ICSA26
Netherlands · Amsterdam
IEEE

TELECOM SudParis
IP PARIS

FORTH
INSTITUTE OF COMPUTER SCIENCE

UNIVERSITY OF PATRAS
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

UNIVERSITY OF CRETE

MEDIATE
PANDORA

AI & Data Engineering session

Architectural Foundations for Collaborative Machine Learning in Federated Data Spaces

Nikolaos Papadakis, Kostas Magoutis, Georgios Bouloukakis

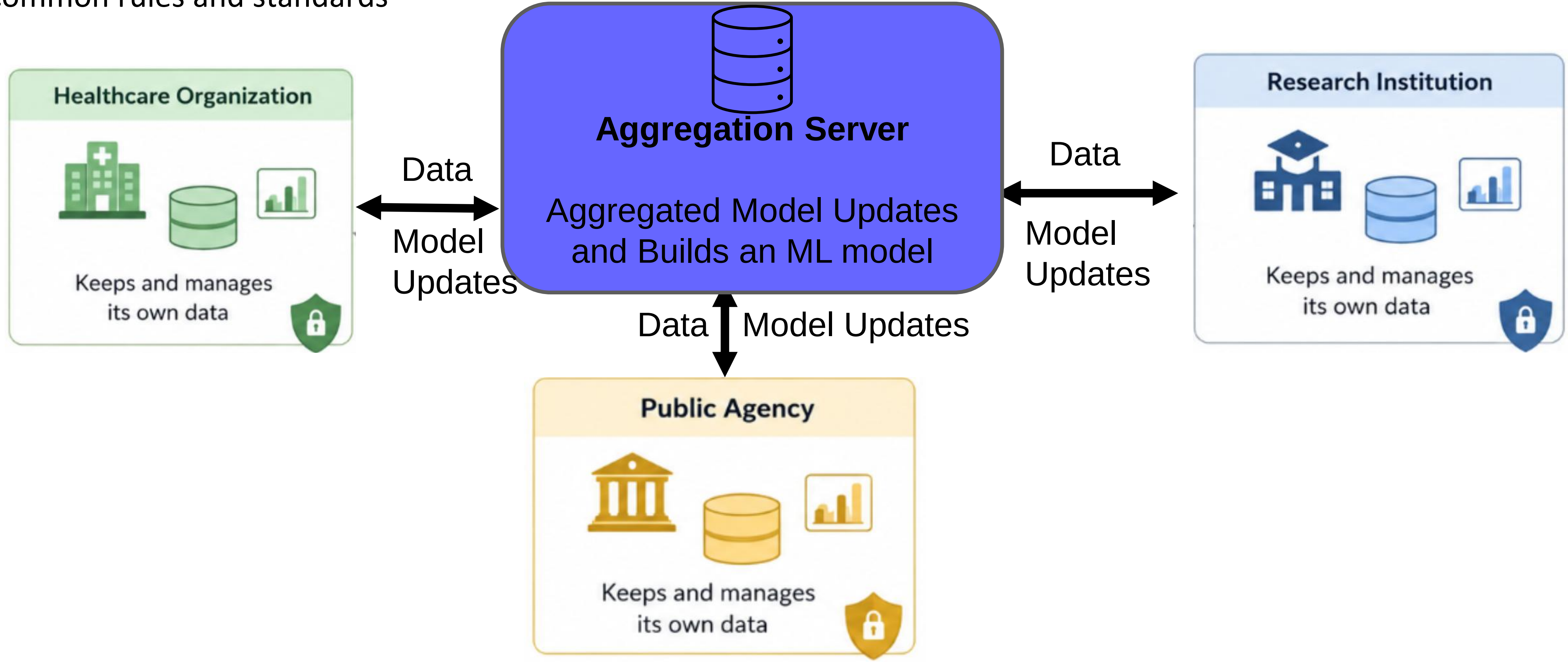
2026 International Conference on Software Architecture (ACM ICSA 2026)

