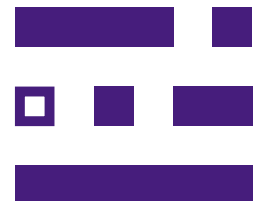
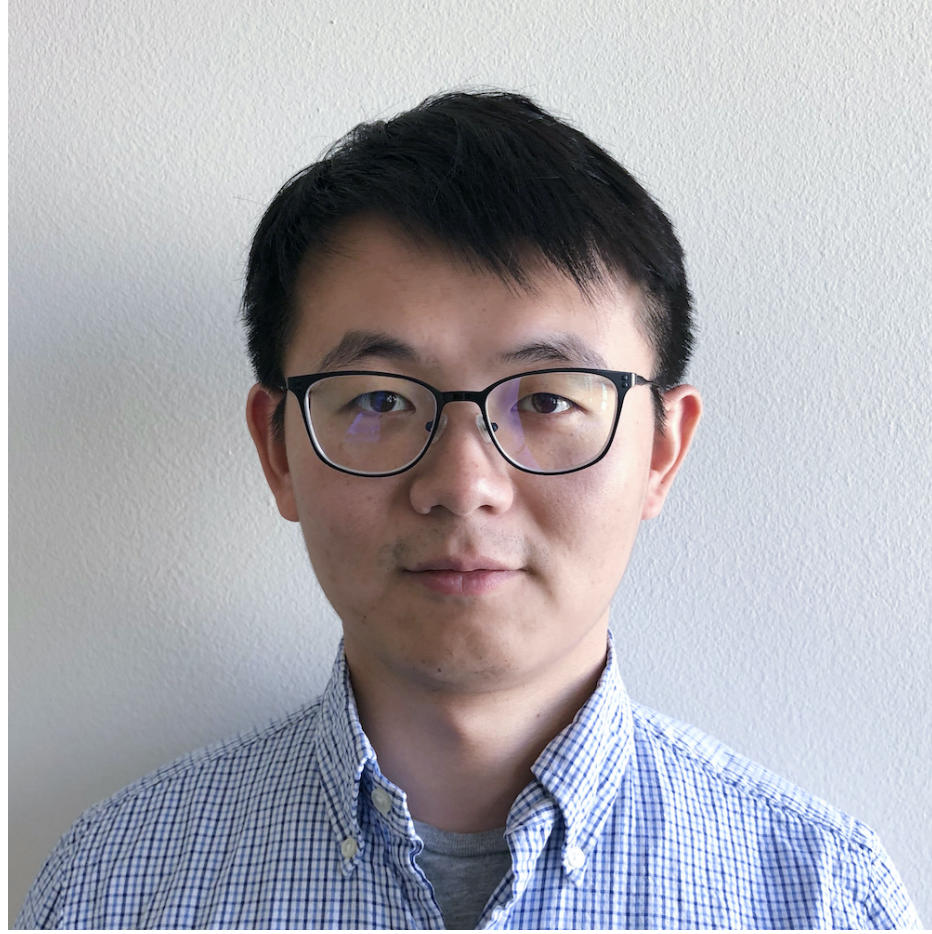


AI + Distributed System

Hao Wang |  IntelliSys Lab



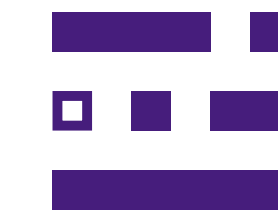
- 09/2015-08/2020
- 01/2021-present



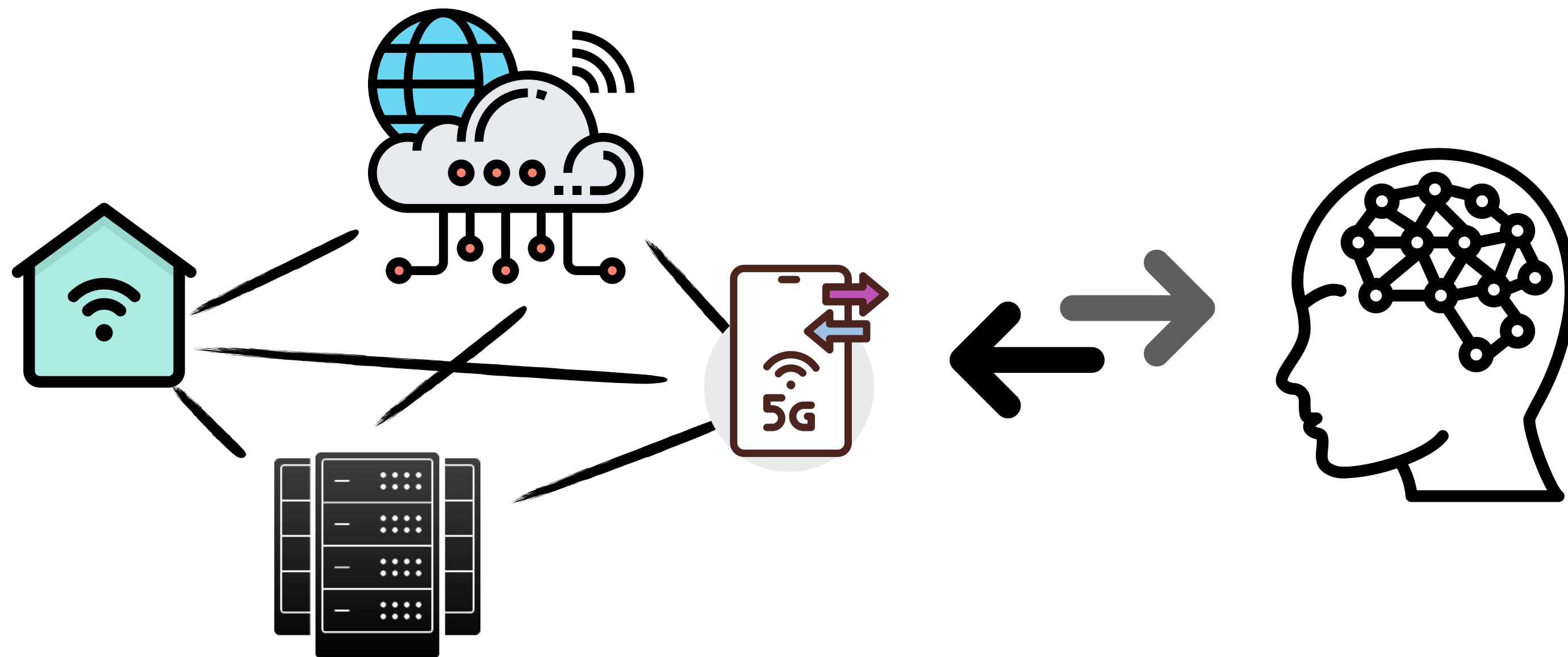
UNIVERSITY OF
TORONTO



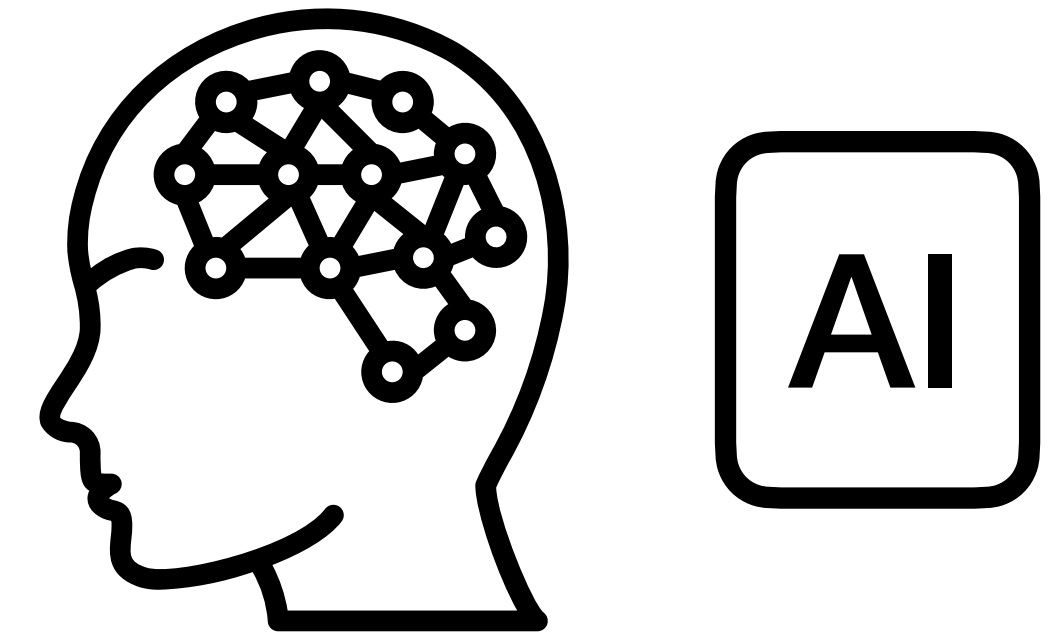
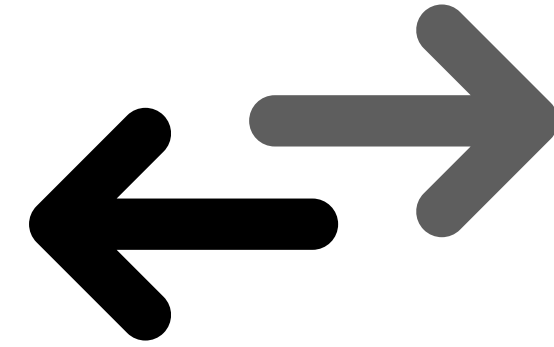
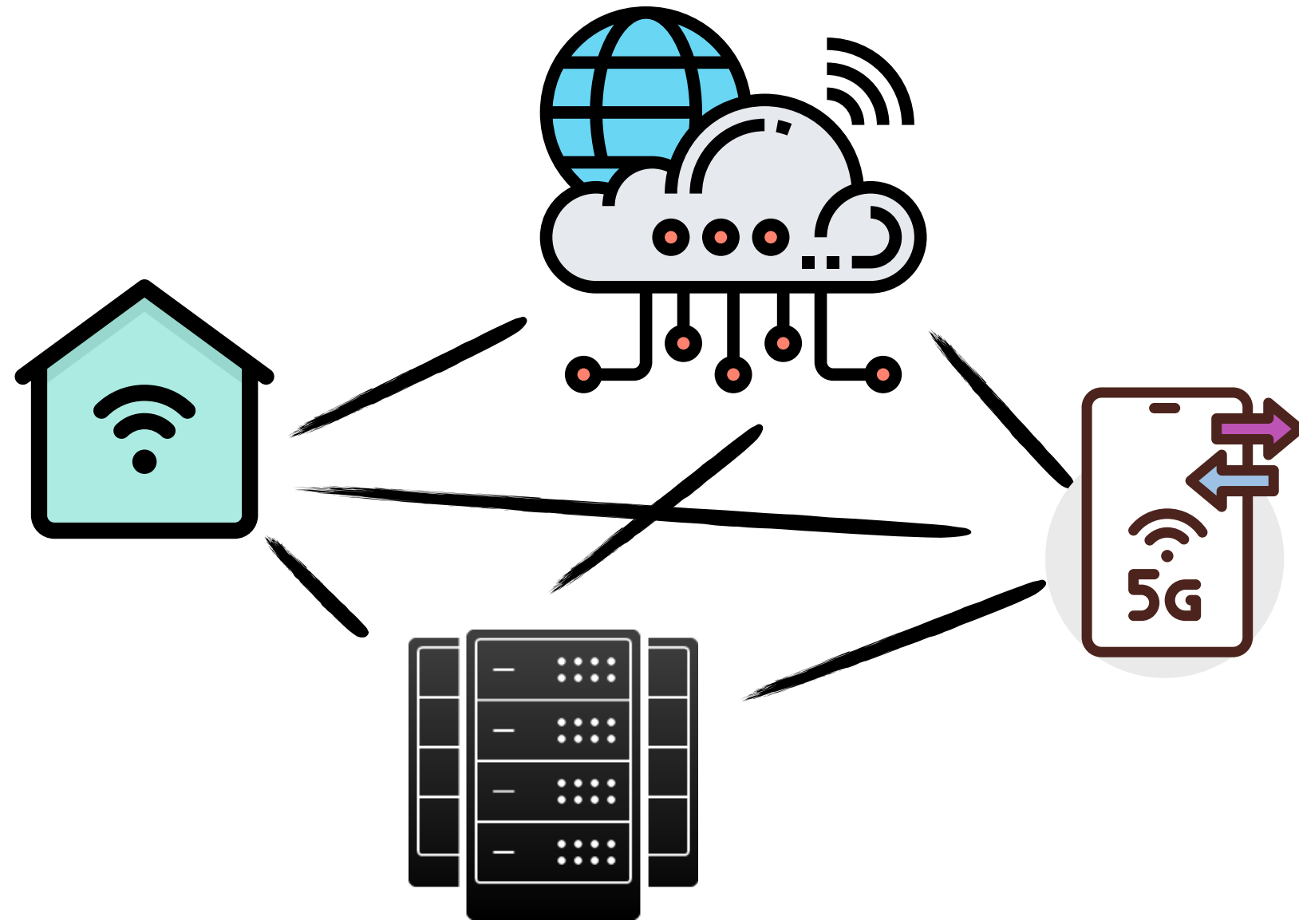
LSU



IntelliSys Lab



- **AI + Distributed System**
- 17 published papers
- 624 citations
- Projects sponsored by
 - NSERC, Huawei Canada, Orbis Investment...



ML Training / Inference

Big Data Analytics

Video Processing

...

Performance

Adaptivity

Robustness

...

Reinforcement Learning

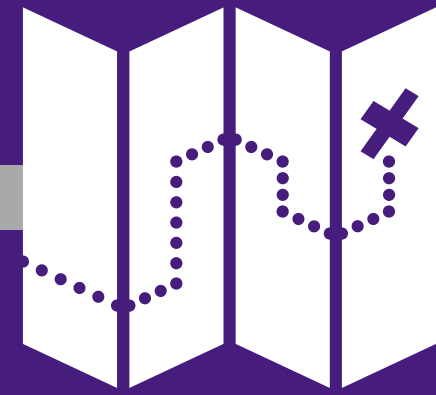
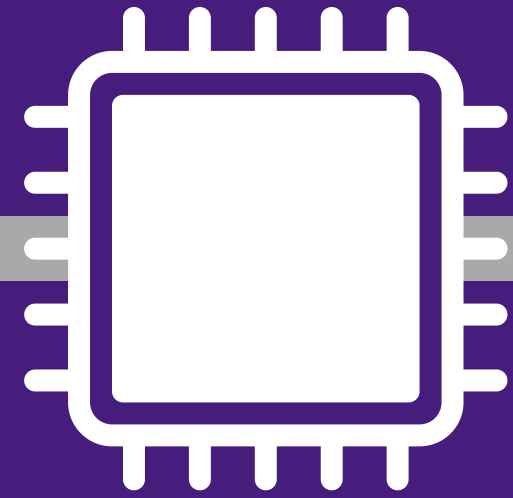
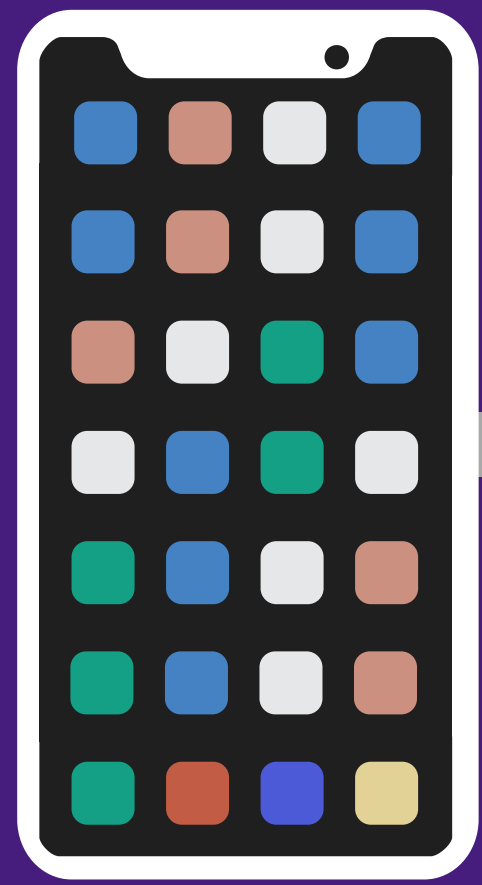
Federated Learning

Transfer Learning

...

