CIoT24
7th Conference on Cloud and Internet of Things
October 29-31, 2024
Montreal, Canada

Advancing IoT and Cloud Security through LLMs, Federated Learning, and Reinforcement Learning

Speaker: Dr Mohamed Chahine GHANEM
Associate Professor
Director- Cyber Security Research Centre
London Metropolitan University (UK)

Agenda

1. Introduction to Cloud & IoT Security Challenges
2. Leveraging Large Language Models (LLMs)
3. Federated Learning (FL) for Decentralized
4. Reinforcement Learning and Deep RL
5. Integrating AI Methods for a better security
6. Future of Cloud and IoT Security
7. Resources

About me

- Engineer Degree in Computer Science, MSc in Digital Forensics and PhD in Cyber Security Engineering form City, University of London
- 15 years+ at Senior-Level in Industry (Law Enforcement and Corporates)
- Certified Expert (CISSP, CPCI, Multi-GIAC…)
- Currently Associate Professor, Director of Cyber Security Research Centre

IoT & Cloud: Rapidly Expanding Ecosystems

Exponential Growth: Over 75 billion IoT devices are expected by 2025, increasing the “attack surface” and thus cyber threat landscape.

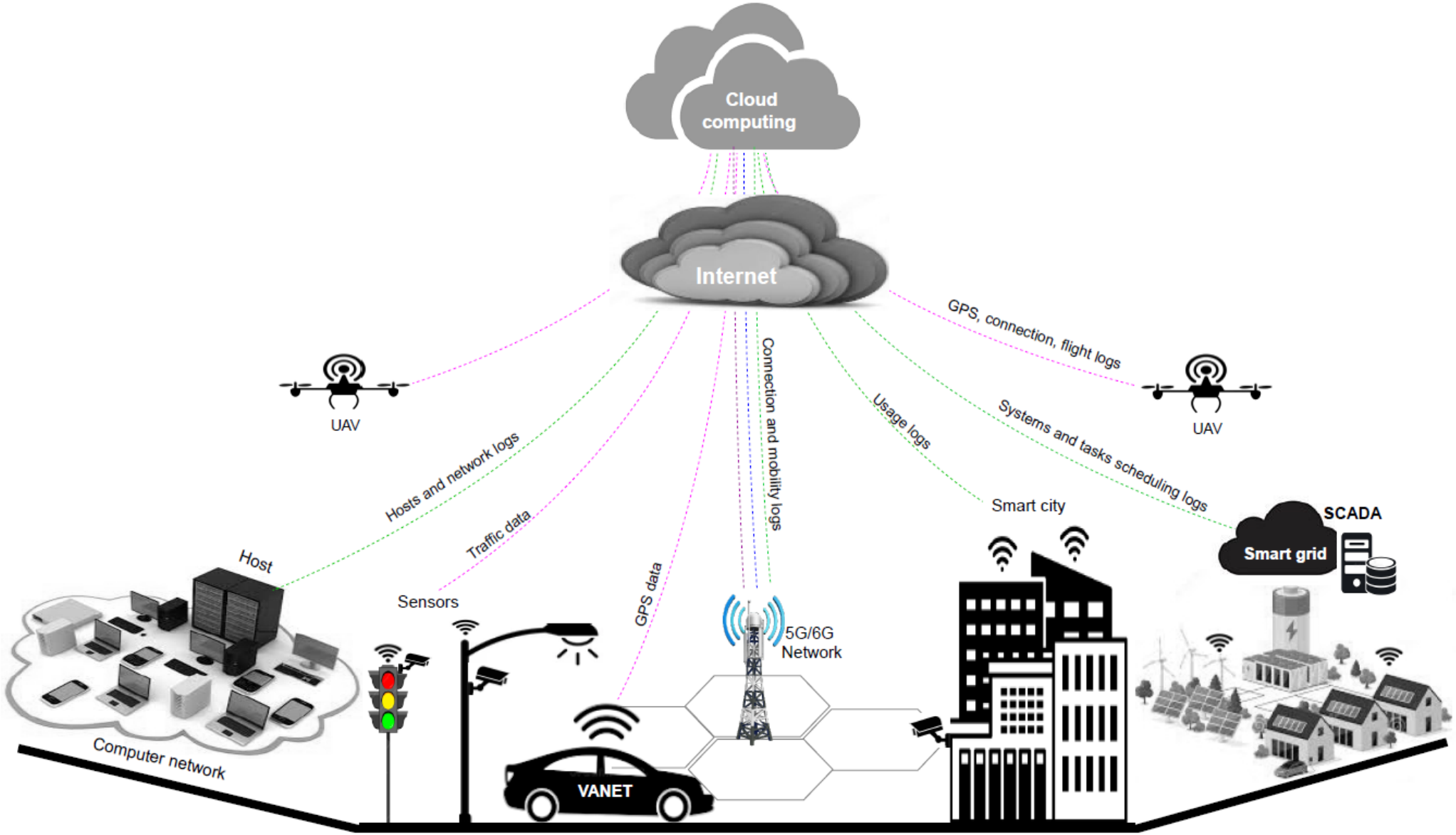
IoT Data Vulnerability: Widespread of cloud adoption elevates data exposure risks and complicates security management.

Targeted Attacks: IoT devices often have minimal security, making them prime targets for cyber-attacks.

Complex Security Challenges: The diverse nature of IoT devices complicates the implementation of effective security measures.

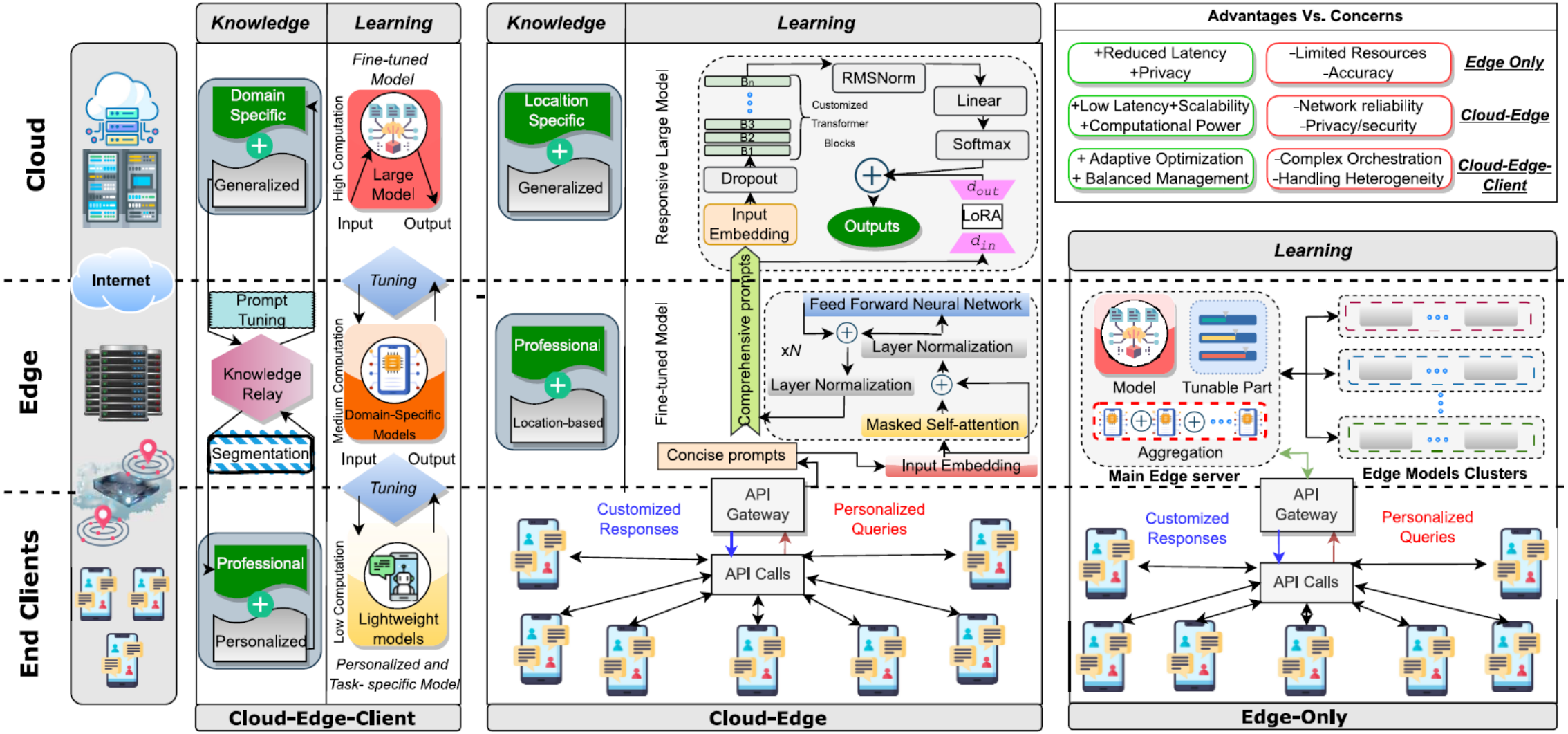
Critical Infrastructure Needs: There is an urgent requirement for robust security frameworks tailored specifically for IIoT environments.