25 June 2026, Amsterdam, Netherlands

Federated data spaces & collaborative ML

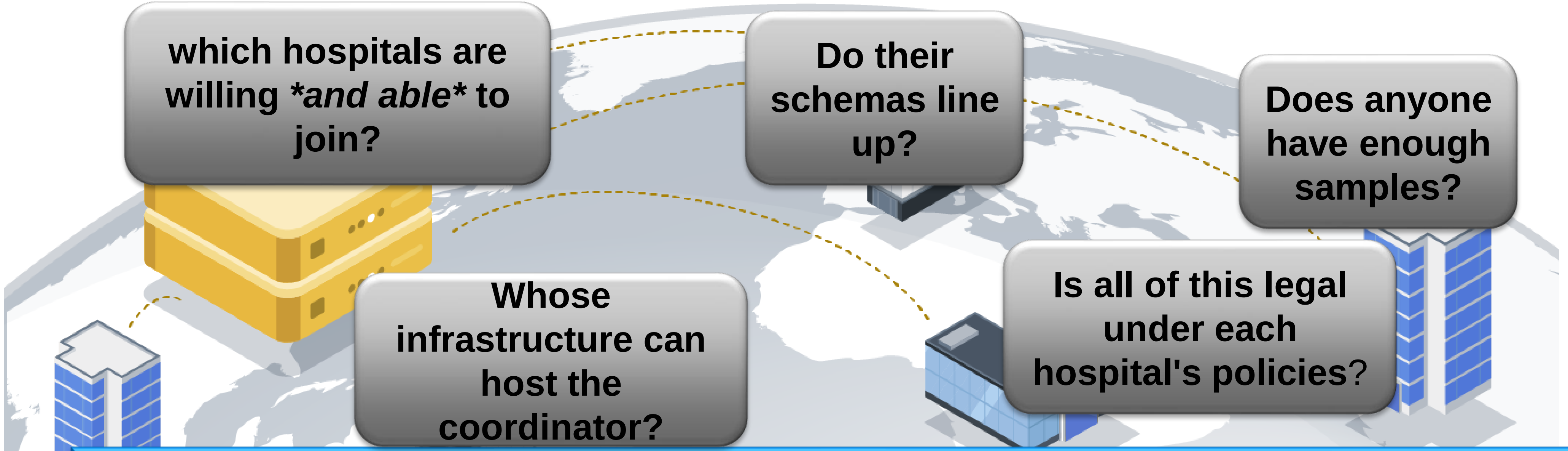
TELECOM SudParis IP PARIS FORTH PANDORA MEDIATE

Federated Data Spaces: Collaborative environments where multiple organizations share and access data using common rules and standards