Topic #1

Federated Learning



Siri



Alexa

Sure. Umami burger?

Yeah. Know the address?

738 E. 3rd St.



The

|

Hi

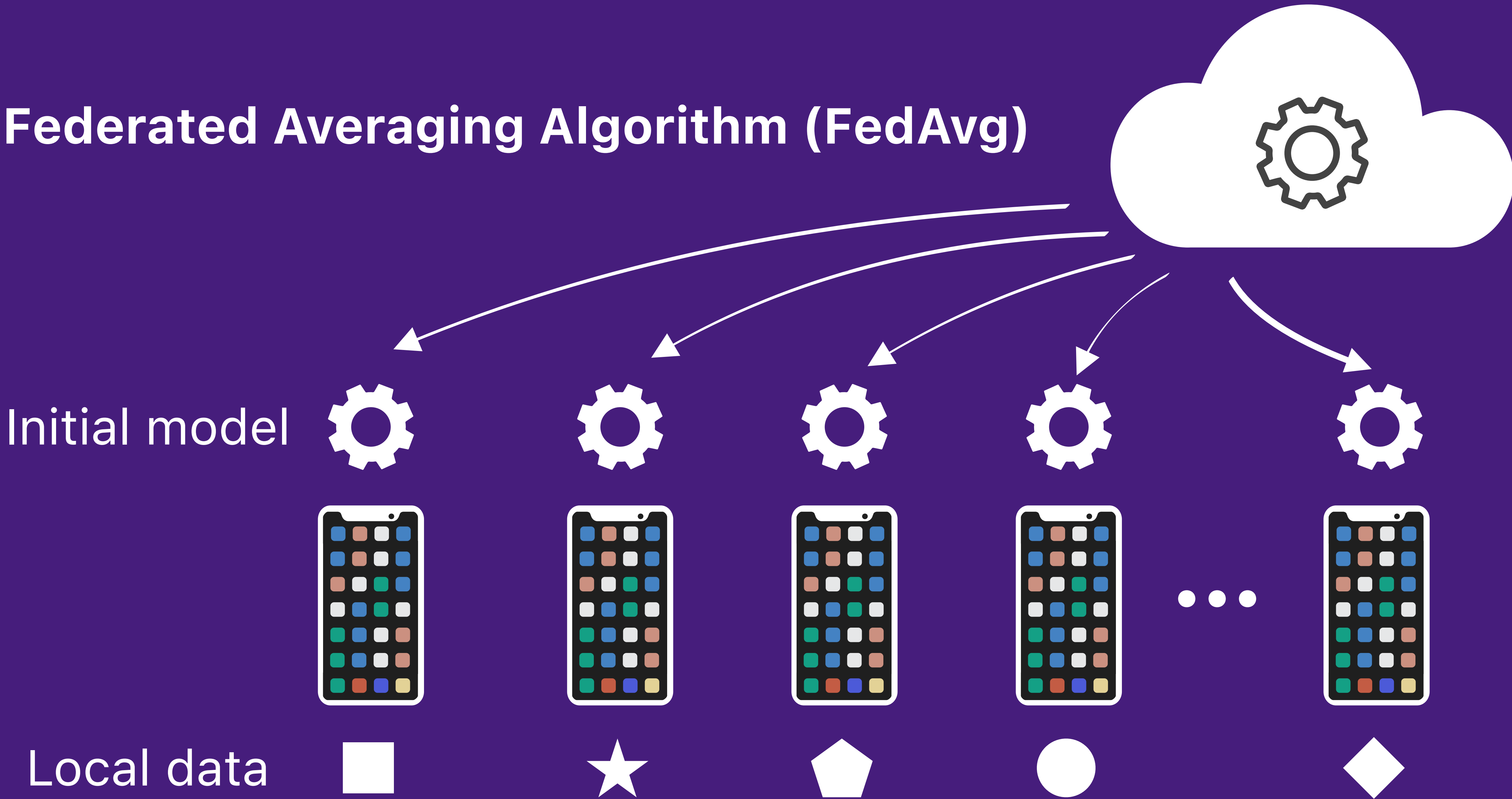


¹q ²w ³e ⁴r ⁵t ⁶y ⁷u ⁸i ⁹o ⁰p

a s d f g h j k l

↑ z x c v b n m ✕

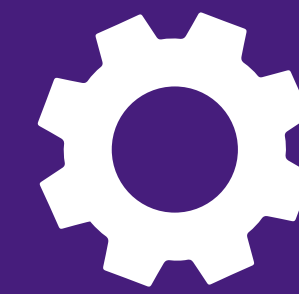
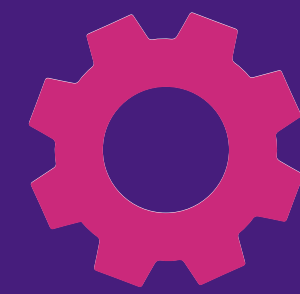
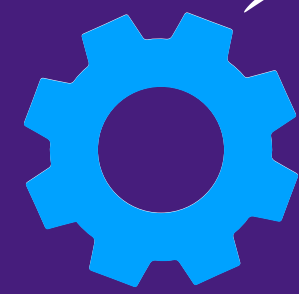
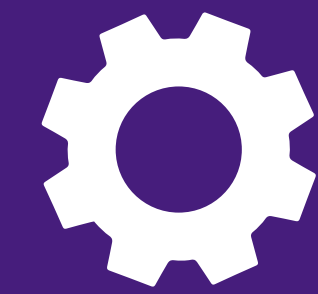
Federated Averaging Algorithm (FedAvg)



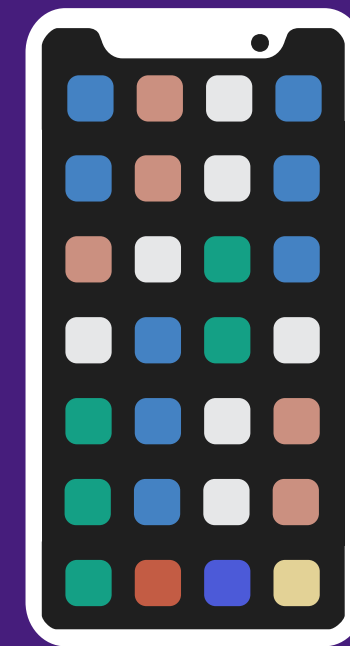
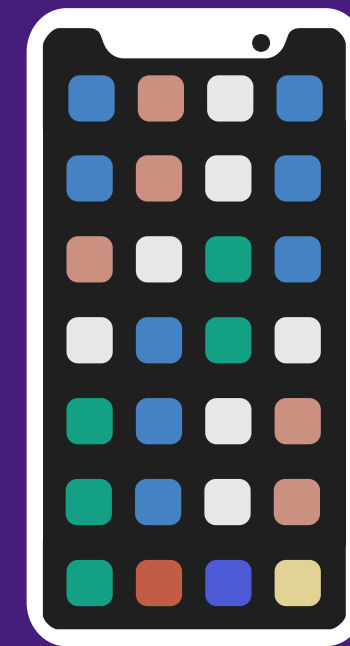
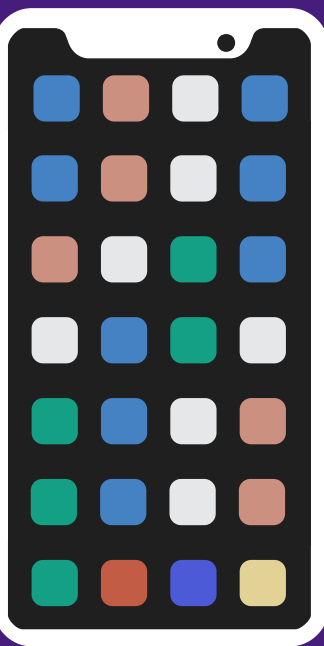
Random selection

Update global model

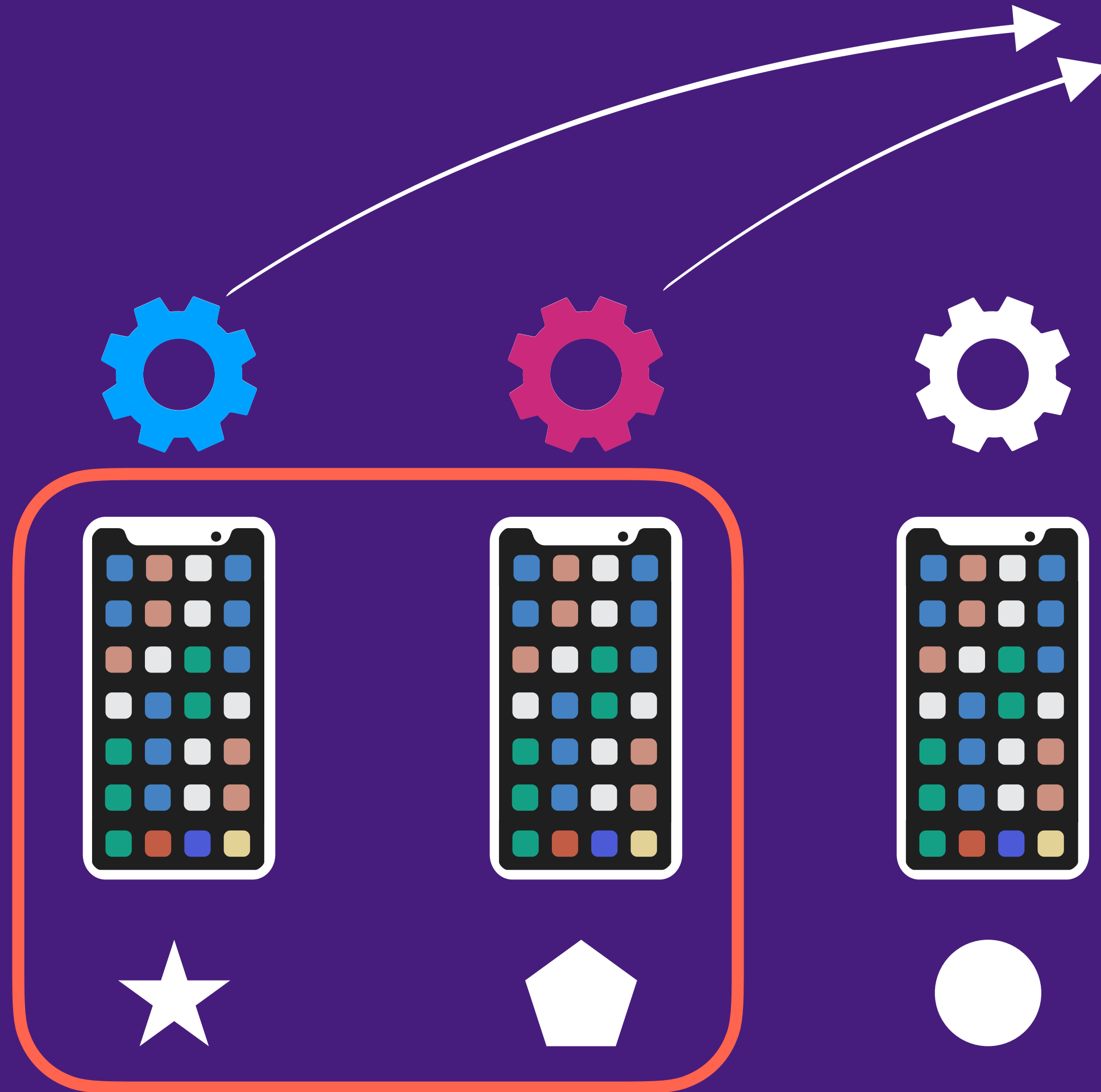
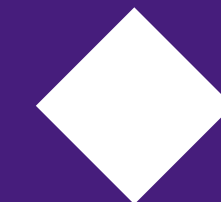
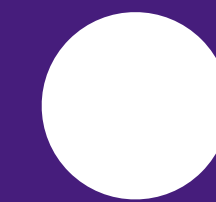
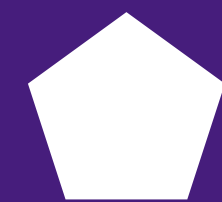
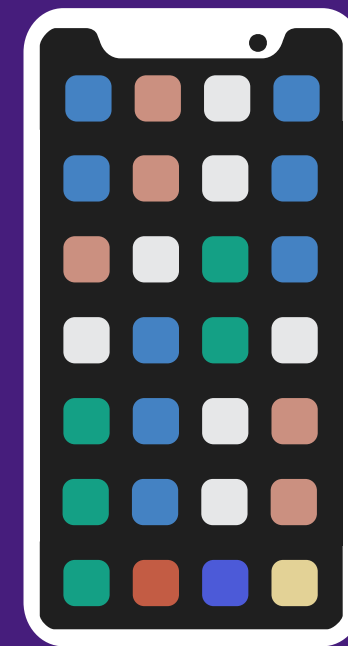
Local model



