Beyond Rule-Based Context Awareness: Large Language Models as Adaptive Cognitive Layers in Cyber-Physical Systems

Md Azher Uddin[*1], **Hanan Salam**[*2†]

[2]Center of AI & Robotics, Social Machines & RoboTics Lab, New York University Abu Dhabi
PO Box 129188, Saadiyat Island, Abu Dhabi, United Arab Emirates
[1]School of Mathematical and Computer Sciences, Heriot-Watt University Dubai
Dubai, United Arab Emirates
m.uddin@hw.ac.uk, hanan.salam@nyu.edu

Abstract

Cyber-physical systems (CPS) have traditionally relied on rule-based mechanisms and machine learning models for context awareness. However, these approaches often struggle with dynamic adaptation, multimodal data integration, and real-time decision-making in complex environments. With the emergence of large language models (LLMs), we argue that CPS should adopt LLMs as adaptive cognitive layers capable of interpreting, reasoning, and responding to real-world contexts in real time. This position paper explores the paradigm shift introduced by LLMs, discusses their advantages and limitations, and presents a vision for their integration into next-generation CPS.

Introduction

Cyber-Physical Systems (CPS) (Gunes et al. 2014; Horvath and Gerritsen 2012) are integrated networks where physical processes interact with computational and communication systems. These systems are widely used in various domains, such as healthcare (Medjahed et al. 2011; Lalwani, Saleh, and Salam 2025), transportation (Xiong et al. 2015), energy (Lu 2018), and manufacturing (Scholze and Barata 2016).

Context in CPS is a dynamically aggregated collection of information defining conditions in which a system operates. It encompasses user context (i.e., identity, preferences, behavior, and intent) (Salam et al. 2023), system context (i.e., device status, operational state, communication status, power consumption, and hardware or software reliability), environmental context (i.e., physical conditions such as temperature, humidity, pressure, and vibration, as well as location-based data and infrastructure state), security context (i.e., threat intelligence, encryption, authentication, and access control policies), temporal context (i.e., time-sensitive operations, historical data, and predictive analytics), and social-organizational context (i.e., collaboration data, business policies, and stakeholder interactions) (Gaur et al. 2019). Contextual information is vital in enhancing CPS functionality since it assists in better decision-making, optimizes processes, and makes systems responsive to varied environmental and running conditions (Sahlab, Jazdi, and Weyrich 2021). For example, in industrial automation, CPS can alter machine operation based on temperature variations in a bid to ensure efficiency (Sahlab, Jazdi, and Weyrich 2021), while in autonomous vehicles, traffic congestion variations influence navigation plans (Mortlock et al. 2024).

Context awareness (Dey 2001) is a fundamental requirement in CPS, allowing systems to adapt their behavior based on real-time environmental conditions, sensor data, and human interactions (Ivanov, Weimer, and Lee 2018; Scholze and Barata 2016). Traditional context-aware approaches in CPS rely on rule-based systems (Scholze and Barata 2016) and machine learning models (Wasim et al. 2021; Mortlock et al. 2024). However, these methods face limitations in: (1) handling dynamic environments where predefined rules become obsolete (Zhang et al. 2024), (2) integrating multimodal data, including numerical sensor inputs, human commands, and environmental changes (Alsamhi et al. 2024; Schirner et al. 2013; Noorani et al. 2024; Medjahed et al. 2011), and (3) scalability and adaptation, particularly in heterogeneous, real-world CPS deployments (Sahlab, Jazdi, and Weyrich 2021).

The emergence of large language models (LLMs) such as GPT-4 (OpenAI 2023), PaLM (Chowdhery et al. 2022), and LLaMA (Touvron et al. 2023) marks a transformative shift in CPS. LLMs can process vast amounts of unstructured information, generate adaptive responses, and generalize across novel contexts with minimal retraining. Researchers have begun exploring LLMs for CPS applications, including CPS requirement extraction (Jin et al. 2024), CPS design support (Choaib et al. 2024), and human behavior analysis during the CPS design phase (Burgueño et al. 2024). Additionally, LLMs are being applied to develop context-specific advisory services, such as agricultural risk assessment tailored to local legislation and regulations (Stoyanov et al. 2023). However, these efforts remain limited in their ability to explicitly address dynamic context adaptation, as current works primarily focus on automating manual processes or providing domain-specific recommendations without fully capturing contextual dynamics. For instance, (Jin et al. 2024) highlight the limitations of LLMs in extract-

[*]These authors contributed equally.
[†]Corresponding author: Hanan Salam.

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

ing precise CPS-specific system requirements, emphasizing challenges in capturing specialized concepts and preventing hallucinations. Similarly, (Choaib et al. 2024) demonstrate how LLMs streamline CPS design through context-aware recommendations, yet their approach does not directly address real-time adaptive context management. Furthermore, (Burgueño et al. 2024) propose ontology-driven methods for identifying uncommon human behaviors in CPS design, showing promise for robust interaction scenarios but not directly tackling real-time contextual adaptation. Likewise, (Stoyanov et al. 2023) present LLM-based advisory services for agriculture, though their approach focuses on static advisory contexts rather than dynamic, adaptive interventions. Thus, explicitly integrating LLMs for adaptive, context-aware interventions in CPS remains a significant yet largely unexplored research opportunity.

