Crimson Publishers
Wings to the Research

Review Article

The Role of Continual Learning in the Cloud-Edge Continuum: A Review on Efficiency and Trustworthiness

Andrés L Suárez-Cetrulo*, Rajnish Rakholia, Miguel Aspis, Jaydeep Samanta, Cristian Bosch and Ricardo Simón Carbajo

Ireland's Centre for Artificial Intelligence (CeADAR), School of Computer Science, University College Dublin (UCD), Ireland

ISSN: 2694-4391

***Corresponding author:** Andrés L Suárez-Cetrulo, Ireland's Centre for Artificial Intelligence (CeADAR), School of Computer Science, University College Dublin (UCD), Ireland

Submission: November 27, 2025
Published: January 09, 2026

Volume 4 - Issue 3

How to cite this article: Andrés L Suárez-Cetrulo*, Rajnish Rakholia, Miguel Aspis, Jaydeep Samanta, Cristian Bosch and Ricardo Simón Carbajo. The Role of Continual Learning in the Cloud-Edge Continuum: A Review on Efficiency and Trustworthiness. Int J Conf Proc. 4(3). ICP. 000586. 2026.
DOI: 10.31031/ICP.2026.04.000586

Abstract

The Cloud-Edge Continuum (CEC) represents a paradigm shift towards a heterogeneous, distributed computing landscape. This environment is characterized by massively dis-tributed data sources, dynamic network conditions, and fluctuating computational loads. Traditional Machine Learning (ML) models, trained offline in a centralized manner, are not suited for this reality. They fail to adapt to the constant stream of new data, making them vulnerable to concept drift. This leads to inevitable performance degradation, creates significant processing bottlenecks, and undermines core Trustworthy AI principles of robustness and reliability. This paper argues that Continual Learning (CL) is a critical and necessary paradigm for robust and efficient intelligence in the CEC. We review the relevance of CL, data stream learning, and integrated concept drift detection as the primary mechanisms for maintaining model robustness and resilience. CL, implemented through a combination of data-centric and model-centric compression and frugal AI techniques, is vital for achieving both the efficiency and trustworthiness demanded by next-generation applications operating in the CEC. These methodologies include iterative fine-tuning, model compression, knowledge distillation, and dynamic neural network growth. This adaptive intelligence is required not only for end-user applications but also for MetaOS-level orchestration across the Cloud–Edge Continuum. This paper concludes by presenting key findings that highlight the essential role of adaptive learning across the continuum and outlines future research directions aimed at enabling scalable, trustworthy, and resource-efficient continual learning for MetaOS-based orchestration and management.

Keywords: Continual learning; Cloud-Edge continuum; Concept drift; Trustworthy AI; Frugal AI; Data stream learning; Distillation; Edge AI

Abbreviations: CEC: Cloud-Edge Continuum; ML: Machine Learning; P2P: Peer-to-Peer; RAG: Retrieval-Augmented Generation; ICL: In-Context Learning; PEFT: Parameter-Efficient Fine-Tuning; MoE: Mixture of Experts; GNG: Growing Neural Gas; FL: Federated Learning; DRL: Distributed Reinforcement Learning; MetaOS: Meta-Operating System

Introduction

The CEC, spanning from centralized cloud data centers to re-source-constrained edge and IoT devices, forms the backbone of modern digital infra-structure. This CEC is not a static entity; it is a highly dynamic and heterogeneous environment. European-funded MetaOS projects, such as ICOS [1], NEMO, aerOS, NEPHELE, NebulOuS, and FLUIDOS [2] (Figure 1), explicitly aim to manage this complexity, addressing dynamic, unpredictable conditions and complex orchestration need [3]. These projects are exploring diverse orchestration models, ranging from centralized and hierarchical 'Agent/Controller' systems to fully decentralized 'Peer-to-Peer' (P2P) networks, adding a further layer of complexity to this management.

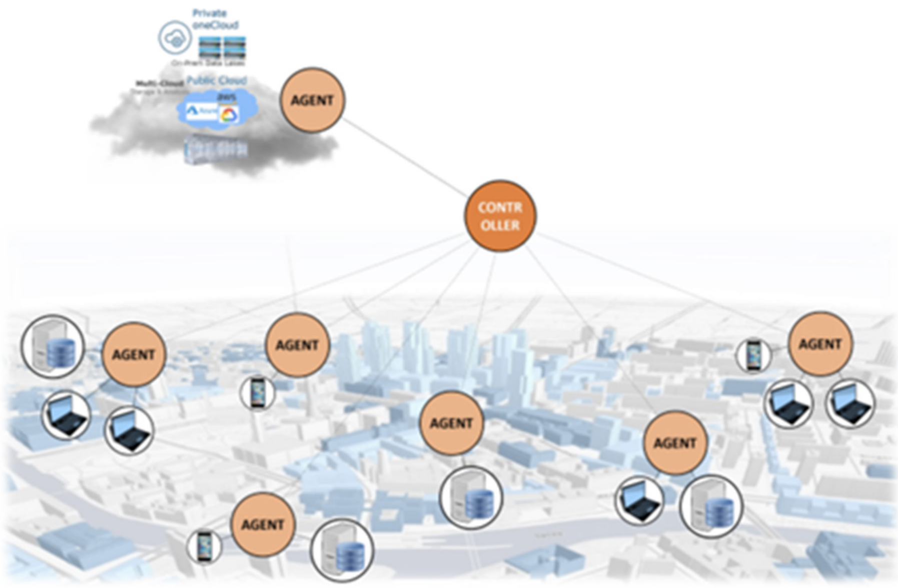

Figure 1: Illustration of an ICOS instance within the Cloud-Edge Continuum
Source: [1].

