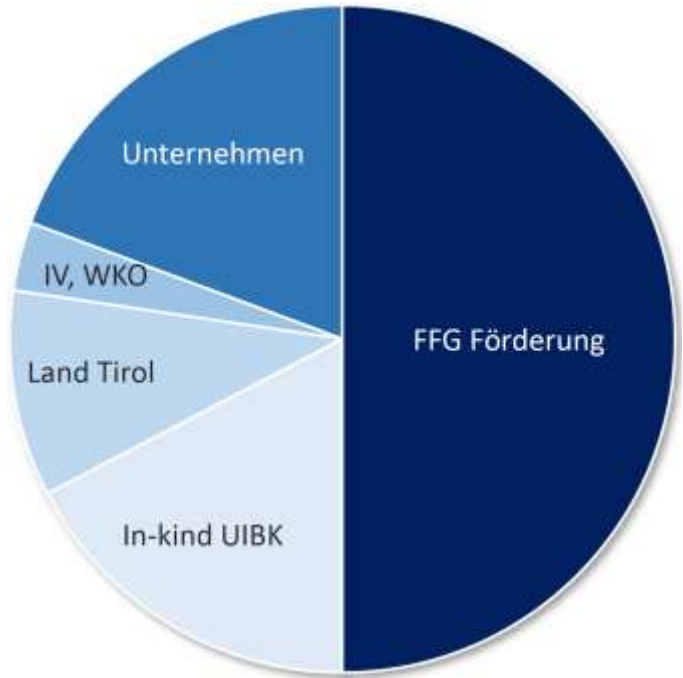


Prof. Radu Prodan

DISTRIBUTE AND LEARN 1 BILLION

Endowed Professorship Edge AI

- AI Mission Austria (AIM AT)
- Department of Computer Science, University of Innsbruck
 - 2025 – 20230



Hybrid AI

Symbolic
AI

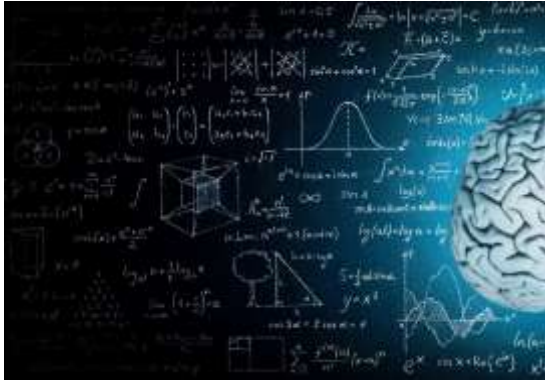


Image courtesy of shutterstock.com

Connectionist
AI

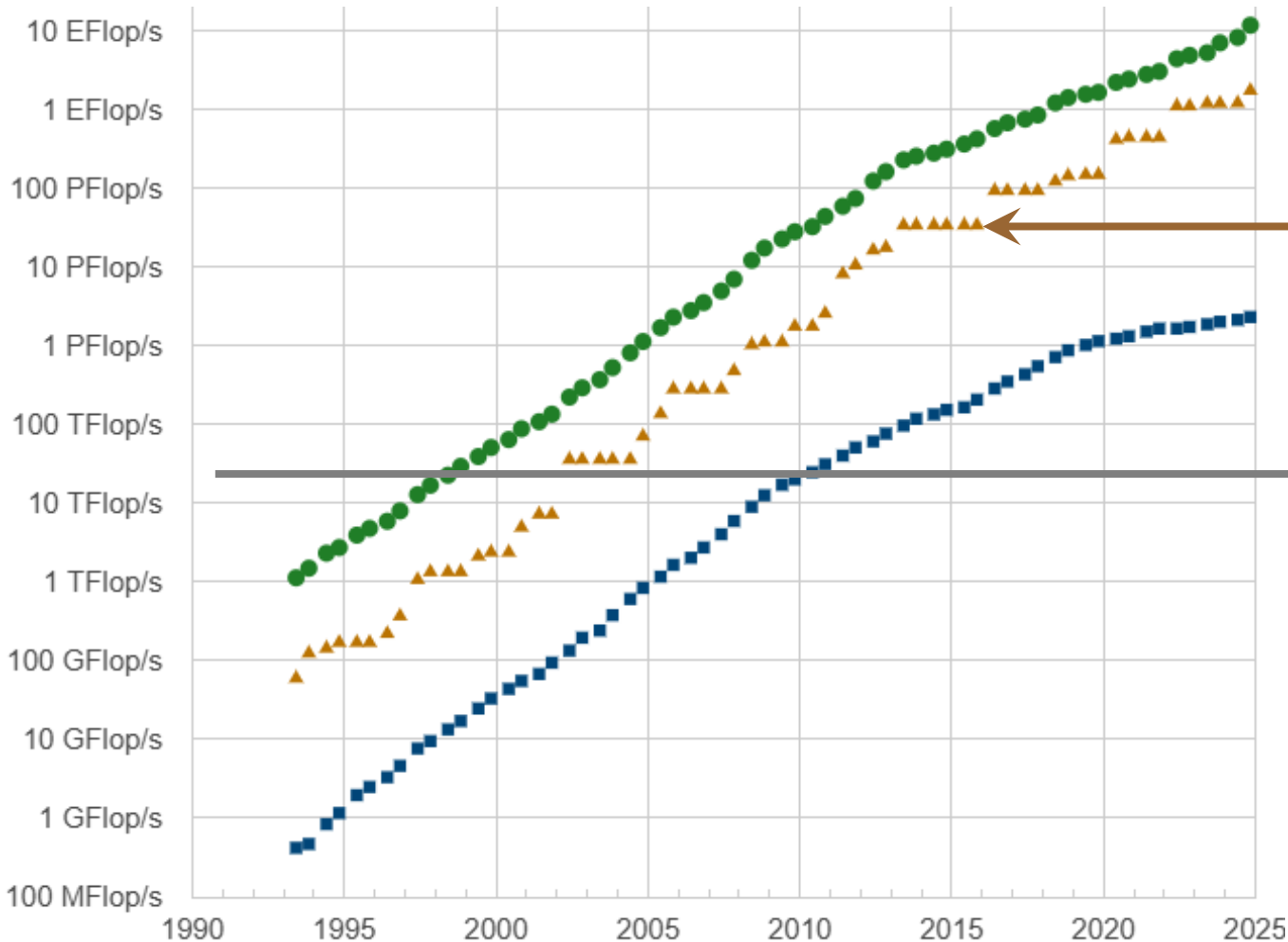


Image courtesy of shutterstock.com



- Symbolic knowledge representation
 - Semantic networks, ontologies, ...
- Automated deductive reasoning and logical inference
 - Declarative languages, rule-based systems, theorem provers, agent-based models, planning, ...
- Simplified mathematical modelling of human brain
- Artificial Neural Networks
 - **Perceptron** (1958)
 - **Deep learning**: FNN, MLP, RBF, RNN, CNN, LSTM, GAN, DNN, GNN, Transformer, ...