In this paper, we propose a new LLM-based context-awareness paradigm for CPS. We propose **LLMs as cognitive layers** in CPS for facilitating real-time adaptation, multimodal reasoning, and human-centric decision-making. This paradigm addresses several critical limitations in traditional rule-based and machine learning-based context-awareness solutions. Specifically, it resolves the inflexibility and lack of scalability associated with predefined rules by benefiting from the zero-shot and few-shot adaptability of LLMs, enabling systems to learn to cope with new and unforeseen situations quickly without requiring substantial retraining. Additionally, as traditional models typically do not deal efficiently with the fusion of numerical sensor inputs, textual commands, and environmental information, the proposed solution successfully addresses this multimodal data fusion issue. LLMs naturally support multimodal fusion, facilitating simultaneous and coherent handling of various input modalities to achieve a holistic contextual understanding. Yet another critical issue overcome by our paradigm is the limited interpretability and human-oriented usability in traditional black-box machine learning models. LLMs has the potential to provide naturally human-oriented explainability, facilitating transparency, trust, and ease of interaction with system operators (Zhao et al. 2024). Finally, the proposed paradigm considerably minimizes the traditional overhead in maintaining and updating rule-based or complicated machine learning models by benefiting from quick deployment and iterative refinement through prompt engineering. These advantages collectively represent a substantial advancement in the robustness, adaptability, and usability of context-aware CPS.

Traditional Context Awareness in CPS

Context awareness within CPS has traditionally employed rule-based systems (Takatsuka et al. 2014; Sol et al. 2018) and machine learning techniques (Wasim et al. 2021; Mortlock et al. 2024) to detect and respond to contextual factors such temperature fluctuations in industrial automation (Sahlab, Jazdi, and Weyrich 2021), traffic density variations in autonomous vehicles (Mortlock et al. 2024), and physiological state shifts in wearable health monitors (Gaur et al. 2019). These approaches integrate sensor-derived information with computational models or rules that trigger automated responses based on predefined conditions. This section reviews relevant research on traditional context awareness in CPS.

Rule-Based Context Awareness

Rule-based systems are commonly applied in context-aware CPS to support decision-making based on pre-defined situations. The systems are based on a rigid if-then-else paradigm, in which responses are activated based on the occurrence of defined situations. For instance, RuCAS by (Takatsuka et al. 2014) utilizes a context-aware management technique with rule-based services supported through cloud computing and machine-to-machine (M2M) communication. It enables users to author context-aware rules with web-service information using an Event-Condition-Action (ECA) model to perform automated actions. Its limitation in utilizing static rules limits adaptability with context changes that must be updated manually, suggesting that it cannot adapt automatically through learning. (Sol et al. 2018) employed context-aware CPS to regulate and monitor energy consumption of lighting and electrical conditions in smart buildings so that energy efficiency would be achieved at minimal user discomfort. In their findings, traditional building integration with CPS technology remained economically infeasible due to upfront costs and incompatibilities with infrastructure. Furthermore, their rule-based system prevented real-time feedback-based adaptive learning. A more user-oriented solution was offered by (Lu 2018) with a game-based, adaptive IoT-based CPS to increase user participation in energy savings through virtual interactions with smart-home decisions. Though effective in triggering user engagement, this approach had scalability challenges with significant user involvement required to incorporate other context factors. User engagement changes also limited automation features of the system. (Gaur et al. 2019) proposed the Context-Aware Programming (CAP) framework to simplify context-aware application programming in CPS using advancements in Wireless Sensor Networks (WSNs) and communication protocols. The CAP framework is dynamically configurable but is not scalable due to its rule-based structure in highly dynamic situations. Later, (Daun and Tenbergen 2023) proposed an ontological paradigm to represent the context in CPS with explicit context dependency modeling. While ontologies offer structured adaptability compared to rule-based approaches, they require frequent updation with new context variables and lack real-time adaptability. In addition, (Asmat et al. 2023) elaborated on context uncertainty by introducing an ontology-based model that seeks to increase reliability and context interpretation in CPS. The system's reliance on pre-defined uncertainty domains limits it to deal with unforeseen or new situations.

Limitations: Rule-based systems ensure predictability and dependability, with the cost of various limitations, making the rule-based methodology less effective in dynamic and intricate situations.

Inability of Dynamic Adaptation. The primary limitation of rule-based systems is that they are inflexible because they cannot handle new or unforeseen situations without hu-

man intervention. This ongoing need to manually update decreases long-term scalability in dynamic settings such as smart cities, industrial automation, and security (Gaur et al. 2019; Takatsuka et al. 2014).

Scalability. Another significant limitation lies in scalability. An increasing number of contextual variables raises complexity in managing and updating rule sets. For instance, in industrial automation, multiple sensors continuously provide real-time feedback from machines, the environment, and human operators, requiring rule-based systems to handle numerous conditions to encompass all possible states (Sahlab, Jazdi, and Weyrich 2021). This complexity grows exponentially, slowing real-time decision-making efficiency. Additionally, conflicting rules arising from interactions among conditions can lead to unintended system behavior, necessitating further manual intervention (Sol et al. 2018).

