OASIS: Open-Source AI-as-a-Service Framework for Federated, Efficient, and Drift-Robust Learning in the Continuum Edge–Cloud
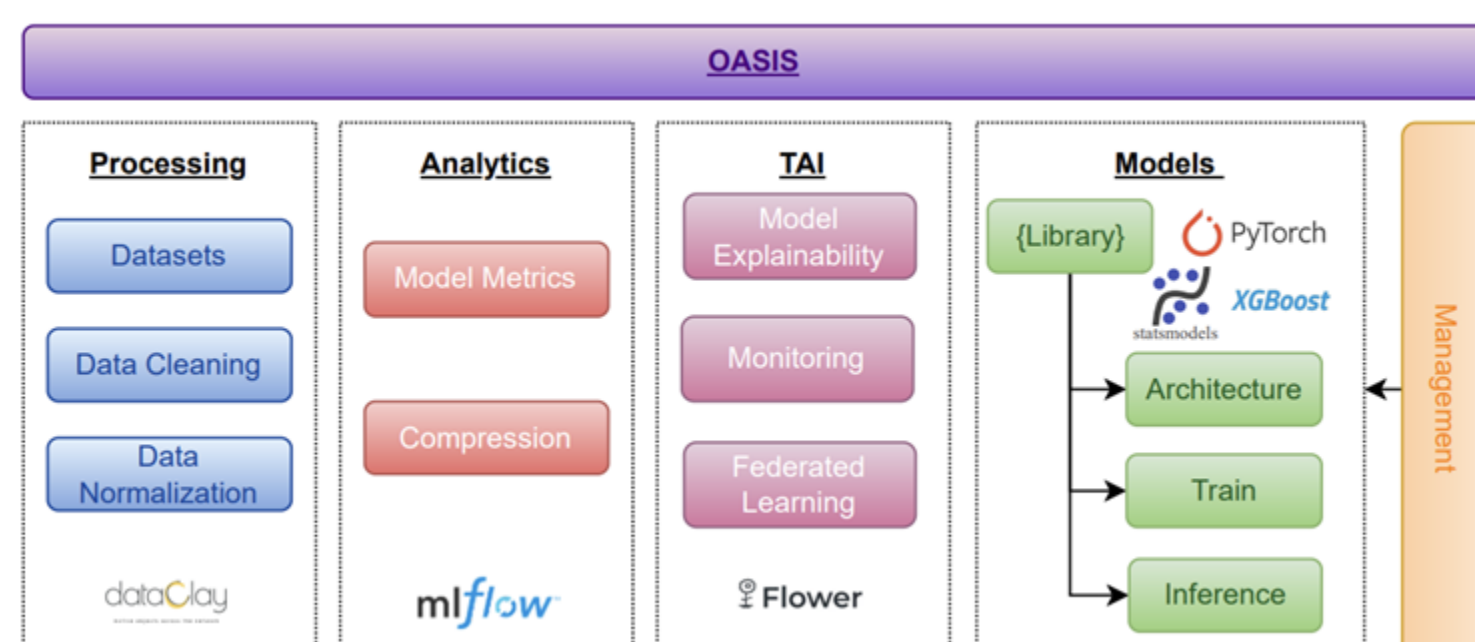
Sebastián Andrés Cajas[1], Jaydeep Samanta[1], Andrés L. Suárez-Cetrulo[1],
Romila Ghosh[1], Anastasios Giannopoulos[2], Alex Barceló[3], and Ricardo Simón Carbajo[1]

[1]CeADAR (Ireland's Centre for Applied AI), University College Dublin, [2]National and Kapodistrian University of Athens, [3]Barcelona Supercomputing Center