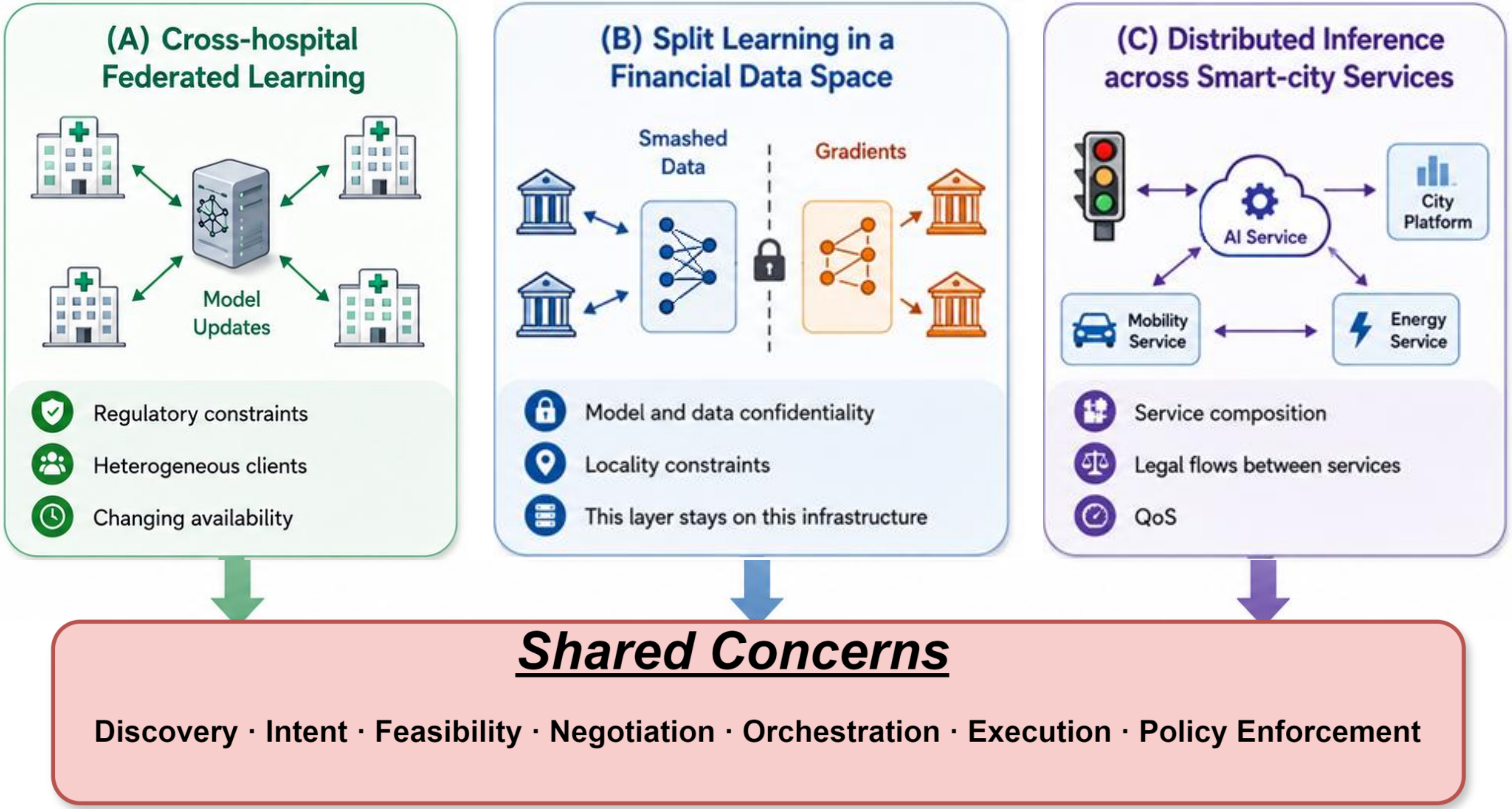
How it actually works today

Imagine **five hospitals** (H_1–H_5) want to collaboratively train a diagnostic model on chest X-rays.