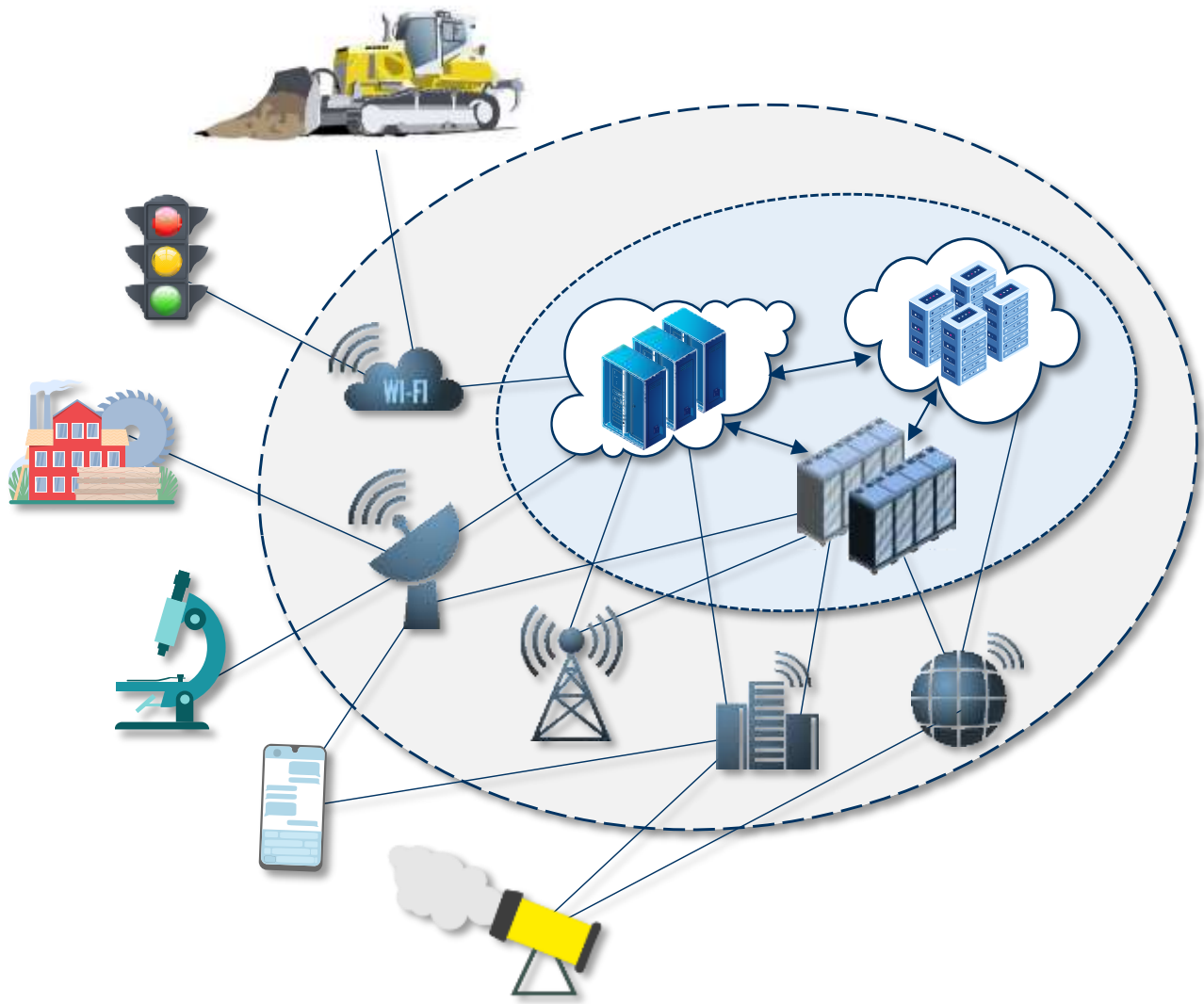
Why Edge AI ?



Fastest
Supercomputer



Edge AI Applications



Graph-Massivizer



- Towards Extreme and Sustainable Graph Processing for Urgent Societal Challenges in Europe
 - € 4,998,062.50
 - 2023 – 2025
- HORIZON-CL4-2022-DATA-01-05
 - Extreme data mining, aggregation and analytics technologies and solutions (RIA)
- 12 partners
 - 4 universities, 2 research centers, 1 HPC center, 5 companies



Real World Graph Structures

- Molecule graph
 - Thousand of nodes and edges
- Transportation networks
 - Tens of thousands of nodes and edges
- Social gaming network
 - Billion vertices
 - 100 billion connections
- Word Wide Web
 - 50 billion pages
 - 1 trillion hyperlinks
- Human brain
 - 100 billion neurons
 - 100 trillion connections



Image courtesy of shutterstock.com



Image courtesy of shutterstock.com

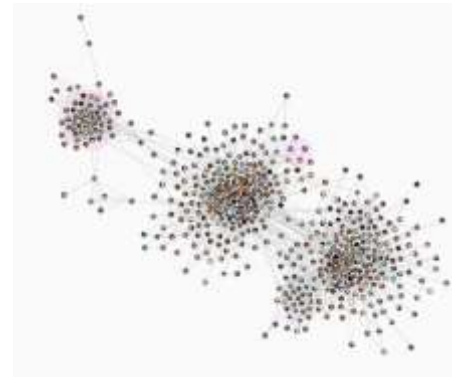
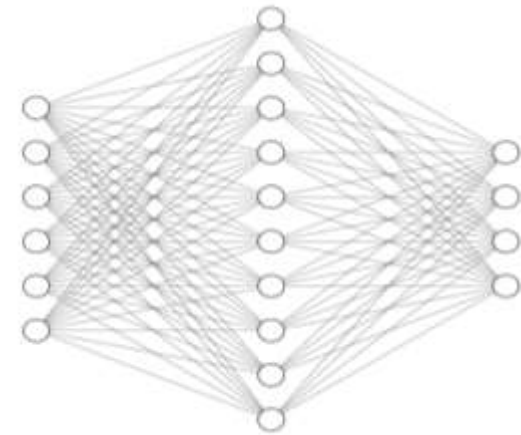
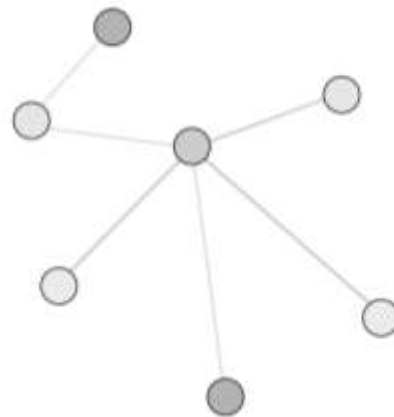
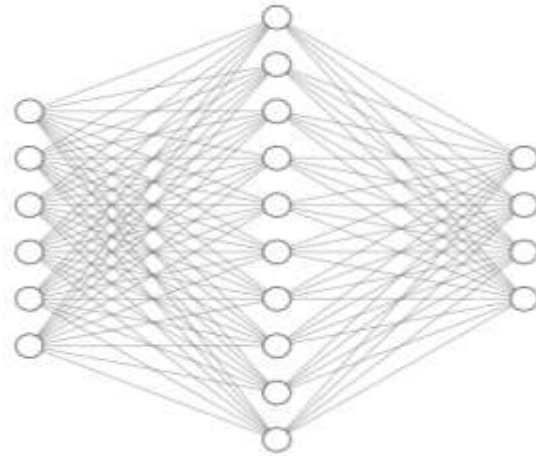