IoT and Cloud Applications



(Kheddar et al., 2024)

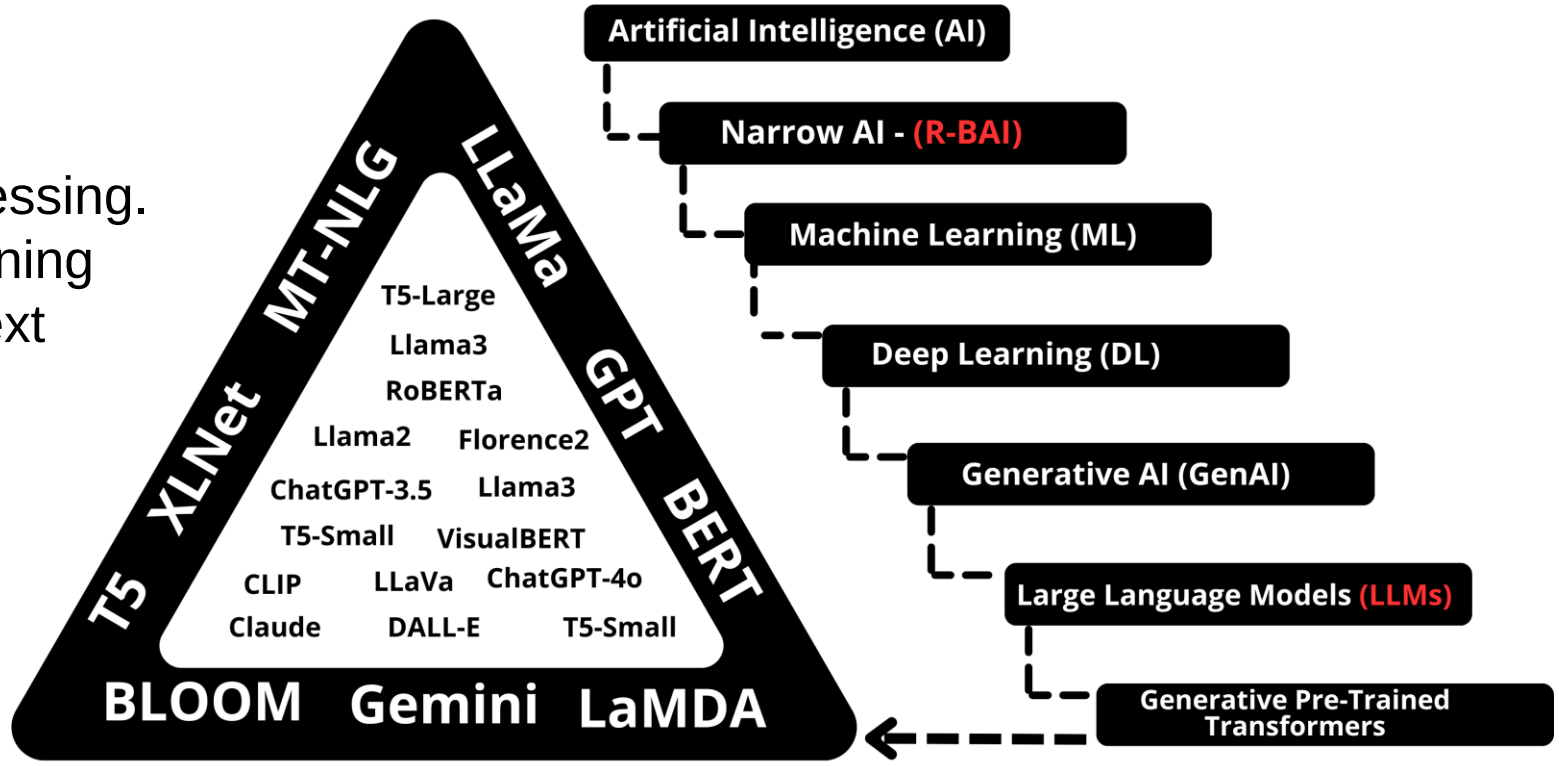
AI in IoT and Cloud



(Ferrag et al., 2022)

Large Language Models (LLMs)

- LLMs are a branch of **Generative AI**
- Focused specifically on language processing.
- LLMs are built on advanced Deep Learning architectures and trained on massive text datasets.



LLMs are meant for **specific tasks** such as text generation, translation, summarization, and question answering by identifying context and patterns within text.

LLMs in IoT and Cloud Security

Advanced Data Analysis: LLMs leverage sophisticated natural language processing to examine vast datasets quickly and accurately.