Noisy and Incomplete Data. Rule-based systems also struggle with dealing with noisy and incomplete data. For instance, sensor readings in healthcare monitoring (Medjahed et al. 2011) and autonomous driving (Mortlock et al. 2024) are usually frequently noisy, incomplete, or erroneous. The traditional rule-based models cannot effectively fill missing values or handle imprecise input and thus make incorrect or suboptimal decisions in some instances (Daun and Tenbergen 2023). Unlike probabilistic and machine learning-based models, rule-based systems lack the capability to dynamically adapt decisions with historical patterns or probabilistic reasoning (Noor et al. 2023).

Maintenance Overhead. Another significant limitation of is maintenance overhead. Rules have to be updated and tuned by hand to maintain pace with system changes and emergent conditions. It is a time-consuming and labor-intensive exercise that requires a high degree of domain-specific knowledge and reduces flexibility as well as increases cost (Takatsuka et al. 2014).

Computational Inefficiency. Handling input via large and complex rules significantly increases system delay and compromises real-time performance in applications with a requirement for rapid responsiveness such as industrial automation, cyber threat detection, and emergency response (Liao et al. 2021). Consequently, modern-day CPS tend to adopt hybrid approaches that combine rule-based reasoning with machine learning, and probabilistic techniques to improve adaptability and efficiency (Wasim et al. 2021; Mortlock et al. 2024).

Machine Learning-Based Context Awareness

Machine learning (ML) enhances context perception in CPS by learning from data to make effective responses in various applications. For instance, (Wasim et al. 2021) introduced a deep learning framework to monitor campus through a Convolutional Neural Network (CNN) to detect academic activity in video streams with high accuracy. In industrial automation, (Sahlab, Jazdi, and Weyrich 2021) proposed a cyber-physical system using semantic tagging and graph modeling to enhance real-time decision-making with limited scalability due to standardization issues among industries and complexity in handling heterogeneous information sources. (Liao et al. 2021) developed an attention-based industrial CPS event recommendation model that, although novel in machine learning applications, suffers from edge deployment due to high processing demands and lack of generalizability due to pre-defined context categories. (Hsieh 2022) explored using Discrete Timed Petri Nets to make cyber-physical production systems (CPPS) deadline-aware and future-state-aware. The model is formalized but not computationally efficient and requires high-level human configuration at the expense of adaptability. (Noor et al. 2023) introduced an intelligent security model to identify cyber threats in real-time in CPS. The model requires large labeled datasets and is not highly scalable because it is computationally intensive. (Mortlock et al. 2024) introduced CASTNet, a spatio-temporal motion prediction model used in autonomous driving. Although it improves real-time decision-making by detecting context changes, it has high processing demands and therefore cannot be implemented in low-resource embedded systems. Finally, (Maity et al. 2024) presented the Data Context-Driven Model Reduction (DCDMR), a two-stage framework that pre-trains context-aware submodels offline and dynamically adapts at run time using regression techniques. Employed to control the Medtronic 670G Artificial Pancreas, DCDMR surpasses traditional methods like RNNs with lower training sample requirements and enables faster as well as safer decision-making in CPS.

Limitations: Machine learning-based CPS, despite the numerous benefits, also suffer from various limitations and challenges.

Data Dependency. One of the major problems in ML-based CPS is the dependence on data. ML models are trained to give accurate predictions from enormous amounts of high-quality labeled data. In most applications in CPS, such as in industrial automation and smart monitoring, the acquisition and labeling of enormous datasets are time-consuming and expensive. The efficacy of ML models also reduces in the case of noisy, incomplete, and biased datasets, leading to unreliable decision-making. This problem is most pronounced in security-critical CPS, in which erroneous decisions cause failure to detect cyber threats and flag legitimate activity as anomalies (Noor et al. 2023).

Computational Complexity. Another significant limitation lies in the problem of the computational complexity (Mortlock et al. 2024). Many current ML techniques, such as deep learning and reinforcement learning, are computationally heavy, and real-time deployment in most situations, particularly in constrained edge devices and IoT-based CPS, is problematic. The greater the model's size, the greater the processing, and the greater the energy and the latency. For instance, in event-based industrial CPS, (Liao et al. 2021) determined group recommendation models based on the use of the attention mechanism provide higher precision, but the heavy processing burden makes them inappropriate in real-time applications.

Interpretability and Explainability. Interpretability and explainability are also major concerns in ML-based CPS. Traditional rule-based models offer transparent, inter-

pretable logic, while deep models are black boxes whose reasoning to specific decisions are difficult to follow. Decision-making transparency concerns in the case of applications in driving and healthcare monitoring could impair trust and slow regulation approval (Wasim et al. 2021).

Deployment Constraints. Deployment constraints further limit ML-based CPS. Many ML models are trained in controlled environments but fail to generalize well in real-world, dynamically evolving CPS applications. (Noor et al. 2023) highlighted this issue in cybersecurity-based CPS, where machine learning models effectively detect known threats but struggle with zero-day attacks and novel cyber threats due to their reliance on historical data. Similarly, (Sahlab, Jazdi, and Weyrich 2021) found that while ML techniques improved context awareness in industrial CPS, they still required periodic retraining to adapt to changing operational conditions.

