Federated learning, secure aggregation

“Sciences that involve human beings rather than elementary mathematics… Economists suffer from physics envy over their inability to neatly model human behavior… we should stop acting as if our goal is to author extremely elegant theories, and instead embrace complexity and make use of the best ally we have: the **unreasonable effectiveness of data**”

Halevy et al, “Unreasonable effectiveness of data”

Access to high quality data

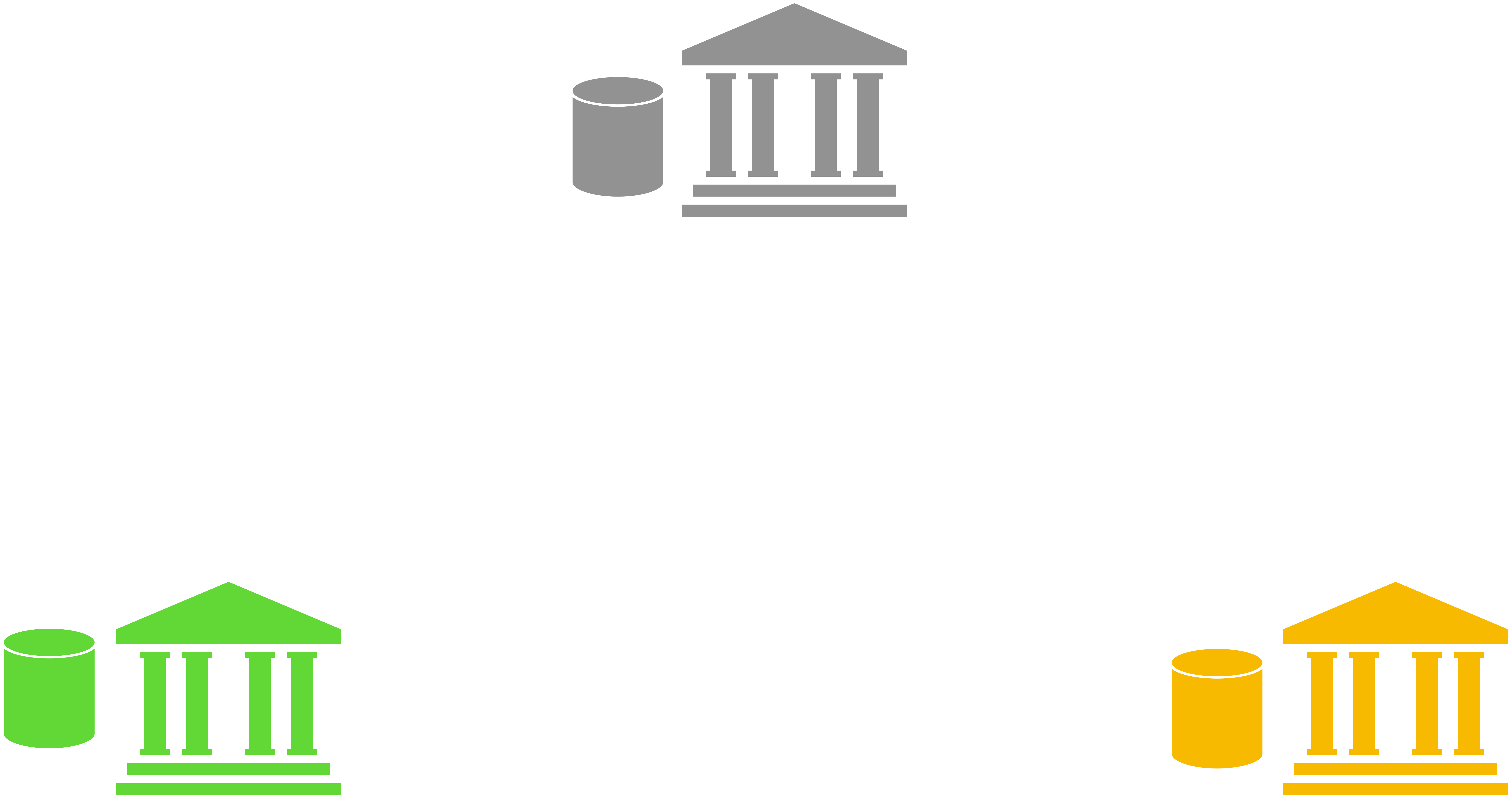
- Often, one party's data is not enough for many applications
  - Customers' mobile devices data for large-scale analytics
  - Banks' customer transaction data to detect money laundering
  - Hospitals' patient data to predict flu outbreaks
- Data is often locked down due to privacy concerns, regulatory constraints, competition

Federated learning

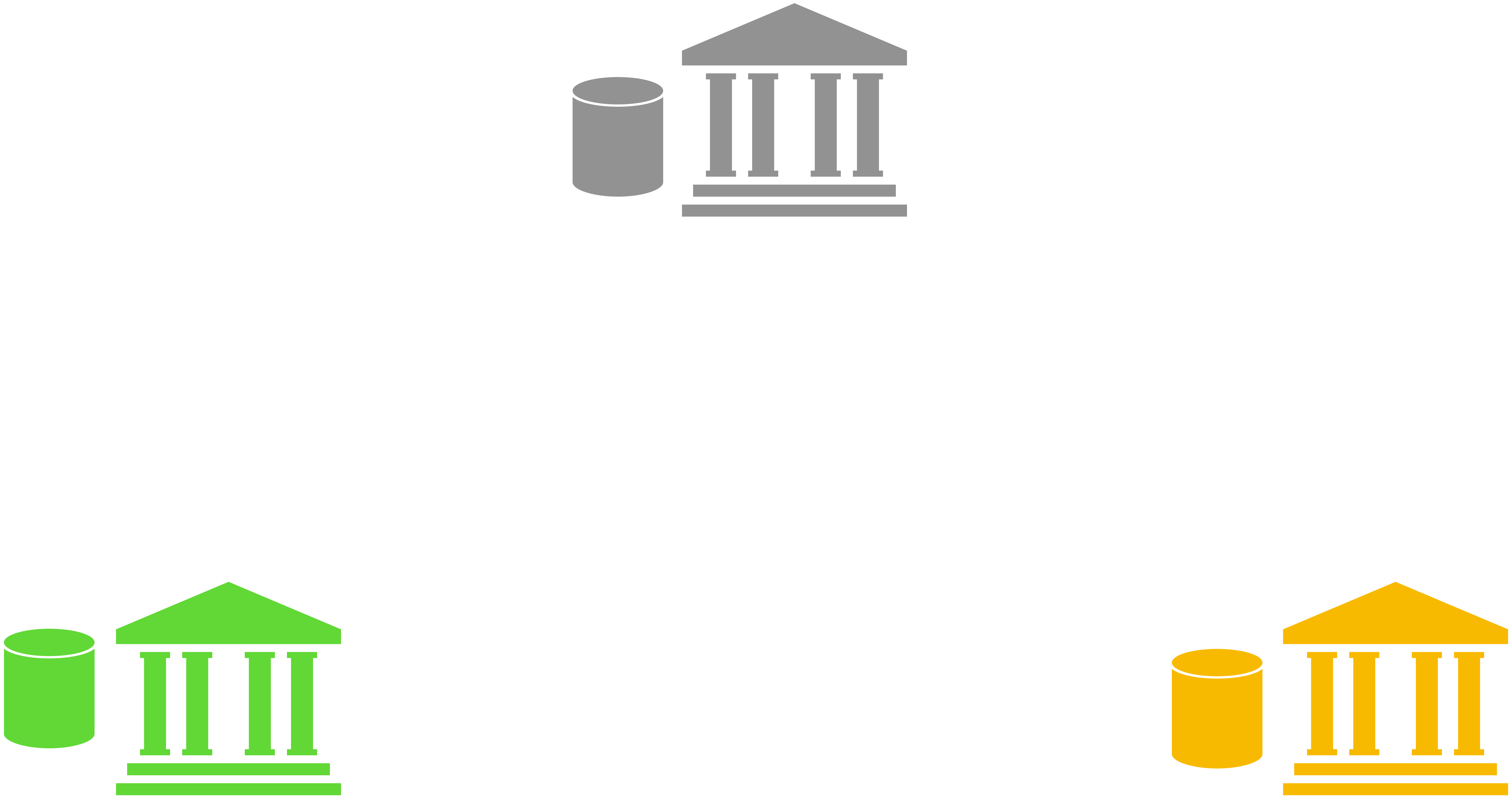
vs.