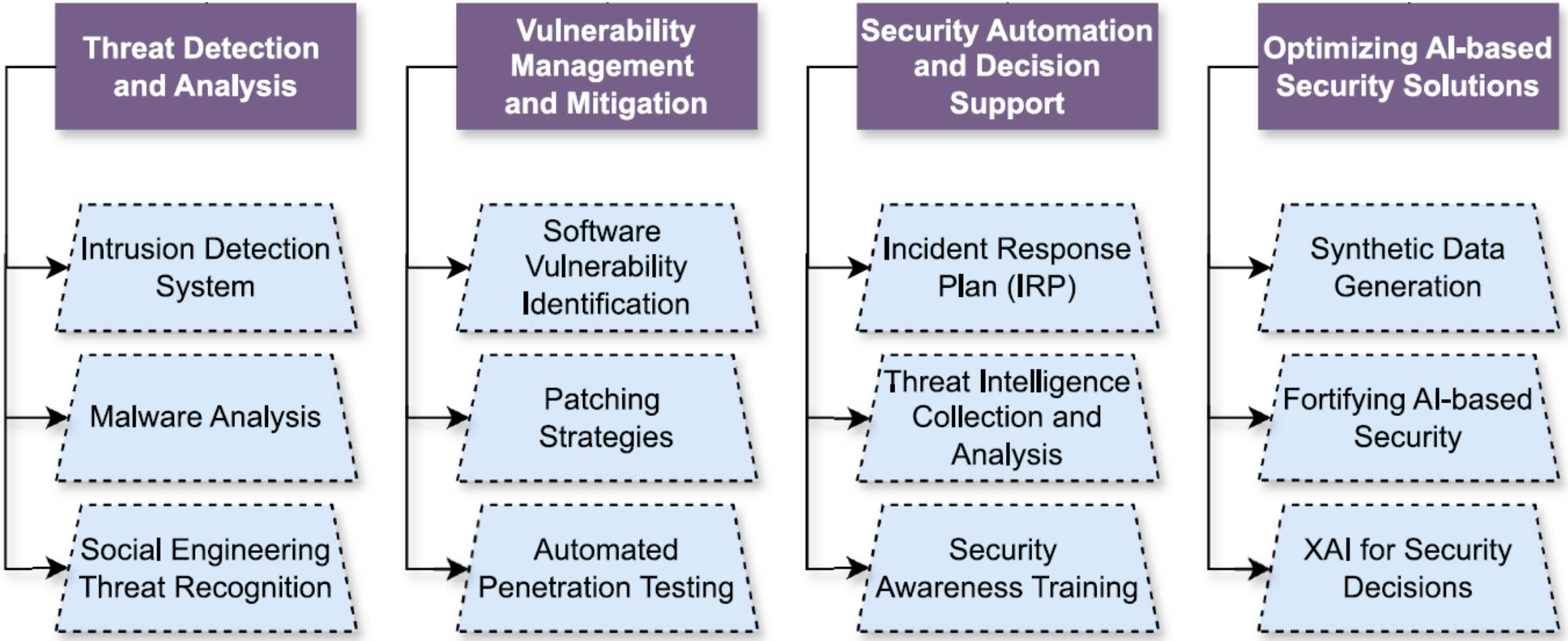
Anomaly Detection: LMs can identify unusual patterns or deviations in data, helping to detect irregularities that could signal threats.

Predictive Capabilities: By analyzing past data trends, LLMs can forecast potential risks, providing early warnings and enabling proactive responses.

Context Understanding: LLMs process information with contextual awareness, improving accuracy in recognizing potential risks across diverse domains.

Decision Support: LLMs insights assist analysts in prioritizing and addressing threats, reducing response times and enhancing security measures.

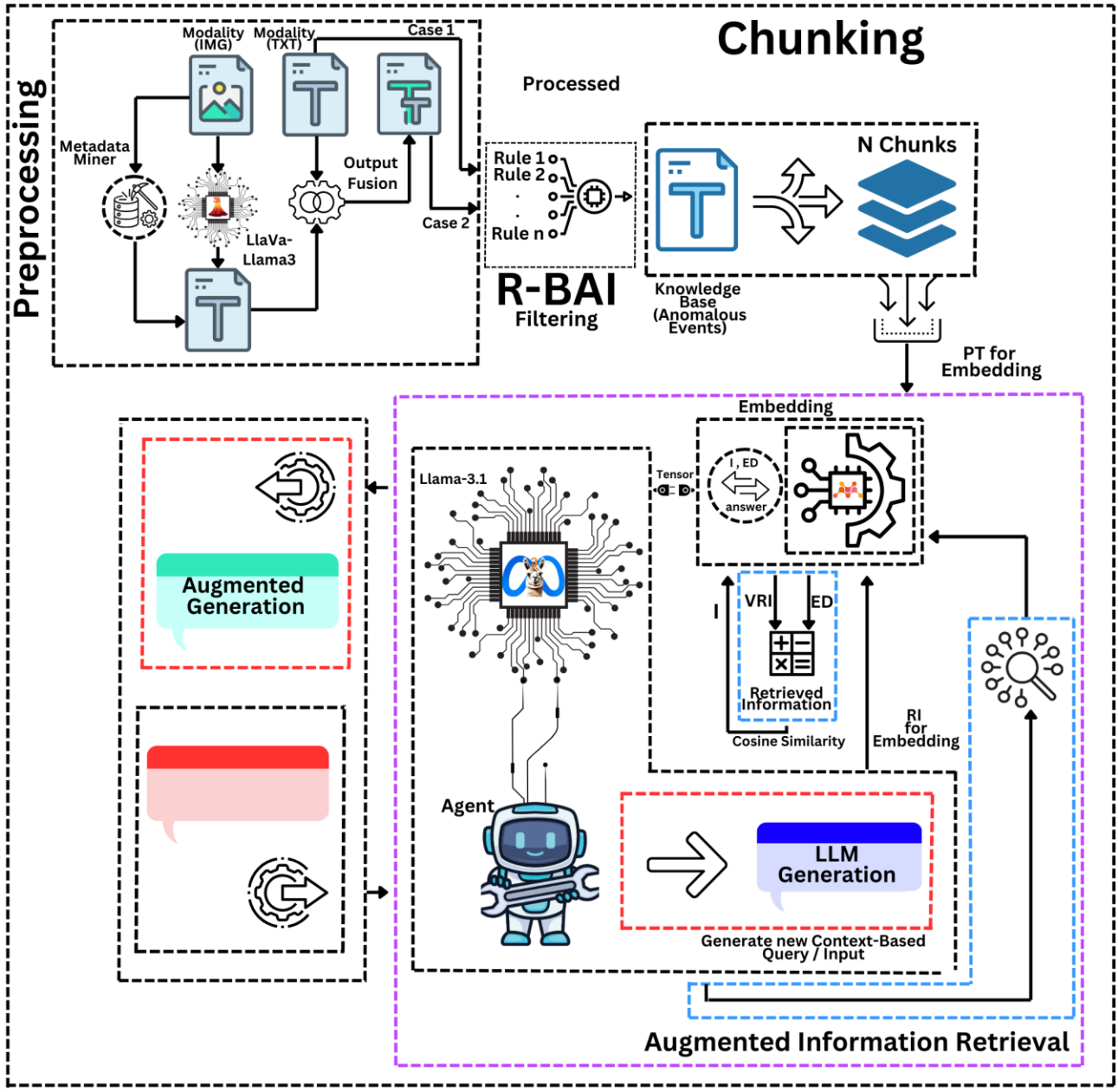
LLMs in IoT and Cloud Security (cont.)



(Ferrag et al., 2022)