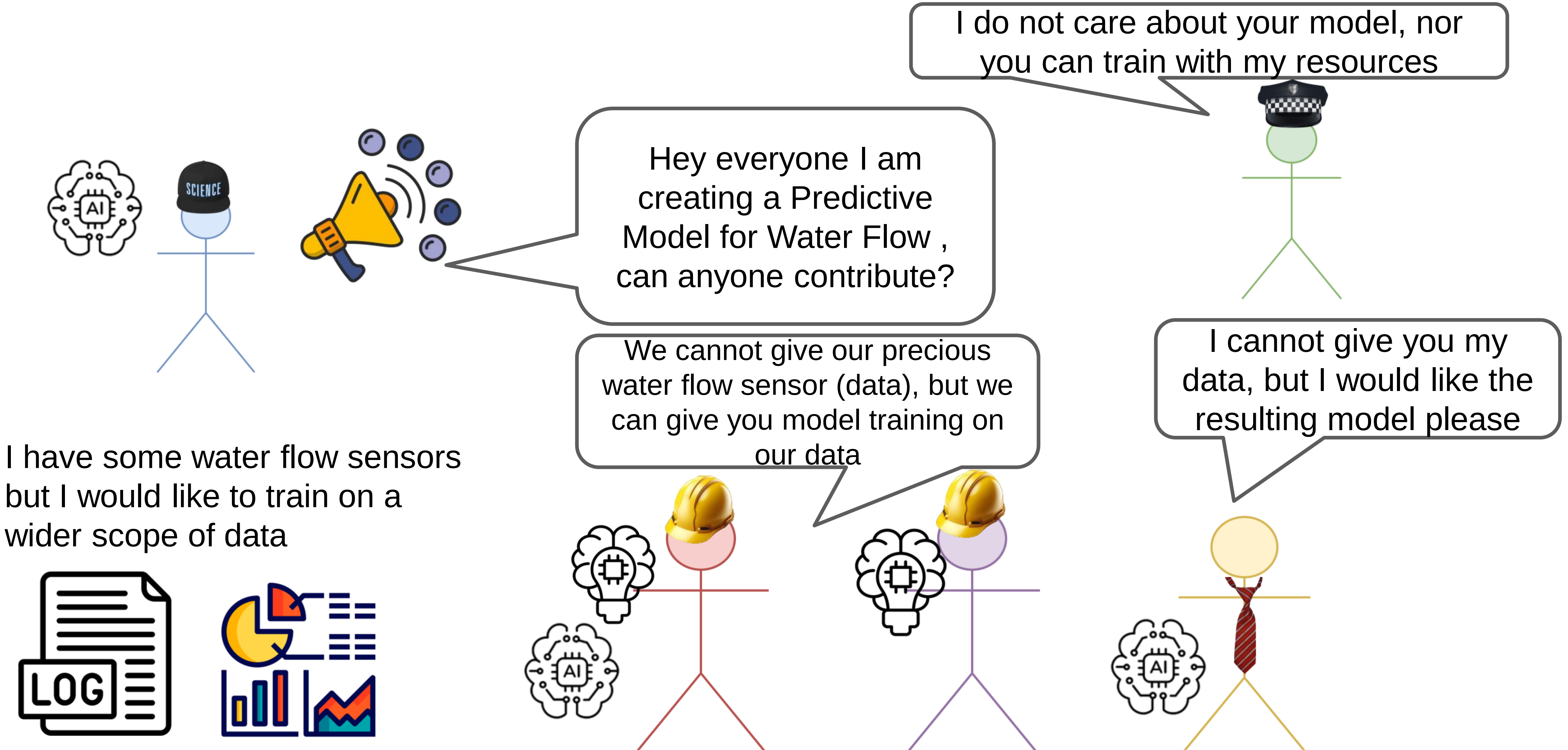


Textbook FL:
The hospitals **agree on a single FL paradigm** (say, FedAvg).
All will participate, have compatible data, and can exchange model weights.
(Usually) A central orchestrator coordinates rounds.
→. You just plug in Flower or TensorFlow Federated, and run.