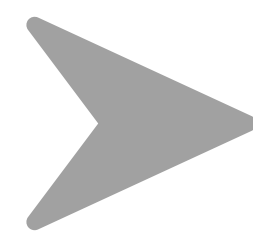
Local data



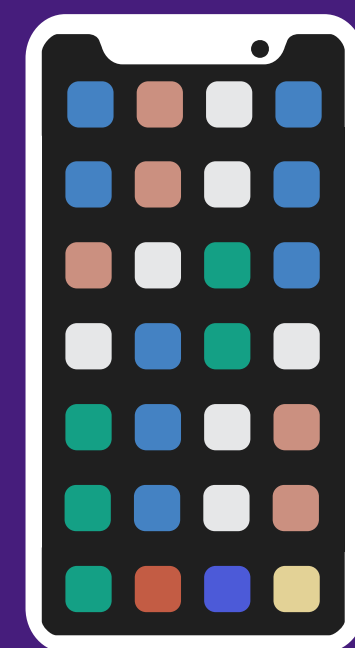
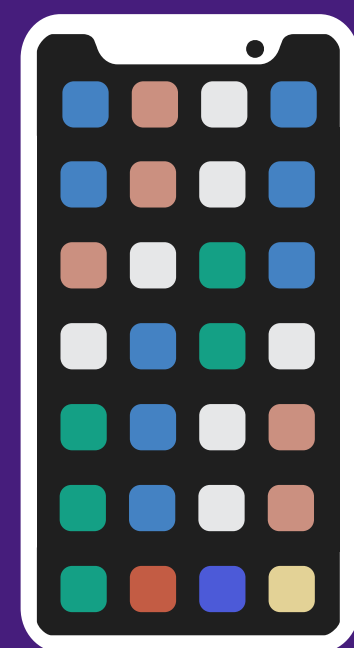
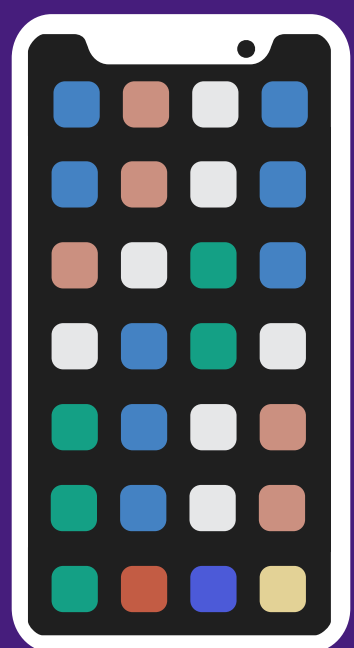
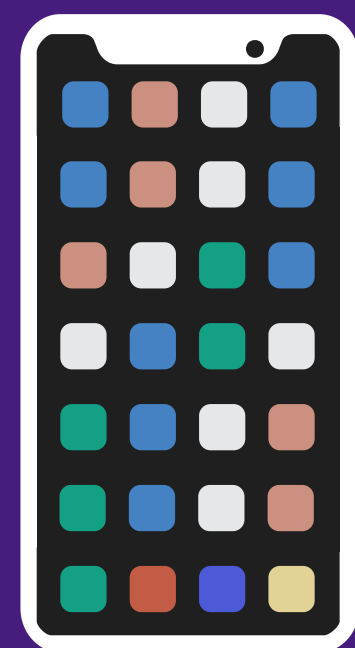
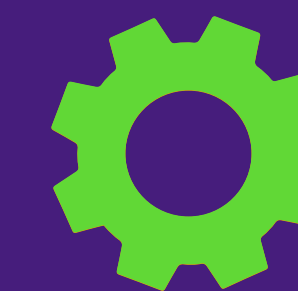
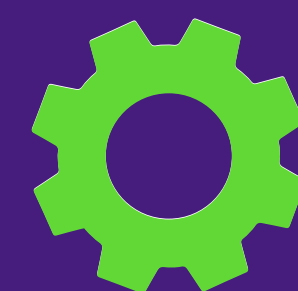
...



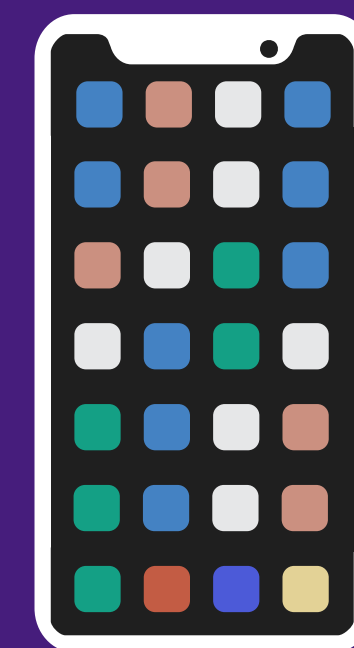
Thank you for the feedback



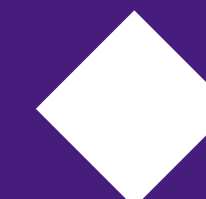
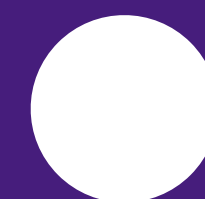
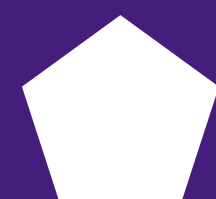
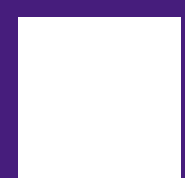
Local model



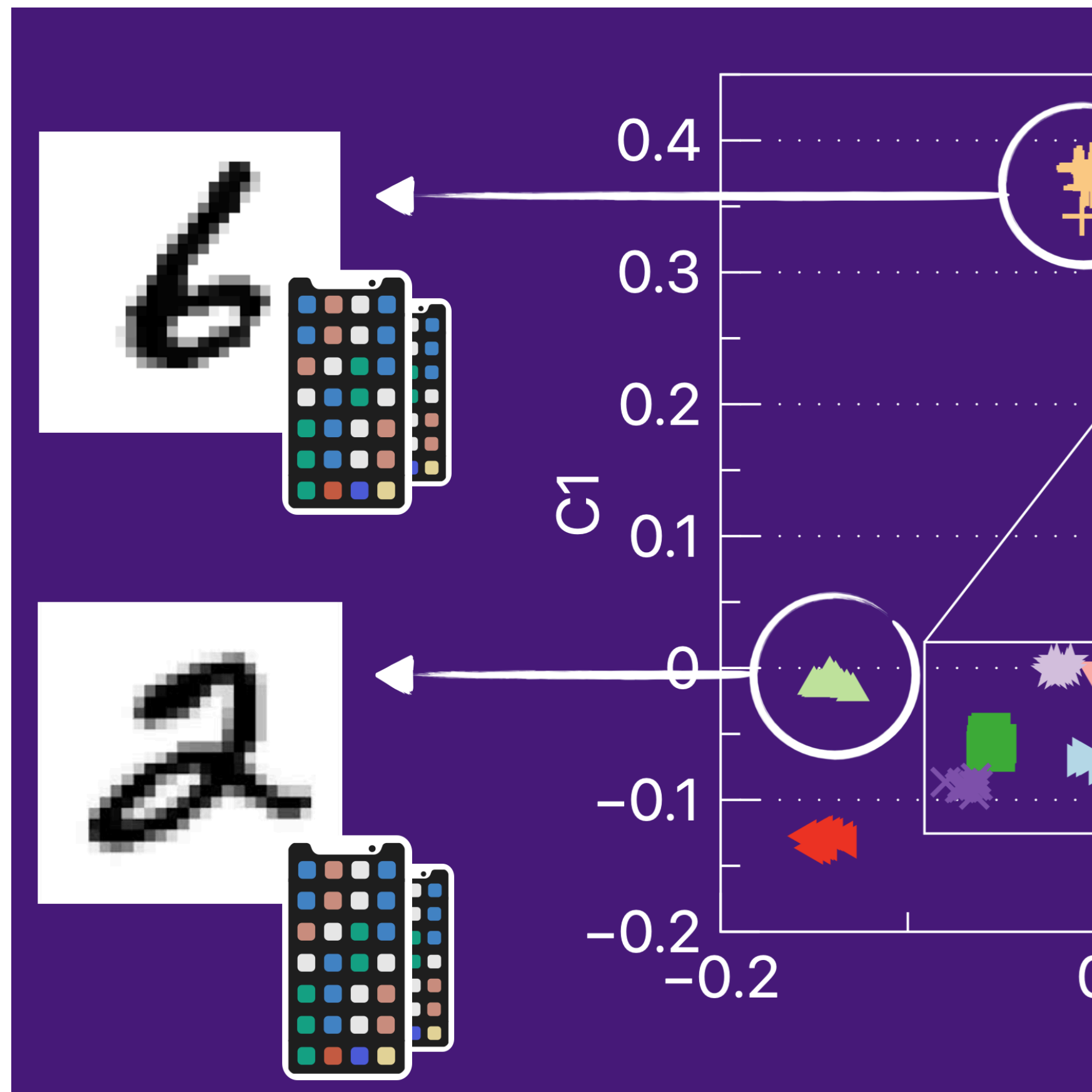
...



Local data



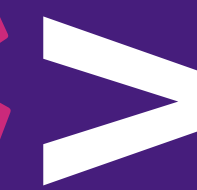
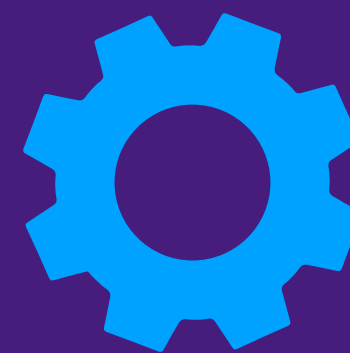
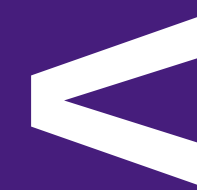
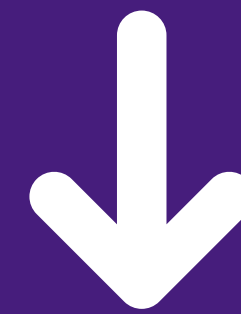
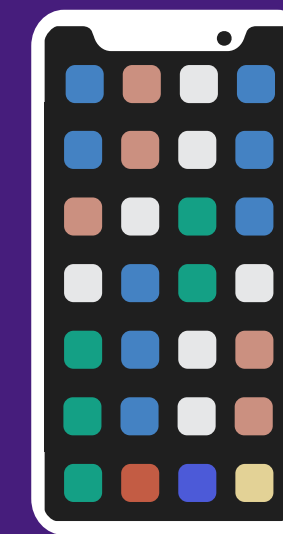
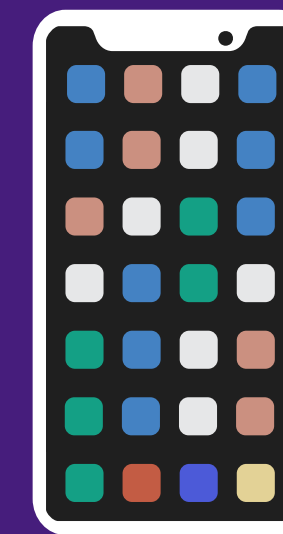
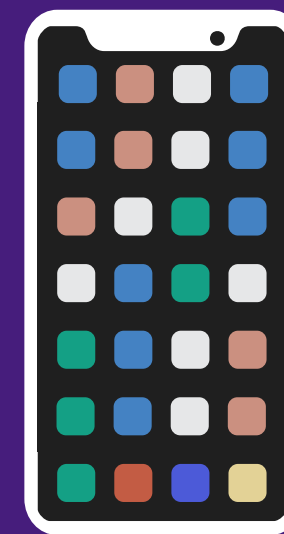
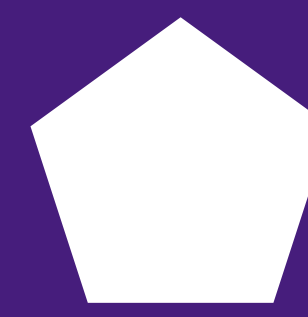
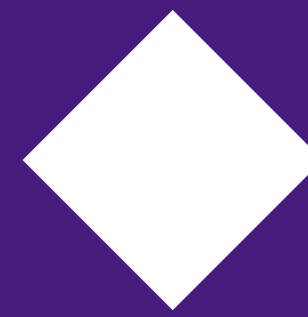
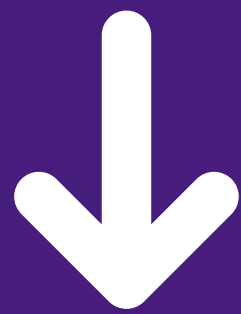
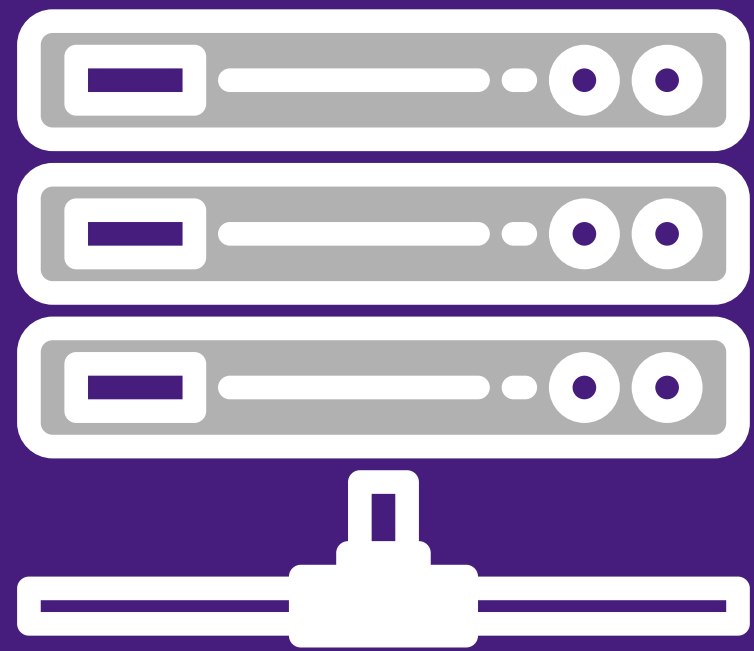
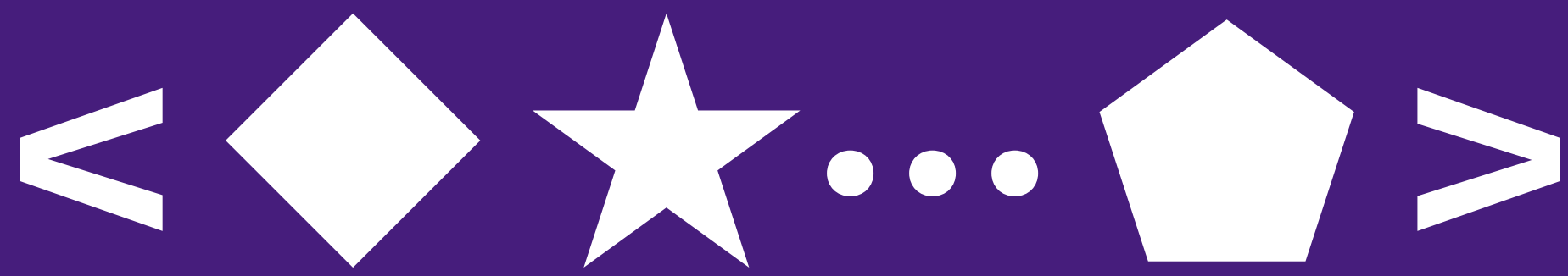
Projects Done