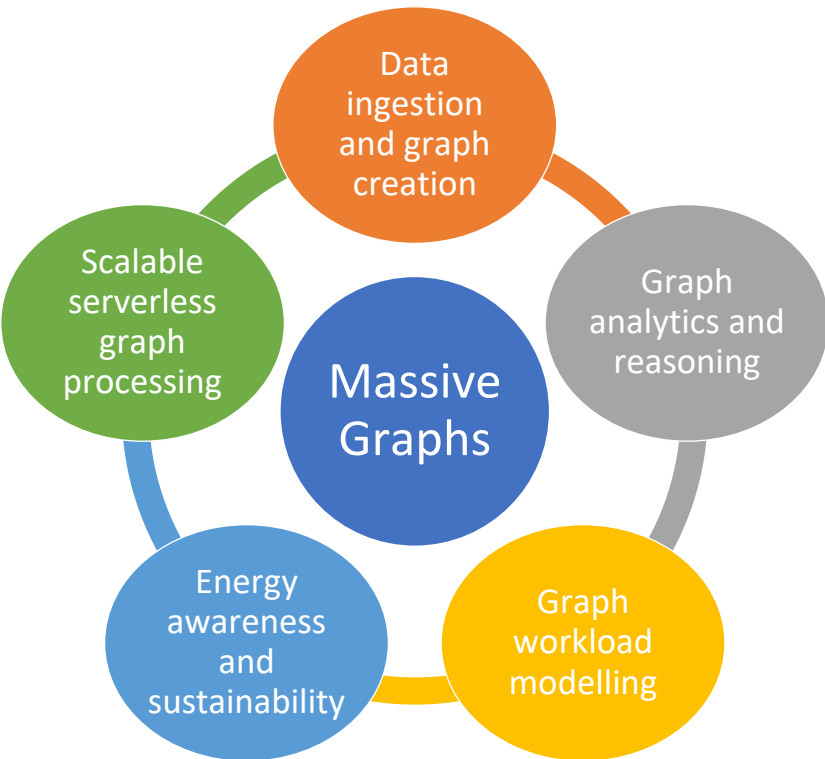
Image courtesy of shutterstock.com

Non-Euclidian AI

1
8
3
5
2
1

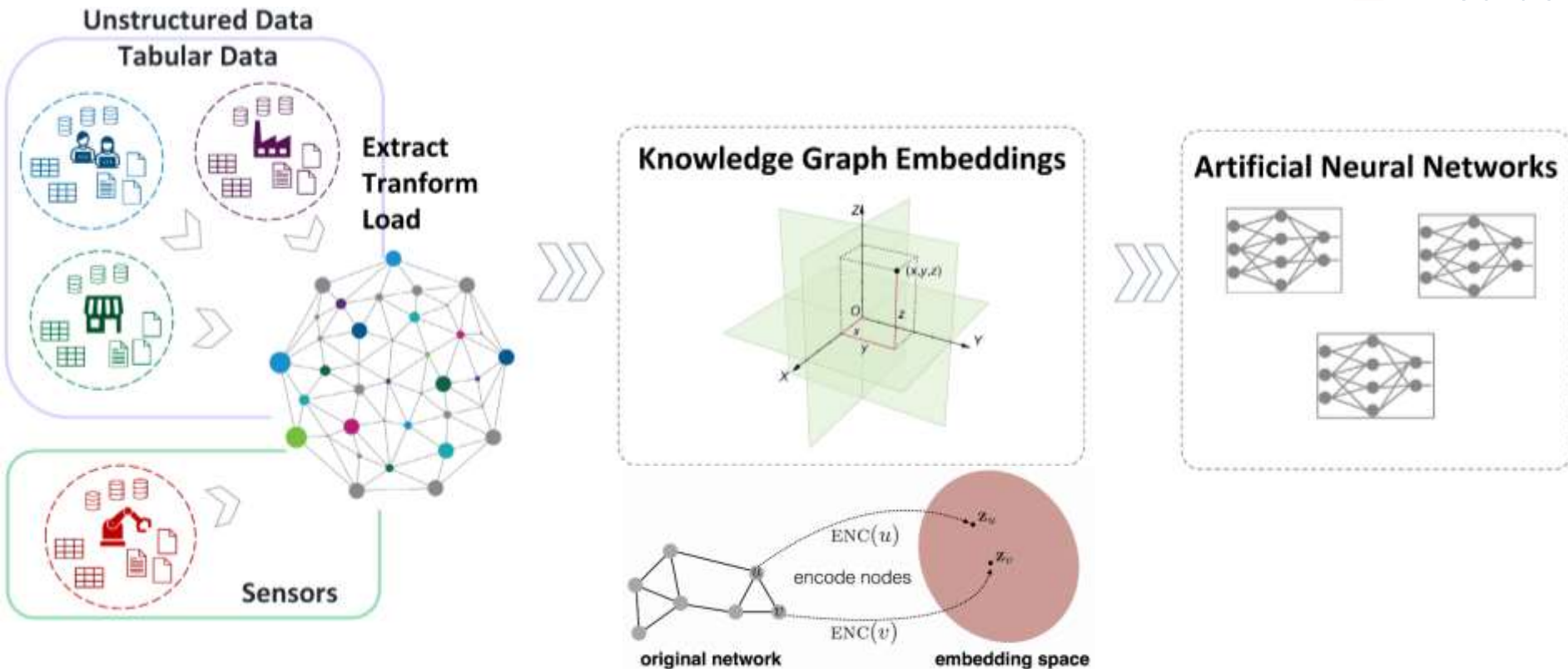


Holistic Graph Processing



	Volume	Velocity	Value	Veracity	Variety	Viscosity	Viridescence
Finance	✓		✓	✓	✓		✓
Event forecast	✓				✓	✓	✓
Manufacturing		✓		✓	✓		✓
Supercomputing	✓	✓			✓		✓

Neuro-Symbolic Graph Operations



- Basic graph operations
 - Creation, load, query, traversal, node/graph properties, sampling, summarization, link prediction, ...

Graph Neural Networks

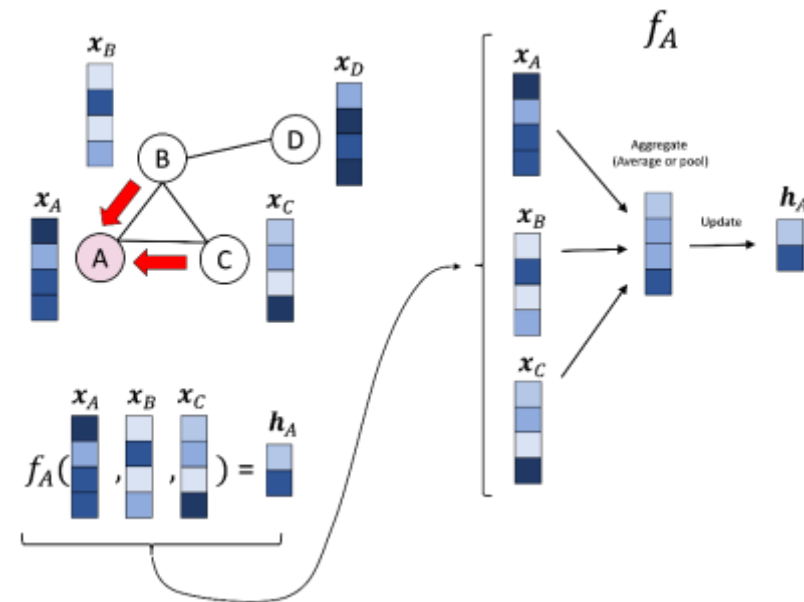
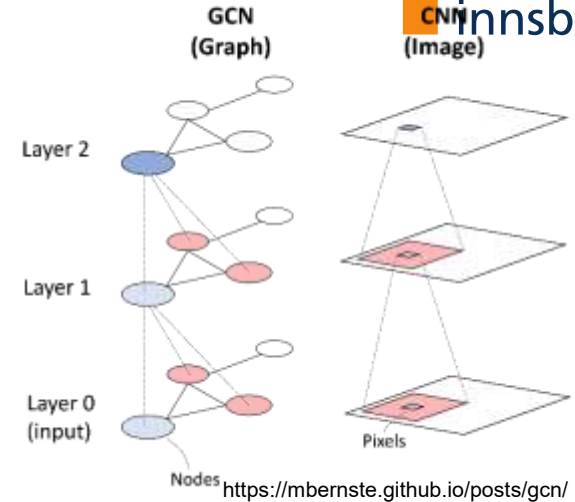
- Input
 - Node feature vectors: $X[N, D]$
 - **Adjacency matrix**: $A[N, N]$

- Graph convolutional layers

- **Aggregation** of neighboring node feature vectors:

$$\overline{A[i]} \cdot X$$

- **Update** of node feature vectors:
- $$f_w(A, X) = \sigma(\overline{A[i]} \cdot X \cdot W)$$



GNN Message Passing Architecture

- n stacked convolution layers

- $\mathbf{Y}_1 = f_{W_1}(X, \bar{A})$

- $\mathbf{Y}_2 = f_{W_2}(\mathbf{Y}_1, \bar{A})$

- ...

- $\mathbf{Y}_{n-1} = f_{W_{n-1}}(\mathbf{Y}_{n-2}, \bar{A})$

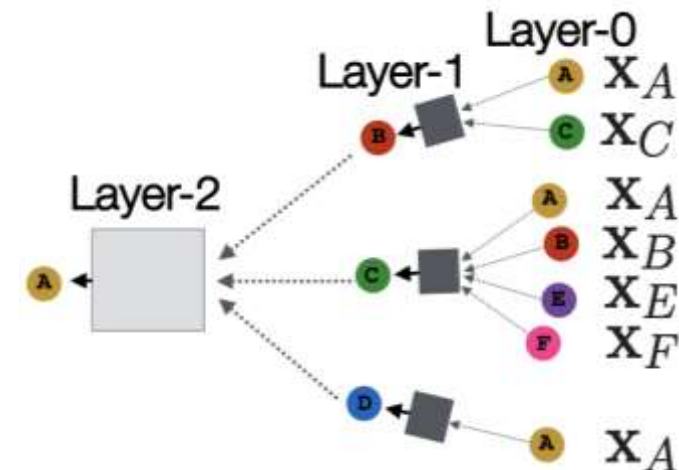
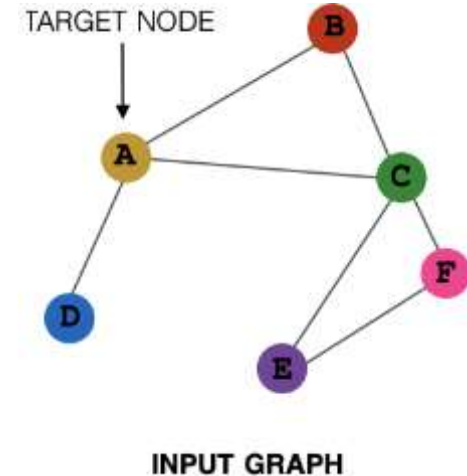
- $\mathbf{Y}_n = f_{W_n}(\mathbf{Y}_{n-1}, \bar{A})$