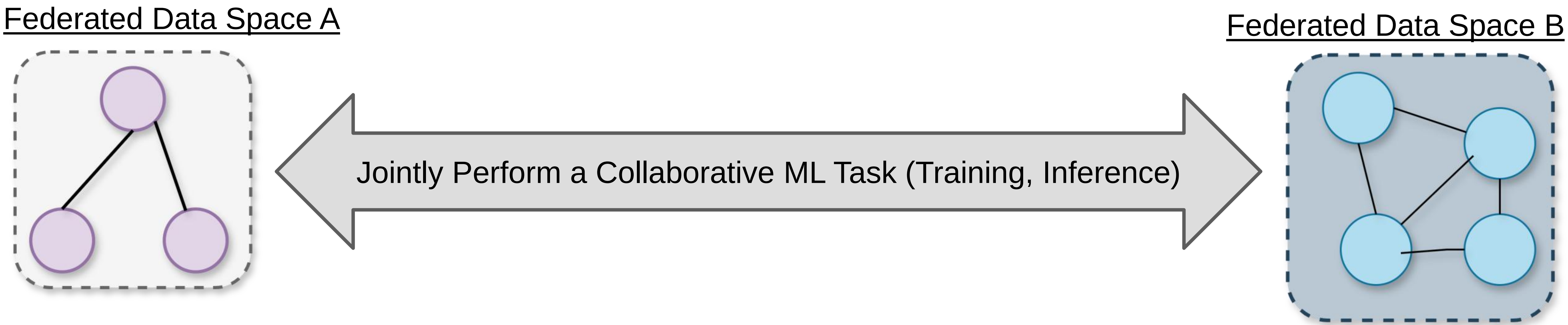
Example scenarios



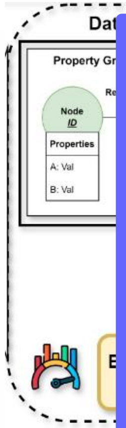
A broader vision



Architecture overview



Collaboration Lifecycle




```
"@id":            "task:hf_diagnosis_fl_v1" ,
"@type":          "TaskIntent" ,
"pattern":        "federated_learning" ,
"objective":      "heart_failure_risk" ,
"data_class":     "clinical.tabular" ,
"min_partners":   3 ,
"residency":      "EU_only" ,
"raw_data_egress": false ,
"target_metric": { "name": "AUROC", "min": 0.82 },
"budget":         { "rounds": 100, "wallclock_s": 600 },
"policies":       ["gdpr_art9", "hospital_irb", "gaiax_label_3" ],
"audit":          "on_every_round"      // → §6 Policy
```