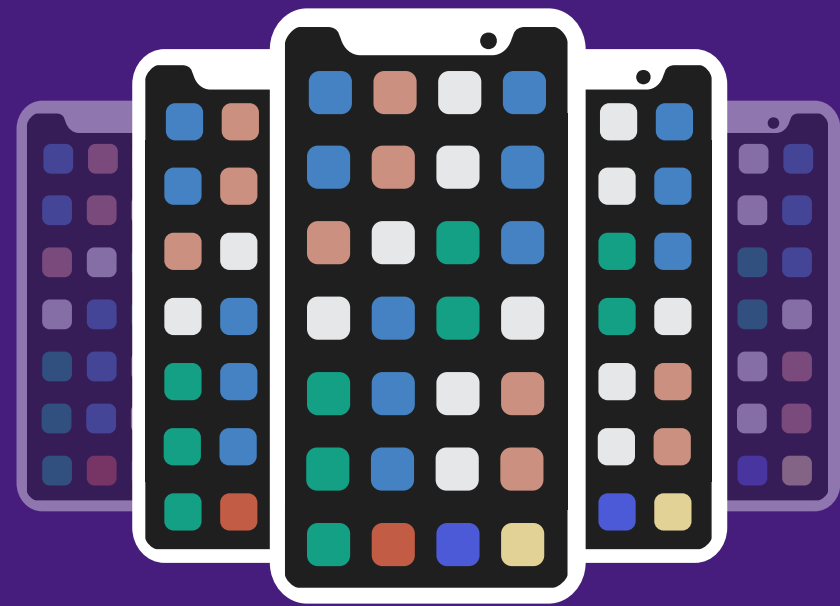
- Optimizing Federated Learning on Non-IID Data with Reinforcement Learning
- [INFOCOM'20]
 - Federated learning
 - Client selection
 - Reinforcement learning

ML algorithms assume the training data is **independent and identically distributed (IID)**

Federated Learning reuses the existing ML algorithms but on **non-IID** data

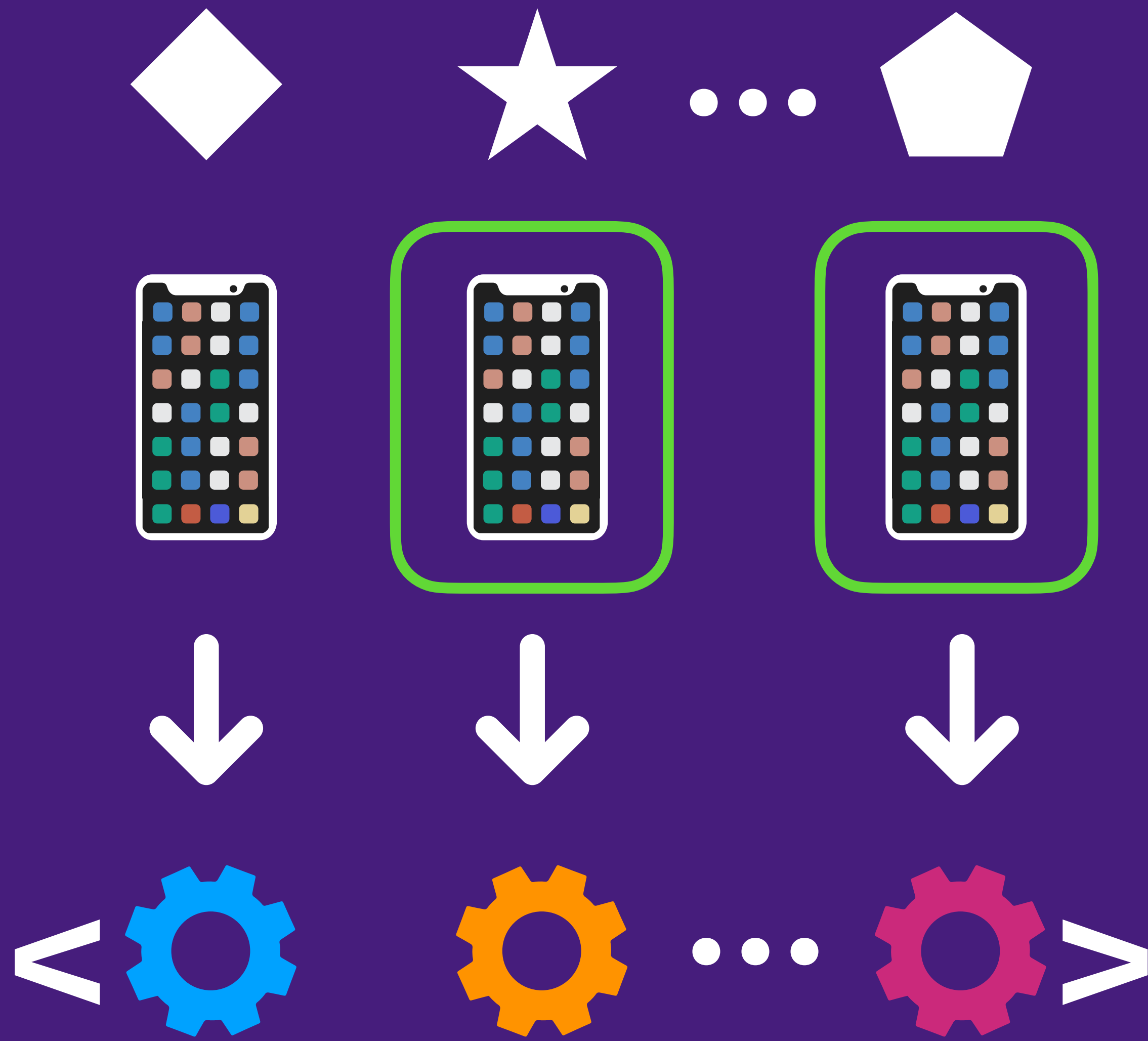


Build IID training data? **No**



Peeking into the data distribution on each device without violating data privacy

Probing the bias of non-IID data



Carefully **select devices** to **balance** the bias introduced by **non-IID data**



100 devices, each has 600 samples

Non-IID data



80% data has the same label, e.g, "6"

Initial model



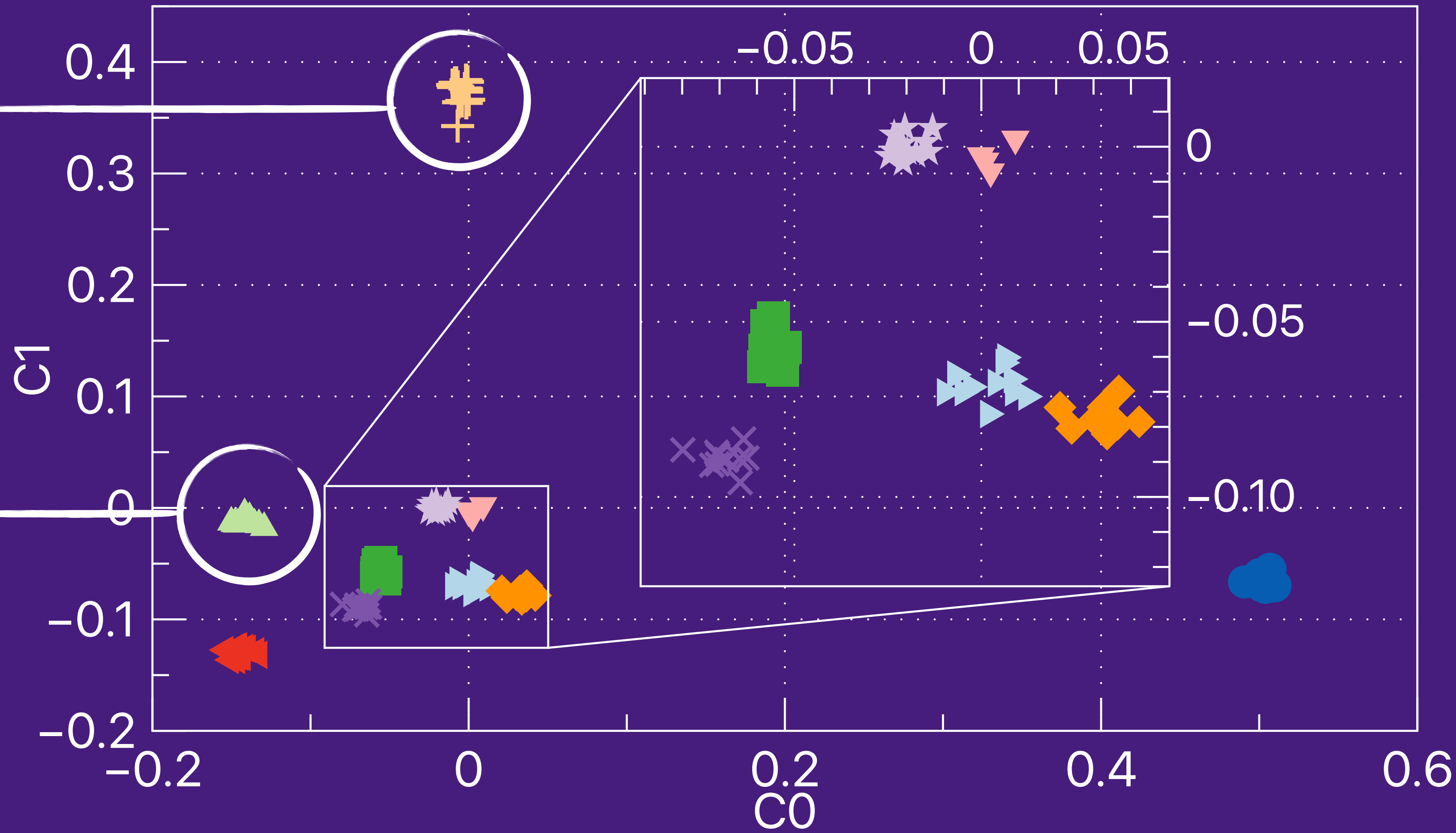
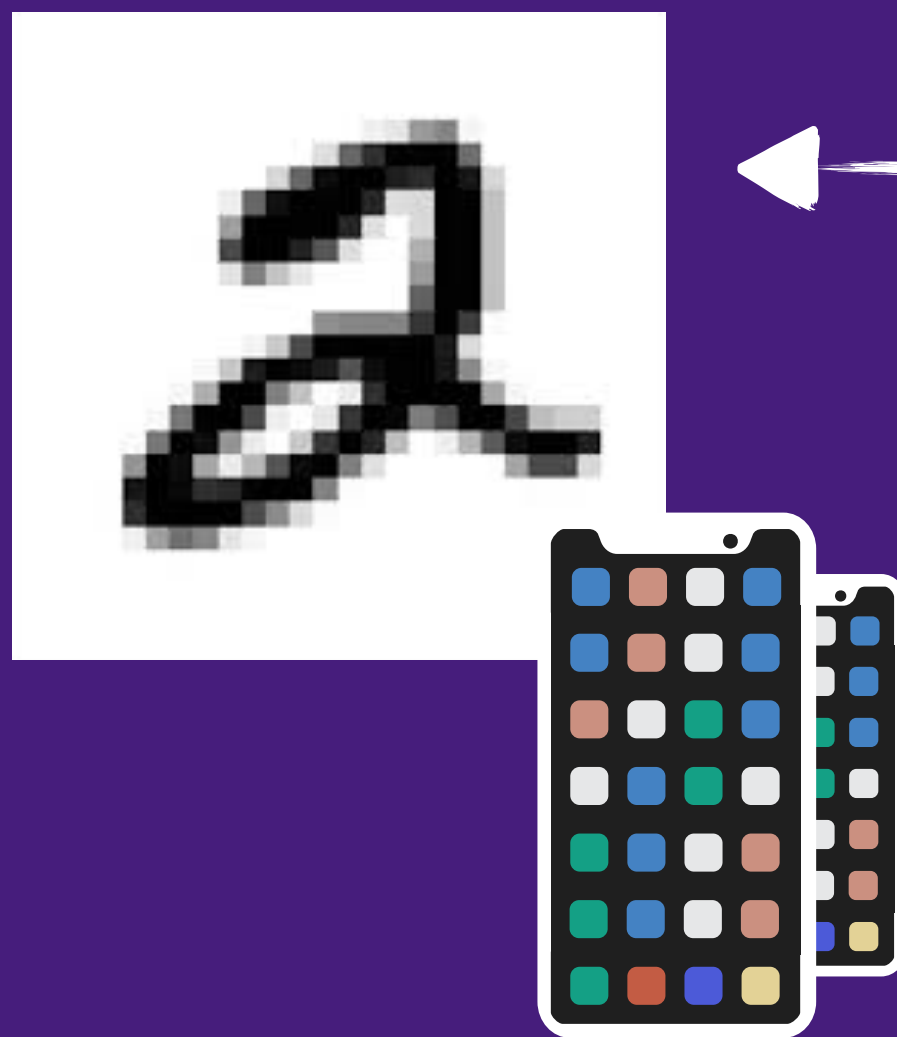
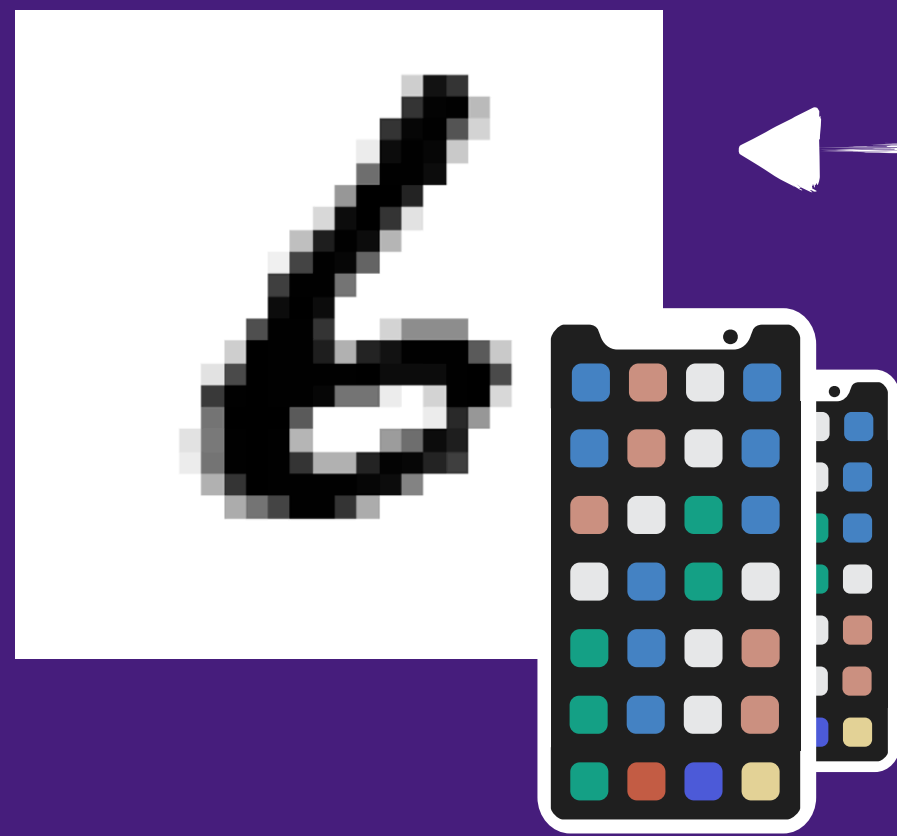
A two-layer CNN model with 431,080 parameters