- $J = -(\hat{Y} \cdot \log(\mathbf{Y}_n) + (1 - \hat{Y}) \cdot \log(1 - \mathbf{Y}_n))$

- Backward propagation: $i = n \rightarrow 1$

- $\nabla \vec{J} = \left(\frac{\delta J}{\delta W_n} \dots \frac{\delta J}{\delta W_1} \right), \quad \frac{\partial J}{\partial W_i} = \frac{\delta J}{\delta Y_n} \cdot \dots \cdot \frac{\delta Y_i}{\delta W_i}$

- $\vec{W} = \vec{W} - \alpha \cdot \nabla \vec{J},$



Anomaly Prediction in HPC

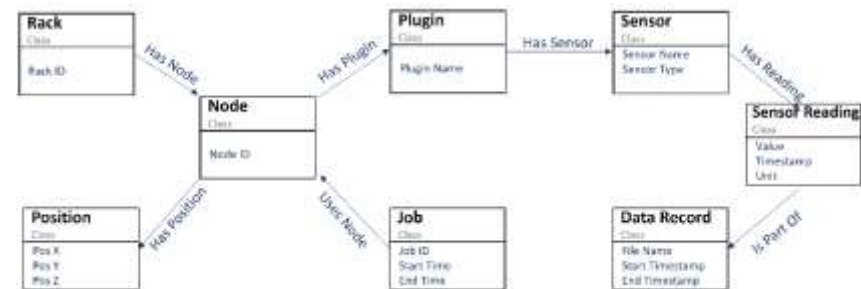
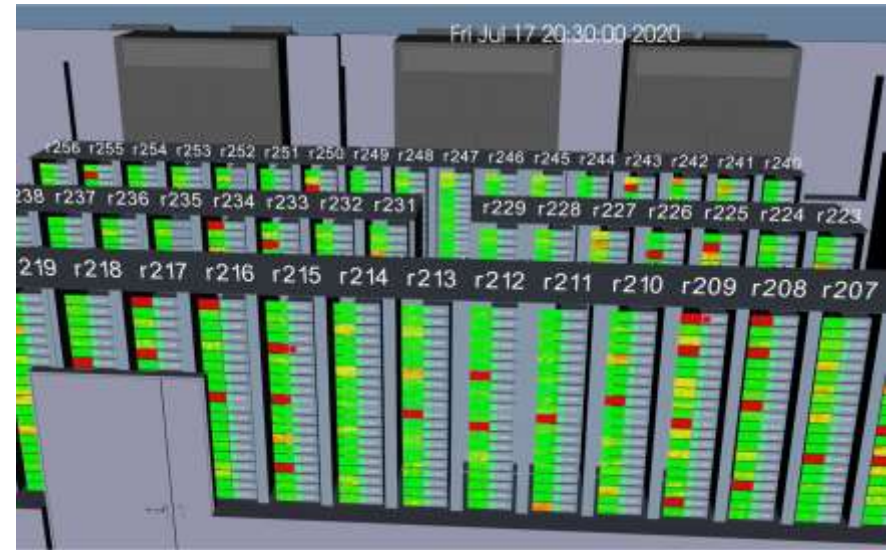
- Significant deviation from normal operation
 - Hardware: component failure
 - Software: drained node
- Reduce IT downtimes and maintenance
 - Avoid unplanned downtime
 - \$5,600 per minute (Gartner, 2022)
 - Improved planning, utilization, and operation
- State-of-the-art in anomaly detection
 - Individual node-level telemetry metrics
 - No spatial interconnection (network topology)

Marconi 100 Supercomputer

- Architecture
 - 49 racks, 980 computing nodes
 - 2x16 cores IBM POWER9 AC922
 - 4 Nvidia Volta V100

- ExaMon monitoring system
 - Slurm, Nagios, Ganglia, IPMI, ...
 - 573 metrics (Avg, Max, Min, SD)
 - 31-month operation (15-minute interval)
 - KairoDB, Cassandra NoSQL database
 - ExaData: 50 TB open data (Parquet, Zenodo)

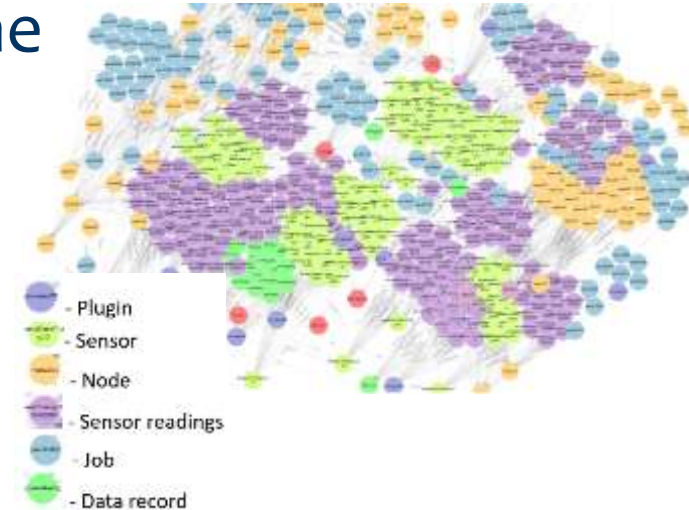
- Operational data analytics (ODA) ontology
 - Rack, node, device, position, job, sensor, metrics, ...



GNN-based Anomaly Prediction

- ETL monitoring data ingestion pipeline

- Mapping time series to knowledge graph
- Network architecture topology
- GraphDB RDF database
- 11.6 billion nodes, 68.6 billion edges

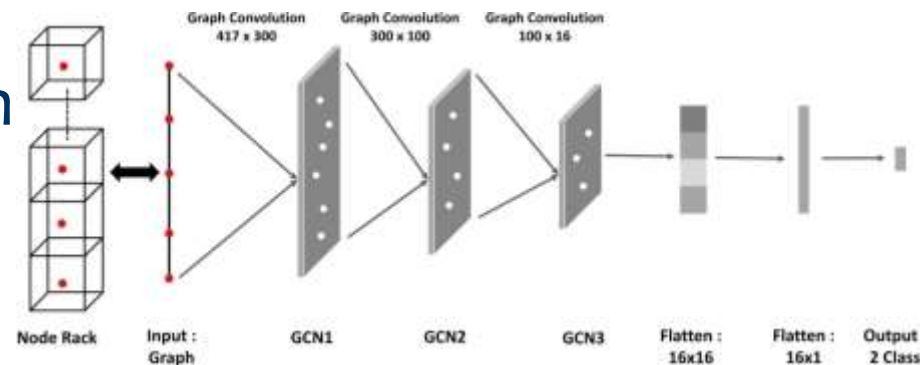


- GNN with 3 convolution layers

- Room, rack, node
- 573 semantic node features (**no feature engineering**)