ML & analytics using secure multiparty computation (MPC)

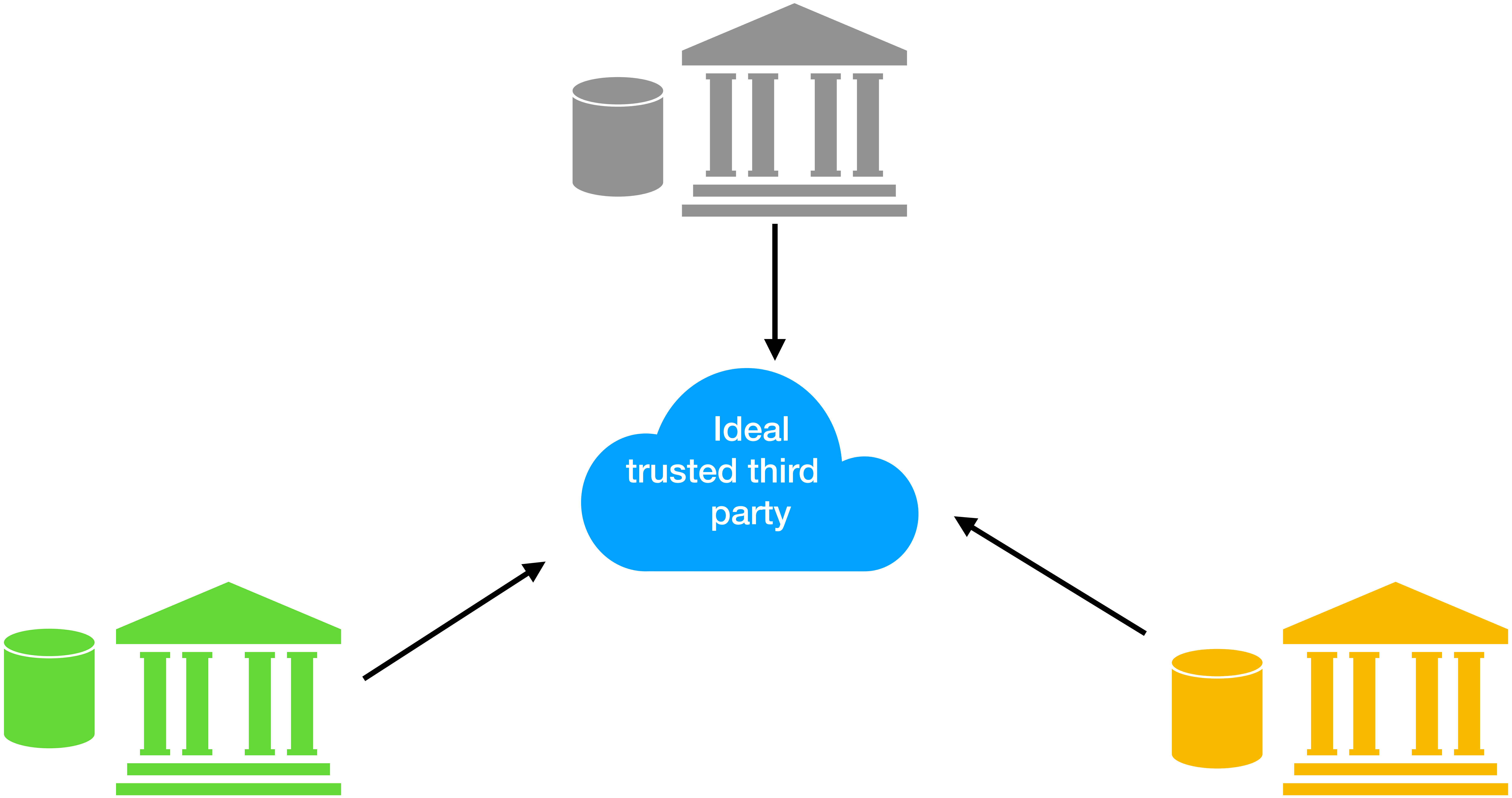
Multiple parties have sensitive local data