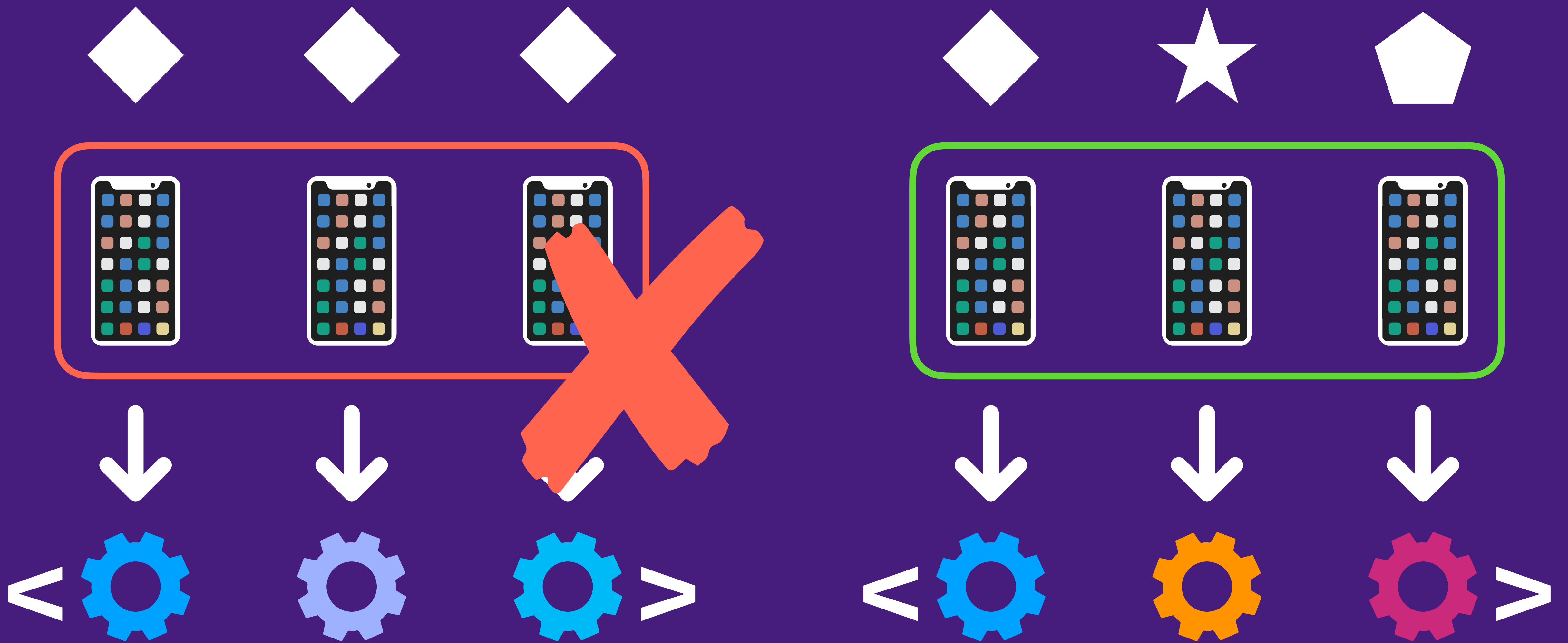
Local model



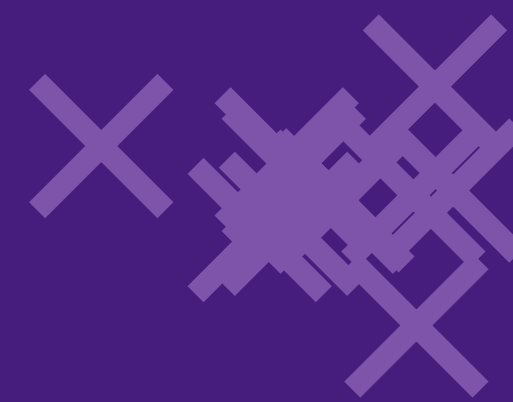
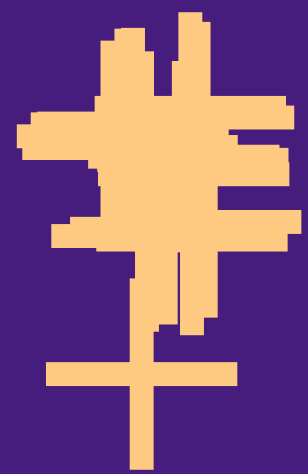
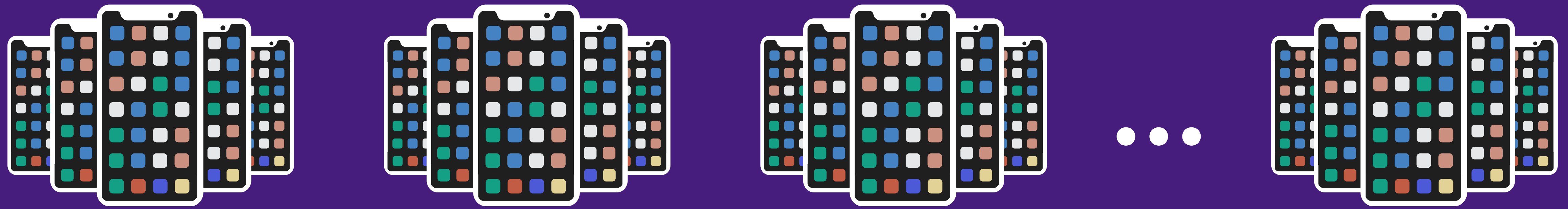
We apply **Principle Component Analysis (PCA)** to reduce dimensionality

431,080-dimension model weight  2-dimension space

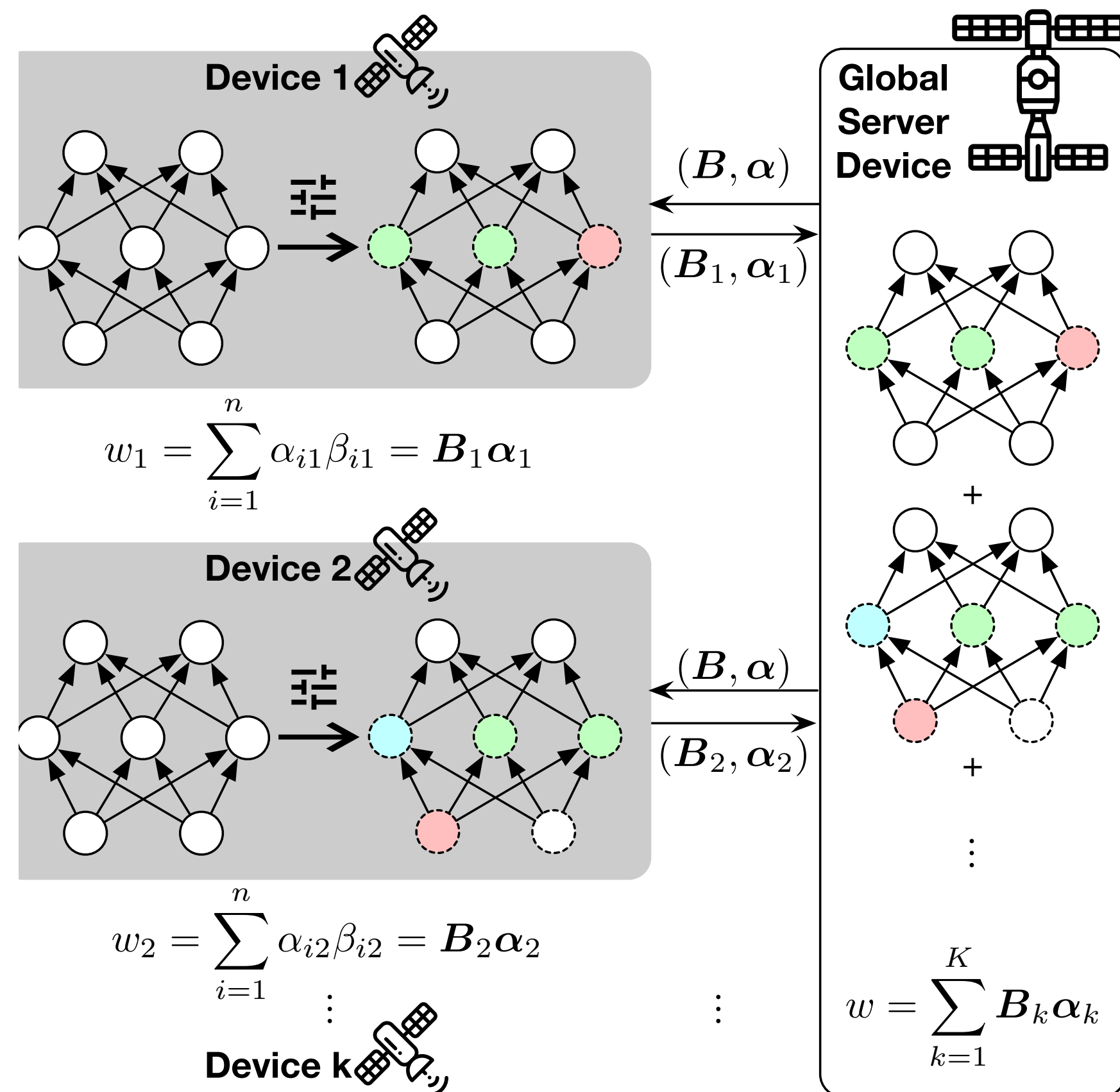




How to select participating clients?