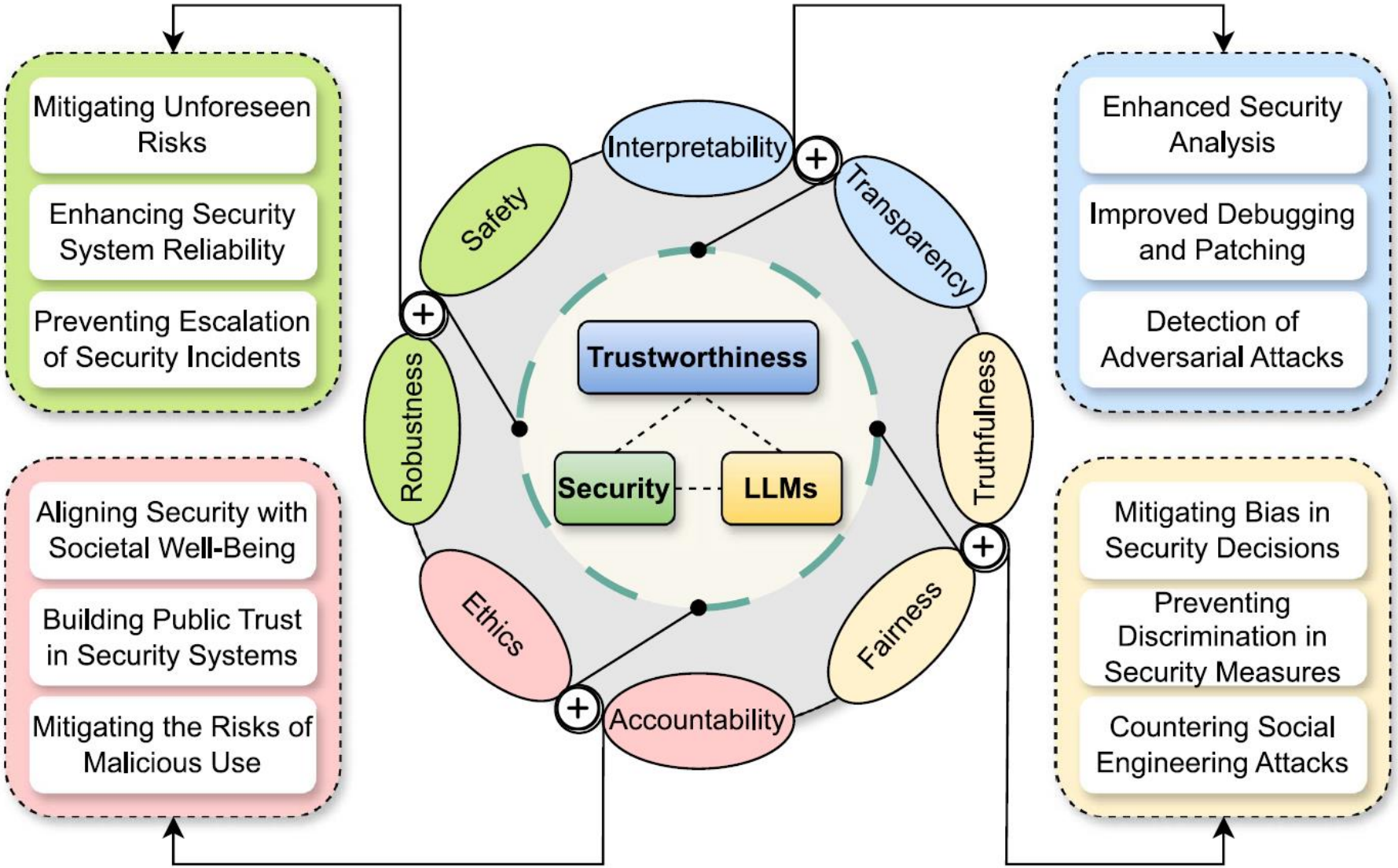
LLMs in Incident Response

"LLMs can **ONLY** be a team player in Cyber Security"

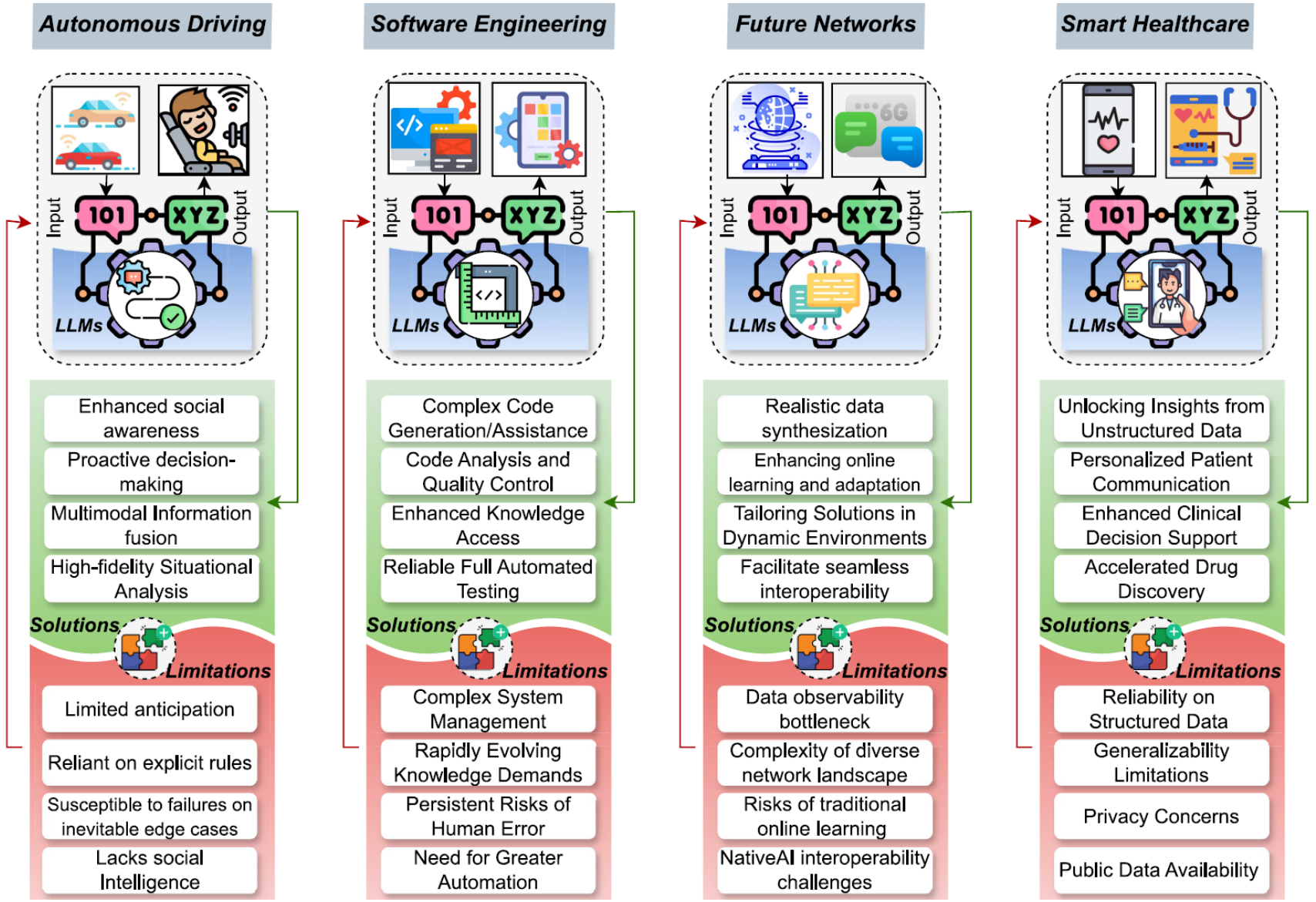


(Loumachi & Ghanem, 2024)

