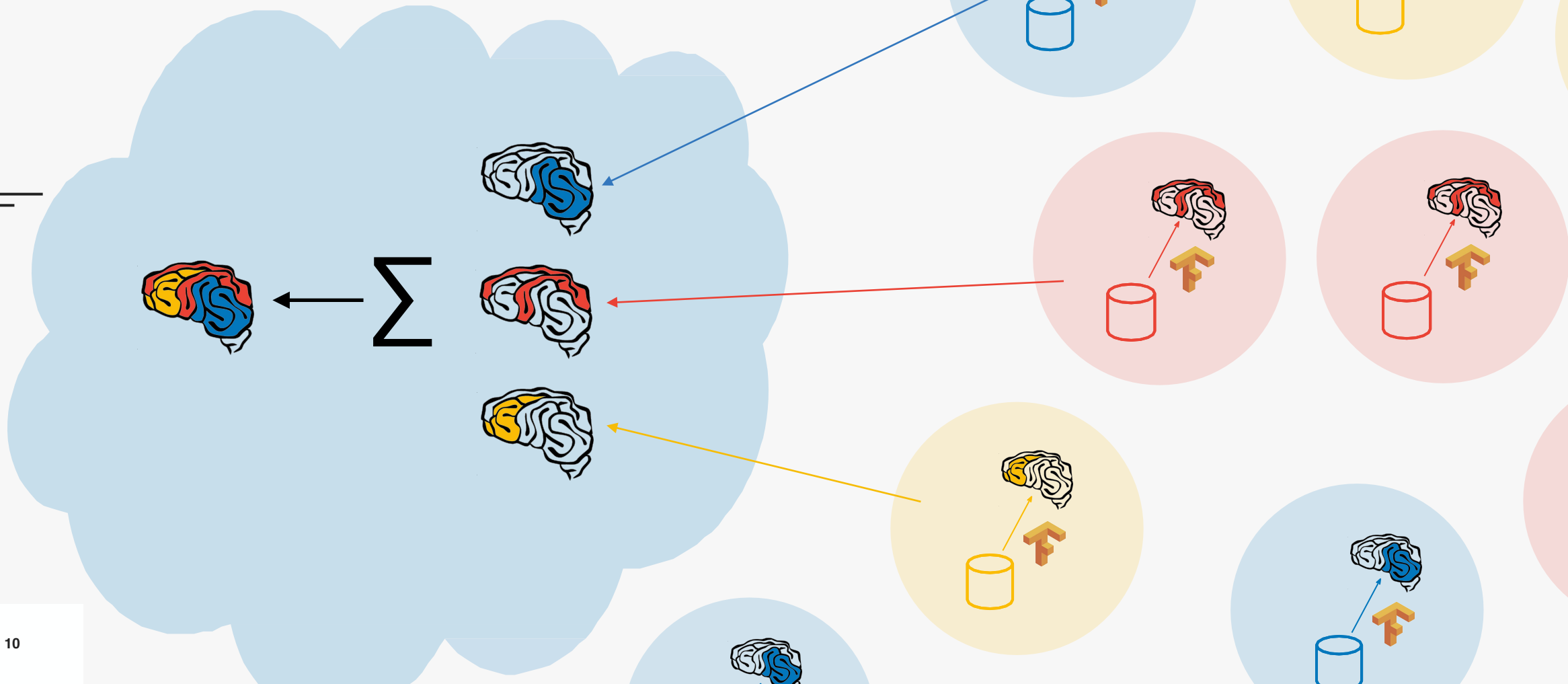
Latency

You can run prediction in Model with virtually **no Latency Penalty**

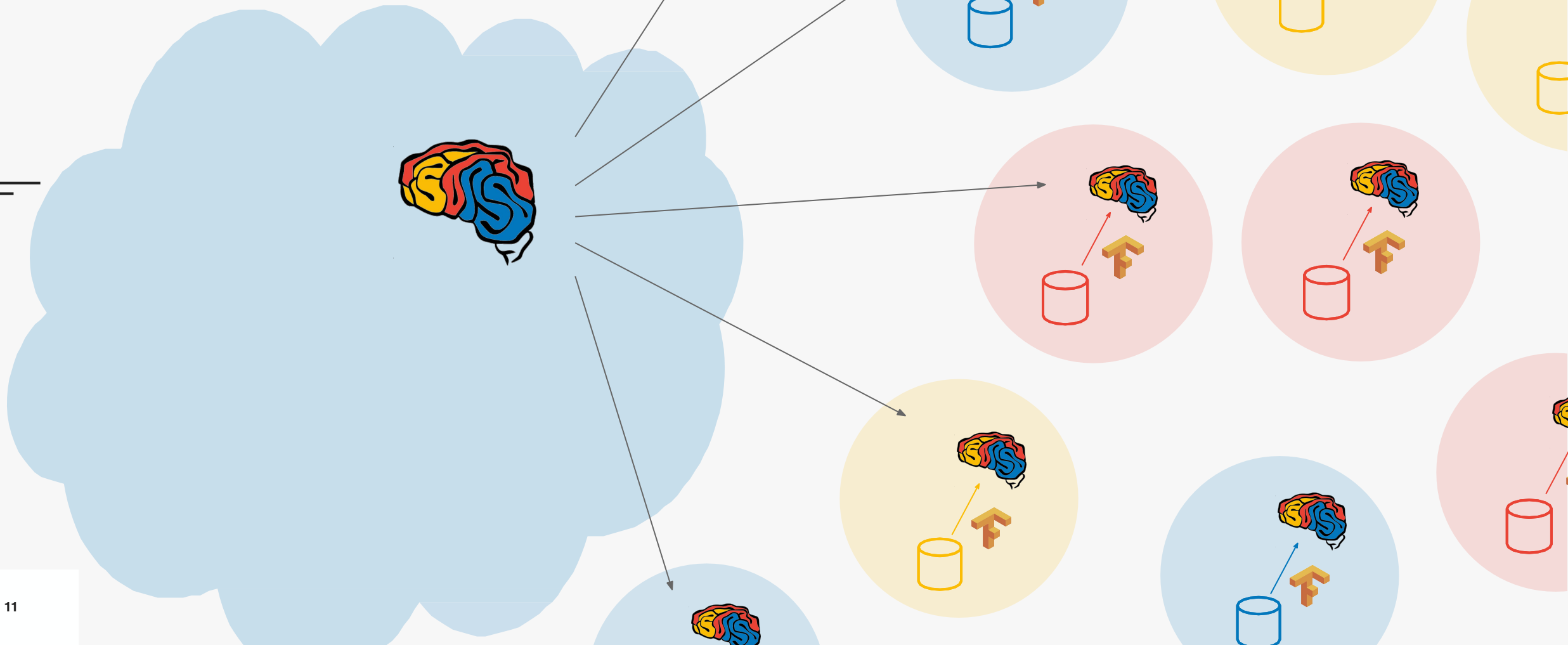
Hospitals compute an update using local training data