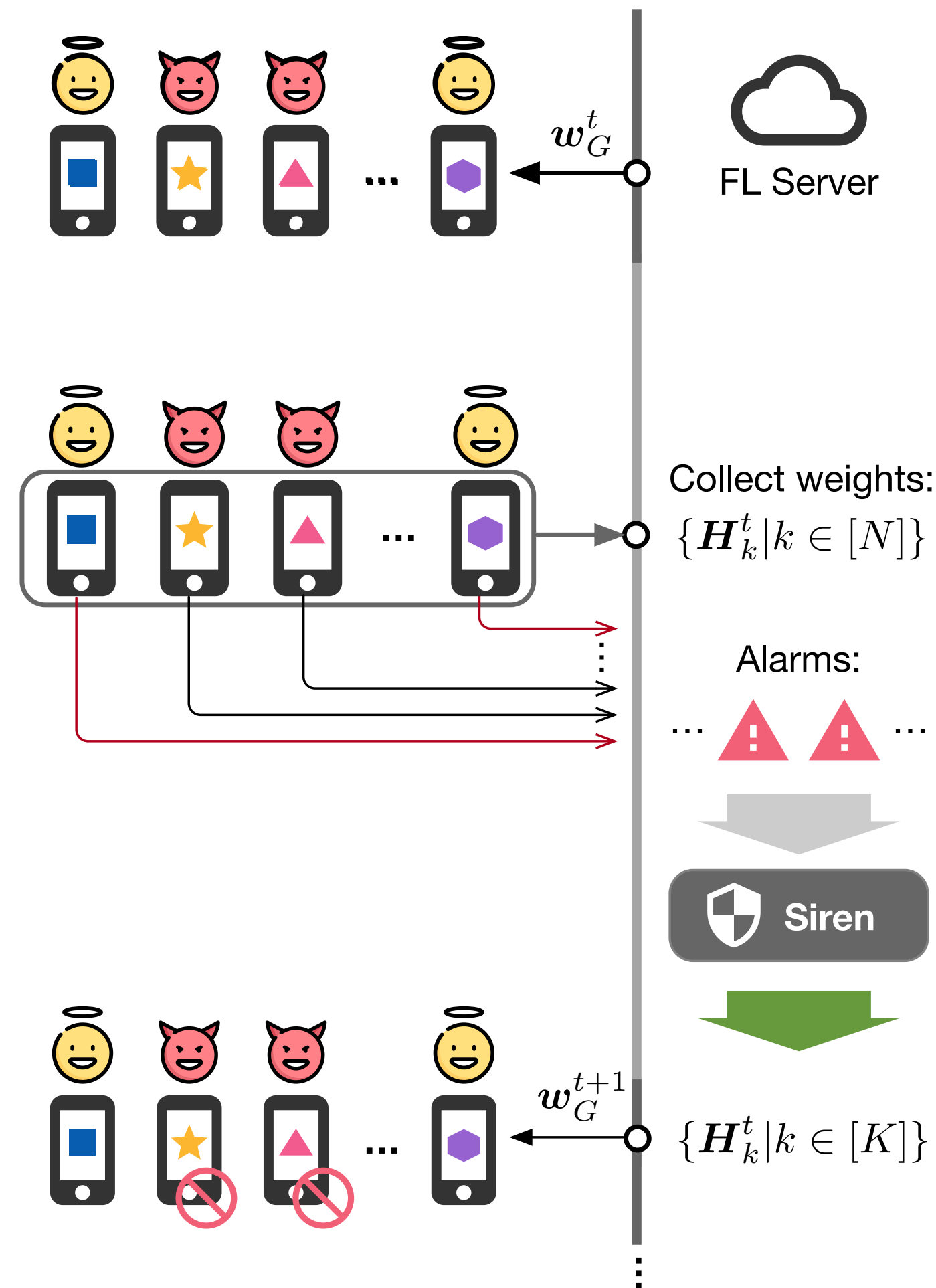


Projects Done

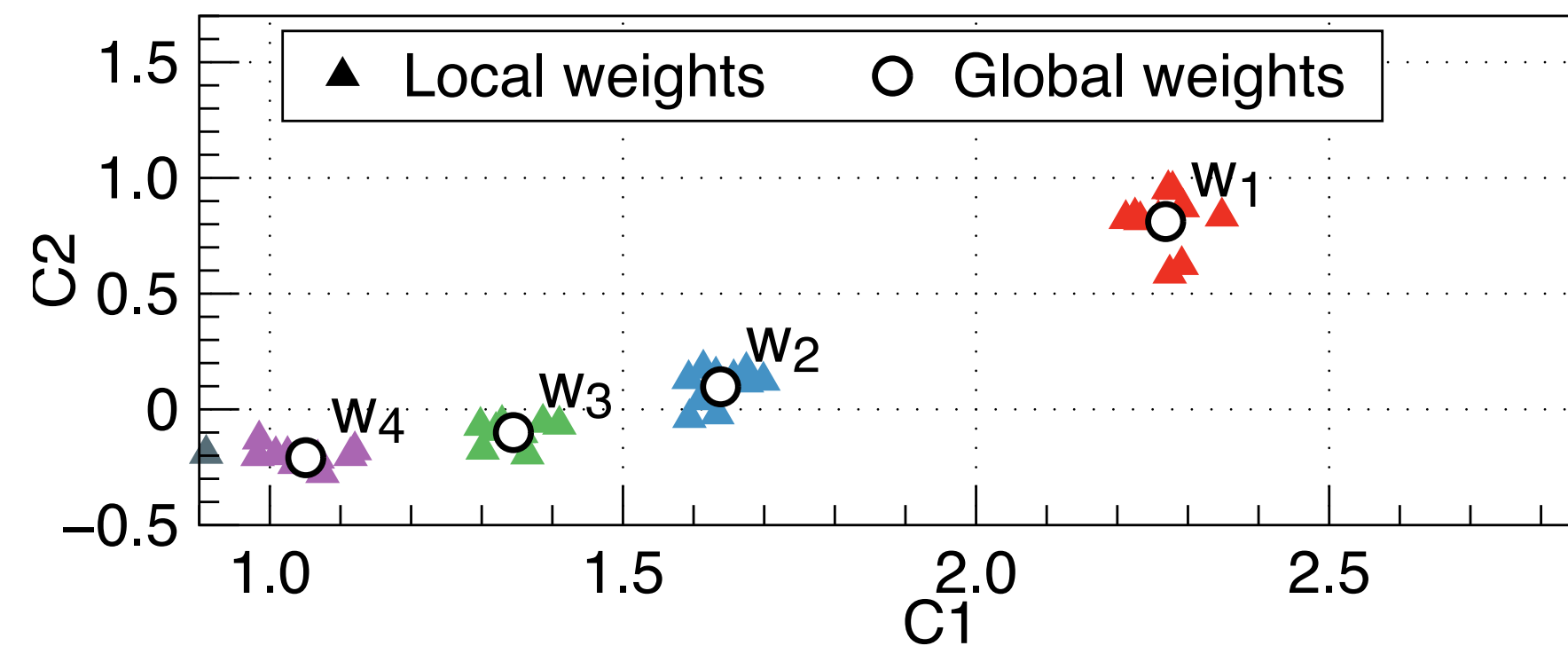
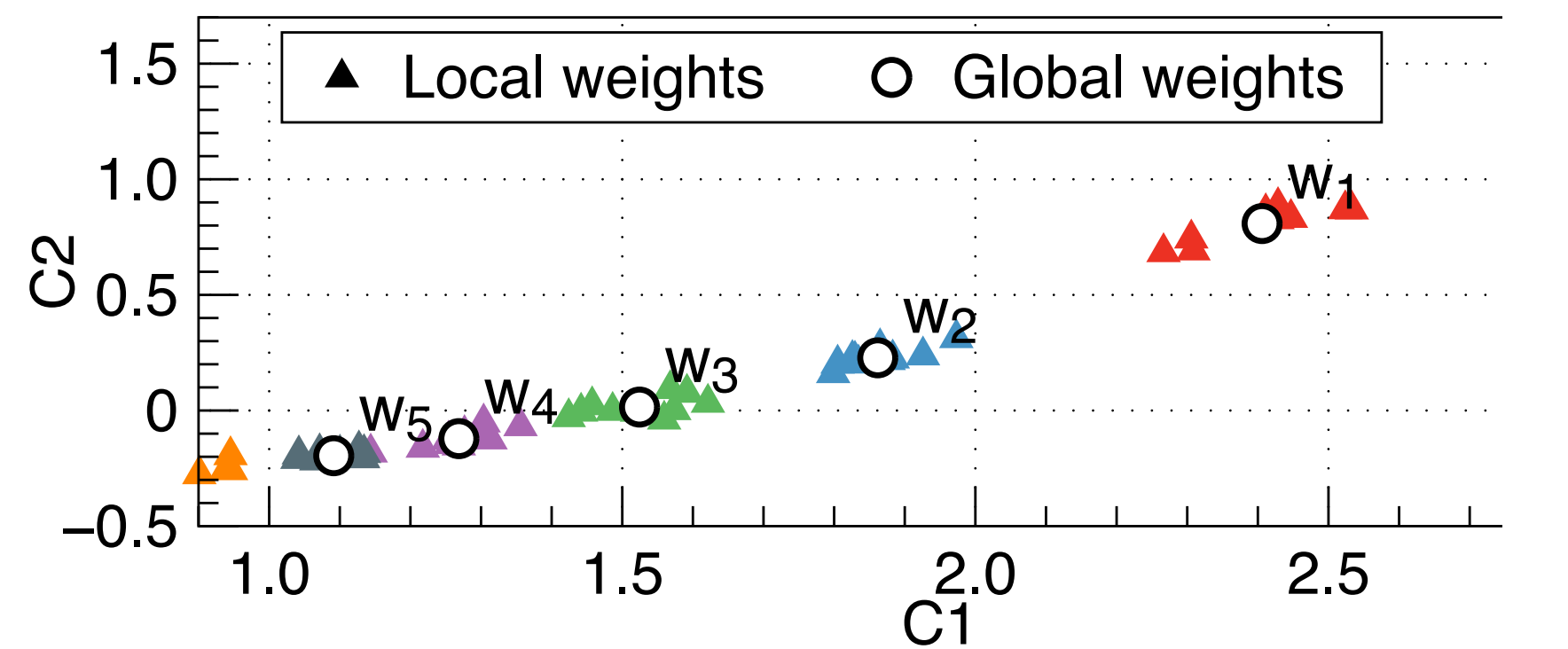


- SpaceDML: Enabling Distributed Machine Learning in Space Information Networks
- [IEEE Network]
 - Federated learning
 - Model compression
 - Space information network

Projects Ongoing



- Byzantine-robust Federated Learning
 - Federated learning
 - Byzantine attacks
 - Security & privacy



- Debugging Federated Learning
 - Mitigating stragglers
 - Model tuning & testing
 - Personalized federated learning

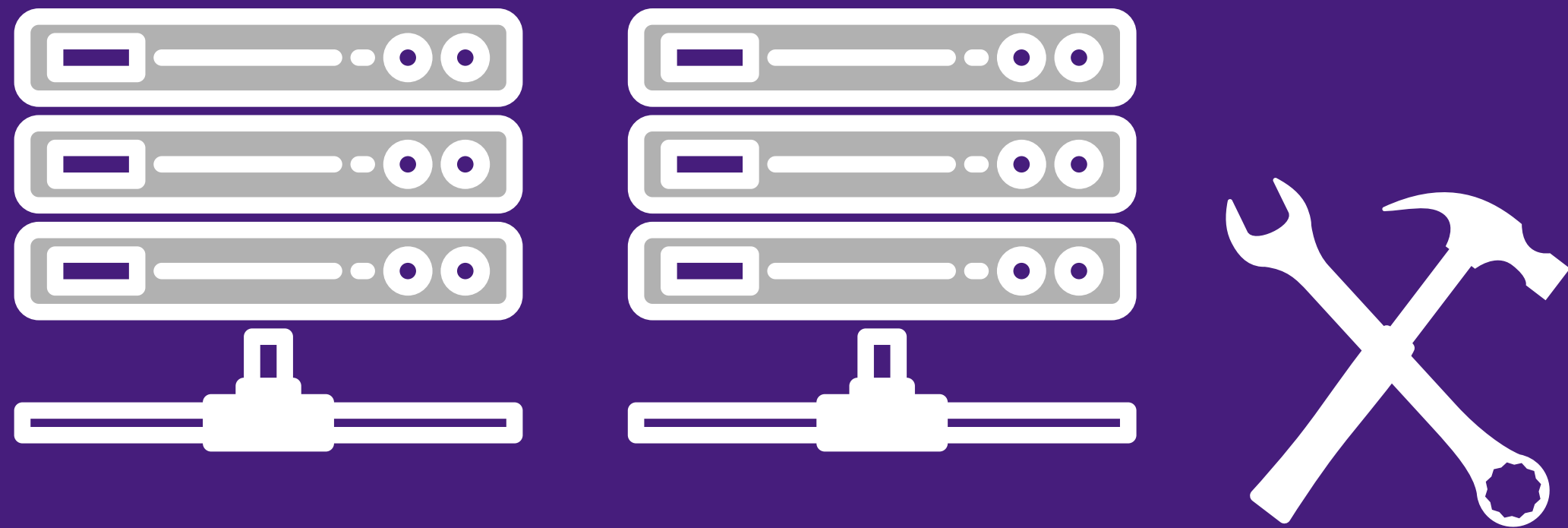
Topic #2

Serverless Computing

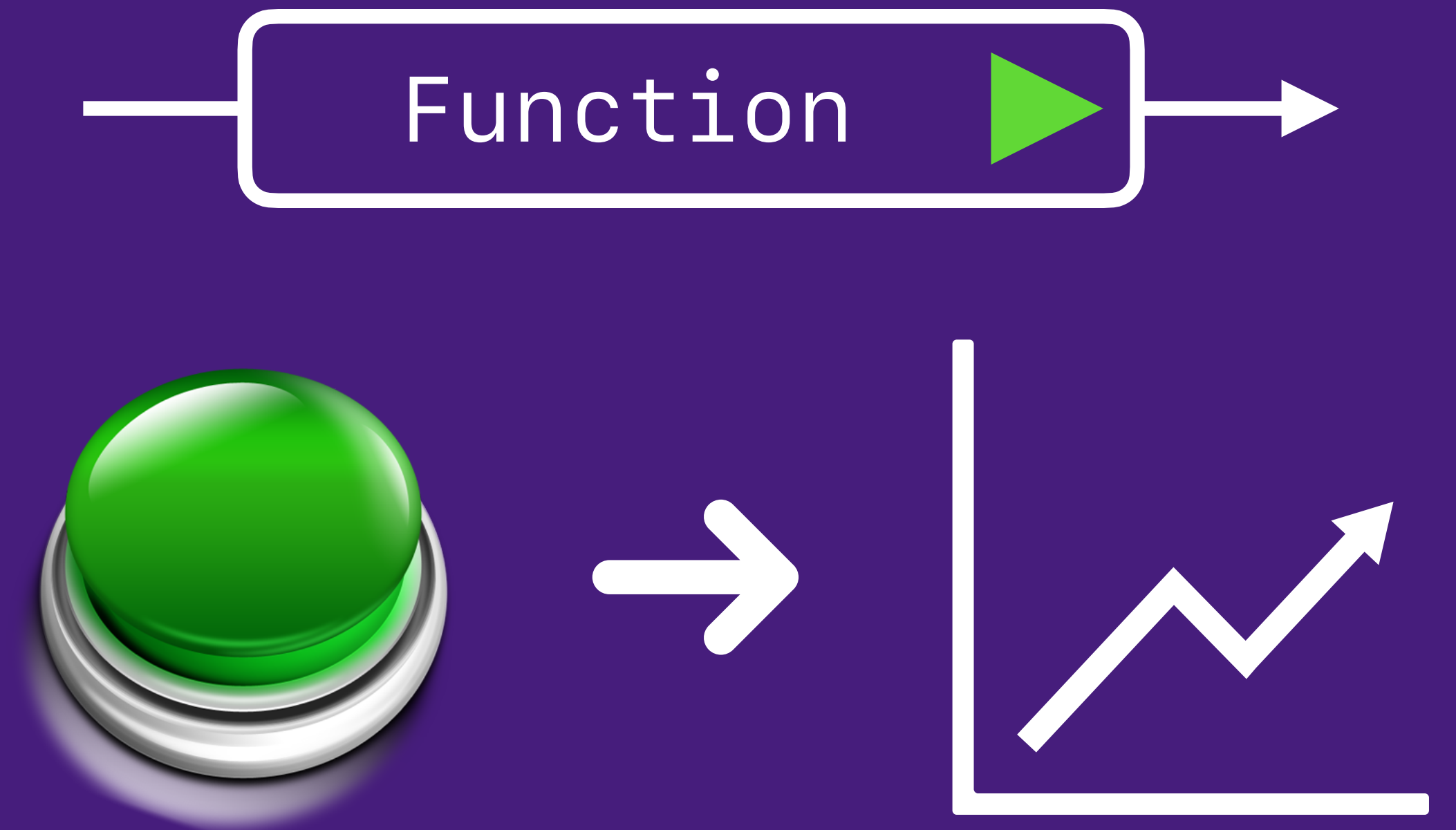
What is Serverless?

	IaaS	Serverless
Pricing	Per hour	Per call
Resources	Virtual Server	Functions
Maintenance	By users	By providers
Examples	AWS EC2	AWS Lambda