LLMs Trustworthiness and Security



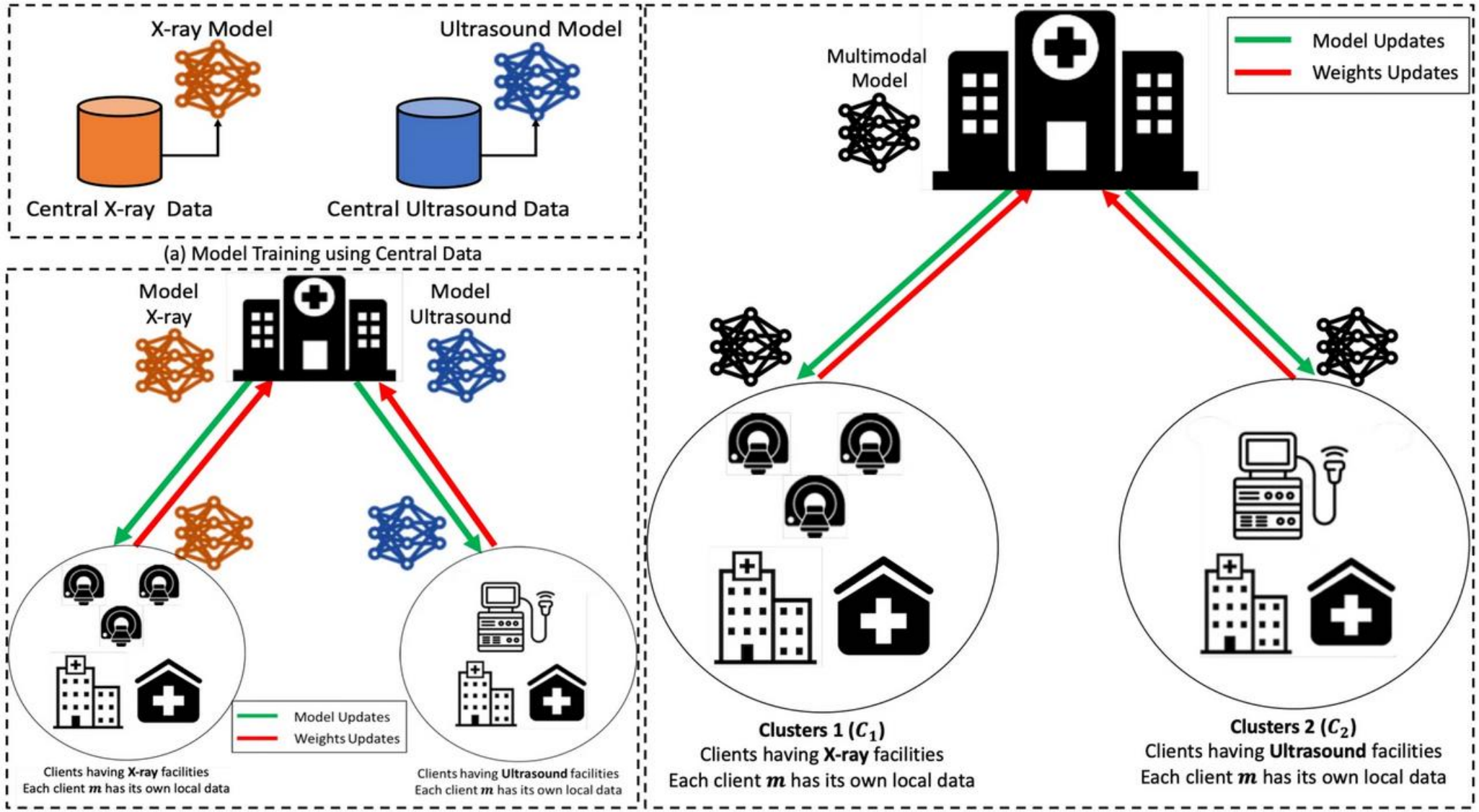
LLMs Opportunities and Challenges



Federated Learning

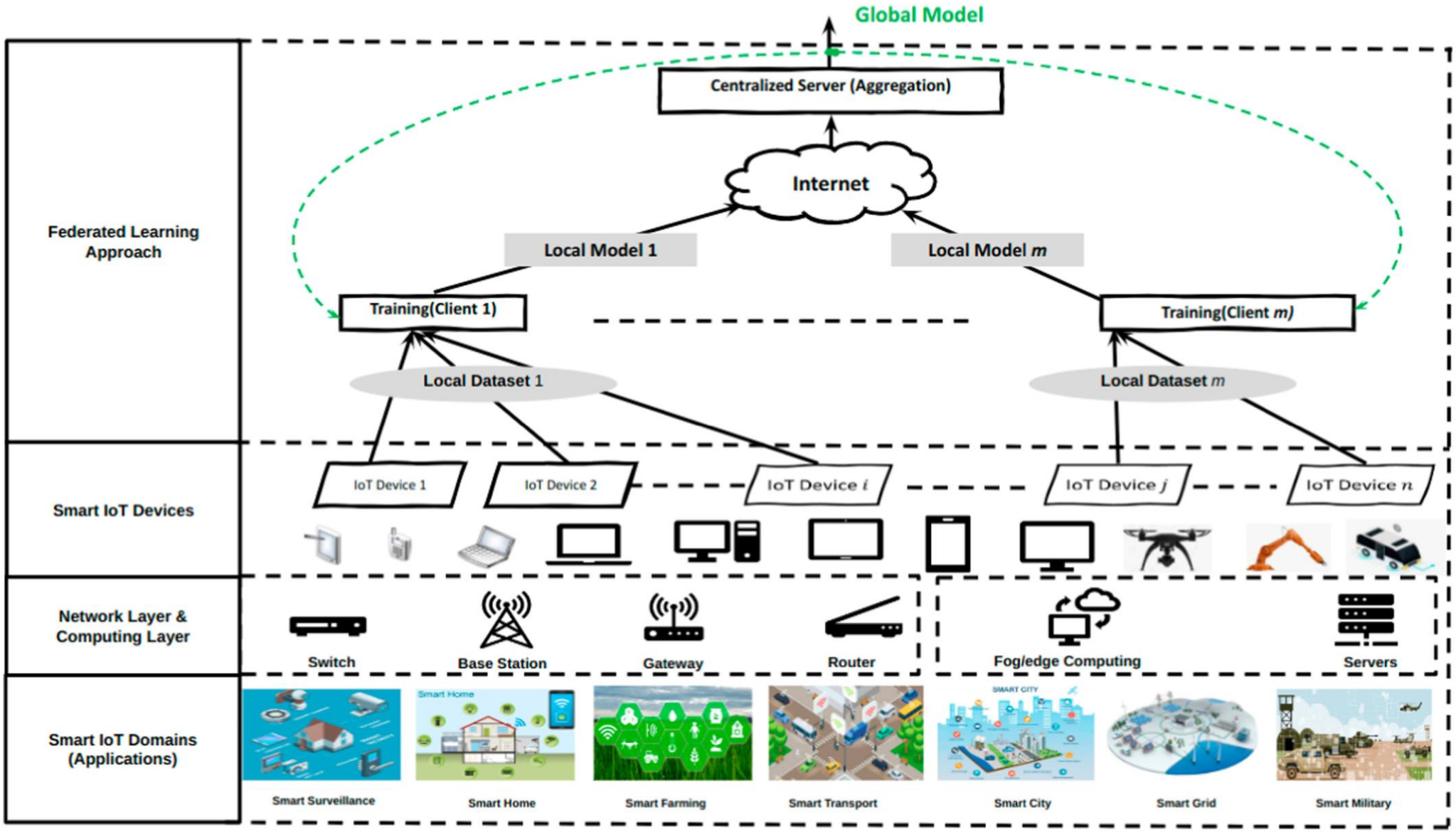
- A decentralized machine learning approach where models are trained locally on devices without transferring raw data, enhancing privacy and reducing central dependency.
- IoT devices can collaboratively learn from local data without exposing sensitive information, significantly enhancing data privacy and reducing security risks.
- FL Supports privacy regulations (like GDPR) by ensuring data stays on local devices, making it ideal for handling sensitive IoT data.
- Optimizes network usage and lowers latency, which is essential for real-time IoT applications (Since data doesn’t need to be transferred to central servers)

FL in Healthcare



(Qayyum et al., 2022)

FL in IoT and Cloud



(Yaacoub et al., 2023)