Secure multiparty computation



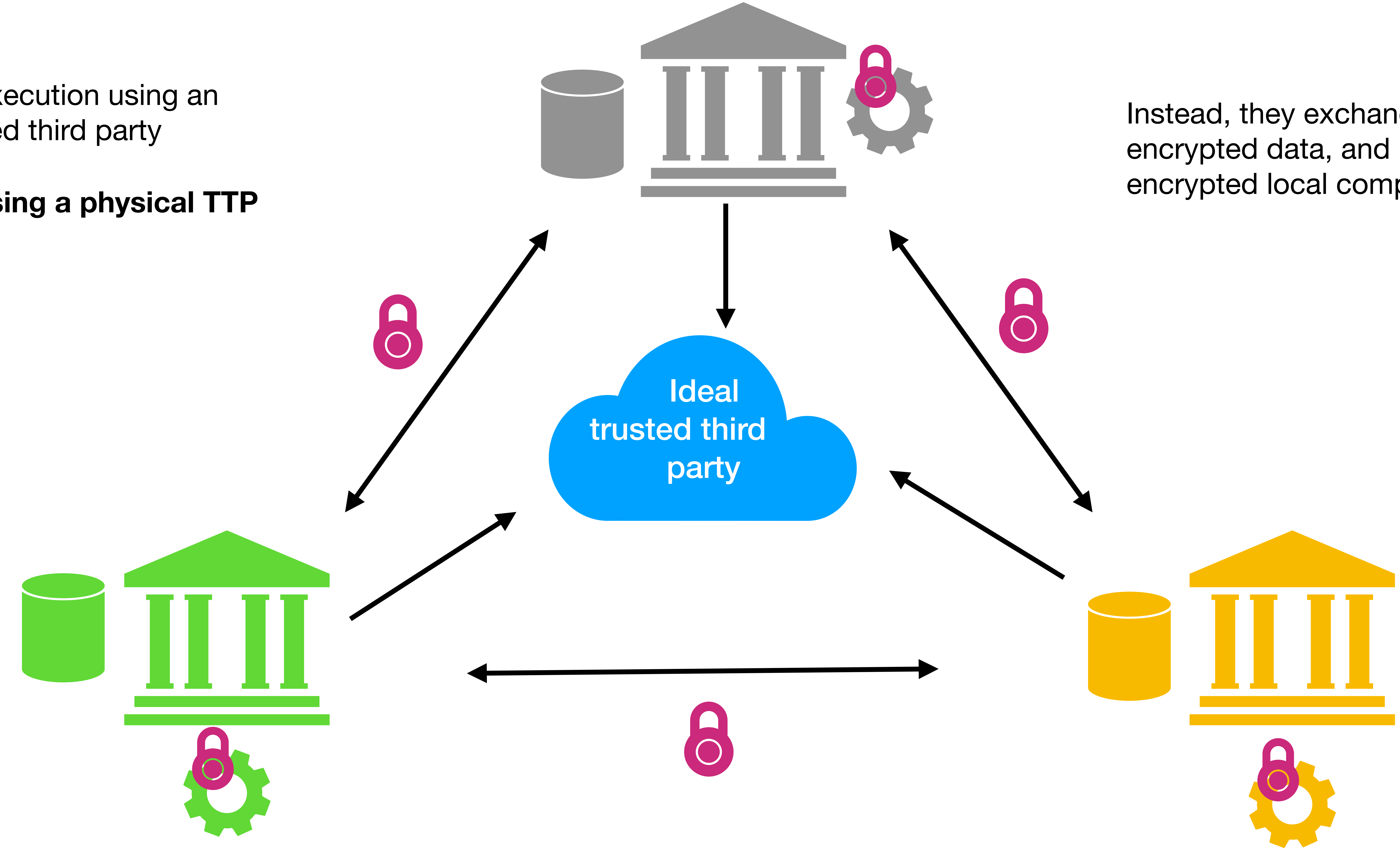
Secure multiparty computation



Secure multiparty computation

Emulate execution using an ideal trusted third party

without using a physical TTP



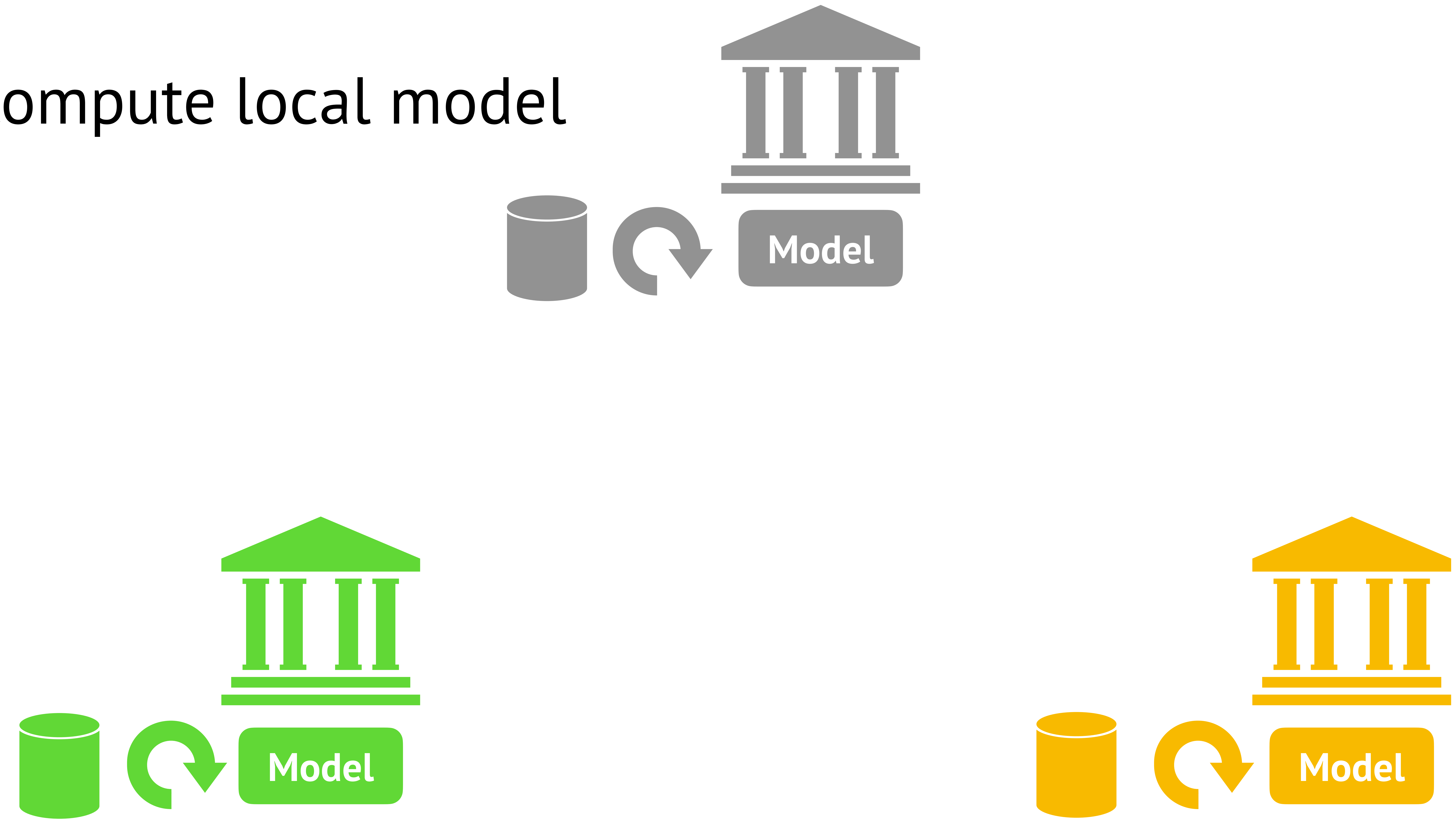
Instead, they exchange encrypted data, and perform encrypted local computation

MPC definitions

- Parties $P_1, \cdots, P_n$ want to securely compute a function $f(x_1, \cdots, x_n)$, where x_i is the private input of P_i
- There exists a trusted party that computes the function f for the parties, given inputs x_i, and the result is given back to every party. Denote $\text{IDEAL}_{f,\mathcal{S}}(x_1, \cdots, x_n)$ as the adversary's view of the parties in the ideal world execution.
- Given a protocol Π that implements f, let $\text{REAL}_{\Pi,\mathcal{A}}(x_1, \cdots, x_n)$ be the view in the real world execution.
- Let f be an n-party functionality and let π be an n-party protocol that computes f. Protocol π is said to securely compute f if for every PPT $\mathcal{A}$ for the real model, there exists a PPT adversary $\mathcal{S}$ for the ideal model such that $\text{IDEAL}_{f,\mathcal{S}}(x_1, \cdots, x_n) \equiv_c \text{REAL}_{\Pi,\mathcal{A}}(x_1, \cdots, x_n)$
  - $\mathcal{S}$ is a simulator in the ideal world that simulates a view for the real world adversary, so that there is nothing an adversary can accomplish in the world that could not also be done in the ideal world