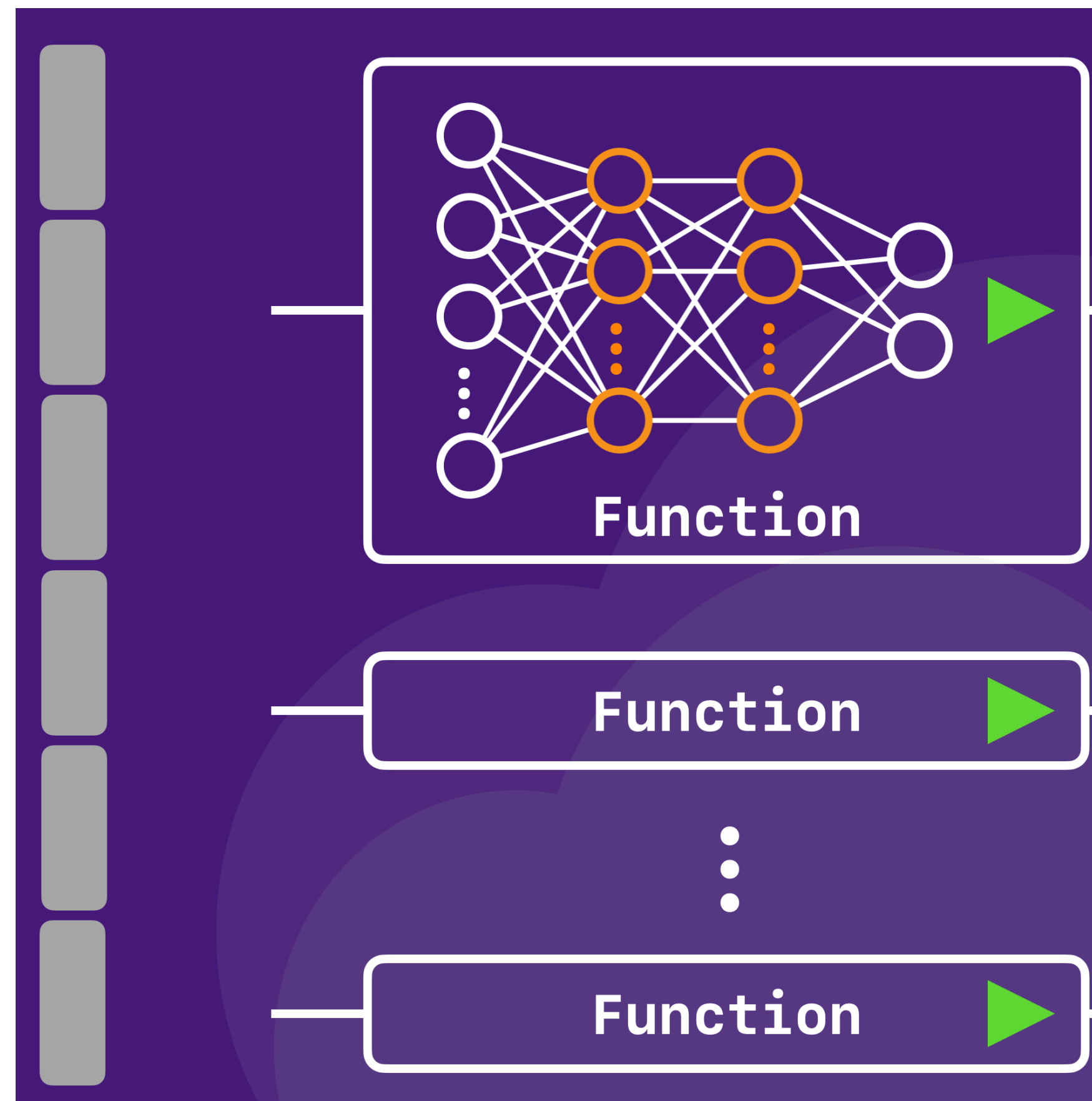
Traditional Cloud



Serverless Cloud

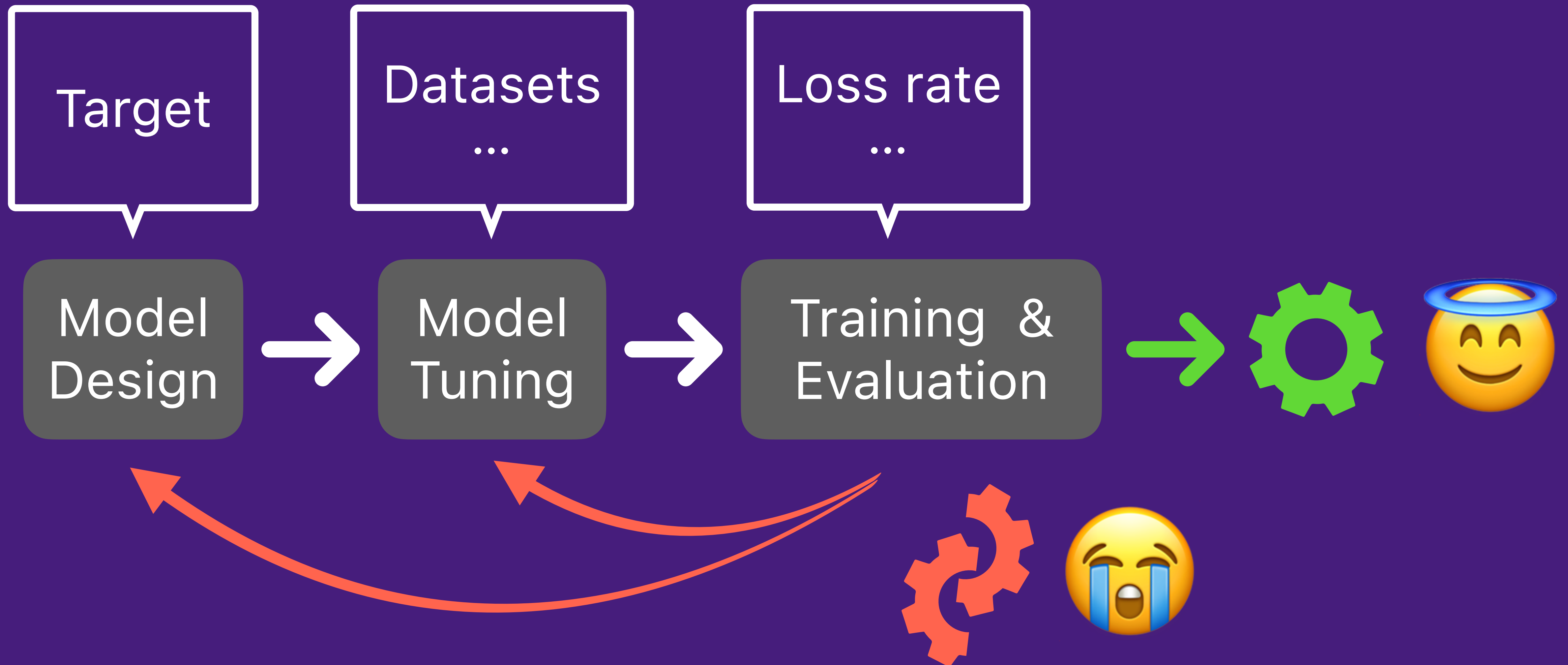


Projects Done

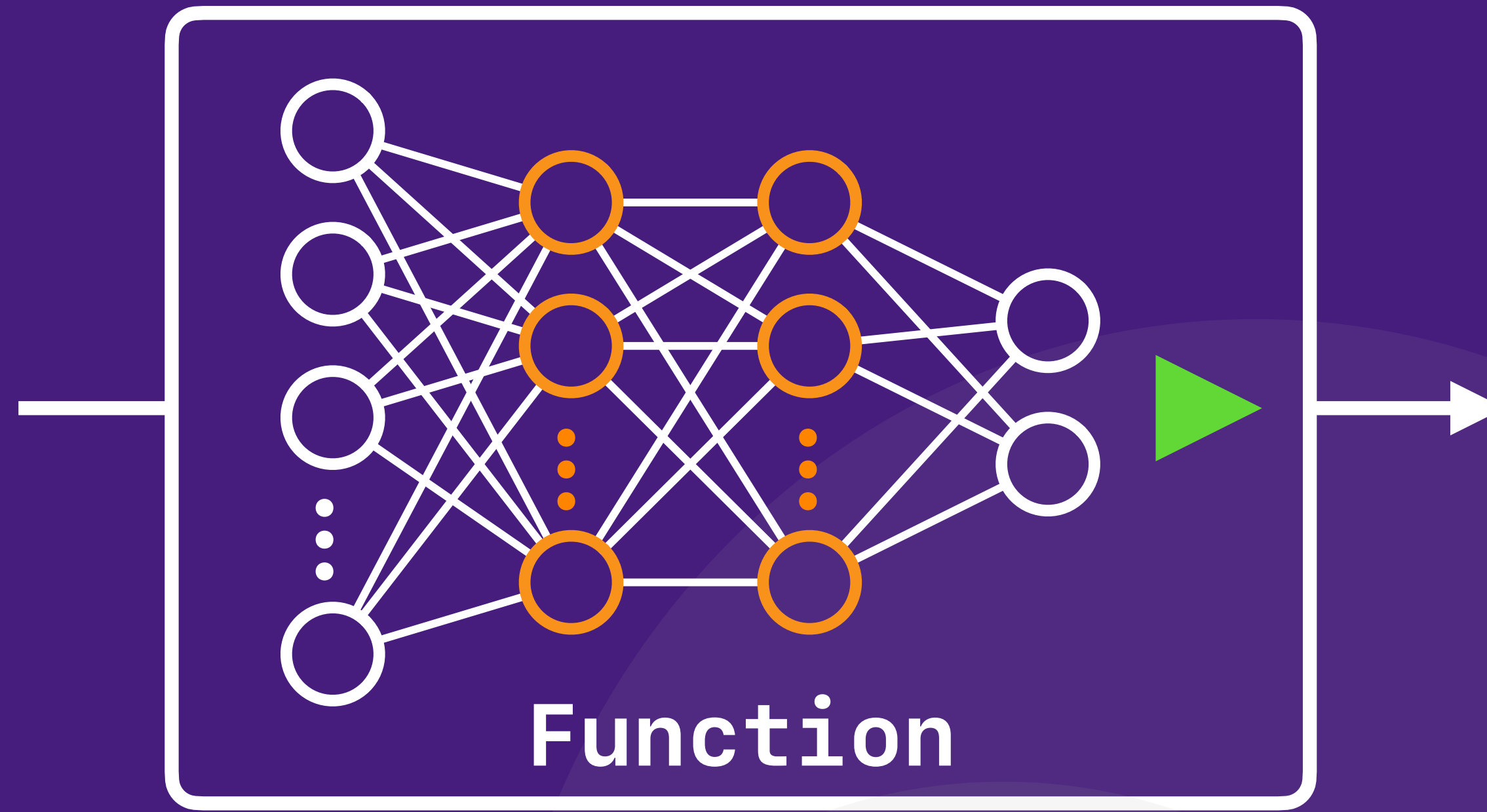


- Distributed Machine Learning with a Serverless Architecture
- [INFOCOM'19]
 - Serverless ML system
 - Resource provisioning