LLMs as Cognitive Layers for Context Awareness

To tackle the limitations of traditional context-awareness in CPS, we propose a new LLM-driven context-awareness paradigm that leverages LLMs (e.g., GPT-4, Gemini, LLaMA) as adaptive context processors to dynamically interpret, predict, and respond to contextual changes in CPS. The LLM acts as a decision support layer, processing multimodal data (text, sensor feeds, historical logs) and generating real-time context-aware recommendations or interventions. In this paradigm, LLMs act as adaptive cognitive layers that enable: (1) Multimodal Integration & Reasoning, (2) Dynamic Real-World Context Interpretation (3) Generation of Adaptive Context-Aware Interventions, and (4) Human-Centric Adaptation. This cognitive layer serves as an intelligent middleware between the physical system (actuators, sensors) and high-level decision-making processes, making CPS more autonomous, context-aware, and responsive (see Figure 1).

Multimodal Integration & Reasoning. LLMs can process and integrate information from multiple modalities, including structured (sensor data) and unstructured (text, speech, images) data to enhance CPS decision-making (Wu et al. 2024). The advantage of LLMs in multimodal data fusion is due to their ability to align multimodal data through shared representations, using architectures such as transformers and cross-attention mechanisms (Tsimpoukelli et al. 2021). Additionally, multimodal embeddings enable CPS to capture cross-domain correlations, improving robustness in uncertain environments. This is useful in various CPS applications such as in autonomous driving where camera feeds, LiDAR signals, GPS data, and textual navigation inputs must be processed simultaneously in order to generate comprehensive driving strategies (Zhou et al. 2024). Similarly, in smart healthcare CPS, this allows LLMs to analyze patient vitals, medical images, and clinical notes efficiently in order to provide holistic diagnostic assistance (Zhang et al. 2015). Recent related work further expands on LLM capabilities by exploring "Penetrative AI," where LLMs leverage their embedded common-sense world knowledge to reason directly over sensory data from IoT devices (Xu et al. 2024). This approach showcases the unique proficiency of LLMs in interpreting and reasoning about real-world physical tasks. It also highlights their potential to incorporate human-level knowledge and intuition into CPS, enabling novel applications beyond traditional text-based scenarios.

Dynamic Real-World Context Interpretation. The capacity and efficiency of LLMs in multimodal integration and reasoning allows them to be used as efficient tools to process inputs to infer real-time context accurately. For instance, contextual embeddings generated by LLMs can be used to maintain historical system states and infer dependencies between events (Enoasmo et al. 2025). This makes them particularly useful in inferring context in industrial CPS, where unexpected system faults can be diagnosed by referencing previous system logs and sensor data. For instance, recent work (Abshari, Fu, and Sridhar 2024) demonstrated that Retrieval-Augmented Generation (RAG) recently introduced for improving LLMs, can automatically extract physical invariants from CPS documentation, significantly enhancing anomaly detection capabilities in industrial CPS by addressing traditional limitations such as manual invariant definition, scalability issues, hallucination, and concept drift.

Generation of Adaptive Context-Aware Interventions. One of the most promising applications of LLMs in CPS is their ability to generate adaptive, context-aware interventions in response to real-time data. Traditional CPS which rely on manually defined rules or pre-trained models for context awareness lack adaptability in dynamic environments (Dey 2001). The ability of LLMs to perform zero-shot and few-shot learning and thus generalize across domains, allows them to adapt to unseen scenarios without extensive retraining (Radford et al. 2019), thus dynamically generating tailored responses based on evolving contextual factors. In high-stakes CPS applications such as autonomous vehicles, healthcare, and industrial automation, real-time adaptation is critical. For example, in smart grids, LLMs can analyze electricity consumption patterns and generate dynamic energy-saving strategies based on demand fluctuations and external environmental conditions. Similarly, in autonomous healthcare systems, LLMs can assist in generating personalized real-time alerts and treatment recommendations by analyzing patient vitals, historical data, and contextual factors (Nazi and Peng 2024). Similarly, an LLM integrated into an autonomous vehicle system can interpret contextual variables like changing road conditions, sensor feedback, and human driver commands to generate adaptive control strategies (Kendall, Gal, and Cipolla 2018; Kendall et al. 2019).

Human-Centric Adaptation. LLMs facilitate natural language interaction in CPS, enabling operators to dynamically query, refine, and adapt system behaviors, thereby improving transparency, usability, and trust (Amershi et al. 2019). Unlike traditional CPS, which limit interactions to predefined interfaces, LLM-driven systems offer context-aware and human-centric adaptation. For instance, in smart manufacturing, human operators can provide high-level in-