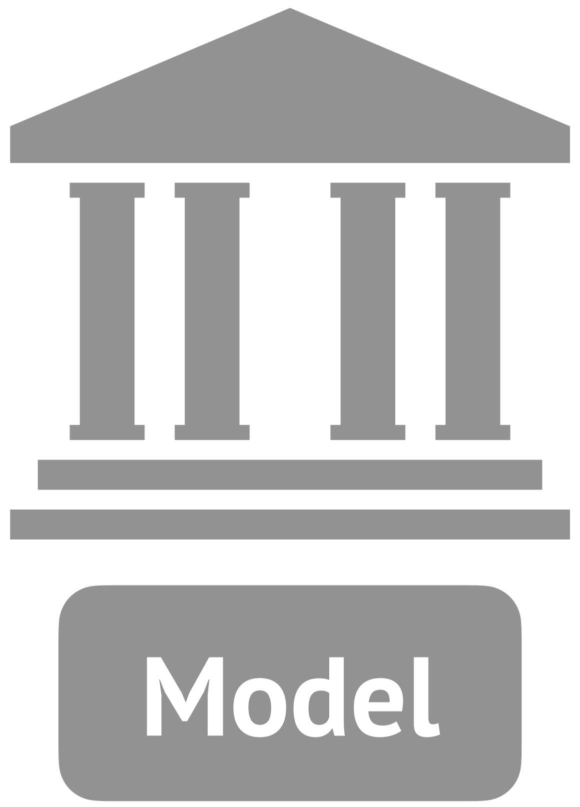
Federated learning

Compute local model



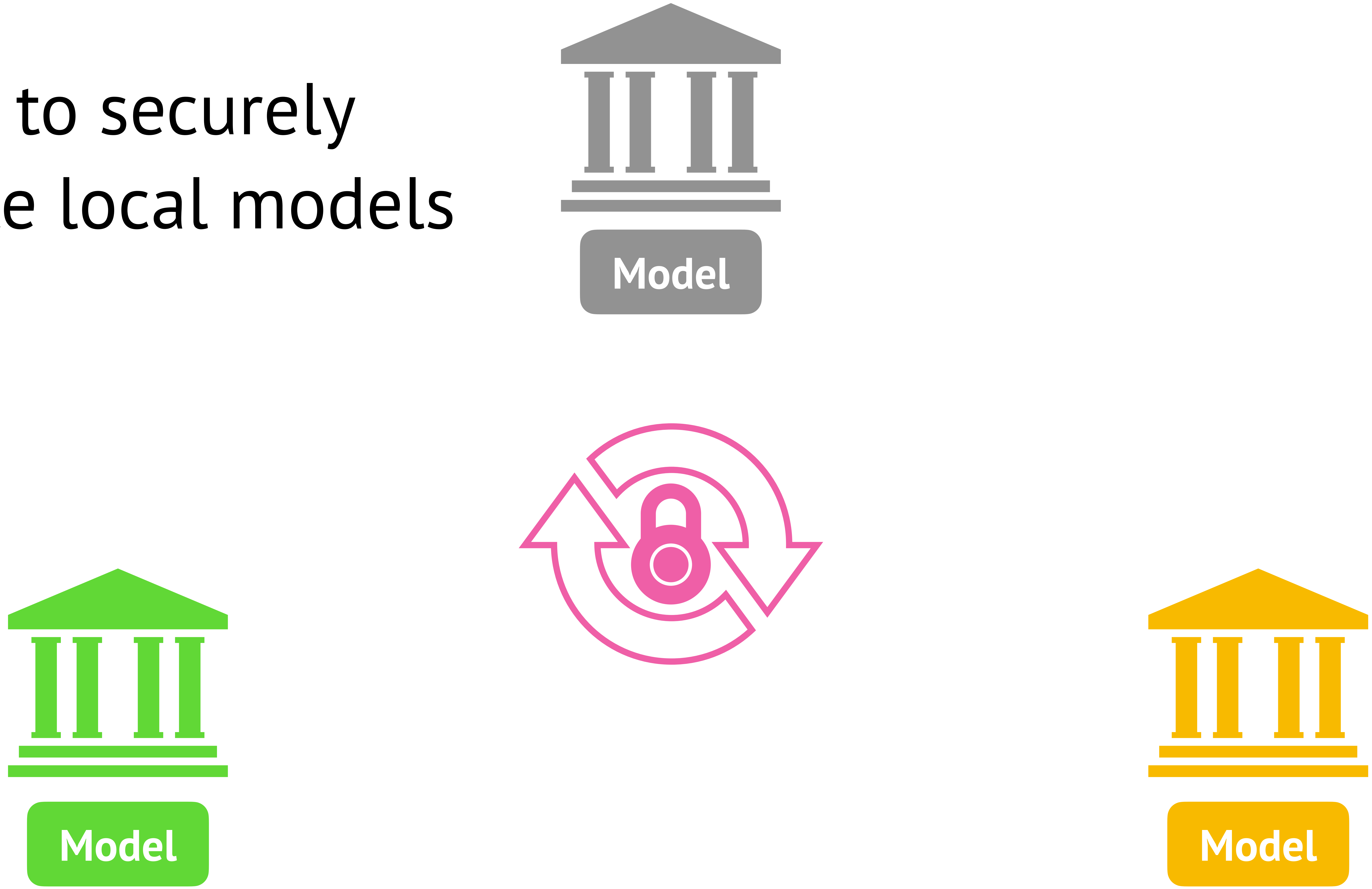
Federated learning

Compute local model



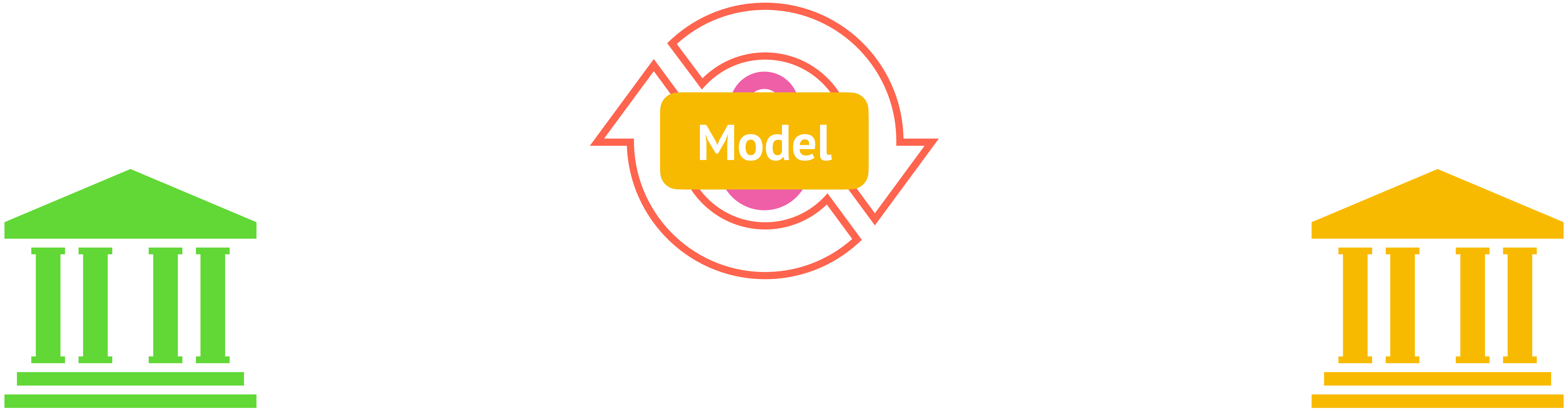
Federated learning

Use MPC to securely aggregate local models



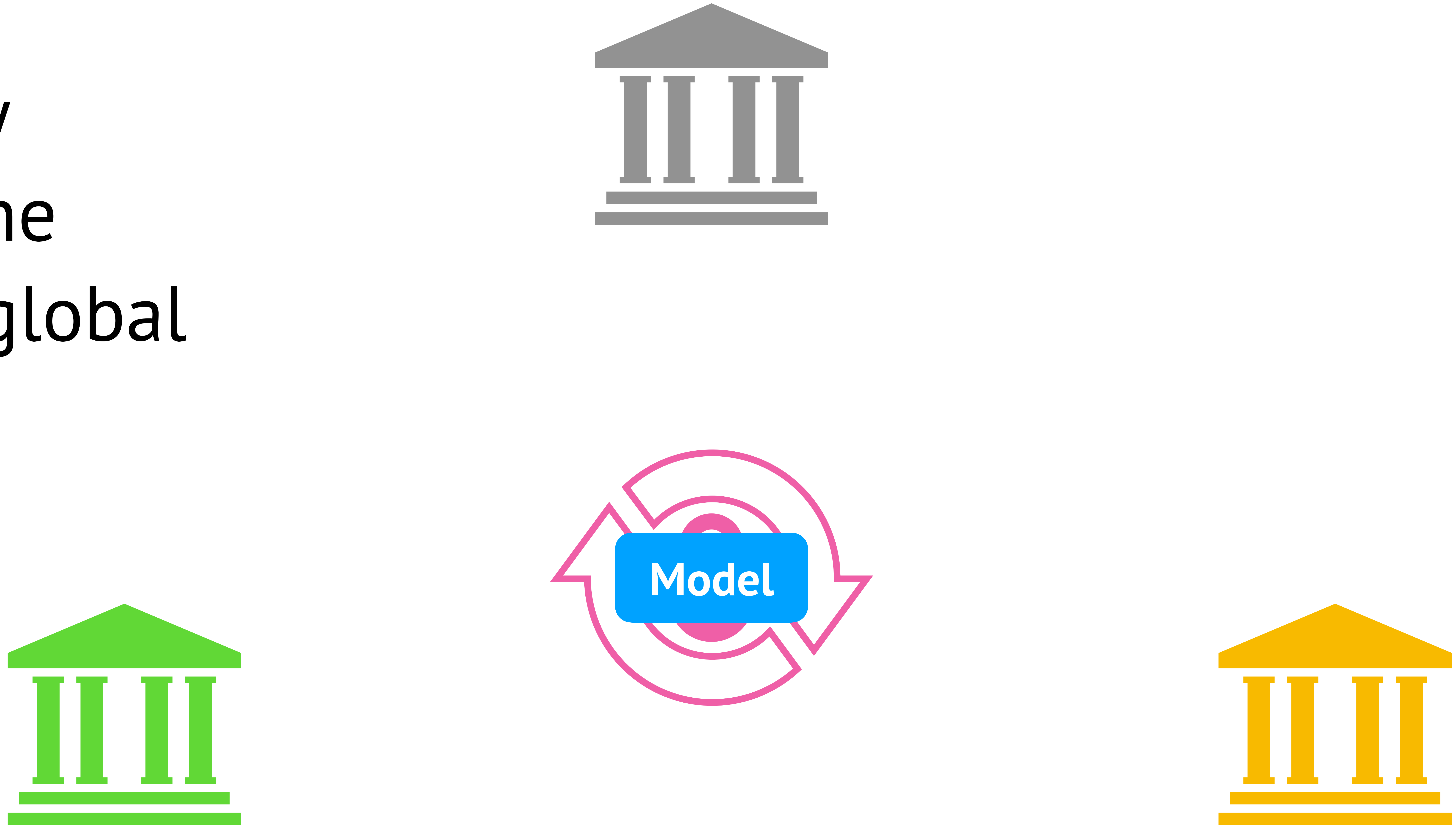
Federated learning

Use MPC to securely
aggregate local models



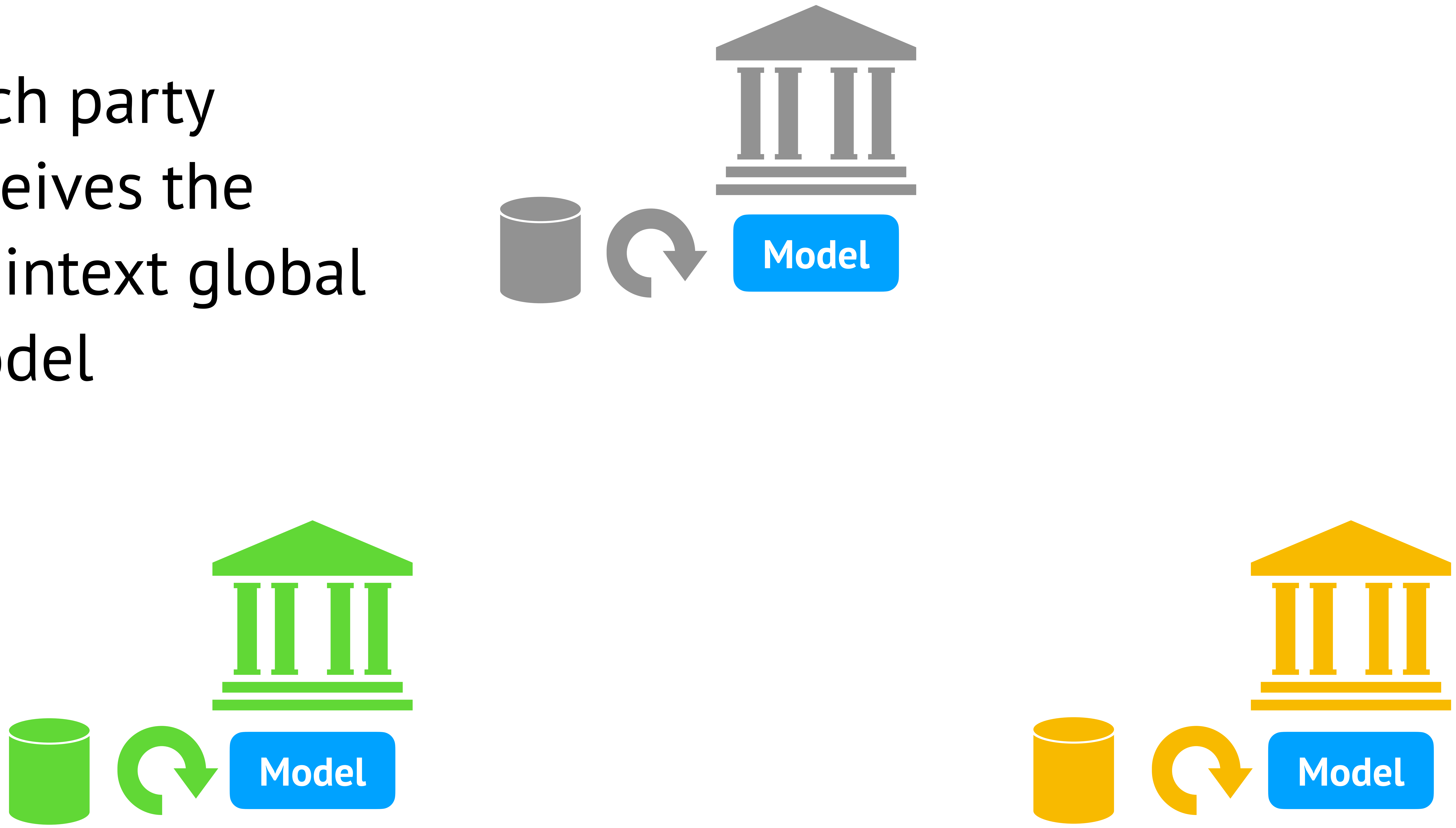
Federated learning