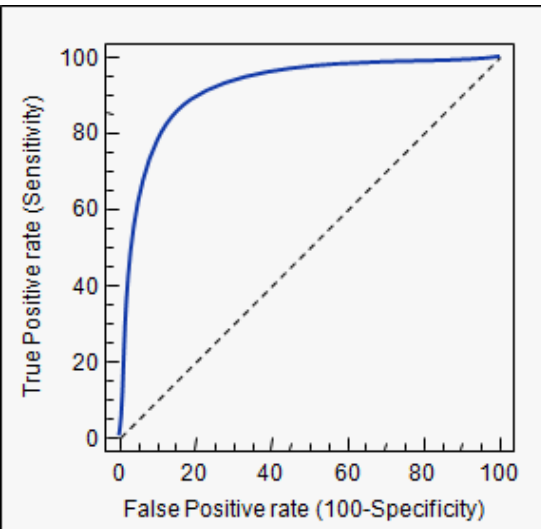
- GNN-based node classification

- Operational or anomalous
- 25 months training
- 6 months testing

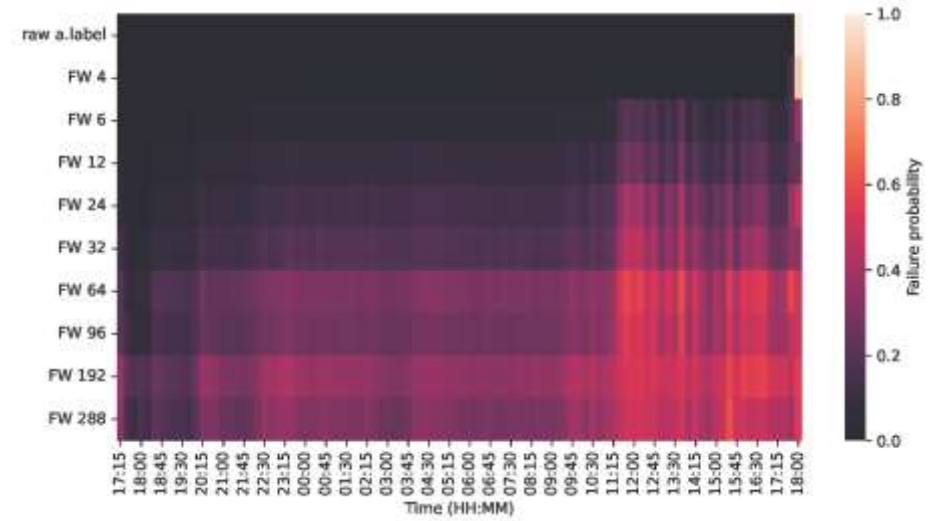


Anomaly Prediction Results

Area Under the Curve



Anomaly: Node 240, 7.05.24, 6:45



FW	rack GNN	room GNN	DNN	GB	RF	DT
4	0,91	0,59	0,64	0,63	0,61	0,51
6	0,89	0,58	0,66	0,64	0,59	0,5
12	0,84	0,47	0,65	0,63	0,59	0,5
24	0,78	0,58	0,62	0,6	0,55	0,5
32	0,75	0,55	0,59	0,58	0,55	0,5
64	0,66	0,42	0,5	0,48	0,49	0,49
96	0,62	0,58	0,55	0,51	0,58	0,51
192	0,55	0,52	0,47	0,48	0,52	0,5
288	0,53	0,50	0,52	0,51	0,52	0,49

DataPACT

- Compliance-by-design of data AI operations and pipelines
 - 2025 – 2027
 - €11 million
- HORIZON-CL4-2024-DATA-01-01
 - AI-driven data operations and compliance technologies (AI, data and robotics partnership) (IA)
- 18 partners
 - 7 universities, 2 research centers, 2 public sector, 7 companies



Compliance in EU

- > 100,000 legislations in 55 years
 - Which laws apply to me?
 - My mandatory reporting and notification obligations are labor-intensive.
 - My data sharing contract might be unfair.
 - When is data sharing compliant?
 - ...

- Compliance as innovation enabler
 - **Market trust** for data sharing, and adoption of digital innovation
 - **Unlock data** and promote altruistic flow, away from large tech companies
 - **Innovation playground** with interoperable regulatory sandboxes

Table 1: Overview of EU Legislations in the Digital Sector

Area of Interest	Consumer Policy	Privacy	Anti-Fraud	IP	Electronic	Law Enforcement	Trust & Safety	Government & Economic Activities	Competition	Media	Energy
Consumer Policy	Consumer Rights Directive (2011/83/EU)	Consumer Protection Cooperation Regulation (2015/2302/EU)	Consumer Protection Against Unfair Trading Practices Directive (2005/29/EC)	Consumer Protection Against Unfair Contract Terms Directive (1993/13/EEC)	Consumer Protection Against Unfair Commercial Practices Directive (2005/29/EC)	Consumer Protection Against Unfair Contract Terms Directive (1993/13/EEC)	Consumer Protection Against Unfair Contract Terms Directive (1993/13/EEC)	Consumer Protection Against Unfair Contract Terms Directive (1993/13/EEC)	Consumer Protection Against Unfair Contract Terms Directive (1993/13/EEC)	Consumer Protection Against Unfair Contract Terms Directive (1993/13/EEC)	Consumer Protection Against Unfair Contract Terms Directive (1993/13/EEC)
Privacy	GDPR (2016/679)	GDPR (2016/679)	GDPR (2016/679)	GDPR (2016/679)	GDPR (2016/679)	GDPR (2016/679)	GDPR (2016/679)	GDPR (2016/679)	GDPR (2016/679)	GDPR (2016/679)	GDPR (2016/679)
Anti-Fraud	Anti-Fraud Directive (2003/78/EC)	Anti-Fraud Directive (2003/78/EC)	Anti-Fraud Directive (2003/78/EC)	Anti-Fraud Directive (2003/78/EC)	Anti-Fraud Directive (2003/78/EC)	Anti-Fraud Directive (2003/78/EC)	Anti-Fraud Directive (2003/78/EC)	Anti-Fraud Directive (2003/78/EC)	Anti-Fraud Directive (2003/78/EC)	Anti-Fraud Directive (2003/78/EC)	Anti-Fraud Directive (2003/78/EC)
IP	IP Directive (2004/48/EC)	IP Directive (2004/48/EC)	IP Directive (2004/48/EC)	IP Directive (2004/48/EC)	IP Directive (2004/48/EC)	IP Directive (2004/48/EC)	IP Directive (2004/48/EC)	IP Directive (2004/48/EC)	IP Directive (2004/48/EC)	IP Directive (2004/48/EC)	IP Directive (2004/48/EC)
Electronic	Electronic Directive (2002/46/EC)	Electronic Directive (2002/46/EC)	Electronic Directive (2002/46/EC)	Electronic Directive (2002/46/EC)	Electronic Directive (2002/46/EC)	Electronic Directive (2002/46/EC)	Electronic Directive (2002/46/EC)	Electronic Directive (2002/46/EC)	Electronic Directive (2002/46/EC)	Electronic Directive (2002/46/EC)	Electronic Directive (2002/46/EC)
Law Enforcement	Law Enforcement Directive (2016/680)	Law Enforcement Directive (2016/680)	Law Enforcement Directive (2016/680)	Law Enforcement Directive (2016/680)	Law Enforcement Directive (2016/680)	Law Enforcement Directive (2016/680)	Law Enforcement Directive (2016/680)	Law Enforcement Directive (2016/680)	Law Enforcement Directive (2016/680)	Law Enforcement Directive (2016/680)	Law Enforcement Directive (2016/680)
Trust & Safety	Trust & Safety Directive (2015/2302)	Trust & Safety Directive (2015/2302)	Trust & Safety Directive (2015/2302)	Trust & Safety Directive (2015/2302)	Trust & Safety Directive (2015/2302)	Trust & Safety Directive (2015/2302)	Trust & Safety Directive (2015/2302)	Trust & Safety Directive (2015/2302)	Trust & Safety Directive (2015/2302)	Trust & Safety Directive (2015/2302)	Trust & Safety Directive (2015/2302)
Government & Economic Activities	Government & Economic Activities Directive (2016/1148)	Government & Economic Activities Directive (2016/1148)	Government & Economic Activities Directive (2016/1148)	Government & Economic Activities Directive (2016/1148)	Government & Economic Activities Directive (2016/1148)	Government & Economic Activities Directive (2016/1148)	Government & Economic Activities Directive (2016/1148)	Government & Economic Activities Directive (2016/1148)	Government & Economic Activities Directive (2016/1148)	Government & Economic Activities Directive (2016/1148)	Government & Economic Activities Directive (2016/1148)
Competition	Competition Directive (2010/13/EU)	Competition Directive (2010/13/EU)	Competition Directive (2010/13/EU)	Competition Directive (2010/13/EU)	Competition Directive (2010/13/EU)	Competition Directive (2010/13/EU)	Competition Directive (2010/13/EU)	Competition Directive (2010/13/EU)	Competition Directive (2010/13/EU)	Competition Directive (2010/13/EU)	Competition Directive (2010/13/EU)
Media	Media Directive (2018/1808)	Media Directive (2018/1808)	Media Directive (2018/1808)	Media Directive (2018/1808)	Media Directive (2018/1808)	Media Directive (2018/1808)	Media Directive (2018/1808)	Media Directive (2018/1808)	Media Directive (2018/1808)	Media Directive (2018/1808)	Media Directive (2018/1808)
Energy	Energy Directive (2019/944)	Energy Directive (2019/944)	Energy Directive (2019/944)	Energy Directive (2019/944)	Energy Directive (2019/944)	Energy Directive (2019/944)	Energy Directive (2019/944)	Energy Directive (2019/944)	Energy Directive (2019/944)	Energy Directive (2019/944)	Energy Directive (2019/944)