PROBLEM & SOLUTION

- **Fragmented AI tools** across IoT-Edge-Cloud continuum
- **Resource constraints** on edge devices limit AI deployment
- **Privacy concerns** with centralized data processing
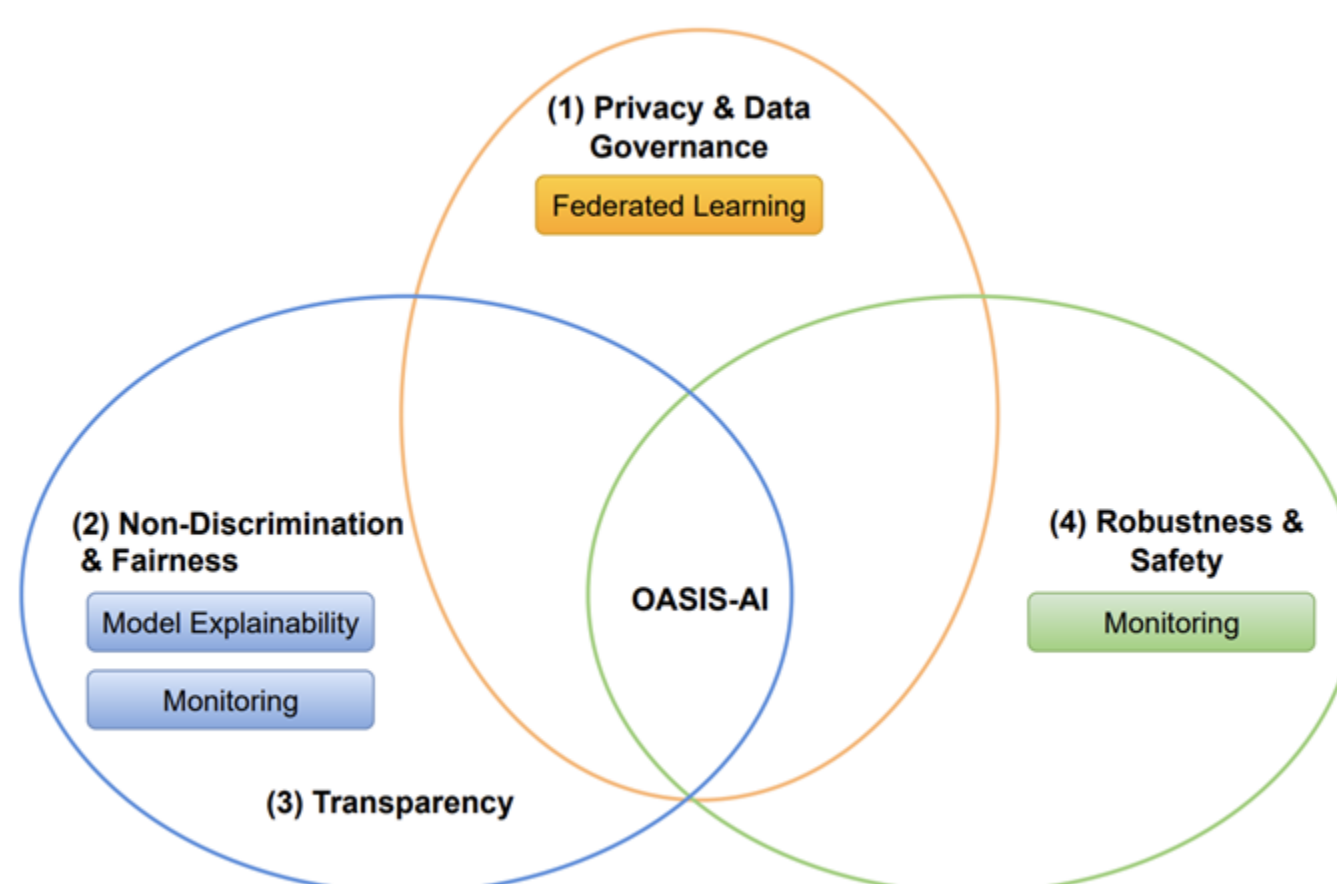- **Complex integration** of ML lifecycle components

OASIS SOLUTION

- **Open-source framework**
- **Library-agnostic**
- **Unified ML pipeline**
- **Edge-to-cloud support**
- **Deployment via Docker and Helm charts**
- **API-Driven**

KEY FEATURES & ARCHITECTURE

- **Federated Learning**: Privacy-preserving distributed training
- **Model Compression**: Quantization + distillation for edge devices
- **Drift Detection**: Real-time monitoring with NannyML
- **Explainability**: SHAP integration for interpretable AI
- **MLOPs**: Model management & performance tracking with MLFlow

FRAMEWORK ARCHITECTURE

API EXAMPLES

```
POST /train
POST /predict
POST /show_models
POST /detect_drift
POST /get_anomalies
POST /remove_model
POST /sync_model_repo
POST /launch_MLFlow_ui
```

RESULTS & IMPACT

Method	Before	After	Reduction	Performance
Quantization	70.66 KB	21.68 KB	3×	Maintained
Distillation	70.36 KB	1.94 KB	36×	2× faster

SUPPORTED LIBRARIES

- River
- BentoML
- MLFlow
- NannyML
- SHAP
- Flower
- PyTorch
- Tensorflow
- XGBoost
- Dataclay
- Transformers
- JupyterLab

REAL WORLD PERFORMANCE

- Deployed on x86 (AMD, Intel) & ARM64 (Nvidia Jetson, Orange Pi, Raspberry Pi, ARM MacBooks)
- API-driven anomaly detection for easy integration
- Supports, Smart farming robots, In-car entertainment streaming & energy management

USE CASES

APPLICATIONS
- Healthcare systems
- Industrial IoT
- Smart cities
- 5G network management

DEPLOYMENT RESULTS
- CPU/RAM forecasting
- Network anomaly detection
- Edge device optimization

CONCLUSIONS

KEY ACHIEVEMENTS
- Unified ML framework
- **3x-36x** model compression
- Privacy-preserving FL

FUTURE DIRECTIONS
- Advanced FL strategies
- Hybrid models
- Streaming support
- Domain-specific apps

AVAILABILITY

Project page