Each party receives the plaintext global model



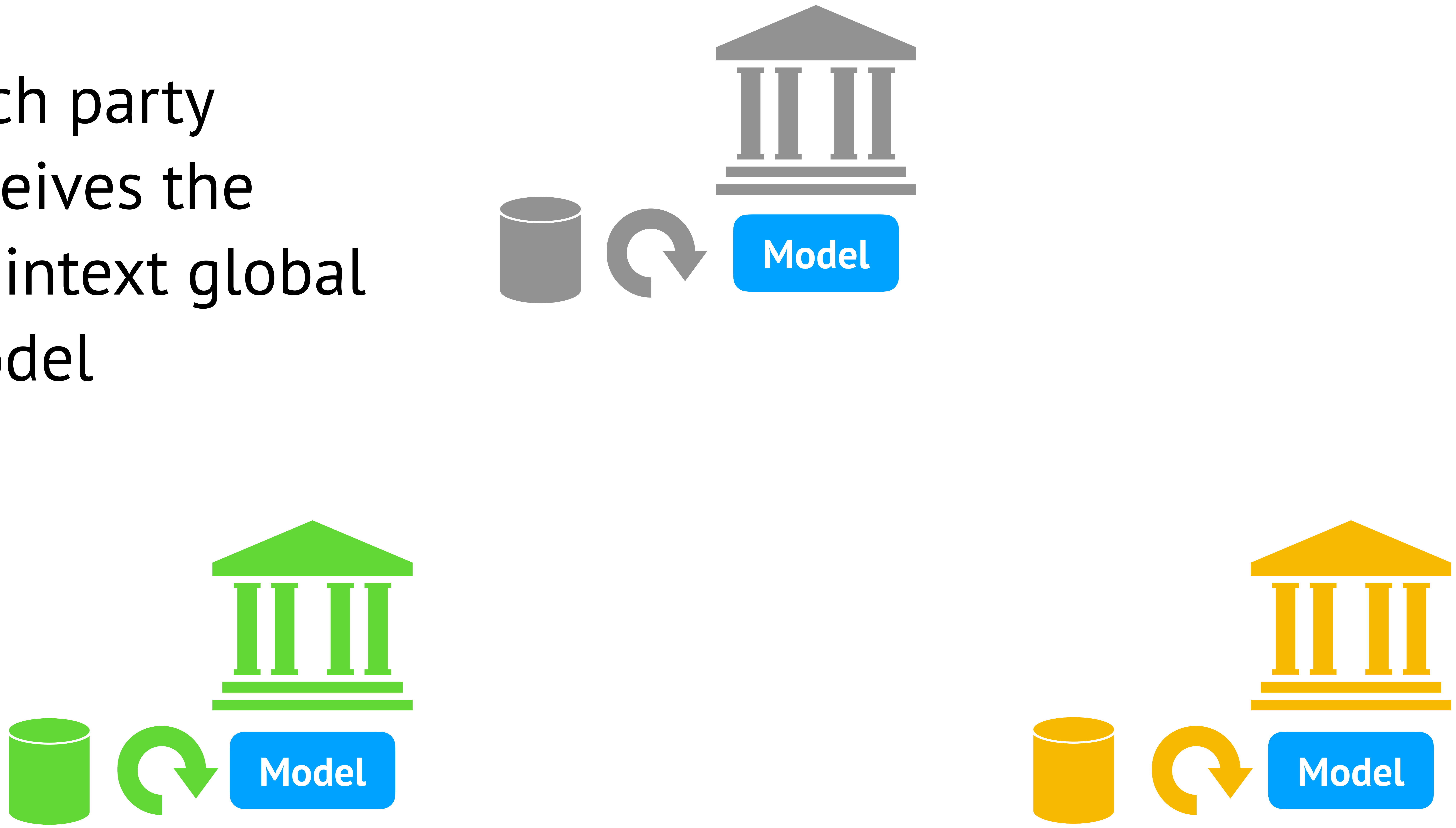
Federated learning

Each party receives the plaintext global model



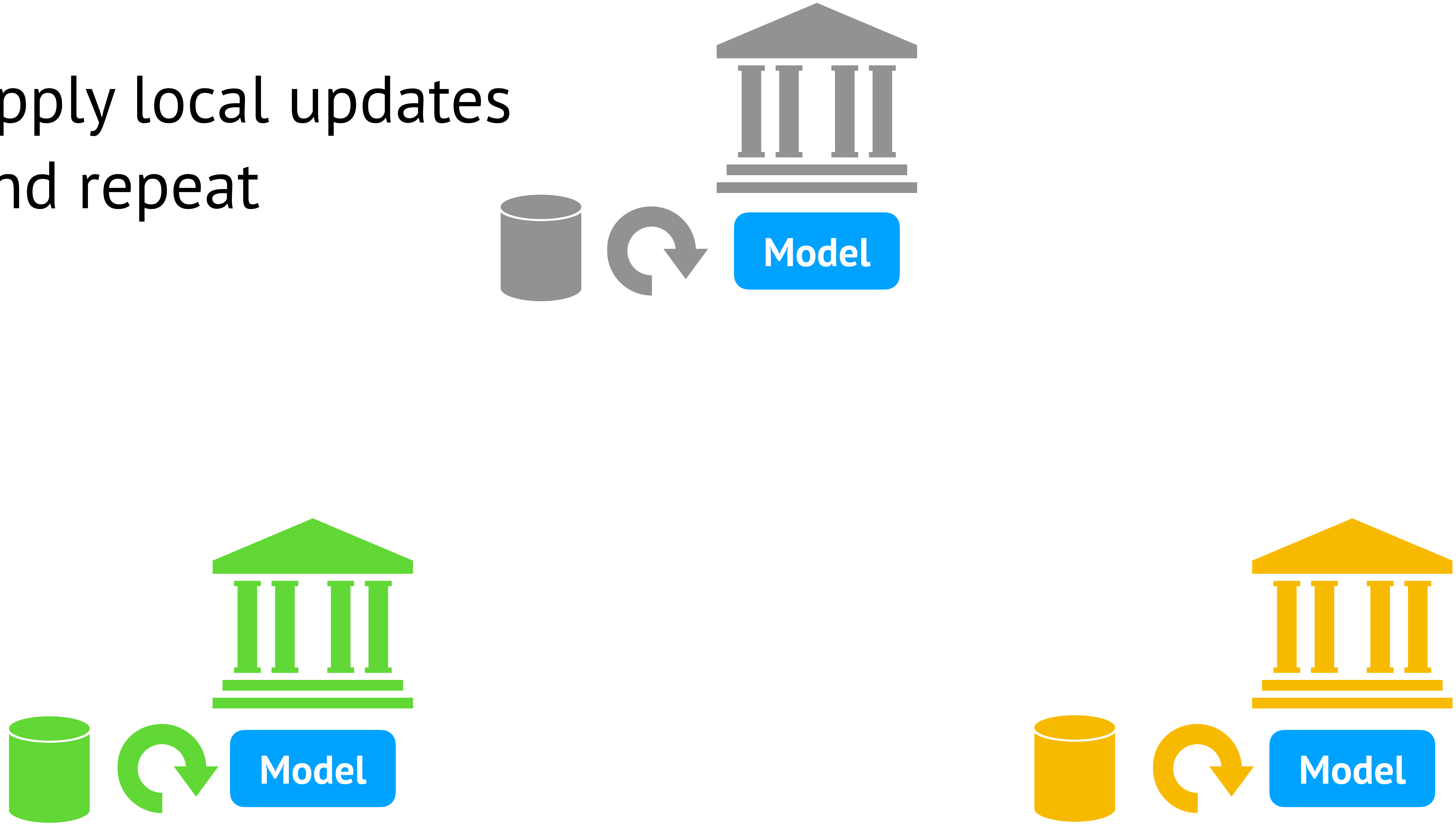
Federated learning

Each party receives the plaintext global model



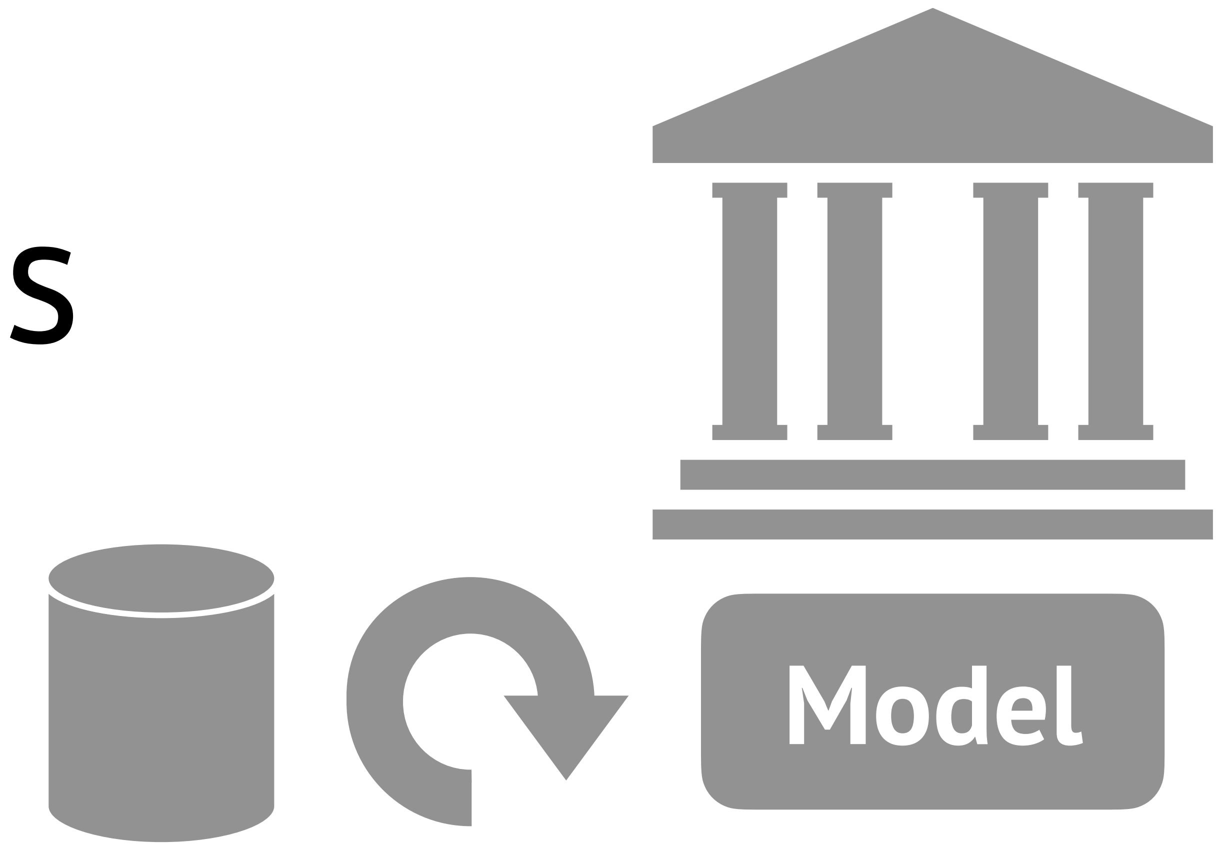
Federated learning

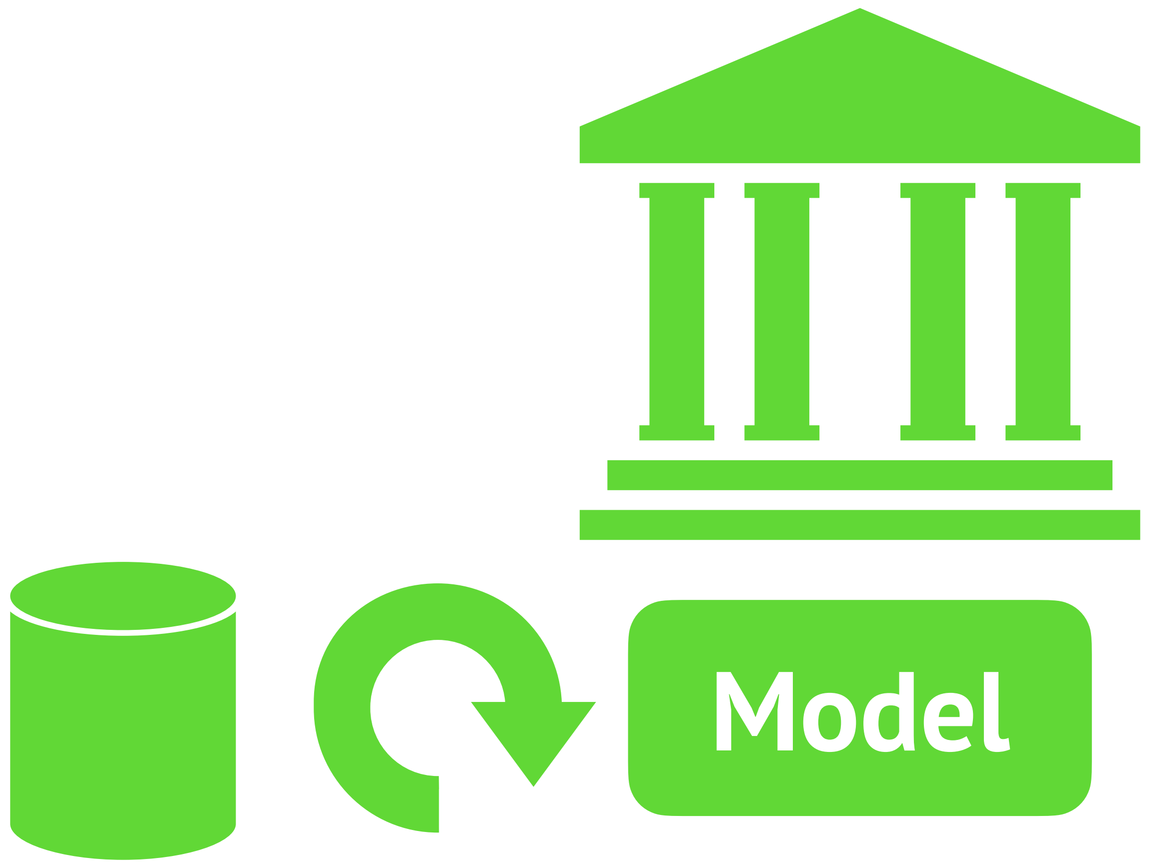
Apply local updates and repeat

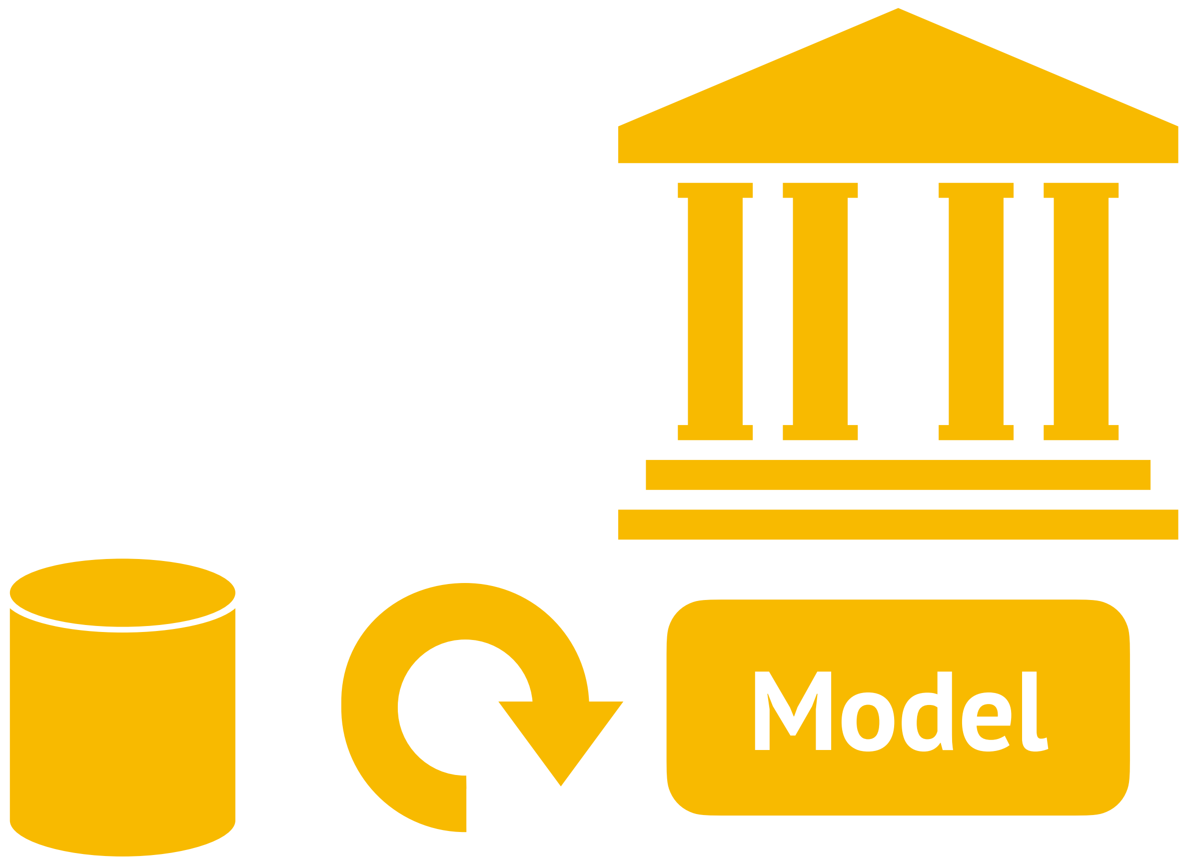


Federated learning

Apply local updates