FL Challenges and Opportunities

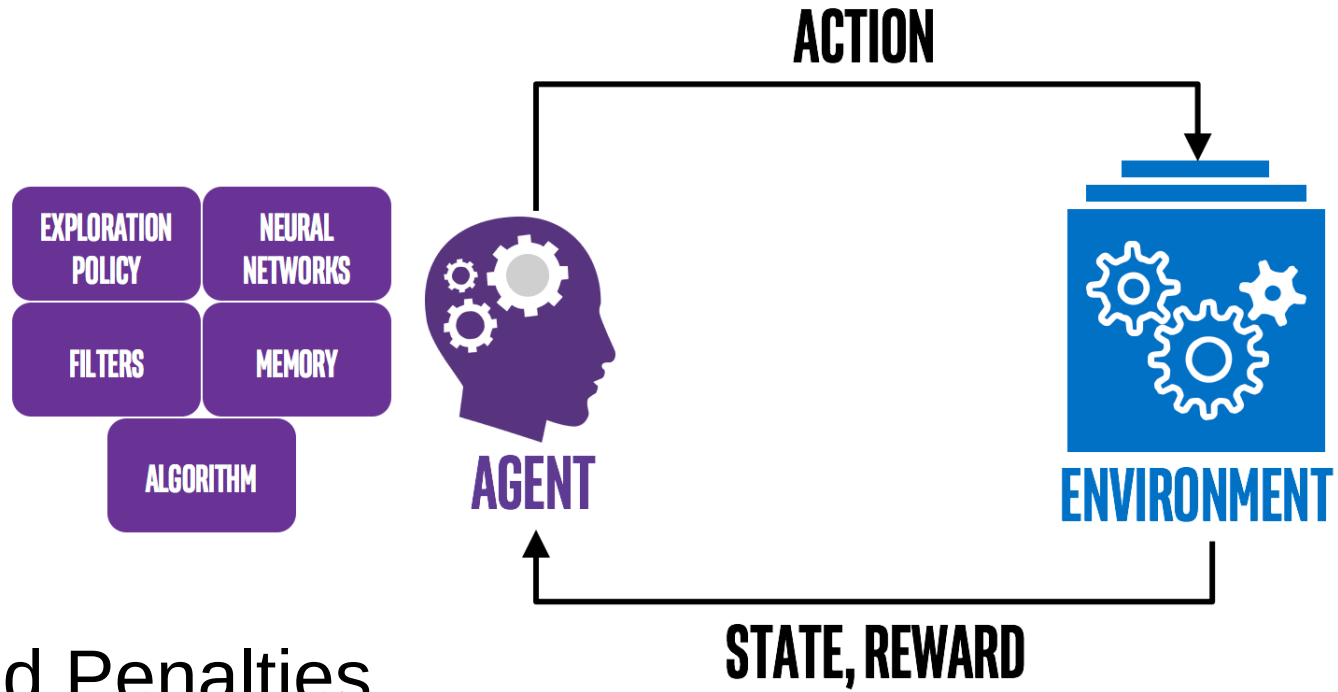
- **Heterogeneity** in IoT devices, thus heterogeneous Data formats, frequencies, and quality, introducing inconsistencies that a federated model can hardly generalize from.

- IoT devices **have limited computing power**, memory, and battery life, which restricts the processing (might change soon)

- Security and **Privacy** risks as attacks such as poisoning (injecting fake data) and inference attacks are still possible

- Model and Quality Ensuring **convergence** is “acceptable” but maintaining model quality across diverse IoT environments is challenging (the asynchronous nature of IoT implies slower or unstable convergence).

Reinforcement Learning (RL) and Deep RL (DRL)



Learning Through Rewards and Penalties

(Lee, 2017)