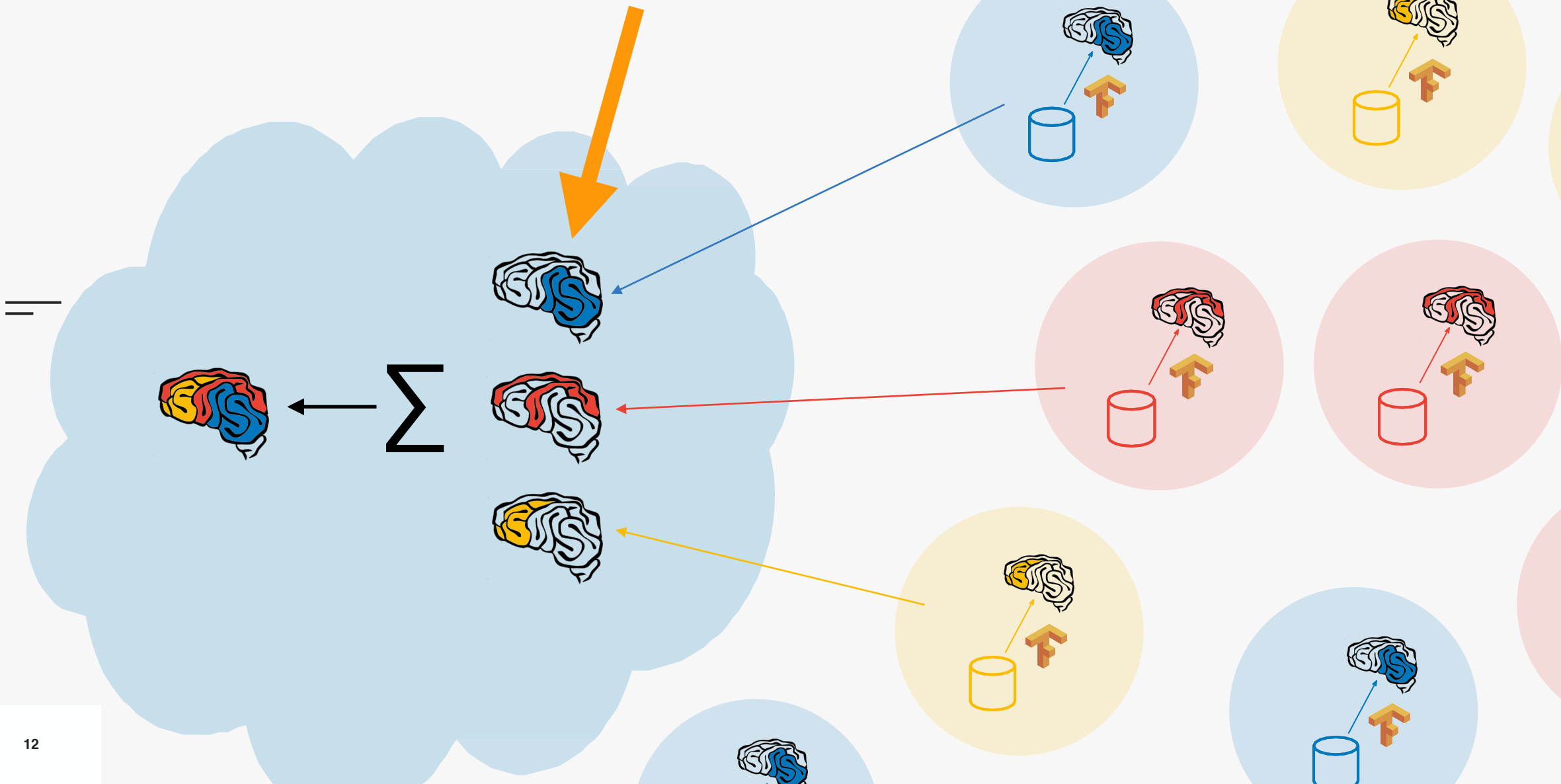
Server aggregates user's updates into a new model.



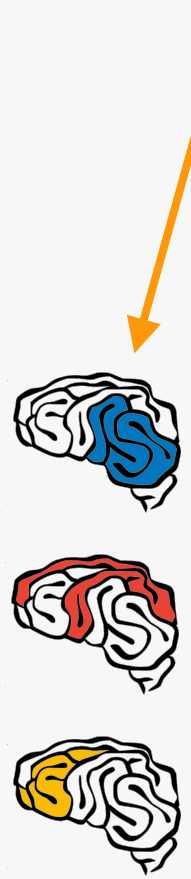
Distribute the Updated model, on which further personal training is computed.



Might these updates contain privacy-sensitive data?