Prototype

TELECOM SudParis

IP PARIS

FORTH

PANDORA

MEDIATE

Distributed Knowledge Base

MQTT

Discovery and Intent

NGSI LD

zenodo

python™

Collaborative ML API

Flower Framework

Orchestration & Execution

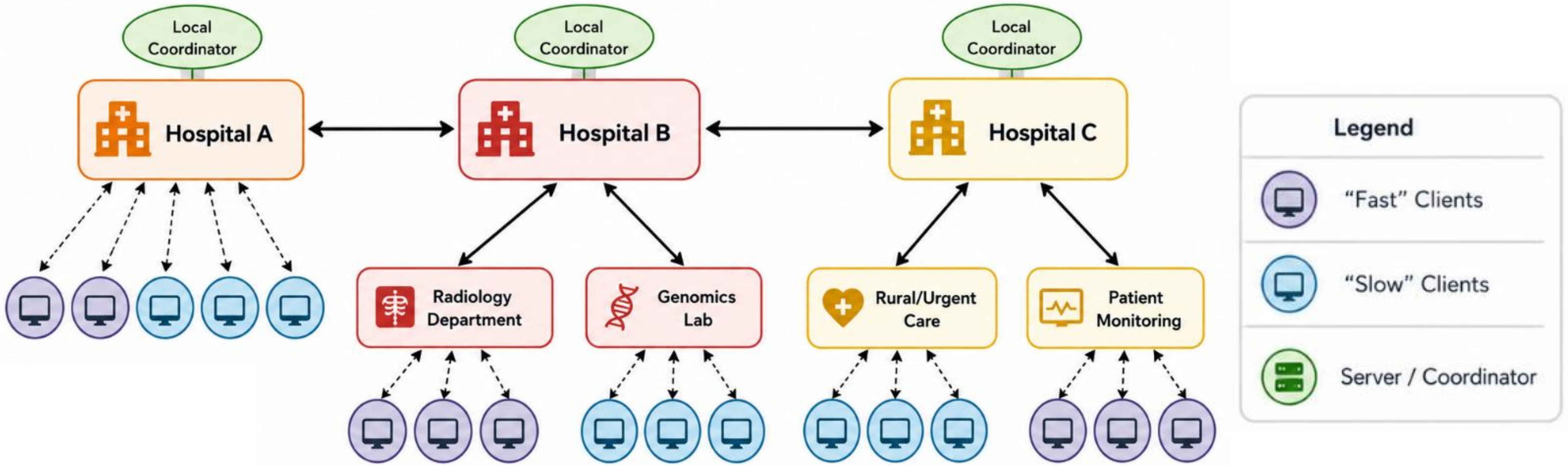
python™

Policy Enforcement

Benchmarking (Stored Past Results)

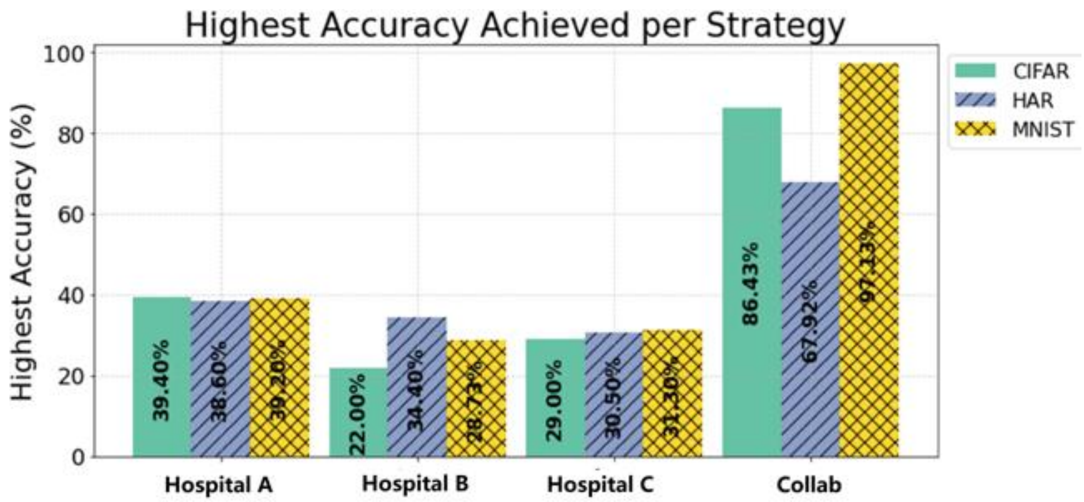
Feasibility Assessment (Manual)