and repeat







Secret sharing

- A very common primitive for secure aggregation & general MPC
- Similar to FSS, except sharing *data* instead of *function*
- Simple example: given a secret s, split $s = s_1 \oplus \cdots \oplus s_5$ and give s_i to party P_i
    - If there are n parties, any $n-1$ subset cannot figure out the value of s
- Also not fault tolerant - if one party fails then no one can construct!

Shamir sharing

- Given n parties, want to share a secret among them such that $t+1$ of them can recover the secret, but no t of them can find out about it
- **Insight: use polynomials!**
- Let F be a finite field. A degree t polynomial is of the form $f(x) = a_0 + a_1 x + \cdots + a_t x^t$ for coefficients $a_0, \cdots, a_t \in F$
- Associate each party with a distinct point $x_i \in F$. Sharer picks $a_1, \cdots, a_t$ at random, and sets $a_0 = s$. Compute $s_i = f(x_i)$ and sends this to P_i
- Security:
    - Given $t+1$ points on the polynomial, can use *polynomial interpolation* to recover the coefficients $a_0, \cdots, a_t$
    - Given any t points on the polynomial, one cannot figure out anything about a_0

Today's reading: privacy preserving aggregation

Next lecture: guest speaker!

- Reading posted!
- No review needed, but still need to send in 1 discussion question by 4 pm the night before
- Please ask your discussion question to the speaker!