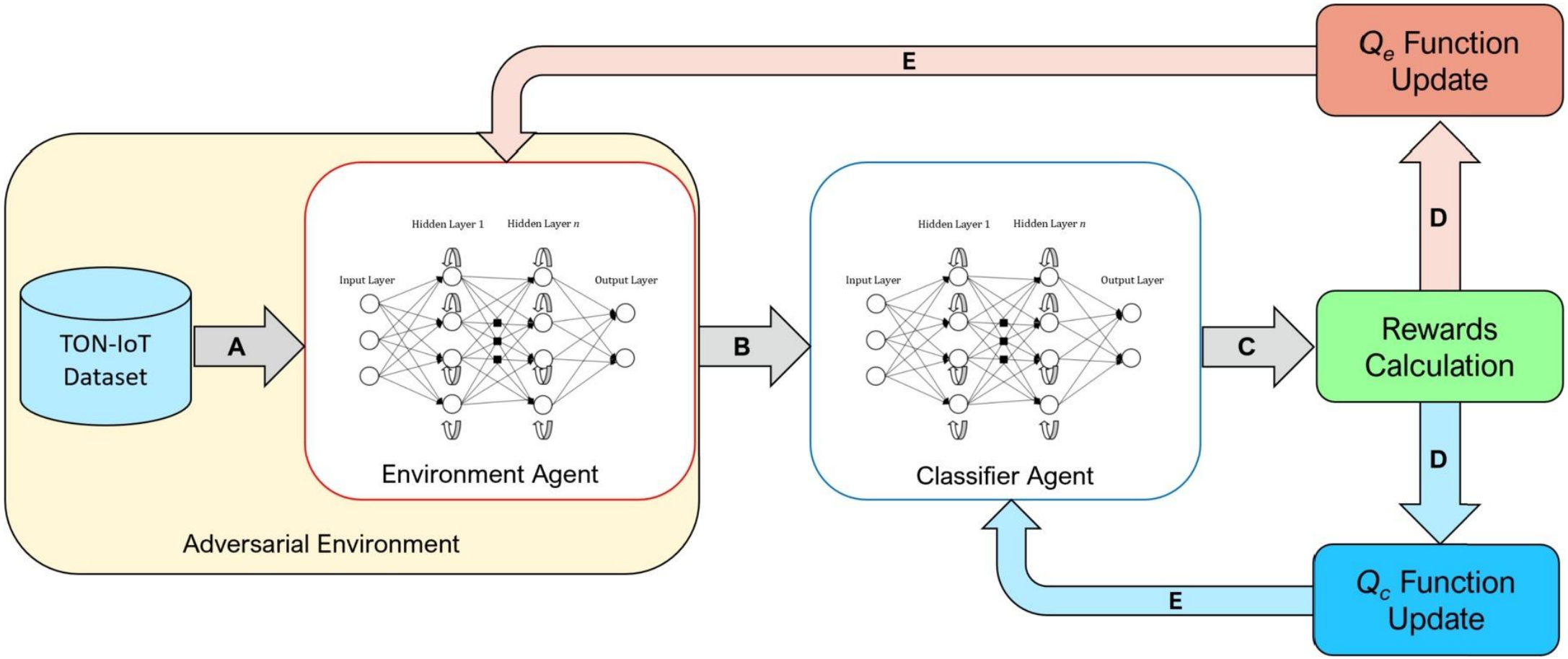
Trial-and-Error Process

Applications in Dynamic Environments

RL and DRL in IoT and Cloud

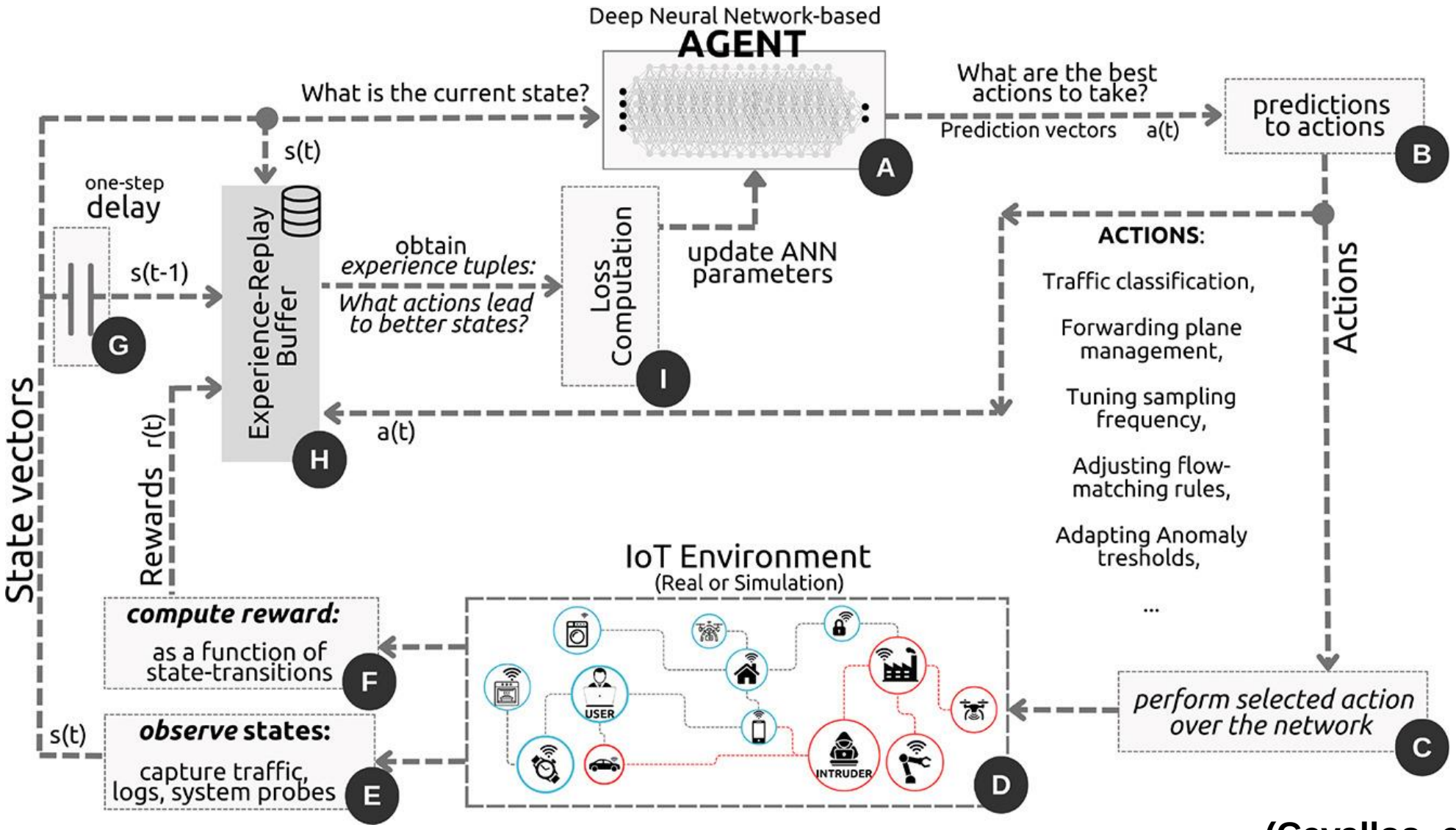
- Automated Intrusion Detection and Prevention
- Adaptive Security Policy Management
- Optimized Resource Allocation
- Continuous Learning from Cyber Threats
- Anomaly Detection with Minimal Human Intervention

RL Threat Detection on IoT devices



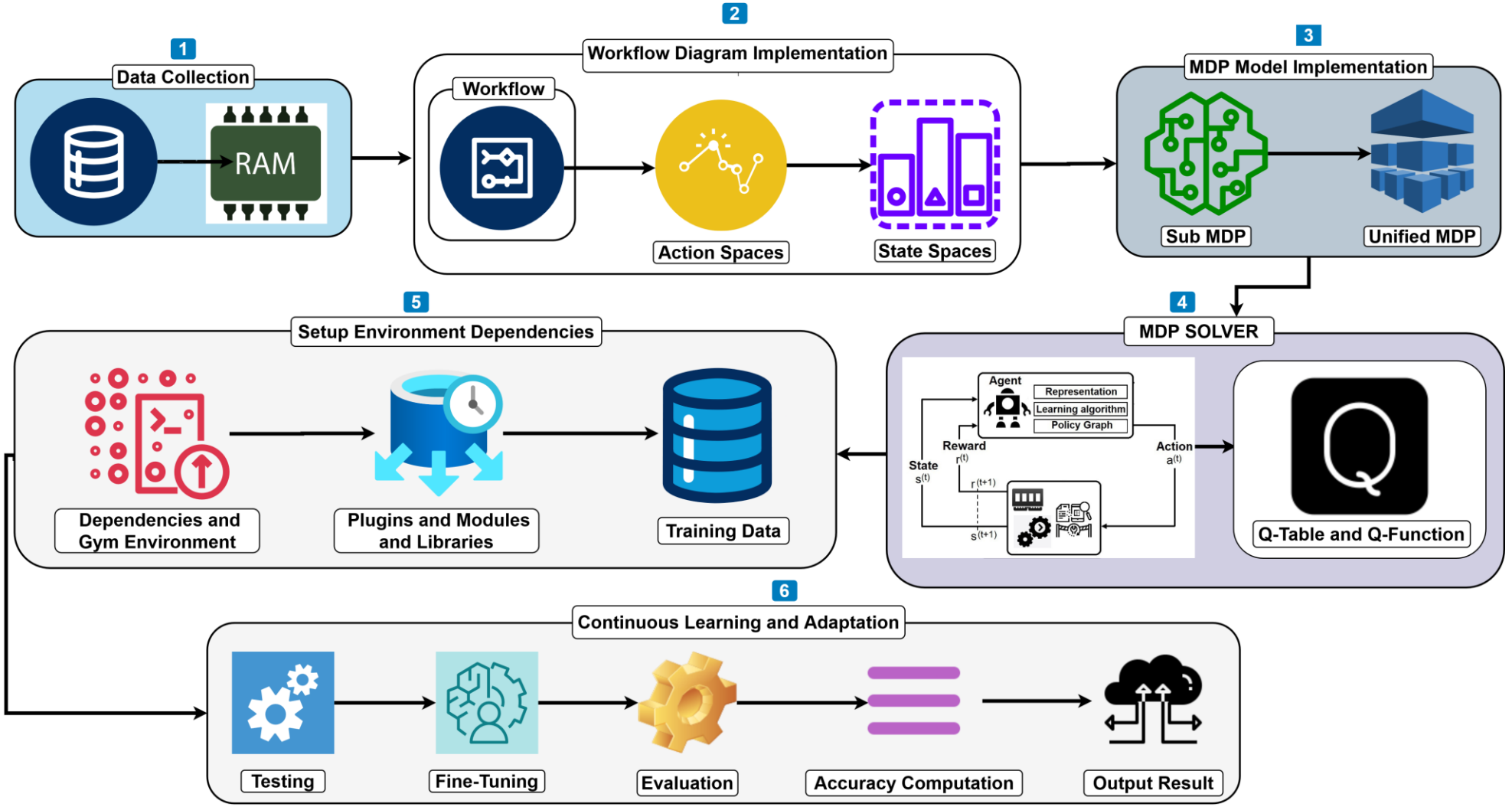
(Rookard & Khojandi, 2024)

DRL Intrusion Detection in IoT



(Cevallos et al., 2023)

DRL In Cyber Incident Response



(Dunsin et al., 2024)

RL and DRL Challenges and Opportunities

Data Scarcity and Quality: High-quality modelling is **essential** for RL and DRL which IoT and Cloud environments often make it difficult and thus **limiting the effectiveness**.

High Computational Demand: DRL requires **significant** computational resources for training and real-time deployment, which can be challenging in resource-constrained IoT devices and restricted cloud usage costs.

RL and DRL are perfect for !

Proactive and Adaptive Security Measures: RL and DRL enable IoT and cloud systems to proactively adapt to new security threats in **real-time.**

Scalability for Large-Scale Environments: DRL can **scale efficiently** to secure large, distributed IoT networks and cloud environments.

The near future …
What needs to be improved (potential ideas!)