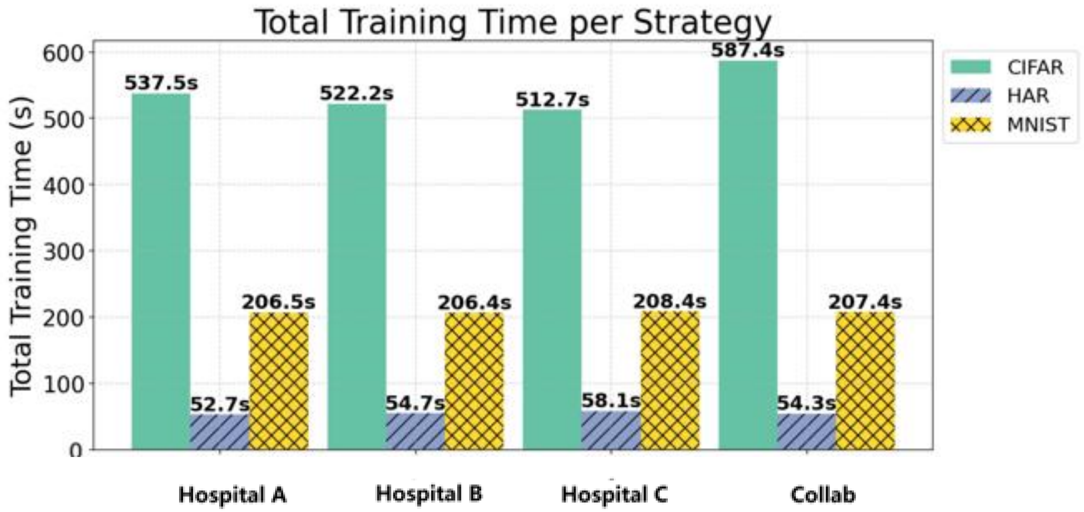
Evaluation setup

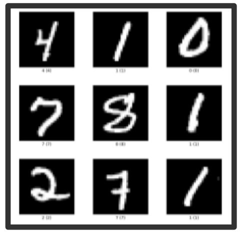


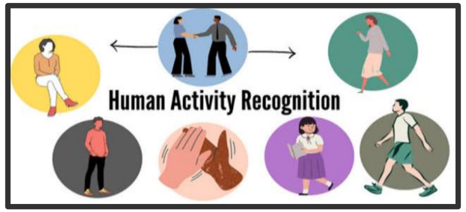
Role	Instance Type	Family	Size	vCPUs	Memory (GiB)
Fast FL clients	c5.large	c5	large	2	4
Slow FL clients	t3.medium	t3	medium	2	4
FL server / aggregator	c5.xlarge	c5	xlarge	4	8

Some Results









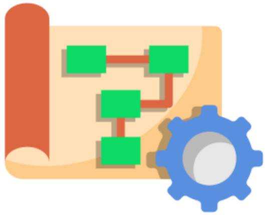
≈ 3× accuracy lift on CIFAR-10, at < 10 % more wall-clock.

Conclusions & Future Work

Key takeaways

- Cross-dataspace ML is mainly a coordination and architecture challenge.
- Six capabilities support the lifecycle: Discovery, Representation, Context, Feasibility, Orchestration, and Policy Enforcement.
- A Distributed Knowledge Base manages shared metadata, constraints, plans, context, and provenance.
- Collaboration can improve model accuracy while maintaining comparable training time.

Future work



- Automate feasibility checks, and context interpretation.
- Evaluate the architecture at larger scale and under realistic network conditions.
- Support richer policies, metadata, and stricter constraints.
- Extend validation to split learning and distributed inference.
- Improve resilience to failures, malicious participants, and adversarial behaviour.

Thank you!

Thank you!

UNIVERSITY OF PATRAS ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

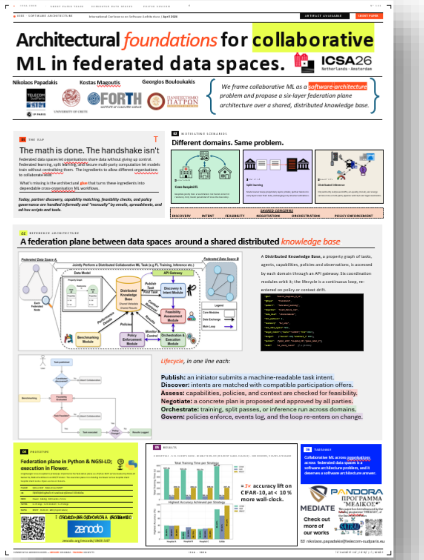
TELECOM SudParis

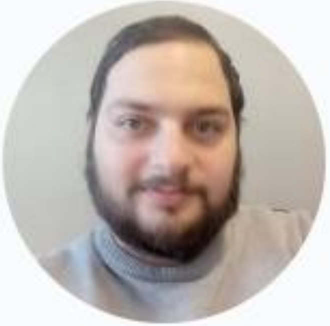
IP PARIS

UNIVERSITY OF CRETE

FORTH INSTITUTE OF COMPUTER SCIENCE

You can also check out our poster





Nikolaos Papadakis

PhD Student

Télécom SudParis / IP Paris

Feel free to contact me at:

nikolaos.Papadakis @telecom-sudparis.eu

PANDORA

MEDIATE

Try out our solution