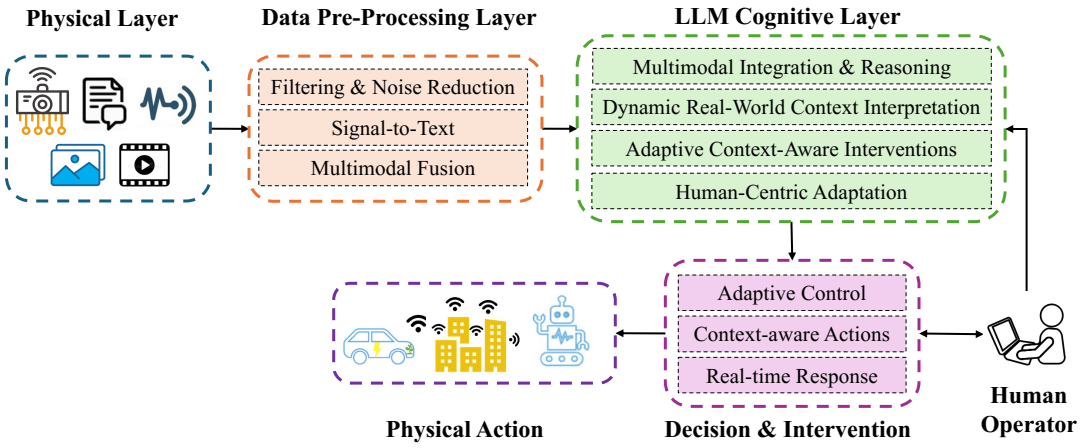

Figure 1: Block diagram of LLM-driven context-awareness in CPS illustrates a structured integration of multimodal sensor inputs, data preprocessing, and cognitive processing via LLMs. It highlights how sensors capture numerical data, IoT readings, textual commands, and visual data, which are preprocessed through filtering and multimodal fusion. Central cognitive processing by LLMs provides adaptive, real-time reasoning and context interpretation. Human interaction further enriches decision-making, ensuring transparency and validation. Finally, actuators respond with appropriate physical actions, forming an interactive loop that enhances the responsiveness, adaptability, and human-centric nature of CPS applications.

structions to dynamically adjust machine parameters, enhancing both efficiency and safety. Similarly, autonomous transportation systems can adapt driving strategies based on passenger input, improving user experience. However, to address challenges such as generating infeasible or unsafe actions, recent approaches like CPS-LLM integrate instruction-tuned LLMs with system dynamics estimation, ensuring that generated action plans remain safe, personalized, and practically viable in critical domains such as automated insulin delivery (Banerjee et al. 2024).

Challenges and Open Research Questions

As LLMs are integrated into CPS as adaptive cognitive layers, some problems are to be addressed to enhance their efficiency, efficacy, and real-world usability. Traditional context-aware CPS, rooted in rule-based and machine learning-based mechanisms, are found to be limited in scalability, adaptability, and real-time responsiveness. Despite the real-time decision-making, semantic reasoning, and integration of multimodality offered by LLMs, deployment in CPS raises major open research issues in terms of computational efficiency, protection of the data, generalization, and integration in real-time CPS.

Computational Overhead and Energy Efficiency: One of the primary challenges is computational overhead and energy efficiency. LLMs, particularly deep transformer-based models, require significant computational resources to process vast amounts of contextual data. This can be prohibitive in resource-constrained environments such as industrial automation systems, edge devices, and IoT networks. Open research questions include: How can LLMs be optimized for efficient real-time execution in CPS? Can knowledge distillation or model compression techniques be leveraged to reduce computational load while preserving decision accuracy?

Adaptability: Another significant challenge lies in real-time adaptability and inference latency. CPS applications, including autonomous driving, smart grids, and industrial automation, demand millisecond-scale decision-making to provide safe and efficient operation. Conventional machine learning-based CPS already suffer from high inference latency, and the addition of LLMs could further aggravate real-time response issues. This poses open questions such as: Are LLMs actually integrated into real-time CPS control loops possible without causing considerable decision delay? How are the adaptability and real-time responsiveness to be balanced in the design of the hybrid architectures (e.g., the integration of the LLM and the use of light-weight predictive models)?

Generalization: Furthermore, context generalization and domain adaptation remain critical issues. CPS environments are highly heterogeneous, spanning domains such as smart healthcare, industrial automation, transportation, and cybersecurity. LLMs, trained on general-purpose knowledge, may not perform well in domain-specific CPS applications without extensive fine-tuning. An open question is: How can LLMs be effectively adapted to CPS-specific contexts without requiring expensive retraining? Can few-shot or continual learning approaches help LLMs incrementally adapt to new CPS environments without catastrophic forgetting?

Data Privacy: Privacy and security concerns are also significant. CPS generates sensitive real-time information, including industrial control commands, users' patterns, and cyber threats. The employment in such applications raises concerns about revealing the data, adversarial manipulation, and regulation compliance. Open research questions are:

how to integrate mechanisms preserving the privacy (e.g., federated learning, homomorphic encryption) into the employment in CPS. How can LLM deployments be made resilient against adversarial attacks, data poisoning, and other threats while maintaining robust context awareness.

Interpretability: Human interpretability and decision transparency are also major issues. Conventional rule-based techniques offer transparent, interpretable decision traces, while LLMs are typically black-box models, and tracing and verifying decisions in safety-critical applications in CPS are problematic. The issues to consider are questions such as the following: What are the ways in which explainable AI (XAI) techniques are to be applied to LLM-based CPS to increase interpretability? Are justifications from the decisions of the LLM possible in a manner consistent with the regulation and operation in CPS?

Safety and Reliability. A major challenge in deploying LLM-driven interventions is ensuring safety and reliability. CPS operate in environments where incorrect interventions can lead to catastrophic failures. Open research questions to mitigate the concerns regarding reliability and trustworthiness that LLM-enabled include: What architectures enable safe and explainable interventions in safety-critical CPS? How can assurance mechanisms be designed to generate human-understandable justifications that increase operator trust in LLM-generated actions? How can human-in-the-loop mechanisms, such as RLHF, be incorporated into verification pipelines while preserving explainability and traceability?

Roadmap: Towards Context-Aware CPS 2.0

The integration of LLMs into CPS opens exciting possibilities for the future. We propose a roadmap toward Context-Aware CPS 2.0, characterized by: (1) Hybrid Architectures, (2) Real-Time Edge Processing, and (3) Human-AI Collaboration.