Lightweight Models: Efficient model architectures and algorithms tailored for IoT constraints, such as lightweight neural networks or adaptive learning algorithms

Secure by Design Communication Protocols: Designing secure communication protocols and compression techniques to reduce security overhead

Enhanced Privacy-Preserving: Advanced privacy mechanisms, e.g. differential privacy and homomorphic encryption enforce data privacy without compromising model accuracy

Robustness: Robust defence mechanisms to detect and mitigate attacks, ensuring the integrity and reliability of federated learning models in hostile environments.

Adaptability: Develop scalable federated learning frameworks that can dynamically adapt to the growth of IoT networks and handle device variability, improving the robustness and applicability of federated learning in diverse IoT scenarios.

Ethical and Regulatory Considerations

Balancing Innovation with Accountability

Data Privacy and Compliance AI must adhere to data protection laws like GDPR to prevent misuse of sensitive information.

Transparency in AI Decision-Making

Ensuring AI models provide clear, understandable outcomes for cybersecurity actions.

Bias and Fairness in AI Algorithms

Preventing bias in threat detection models to maintain fair and effective security practices.

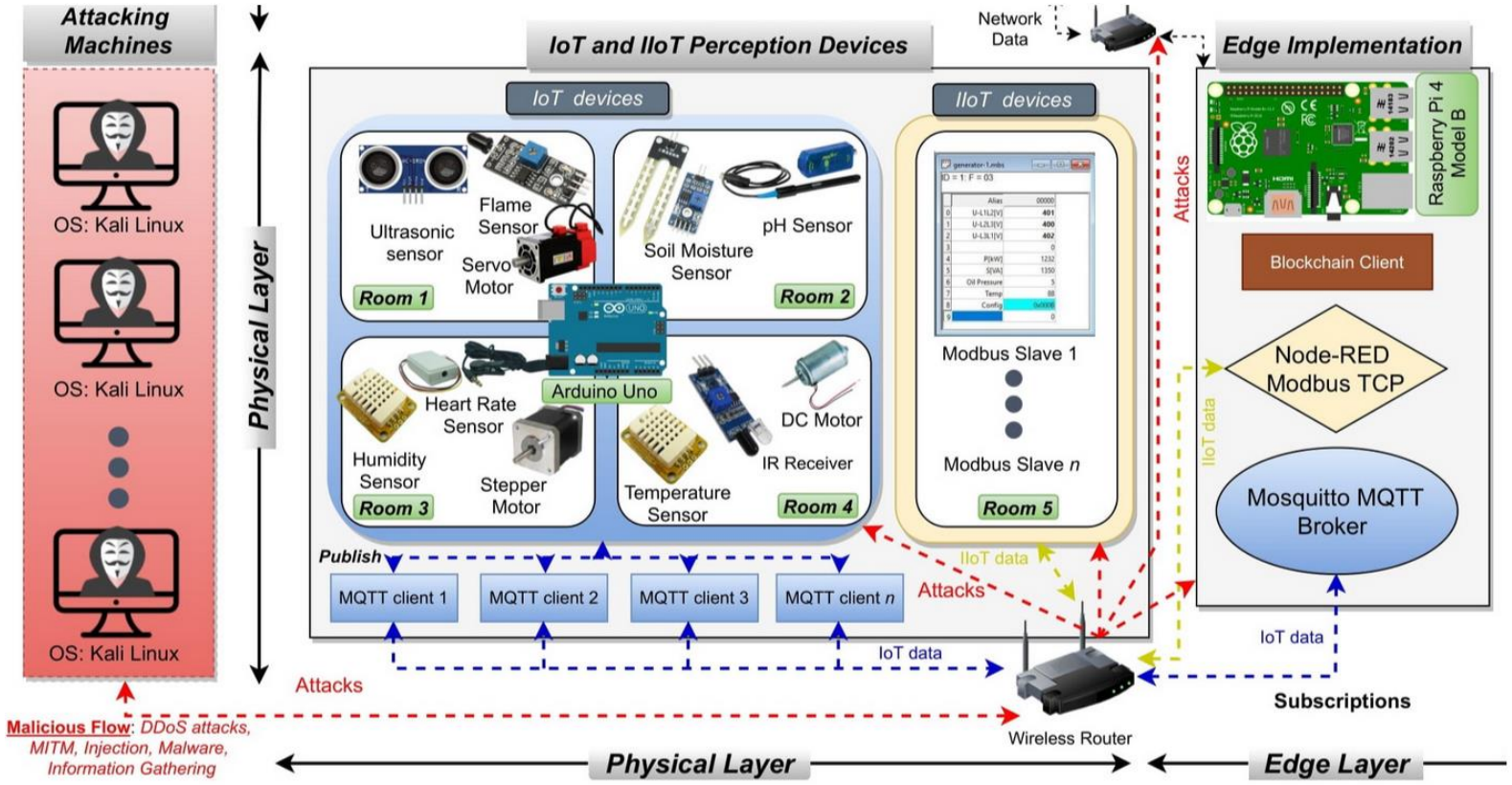
Some Resources

- **Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications: Centralized and Federated Learning**

Last updated:
Thu, 04/27/2023 -
DOI:
10.21227/mbc1-1h68
Data Format:
*.csv; *.pcap; *.txt; *.zip

- **IIoTset**

- **CIC IoT dataset 2023**



References

Yaacoub, J.P.A., Noura, H.N. and Salman, O., 2023. Security of federated learning with IoT systems: Issues, limitations, challenges, and solutions. *Internet of Things and Cyber-Physical Systems*, *3*, pp.155-179.

Friha, O., Ferrag, M.A., Kantarci, B., Cakmak, B., Ozgun, A. and Ghoualmi-Zine, N., 2024. Llm-based edge intelligence: A comprehensive survey on architectures, applications, security and trustworthiness. *IEEE Open Journal of the Communications Society*.

Mohamed Amine Ferrag, Othmane Friha, Djallel Hamouda, Leandros Maglaras, Helge Janicke, January 17, 2022, "Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications: Centralized and Federated Learning", IEEE Dataport,.

A. Qayyum, K. Ahmad, M. A. Ahsan, A. Al-Fuqaha and J. Qadir, "Collaborative Federated Learning for Healthcare: Multi-Modal COVID-19 Diagnosis at the Edge" in IEEE Open Journal of the Computer Society, vol. 3, no. 01, pp. 172-184, null 2022.

Kheddar; H., Dawoud D., Awad, A. Himeur, Y.,Khan, M 2024. Reinforcement-Learning-Based Intrusion Detection in Communication Networks: A Review. IEEE Communications Surveys & Tutorials.

Hamouda, D., Ferrag, M.A., Benhamida, N., Seridi, H. and Ghanem, M.C., 2024. Revolutionizing intrusion detection in industrial IoT with distributed learning and deep generative techniques. *Internet of Things*, *26*, p.101149.

Loumachi, F.Y. and Ghanem, M.C., 2024. Advancing Cyber Incident Timeline Analysis Through Rule-Based AI and Large Language Models. *Forensic Science International: Digital Investigation*, preprint.

Rookard, C. and Khojandi, A., 2024. RRIoT: Recurrent reinforcement learning for cyber threat detection on IoT devices. *Computers & Security*, *140*, p.103786.

Rizzardi, A., Sicari, S. and Porisini, A.C., 2023. Deep Reinforcement Learning for intrusion detection in Internet of Things: Best practices, lessons learnt, and open challenges. *Computer Networks*, *236*, p.110016.

Dunsin, D., Ghanem, M.C., Ouazzane, K. and Vassilev, V., 2024. Reinforcement Learning for an Efficient and Effective Malware Investigation during Cyber Incident Response. *High Confidence Computing. In Press.*

Thank you!

Questions