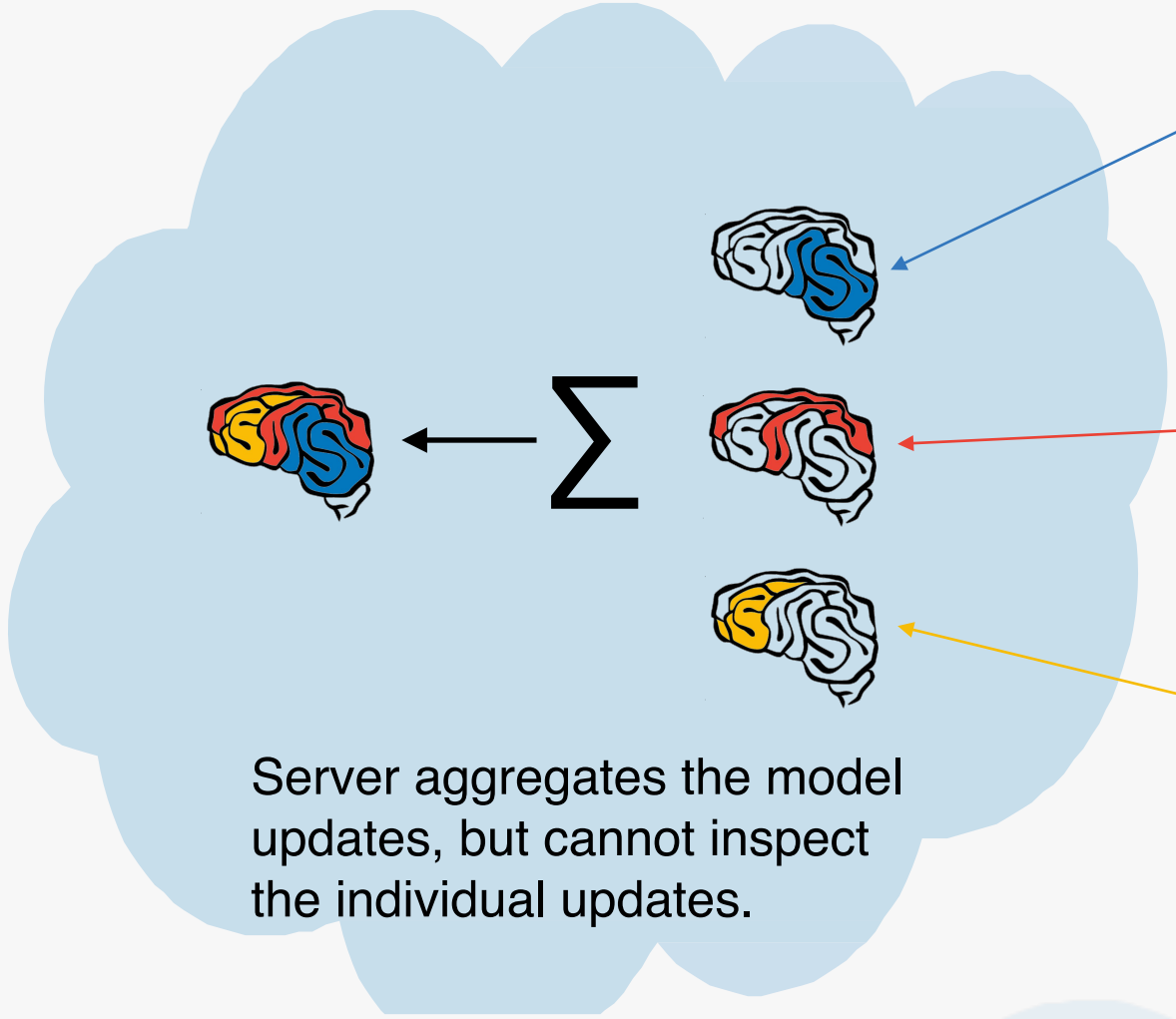
Might these updates contain privacy-sensitive data?