Hybrid Architectures. The successful integration of LLMs into CPS requires hybrid architectures that integrate multiple AI paradigms. This means potentially combining LLMs with symbolic AI, reinforcement learning, and real-time adaptive control to optimize decision-making. Symbolic AI relies on rule-based logic and ontologies. This allows it to provide structured reasoning that complements the flexible adaptability of LLMs (Russell and Norvig 2020). Reinforcement learning (RL) on the other hand could allow CPS systems to learn from interactions and refine decisions based on human feedback (Sutton and Barto 2018). For example, autonomous vehicles can benefit from a hybrid system where LLMs generate contextual responses while RL algorithms optimize real-time navigation based on sensor inputs (Kendall et al. 2019). Such hybrid architectures have the potential to improve trust and interpretability in CPS. For instance, in smart manufacturing, one can think of an LLM that interprets operational logs and generate human-readable explanations for anomalies, while a rule-based system would enforce predefined safety constraints. Additionally, neurosymbolic AI–a fusion of neural networks and symbolic reasoning–have the potential to provide structured contextual reasoning that enhances the transparency of AI-driven CPS (Garnelo and Shanahan 2019).

Real-Time Edge Processing. Another requirement for the successful deployment of LLMs in CPS is real time edge processing. As discussed in Section , deploying LLMs in CPS presents challenges related to latency and computational overhead, particularly in real-time applications such as industrial automation and autonomous systems. To address this, real-time edge processing would allow to leverage lightweight LLM variants and model compression techniques such as quantization, pruning, and knowledge distillation (Han et al. 2015). These techniques enable LLMs to run efficiently on edge devices, ensuring fast response times while minimizing power consumption. For instance, in IoT-driven smart grids, edge-based LLMs would analyze localized sensor data and generate adaptive control policies to optimize energy distribution in real time. In healthcare CPS, wearable devices with compressed LLM models would provide personalized health insights without relying on cloud-based processing, ensuring low-latency feedback for users (Esteva et al. 2019).

Human-AI Collaboration. Integrating human-AI collaboration is crucial to ensure trust, transparency, and effective decision-making in context-aware CPS 2.0. LLM-driven CPS can enable natural language interaction, allowing human operators to query, interpret, and refine AI-generated recommendations (Amershi et al. 2019). This is particularly valuable in high-stakes environments such as aviation, autonomous driving, and healthcare, where human oversight remains essential. One approach is to incorporate explainable AI (XAI) techniques, where LLMs provide natural language justifications for their recommendations. For example, in cybersecurity CPS, an LLM-enhanced intrusion detection system can explain why certain network behaviors are flagged as anomalies, improving operator trust and enabling informed decision-making (Molnar 2020). Furthermore, human-in-the-loop AI mechanisms allow CPS to continuously learn from human feedback. For instance, in robot-assisted manufacturing, workers can interact with an LLM-powered system via voice or text commands to dynamically adjust robotic behavior based on real-time needs (Mosqueira-Rey et al. 2023; Wu et al. 2022).

Conclusion

Large Language Models present a paradigm shift in context awareness for cyber-physical systems, moving beyond rigid rule-based approaches toward adaptive, multimodal, and human-centric decision-making. While challenges exist in real-time performance, bias mitigation, and energy efficiency, continued research into LLM-CPS integration could unlock unprecedented levels of intelligence and adaptability in next-generation systems. We advocate for interdisciplinary research collaborations to address these challenges and pave the way for LLM-powered CPS architectures that redefine context-aware autonomy.

Acknowledgments

This work is supported in part by the NYUAD Center for Artificial Intelligence and Robotics, funded by Tamkeen under the NYUAD Research Institute Award CG010.

References

Abshari, D.; Fu, C.; and Sridhar, M. 2024. LLM-assisted Physical Invariant Extraction for Cyber-Physical Systems Anomaly Detection. *arXiv preprint arXiv:2411.10918*.

Alsamhi, S. H.; Kumar, S.; Hawbani, A.; Shvetsov, A. V.; Zhao, L.; and Guizani, M. 2024. Synergy of human-centered ai and cyber-physical-social systems for enhanced cognitive situation awareness: applications, challenges and opportunities. *Cognitive Computation*, 16(5): 2735–2755.

Amershi, S.; Weld, D.; Vorvoreanu, M.; Fourney, A.; Nushi, B.; Collisson, P.; Suh, J.; Iqbal, S.; Bennett, P. N.; Inkpen, K.; Teevan, J.; Kikin-Gil, R.; and Horvitz, E. 2019. Guidelines for Human-AI Interaction. In *CHI Conference on Human Factors in Computing Systems*, 1–13.

Asmat, M. N.; Khan, S. U. R.; Mashkoor, A.; and Inayat, I. 2023. A Context Ontology-Based Model to Mitigate Root Causes of Uncertainty in Cyber-Physical Systems. In *International Conference on Database and Expert Systems Applications*, 45–56. Springer.

Banerjee, A.; Maity, A.; Kamboj, P.; and Gupta, S. K. 2024. CPS-LLM: Large Language Model based Safe Usage Plan Generator for Human-in-the-Loop Human-in-the-Plant Cyber-Physical System. *arXiv preprint arXiv:2405.11458*.