A General ML Workflow



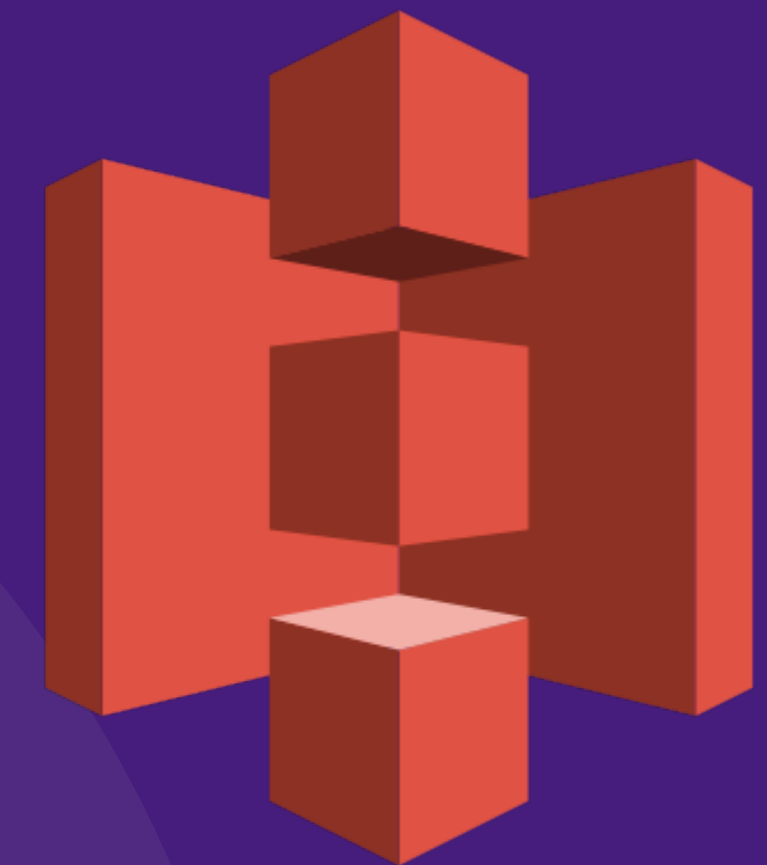
Input samples



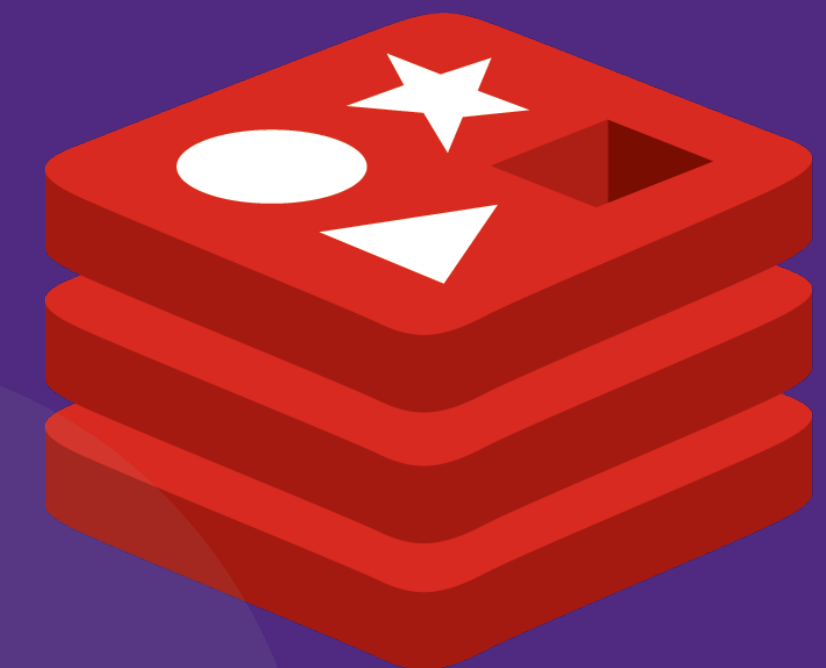
⋮



KV Storage

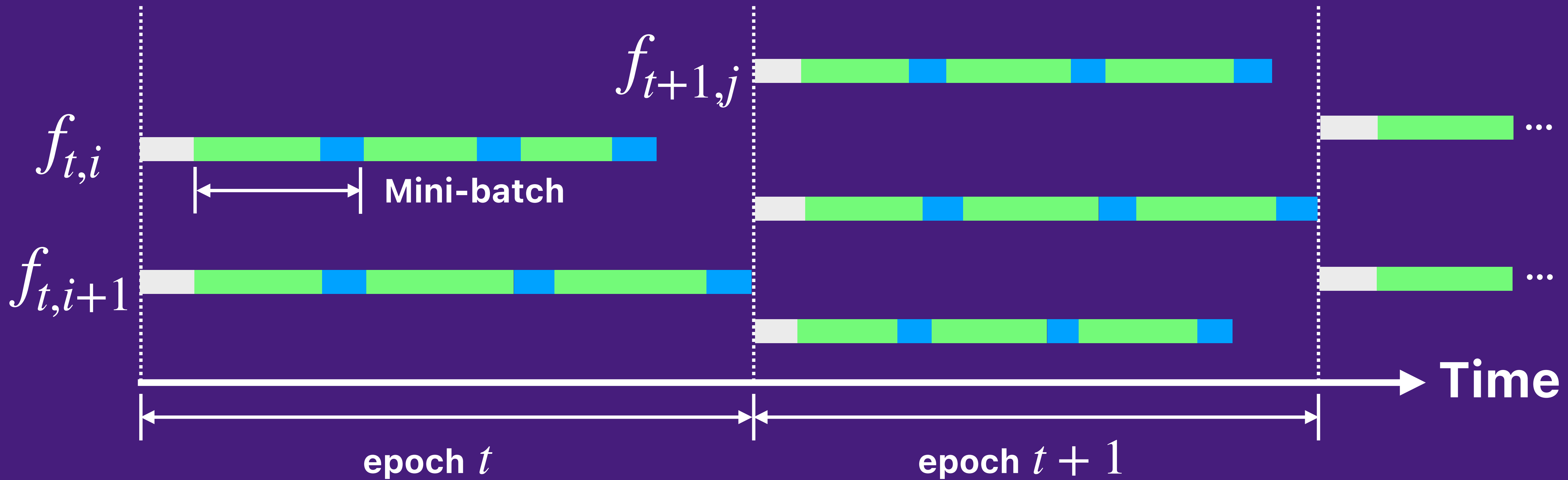


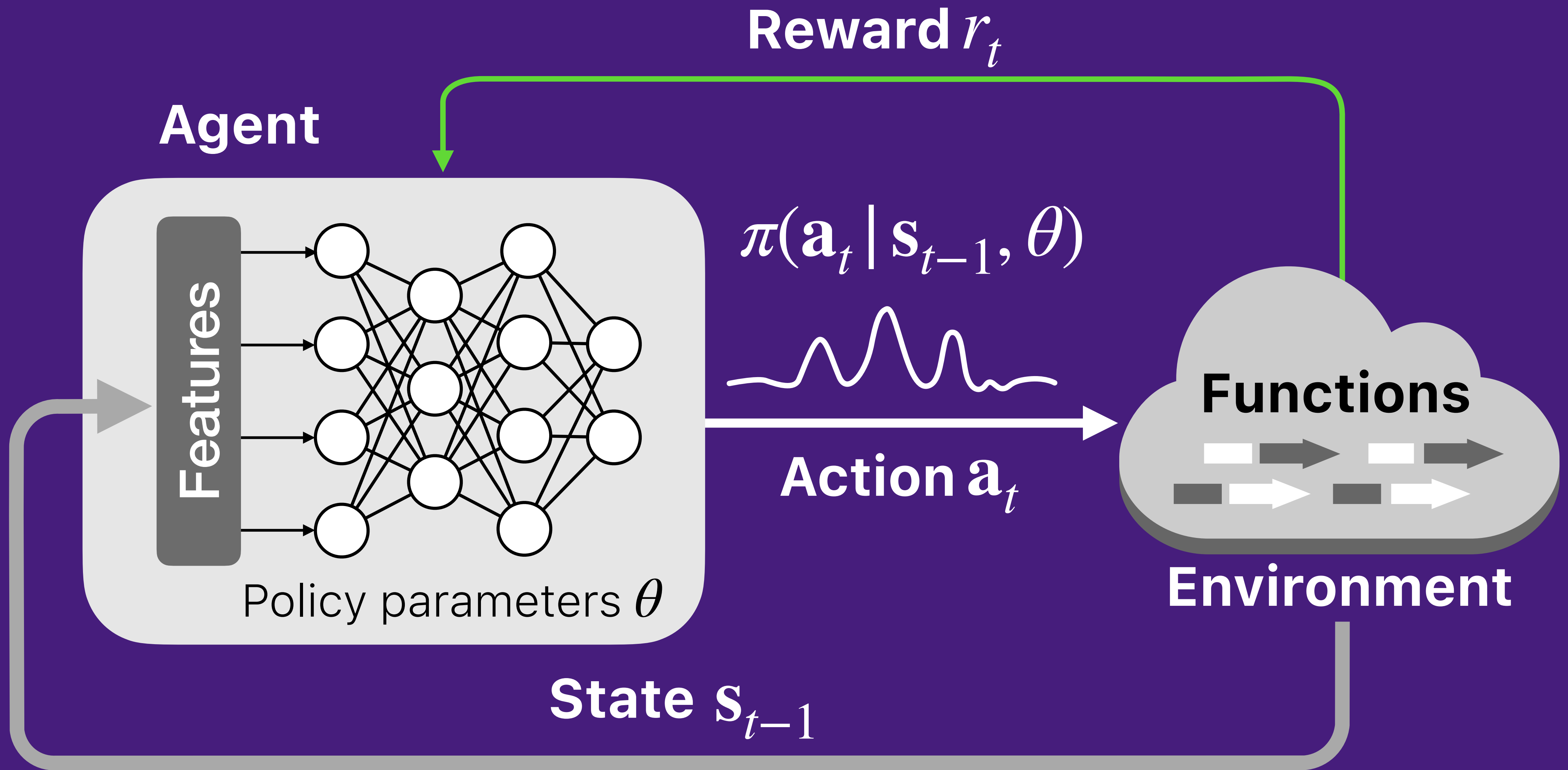
AWS S3

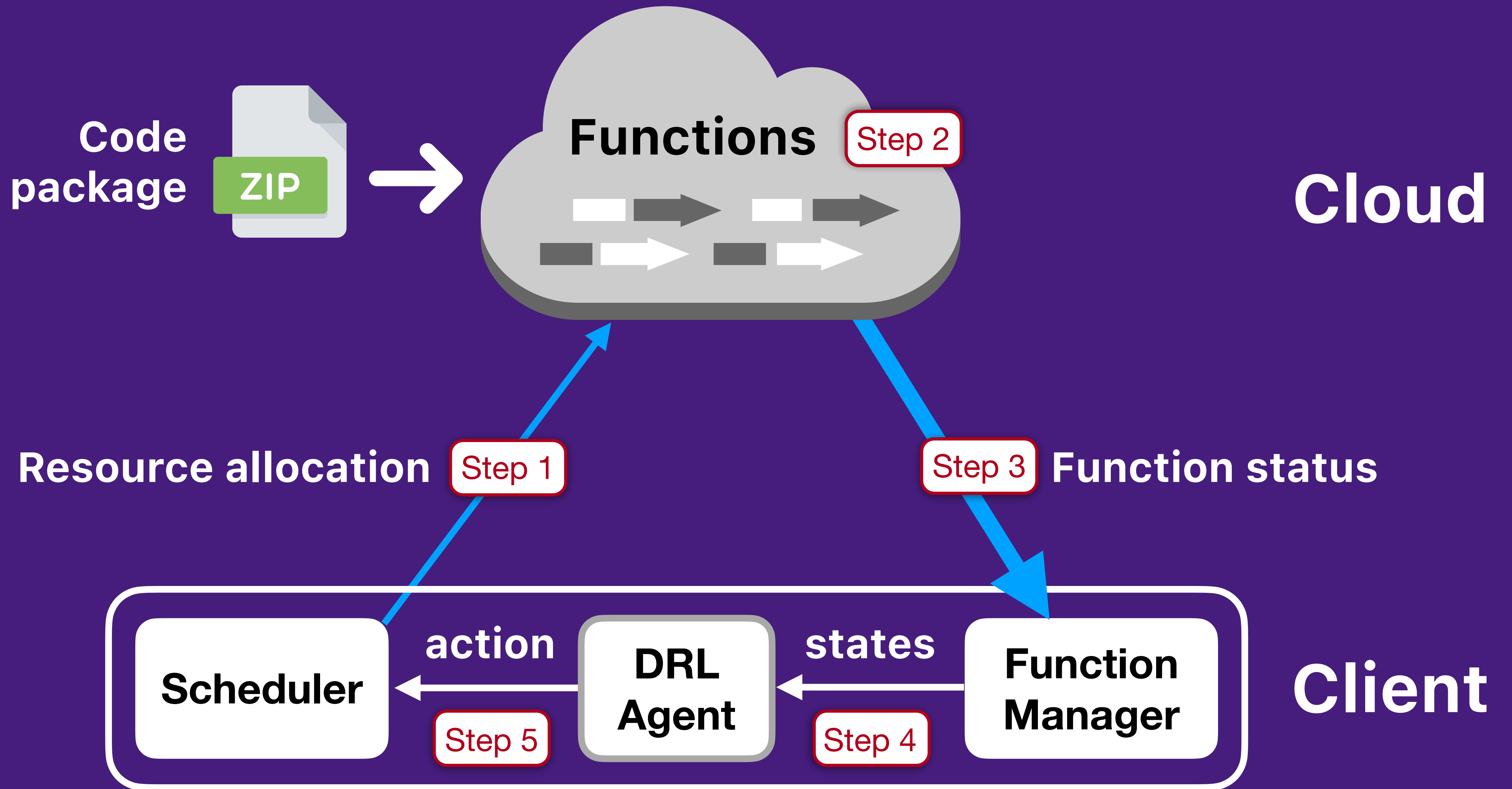


Redis

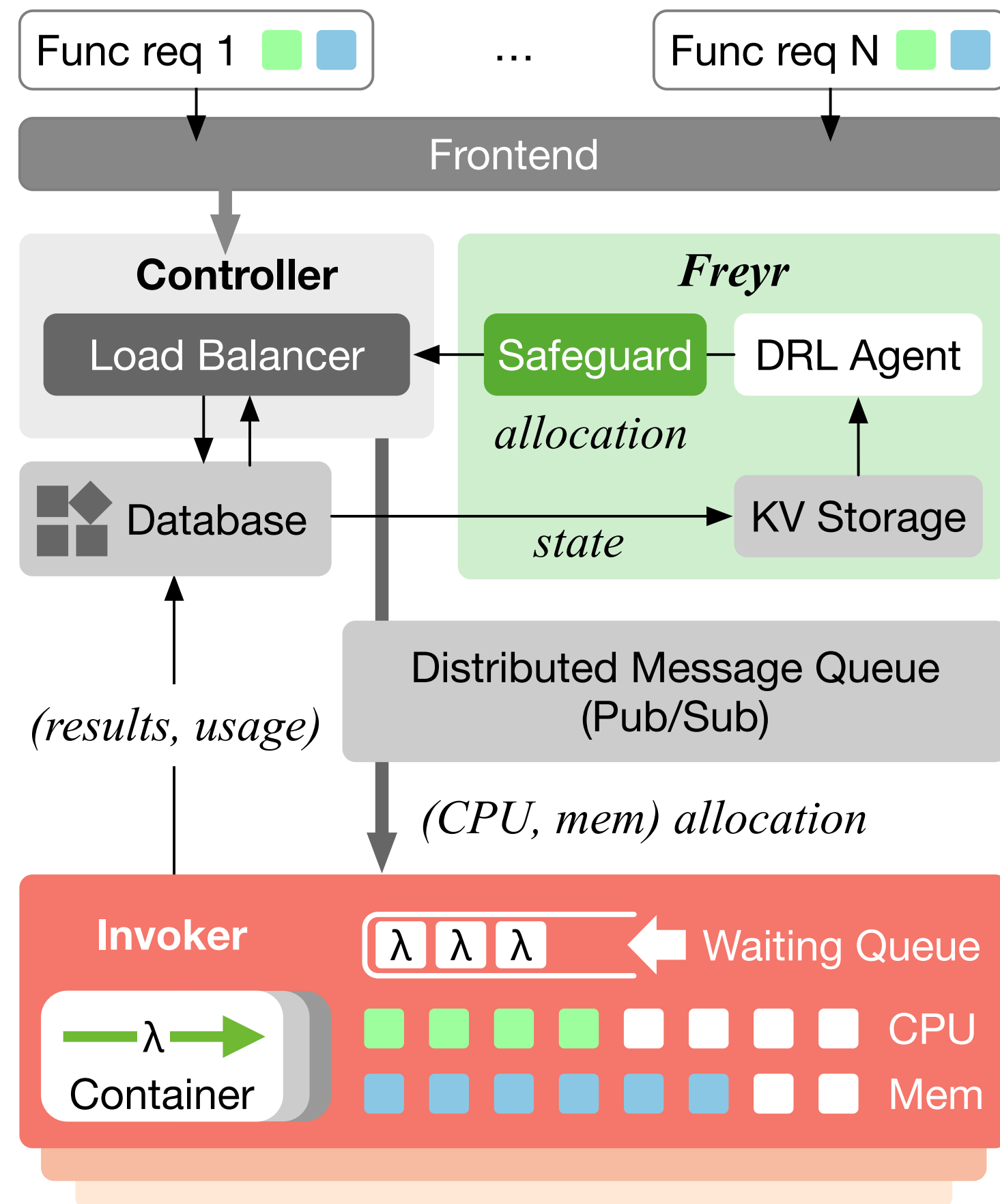
■ Fetching data ■ Computing ■ Updating Parameters





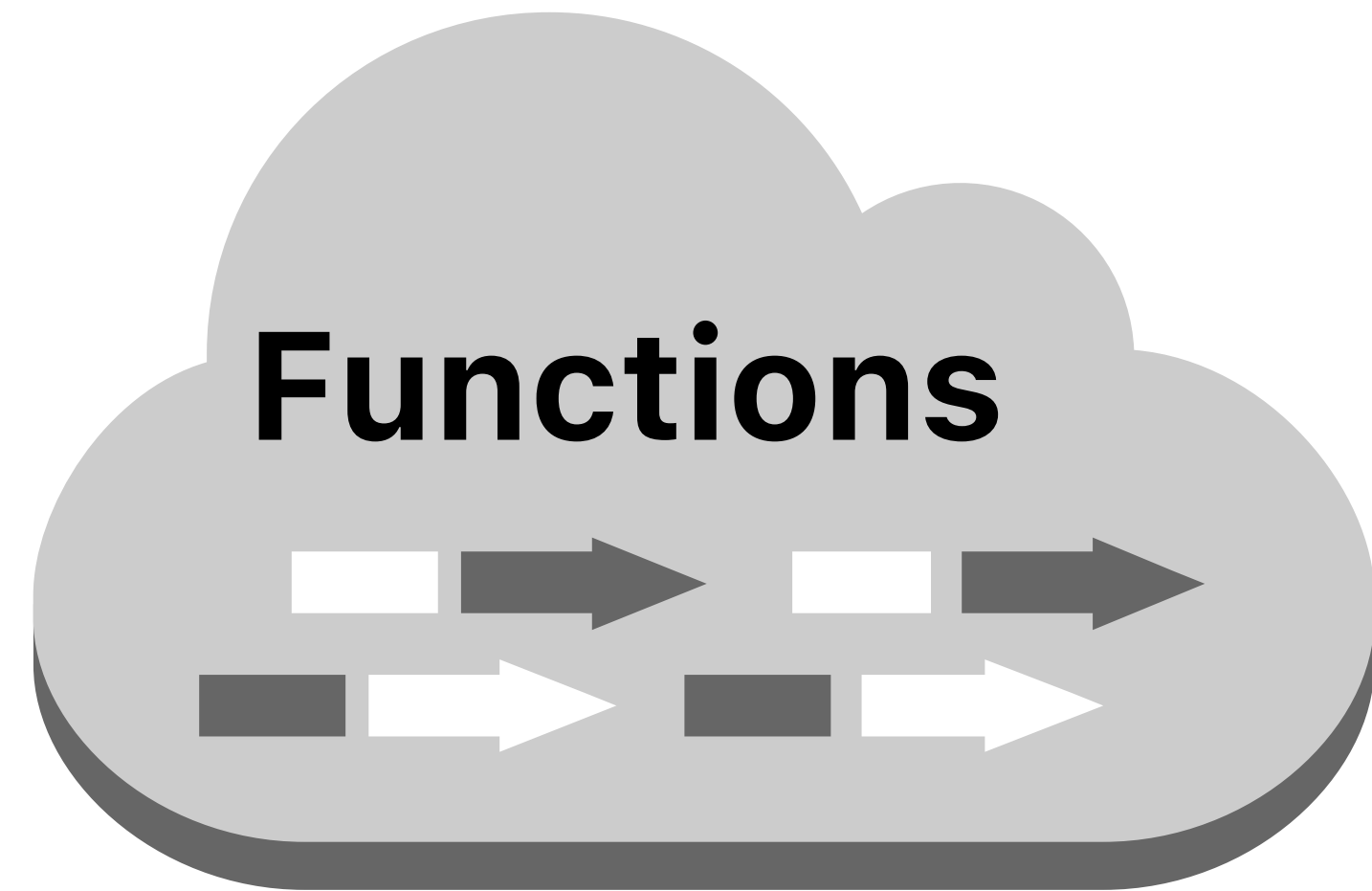


Projects Ongoing



- Resource Provisioning for Serverless
 - Resource rebalancing
 - Function acceleration

Projects Ongoing



- Optimizing Serverless
 - Function cold start
 - Resource provisioning
- New Serverless Applications
 - Serverless scientific computing
 - Serverless high performance computing

Wrap-up

- Federated Learning
 - Performance: stragglers, debugging
 - Security: Byzantine attacks & defense
- Serverless Computing
 - Cold start & other performance issues
 - New serverless applications

A wide, tree-lined street at Louisiana State University. The street is paved and has a yellow center line. Large, mature trees with dense green foliage line both sides of the street, creating a canopy effect. In the distance, a white car and a person on a bicycle are visible on the road. To the left, a person is walking on the sidewalk. To the right, a modern building with large windows is partially visible. The sky is blue with some light clouds. The text "LSU | Louisiana State University" is overlaid in white, bold, sans-serif font in the center of the image.

LSU | Louisiana State University