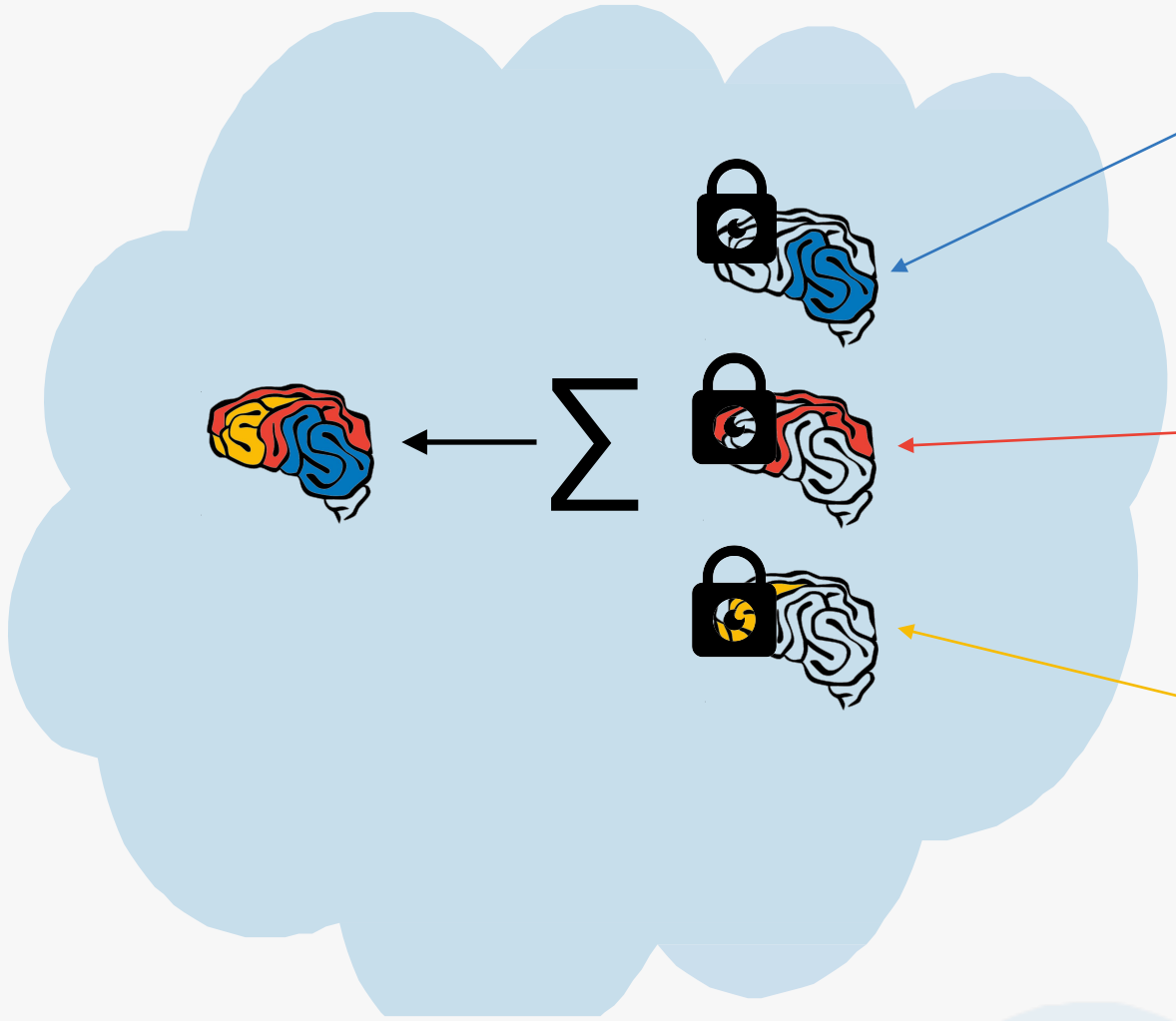
Ephemeral

Updates are incremental & ephemeral, we don't need to store these updates and we only need to receive them and process them and throw them away there is no requirements.