Burgueño, L.; Keet, M.; Kienzle, J.; Michael, J.; and Babur, Ö. 2024. A Human Behavior Exploration Approach Using LLMs for Cyber-Physical Systems. In *Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems*, 578–586.

Choaib, M.; Garouani, M.; Bouneffa, M.; and Mohanna, Y. 2024. IoT Sensor Selection in Cyber-Physical Systems: Leveraging Large Language Models as Recommender Systems. In *10th International Conference on Control, Decision and Information Technologies (CoDIT)*, 2516–2519.

Chowdhery, A.; Narang, S.; Devlin, J.; Bosma, M.; Mishra, G.; Roberts, A.; Barham, P.; Chung, H. W.; Hou, Z.; Lester, B.; et al. 2022. PaLM: Scaling Language Modeling with Pathways. *arXiv preprint arXiv:2204.02311*.

Daun, M.; and Tenbergen, B. 2023. Context modeling for cyber-physical systems. *Journal of Software: Evolution and Process*, 35(7).

Dey, A. K. 2001. Understanding and Using Context. *Personal and Ubiquitous Computing*, 5(1): 4–7.

Enoasmo, V.; Featherstonehaugh, C.; Konstantinopoulos, X.; and Huntington, Z. 2025. Structural Embedding Projection for Contextual Large Language Model Inference. *arXiv preprint arXiv:2501.18826*.

Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; and Dean, J. 2019. A Guide to Deep Learning in Healthcare. *Nature Medicine*, 25(1): 24–29.

Garnelo, M.; and Shanahan, M. 2019. Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. *Current Opinion in Behavioral Sciences*, 29: 17–23.

Gaur, S.; Almeida, L.; Tovar, E.; and Reddy, R. 2019. Cap: Context-aware programming for cyber physical systems. In *24th IEEE International Conference on Emerging Technologies and Factory Automation*, 1009–1016.

Gunes, V.; Peter, S.; Givargis, T.; and Vahid, F. 2014. A survey on concepts, applications, and challenges in cyber-physical systems. *Journal of Computing and Information Systems*.

Han, S.; Pool, J.; Tran, J.; and Dally, W. J. 2015. Learning both weights and connections for efficient neural networks. In *NeurIPS*, volume 28, 1135–1143.

Horvath, I.; and Gerritsen, B. H. M. 2012. Cyber-physical systems: Concepts, technologies, and implementation principles. *Proceedings of TMCE*.

Hsieh, F.-S. 2022. A theoretical foundation for context-aware cyber-physical production systems. *Applied Sciences*, 12(10): 5129.

Ivanov, R.; Weimer, J.; and Lee, I. 2018. Towards context-aware cyber-physical systems. In *IEEE Workshop on Monitoring and Testing of Cyber-Physical Systems (MT-CPS)*, 10–11.

Jin, D.; Zhao, S.; Jin, Z.; Chen, X.; Wang, C.; Fang, Z.; and Xiao, H. 2024. An evaluation of requirements modeling for cyber-physical systems via llms. *arXiv preprint arXiv:2408.02450*.

Kendall, A.; Gal, Y.; and Cipolla, R. 2018. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. In *CVPR*, 7482–7491.

Kendall, A.; Hawke, J.; Janz, D.; Mazur, P.; Reda, D.; Allen, J.-M.; Lam, V.-D.; Bewley, A.; and Shah, A. 2019. Learning to drive in a day. In *ICRA*, 8248–8254.

Lalwani, H.; Saleh, M.; and Salam, H. 2025. A Study Companion for Productivity: Exploring the Role of a Social Robot for College Students with ADHD. In *20th ACM/IEEE International Conference on Human Robot Interaction (HRI)*, 1438–1442.

Liao, G.; Huang, X.; Xiong, N.; Wan, C.; and Mao, M. 2021. Softwarized attention-based context-aware group recommendation technology in event-based industrial cyber-physical systems. *IEEE Transactions on Industrial Informatics*, 17(10): 6894–6905.

Lu, C.-H. 2018. IoT-enabled adaptive context-aware and playful cyber-physical system for everyday energy savings. *IEEE Transactions on Human-Machine Systems*, 48(4): 380–391.

Maity, A.; Banerjee, A.; Lamrani, I.; and Gupta, S. K. 2024. Context aware model learning in cyber physical systems. In *7th International Conference on Industrial Cyber-Physical Systems*, 1–6.

Medjahed, H.; Istrate, D.; Boudy, J.; Baldinger, J.-L.; and Dorizzi, B. 2011. A pervasive multi-sensor data fusion for smart home healthcare monitoring. In *International conference on fuzzy systems*, 1466–1473.

Molnar, C. 2020. *Interpretable Machine Learning*. Leanpub.

Mortlock, T.; Malawade, A.; Tsujio, K.; and Al Faruque, M. 2024. CASTNet: A Context-Aware, Spatio-Temporal Dynamic Motion Prediction Ensemble for Autonomous Driving. *ACM Transactions on Cyber-Physical Systems*, 8(2): 1–20.

Mosqueira-Rey, E.; Hernández-Pereira, E.; Alonso-Ríos, D.; Bobes-Bascarán, J.; and Fernández-Leal, Á. 2023. Human-in-the-loop machine learning: a state of the art. *Artificial Intelligence Review*, 56: 3005–3054.

Nazi, Z. A.; and Peng, W. 2024. Large language models in healthcare and medical domain: A review. In *Informatics*, volume 11, 57.