Home Healthcare Pipeline

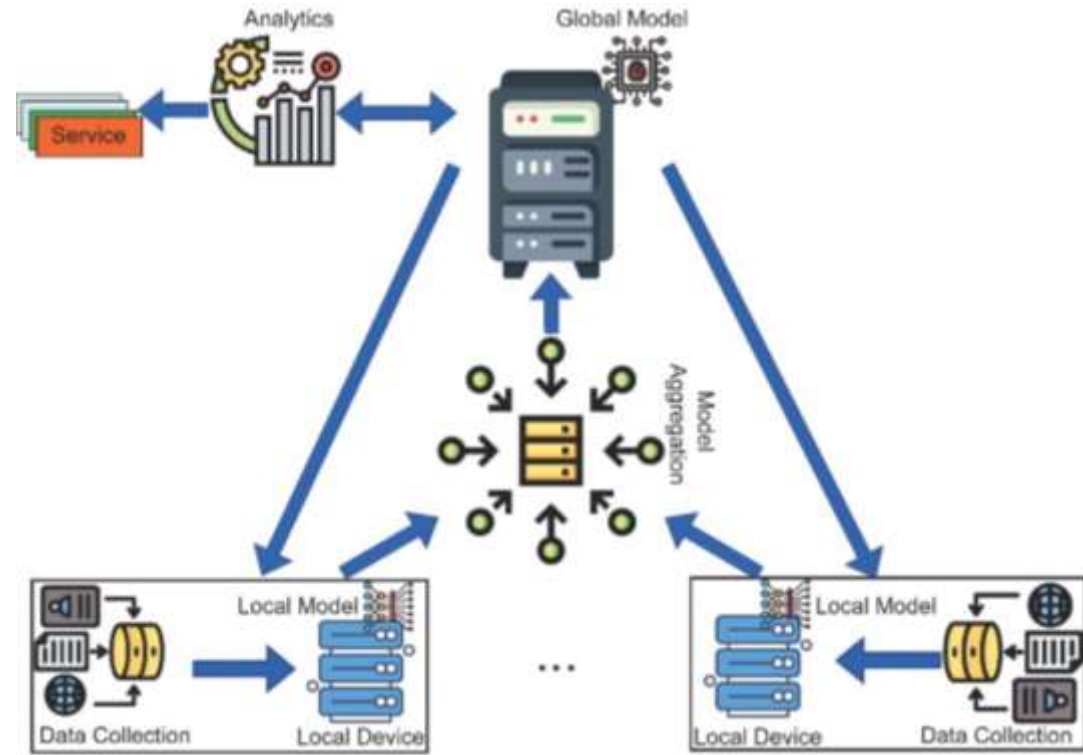
- Patient health monitoring
 - Edge IoT healthcare devices
- Health data collection, preprocessing and routing (MQTT)
 - Consent management (Edge-based, GDPR)
- Medical information system
 - Electronic health records (Cloud-based, Data Act)
- Application front-end
 - Health monitoring against treatment plan and notifications
 - AI Act, sectorial legislation
- European Health Dataspaces
 - Fast Healthcare Interoperability Resources standard
 - Data Governance Act



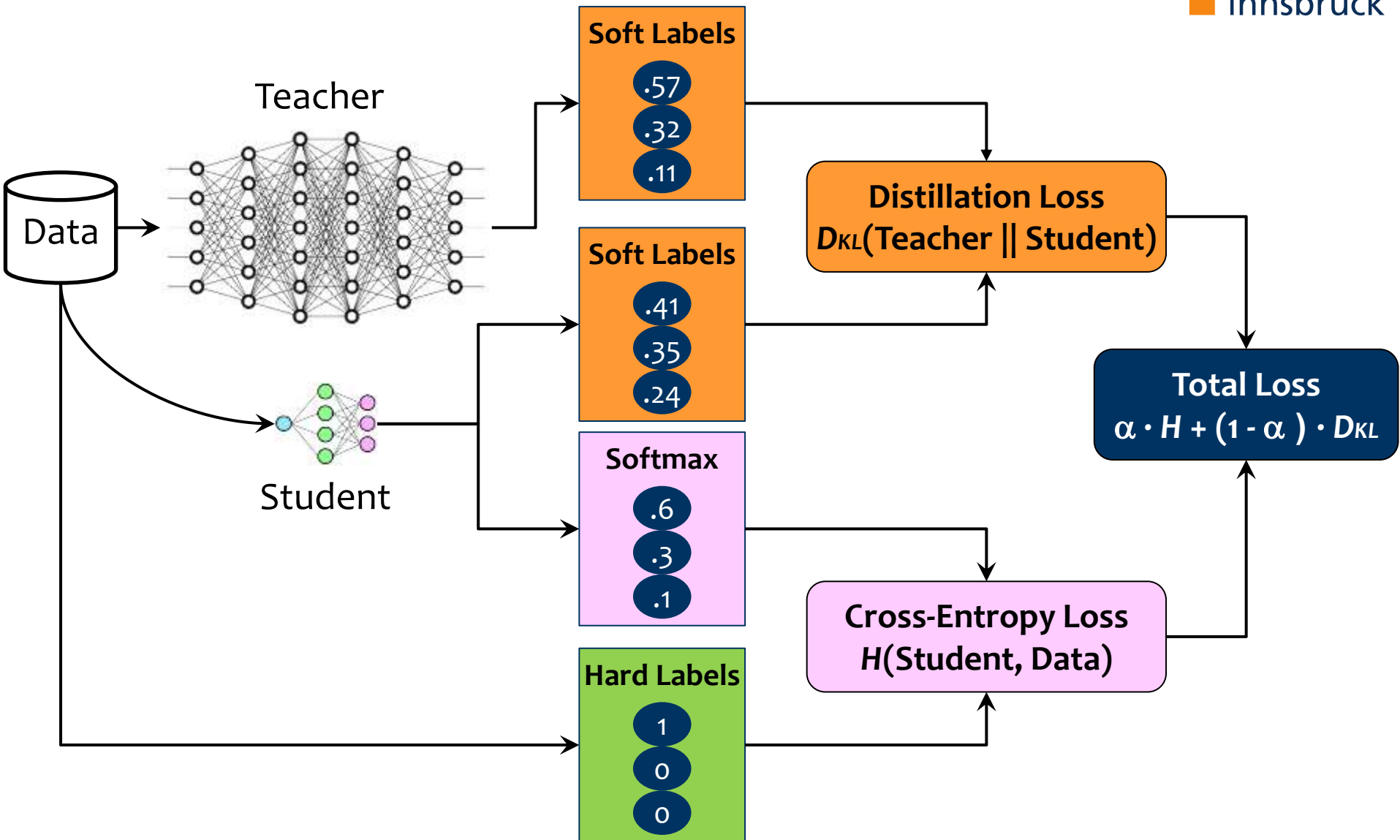
Federated Learning

- Share global knowledge, keep local data
 - Train AI models locally on private data
 - Share the model for global aggregation