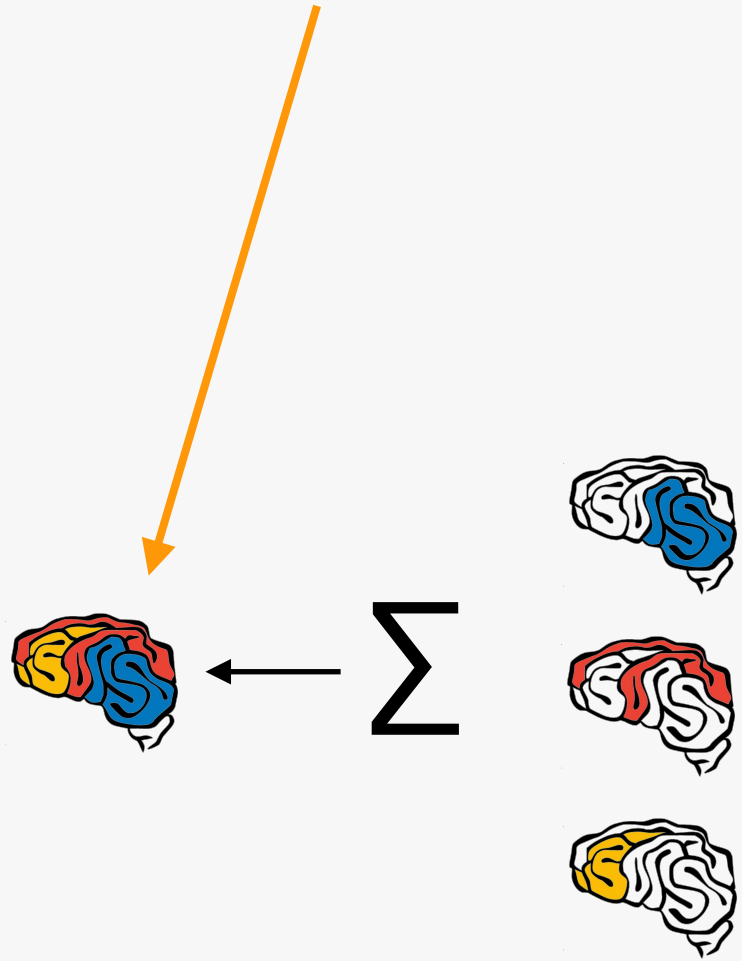
Only in Aggregate



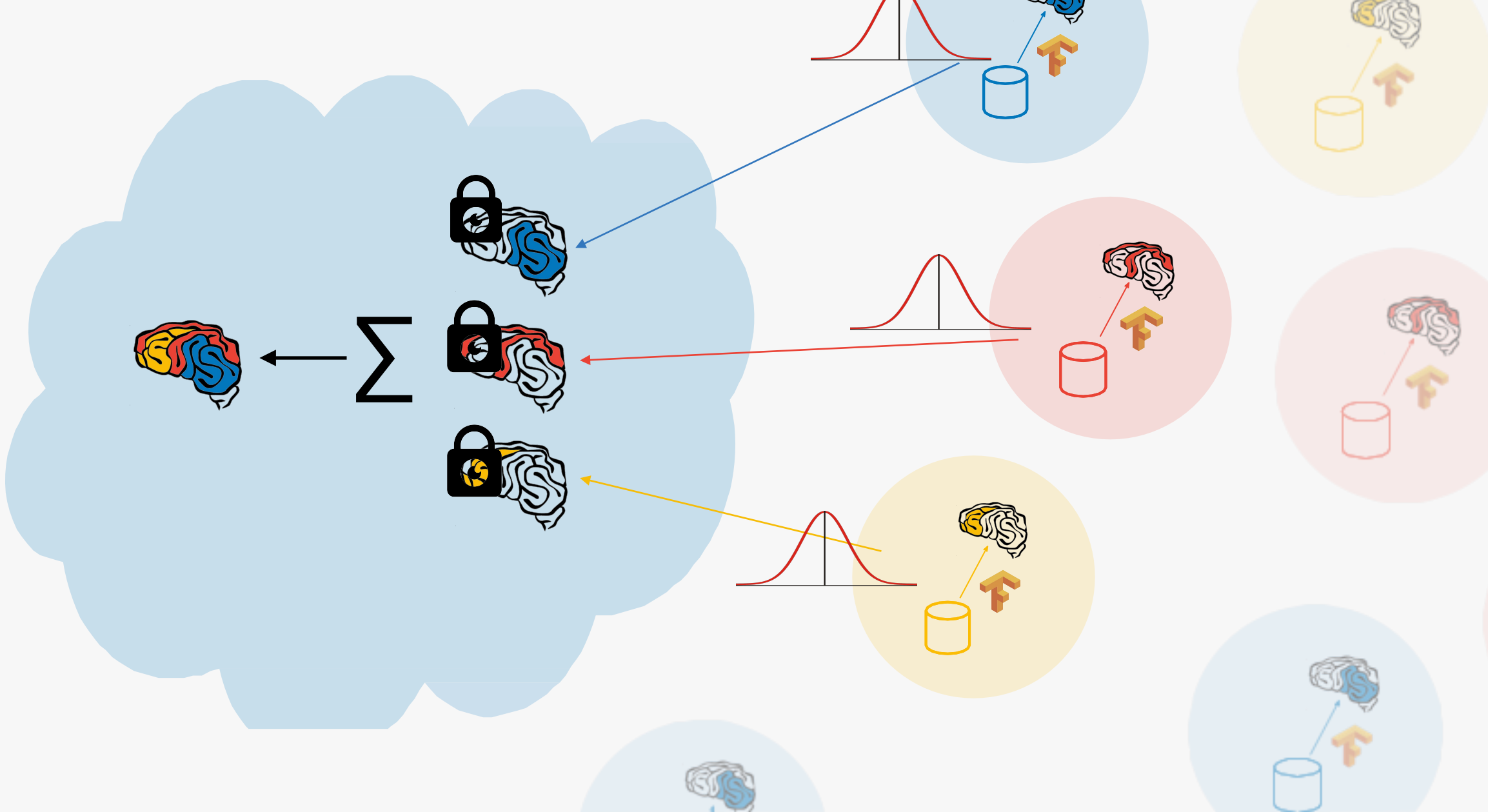
Secure Aggregation



Might the final model memorize a patient's personal data?



Differential Privacy



Applications of Federating learning

What makes a good application?

- Language modeling (e.g., next word prediction) for mobile keyboards.
- Image classification for predicting which photos people will share.
- Safe use of hypersensitive data, like Medical Data.
- Models with Physical Context.
- Intrusion Detection.
- Spam Filters.

Challenges of Federated Learning

Massively Distributed

Training data is stored across a very large number of hospitals

Limited Communication

Only a handful of rounds of unreliable communication with each hospital

Unbalanced Data

Some devices have few examples, some have orders of magnitude more

Highly Non-IID Data

Data on each device reflects one individual's usage pattern

Unreliable Compute Nodes

Devices go offline unexpectedly; expect faults and adversaries

Dynamic Data Availability

The subset of data available is non-constant, e.g. time-of-day vs. country