Noor, Z.; Hina, S.; Hayat, F.; and Shah, G. A. 2023. An intelligent context-aware threat detection and response model for smart cyber-physical systems. *Internet of Things*, 23: 100843.

Noorani, M.; Puthanveettil, T. V.; Zoulkarni, A.; Mirenzi, J.; Grody, C. D.; and Baras, J. S. 2024. Multimodal Anomaly Detection for Autonomous Cyber-Physical Systems Empowering Real-World Evaluation. In *International Conference on Decision and Game Theory for Security*, 306–325.

OpenAI. 2023. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774*.

Radford, A.; Wu, J.; Child, R.; et al. 2019. Language Models are Unsupervised Multitask Learners. *OpenAI Blog*, 1(8).

Russell, S.; and Norvig, P. 2020. *Artificial Intelligence: A Modern Approach*. Pearson, 4th edition.

Sahlab, N.; Jazdi, N.; and Weyrich, M. 2021. An approach for context-aware cyber-physical automation systems. *IFAC-PapersOnLine*, 54(4): 171–176.

Salam, H.; Celiktutan, O.; Gunes, H.; and Chetouani, M. 2023. Automatic context-aware inference of engagement in hmi: A survey. *IEEE transactions on affective computing*.

Schirner, G.; Erdogmus, D.; Chowdhury, K.; and Padir, T. 2013. The future of human-in-the-loop cyber-physical systems. *Computer*, 46(1): 36–45.

Scholze, S.; and Barata, J. 2016. Context awareness for flexible manufacturing systems using cyber physical approaches. In *Technological Innovation for Cyber-Physical Systems: 7th IFIP WG 5.5/SOCOLNET Advanced Doctoral Conference on Computing, Electrical and Industrial Systems*, 107–115.

Sol, D. C.; Devidas, A. R.; Anjana, M.; and Ramesh, M. V. 2018. Design and implementation of context aware cyber physical system for sustainable smart building. In *International Conference on Smart Grid and Clean Energy Technologies*, 162–167.

Stoyanov, S.; Kumurdjieva, M.; Tabakova-Komsalova, V.; and Doukovska, L. 2023. Using LLMs in Cyber-Physical Systems for Agriculture-ZEMELA. In *International Conference on Big Data, Knowledge and Control Systems Engineering*, 1–6.

Sutton, R. S.; and Barto, A. G. 2018. *Reinforcement Learning: An Introduction*. MIT Press, 2nd edition.

Takatsuka, H.; Saiki, S.; Matsumoto, S.; and Nakamura, M. 2014. Design and implementation of rule-based framework for context-aware services with web services. In *16th International Conference on Information Integration and Web-based Applications & Services*, 233–242.

Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; Rodriguez, A.; Joulin, A.; Grave, E.; and Lample, G. 2023. LLaMA: Open and Efficient Foundation Language Models. *arXiv preprint arXiv:2302.13971*.

Tsimpoukelli, M.; Menick, J.; Cabi, S.; Eslami, S. M. A.; Vinyals, O.; and Hill, F. 2021. Multimodal Few-Shot Learning with Frozen Language Models. In *NeurIPS*.

Wasim, M.; Ahmed, I.; Ahmad, J.; and Hassan, M. M. 2021. A novel deep learning based automated academic activities recognition in cyber-physical systems. *IEEE Access*, 9: 63718–63728.

Wu, S.; Fei, H.; Qu, L.; Ji, W.; and Chua, T.-S. 2024. Next-gpt: Any-to-any multimodal llm. In *41st International Conference on Machine Learning*.

Wu, X.; Xiao, L.; Sun, Y.; Zhang, J.; Ma, T.; and He, L. 2022. A Survey of Human-in-the-loop for Machine Learning. *Future Generation Computer Systems*.

Xiong, G.; Zhu, F.; Liu, X.; Dong, X.; Huang, W.; Chen, S.; and Zhao, K. 2015. Cyber-physical-social system in intelligent transportation. *IEEE/CAA Journal of Automatica Sinica*, 2(3): 320–333.

Xu, H.; Han, L.; Yang, Q.; Li, M.; and Srivastava, M. 2024. Penetrative ai: Making llms comprehend the physical world. In *25th International Workshop on Mobile Computing Systems and Applications*, 1–7.

Zhang, L.; Hang, L.; Zu, K.; and Wang, Y. 2024. A Smart Contract-Based Algorithm for Offline UAV Task Collaboration: A New Solution for Managing Communication Interruptions. *Preprints*.

Zhang, Y.; Qiu, M.; Tsai, C.-W.; Hassan, M. M.; and Alamri, A. 2015. Health-CPS: Healthcare cyber-physical system assisted by cloud and big data. *IEEE Systems Journal*, 11(1): 88–95.

Zhao, H.; Chen, H.; Yang, F.; Liu, N.; Deng, H.; Cai, H.; Wang, S.; Yin, D.; and Du, M. 2024. Explainability for large language models: A survey. *ACM Transactions on Intelligent Systems and Technology*, 15(2): 1–38.

Zhou, X.; Liu, M.; Yurtsever, E.; Zagar, B. L.; Zimmer, W.; Cao, H.; and Knoll, A. C. 2024. Vision language models in autonomous driving: A survey and outlook. *IEEE Transactions on Intelligent Vehicles*.