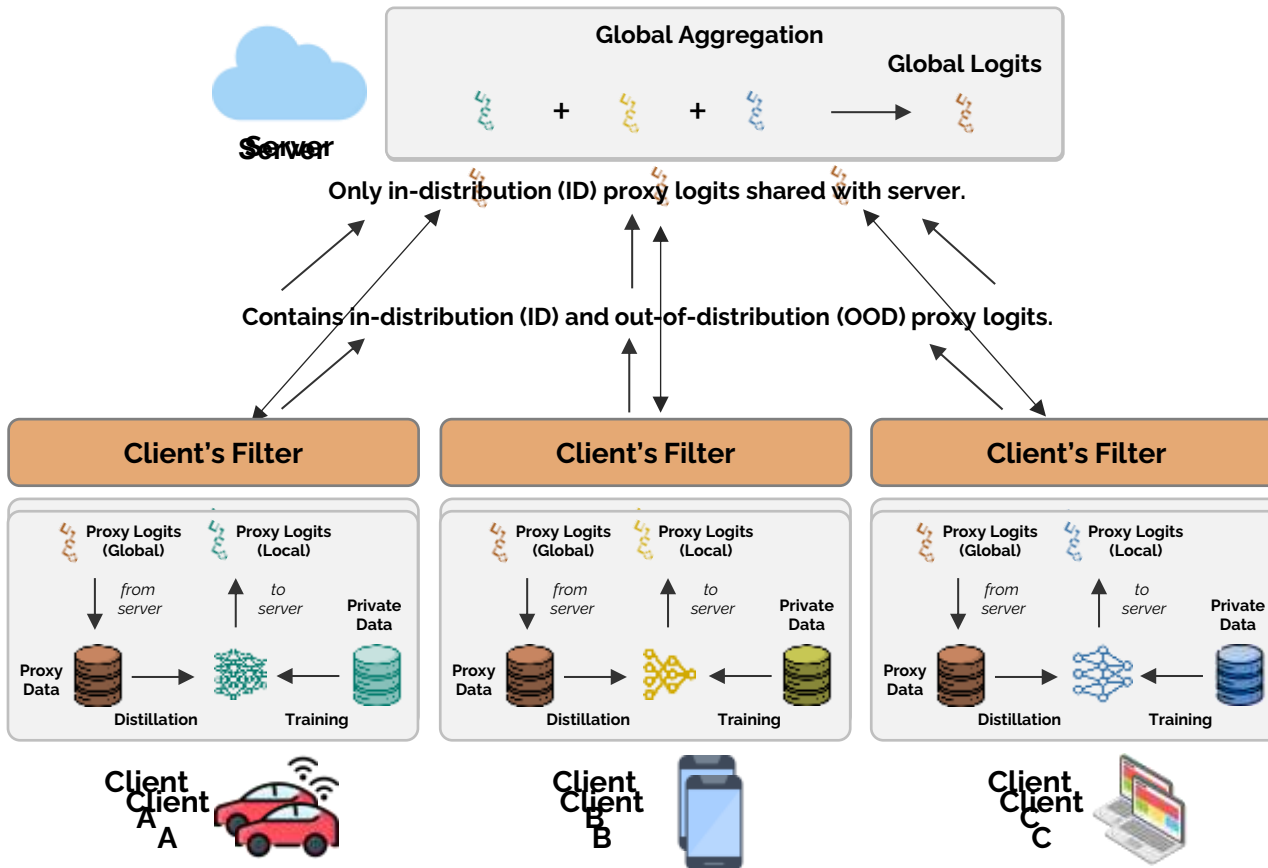
- Disadvantages
 - Homogeneous model
 - Large gradient communication
 - Model inversion
 - Membership inference
 - Non-IID data inaccuracies



Knowledge Distillation



Federated Distillation



Advantages

- Low logits communication overhead
- Heterogeneous models

Disadvantages

- Proxy dataset
- Low accuracy on non-IID (out-distribution) data

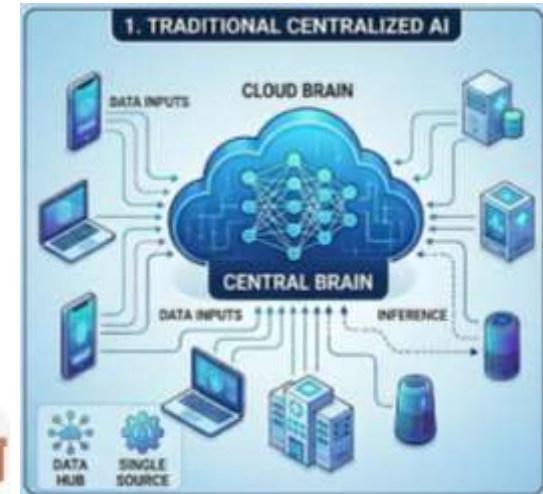
Client-side filtering

- K-Means algorithm
- In-distribution data

Agentic AI

- Foundational models dominate AI but do not cut the edge

- Outperform humans in computer vision, natural language processing, ...
- Oversized, costly, unsustainable, unfocused, unreliable, unexplainable (black-box), centralized, passive, ...



- Agentic AI

- Small fine-tuned, focused and active edge AI models
- Federated learning on private non-IID data
- Symbolic data representation
- Single step inductive inference
- Hybrid, neuro-symbolic multi-step deductive reasoning



CoAgent

- Cognitive Computing Continuum for Large-Scale Distributed GenAI Agents
 - Cognitive Computing Continuum for Large-Scale Distributed GenAI Agents
 - 2026 – 2029
 - €8 million

- HORIZON-CL4-2025-04
 - Empowering AI/generative AI along the Cognitive Computing continuum (RIA)

- Partners
 - 17 partners, 9 countries
 - 5 universities, 1 research center, 1 public body, 10 companies



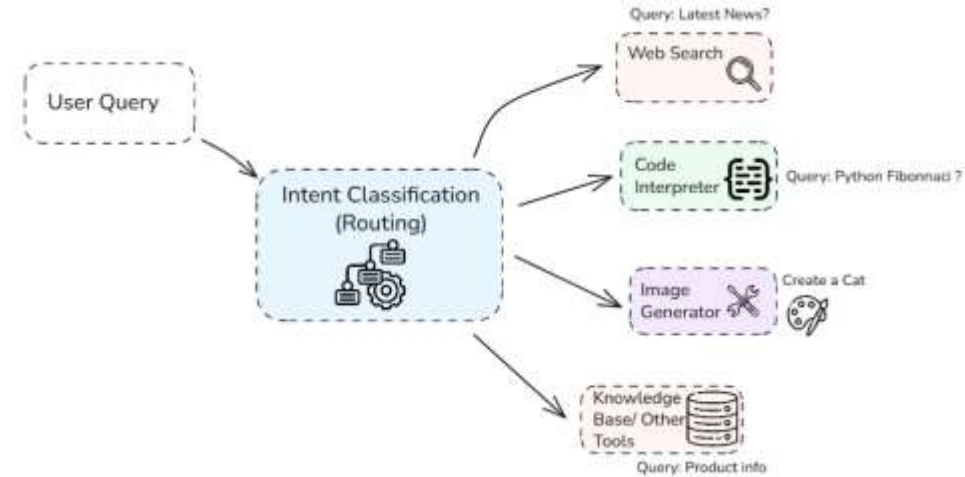
AI Agent

- **Intelligence**
 - **Intent-based SLMs:** small specialized, active
 - **Edge hosting:** quantization, pruning, knowledge distillation, factorization
 - **Edge fine-tuning:** federated, parameter-efficient, compliant
- **Memory**
 - **Short-term:** ephemeral, capacity constrained (bottleneck)
 - **Episodic:** edge bottleneck
 - **Long-term:** symbolic, semantic data representation, RAG
- **Persona**
 - **Personality:** *You are 'Aria,' a knowledgeable AI developer*
 - **Role:** *Help user debug Python code*
 - **Tone:** *Friendly, concise, and slightly sarcastic*
 - **Verbosity:** *Explain bugs in 2 sentences*
 - **Constraints:** *Write only the corrected function, not all code*
 - **Focus:** *Refuse questions other than Python*



Agentic Intent-based Planning

- Intent classification
- Intent recognition
 - Get to the airport by 10 AM
- Intent decomposition
 - Alarm clock agent: wakeup call
 - Router agent: Google Maps API
 - Traffic agent: real-time sensor data
 - Train scheduling agent: CFR Calatori API
 - Taxi agent: Uber/Bolt API
 - Payment agent: Revolut API
- Intent-based agent orchestration
 - Symbolic rule-based guardrails for compliance
 - Small context windows, fault tolerance, parallelism, ...





Use Cases

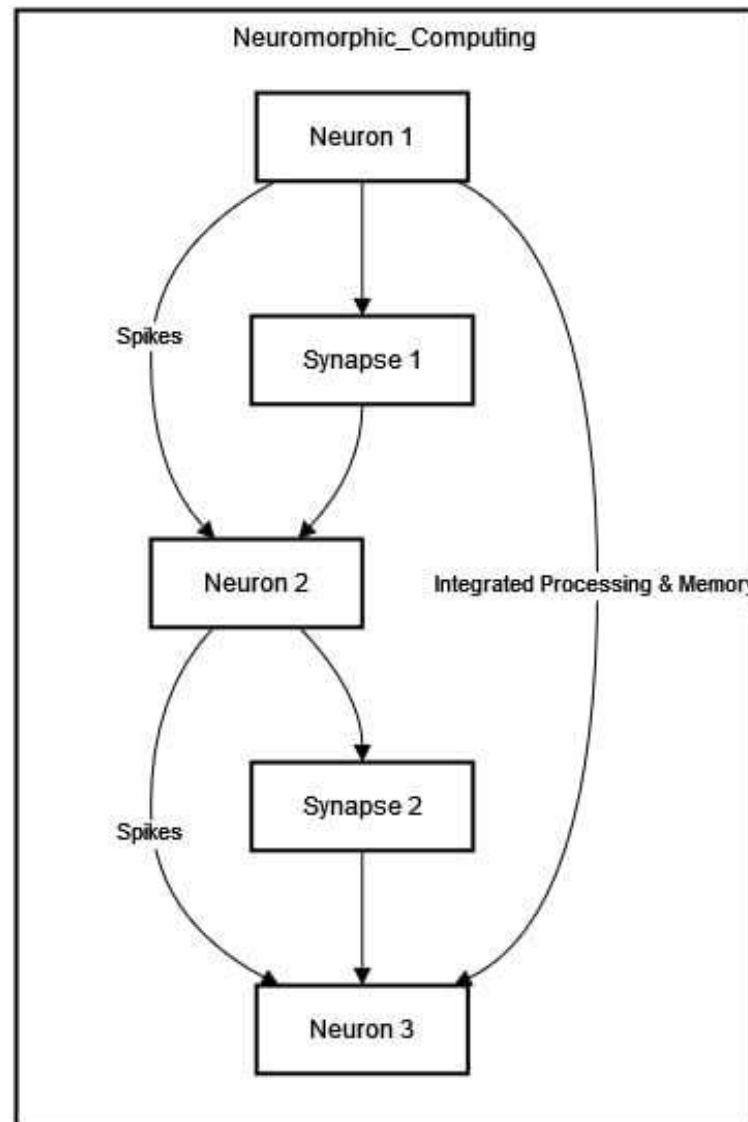
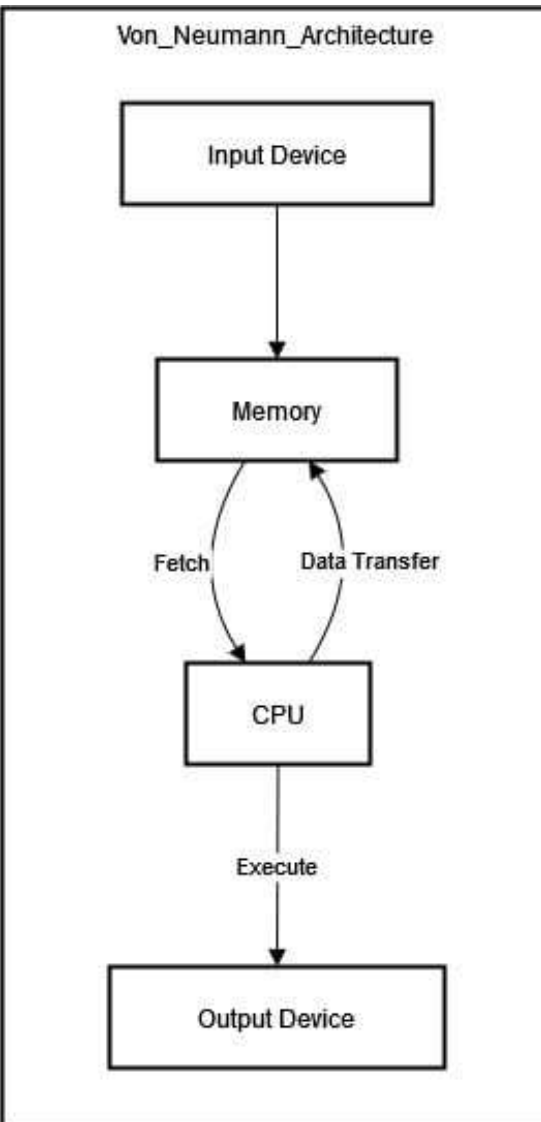
- Smart mobility, road management and traffic control (Indra Sistemas)
 - IF main road traffic IS high AND side road traffic IS low THEN maximize green light
 - IF emergency vehicle IS on road THEN turn on green light
 - IF pedestrian button IS pushed THEN turn on red light

- Drone mission control in precision agriculture (ETRA)
 - IF area IS under-covered THEN scan area ELSE skip area
 - IF crop stress dense formation AND spray pesticides ELSE scanning formation
 - IF drones share the same air corridor THEN activate vertical separation

- Predictive medical device maintenance (Philipps)
 - IF chiller water temperature $> 22^{\circ}\text{C}$ AND vibration sensor on magnet pump $> 0.5 \text{ mm/s}$ THEN generate high-priority maintenance alert for imminent cooling failure
 - IF device IS in docking station AND charge cycle > 500 AND max capacity $< 80\%$ THEN replace battery

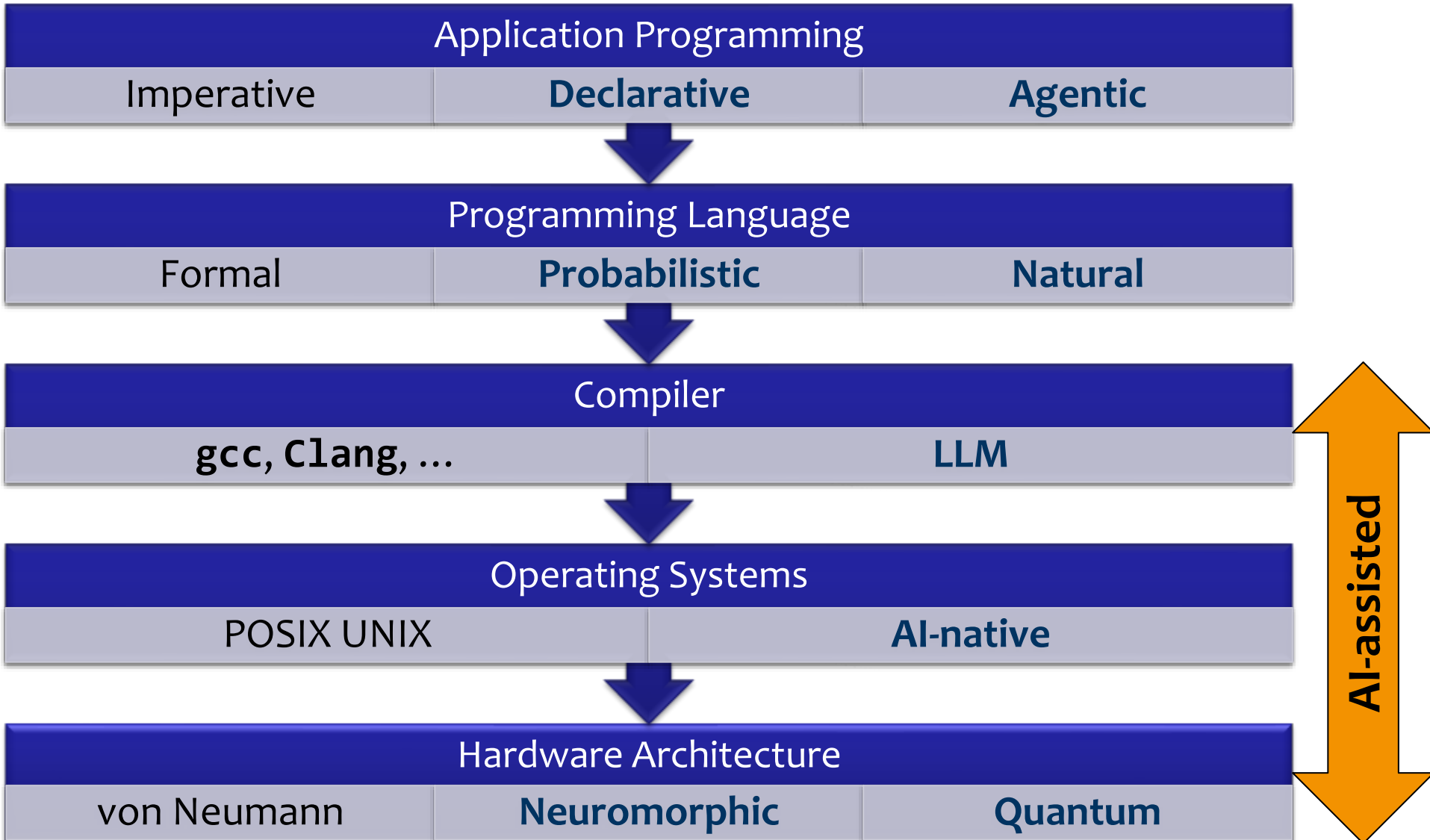
- Autonomous telecommunication network operation (Telenor)
 - IF $> X$ packets per second from an unknown IP THEN adds IP to a local blacklist
 - IF traffic is normal THEN forward packets to the next node
 - IF primary link latency $> 50\text{ms}$ THEN switch to secondary fiber link

Neuromorphic Computing



- Event-based processing
- Spiking neural networks
- Energy efficiency
- Parallelism and adaptability

Research Outlook





Thank you!

<http://edgeai-informatik.uibk.ac.at/>

radu.prodan@uibk.ac.at