A primary challenge in the CEC is the nature of data itself. Data is no longer a static batch to be processed; it is an incremental, continuous stream, often generated at the extreme edge [4]. Processing this data on the go is essential for efficiency and low-latency decision-making. However, static ML models deployed in this continuum are inherently brittle. Their performance degrades as the statistical properties of the data streams change over time, a phenomenon known as concept drift [5]. This is a central challenge in re-al-world operations and changing environments, where detecting data drift situations where the incoming data starts to differ from the data the model was trained on is critical. Projects like MANOLO [6] are actively developing robust data inspection techniques, including noise detection and concept drift detection algorithms, to identify these shifts.

Apart from this, processing information in large batches and training AI models through iterative processes involving substantial amounts of data creates a significant bottleneck, not only affecting the efficiency of these models but also the reliability of AI-driven services. Static, non-adaptive models undermine many of the core Trustworthy AI principles [7], such as robustness and reliability, as their performance decays unpredictably over time.

The Figure 1 depicts a controller-centric MetaOS architecture in which a central controller coordinates multiple distributed agents deployed across cloud, edge, and IoT resources. This architecture highlights where intelligence and continual learning can be applied for adaptive orchestration under dynamic and heterogeneous conditions

This paper proposes that training models sequentially within the continual learning paradigm, combined with other frugal AI techniques, compression, and meta-learning, is the key to unlocking robust and efficient intelligence in the CEC. By learning incrementally and adapting to data streams in an online manner, CL provides a direct solution to the dual challenges of concept drift and resource inefficiency. This approach enables the development of frugal, adaptive models and resilient intelligent systems, aligning perfectly with the vision of a truly adaptive and trustworthy CEC.

Literature review methodology

This review follows a structured, narrative literature review methodology to identify and analyze relevant research on Continual Learning in the Cloud–Edge Continuum. Scholarly articles were primarily retrieved from established academic databases, including IEEE Xplore, Scopus, and arxiv, complemented by selective searches on Google Scholar to capture emerging and interdisciplinary work.

The review focuses on publications from 2017 to 2025, reflecting the period during which continual learning, edge AI, and cloud–edge architectures have matured as active research areas. Earlier foundational works were included selectively where necessary to provide conceptual grounding.

Studies were included if they addressed the following themes:

(i) continual or online learning, (ii) data stream learning and concept drift detection, (iii) frugal or re-source-efficient AI, (iv) learning in cloud-edge or distributed computing environments, or (v) trustworthiness aspects such as robustness, reliability, or adaptability. Works focusing exclusively on static, batch-trained models without relevance to dynamic or distributed settings were excluded.

The selected literature was analyzed thematically to synthesize current trends, extract key findings, and identify open challenges and future research directions relevant to efficiency and trustworthiness in the Cloud-Edge Continuum.

Continual Learning in the CEC

Addressing data streams and concept drift

Continual Learning is a methodology designed for scenarios where data is not available all at once, but instead arrives sequentially [8]. This aligns perfectly with the reality of IoT and edge devices, which generate data incrementally [4,9]. In the CEC, data streams are non-stationary by default. Network conditions fluctuate, user behavior changes, and sensors experience new environmental states. The field of data stream learning, related to CL as it focuses on online lifelong learning, has long focused on concept drift detection [5]. Concept drift occurs when the under-lying data distribution changes, causing the trained model's predictions to become inaccurate. In the context of the CEC, a drift (e.g., a new type of network threat or a change in factory-floor sensor readings [10]) that goes undetected can compromise the robustness and trustworthiness of the entire system. CL, particularly online learning variants, pro-vides the mechanisms to detect these drifts and adapt the model accordingly, ensuring that the AI remains robust and reliable (Figure 2).

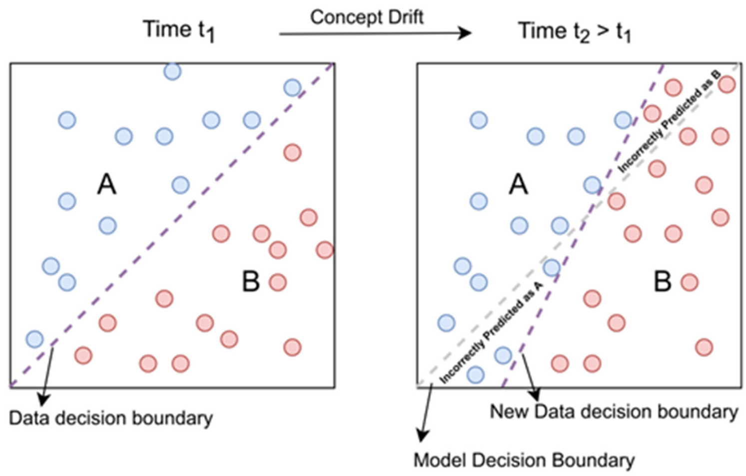

Figure 2: Illustration of concept drift in a data stream over time
Source: [5].

The Figure 2 shows how changes in the underlying data distribution between time t1t_1t1 and time t2t_2t2 shift the optimal decision boundary, leading to increasing mis-classification when a static model is used. This example illustrates real concept drift in non-stationary environments and motivates the need for continual learning and online model adaptation.

This challenge is a central concern for assessing Trustworthy AI. For instance, the MANOLO project explicitly identifies the need to detect data drift situations where the in-coming data starts to differ from the data the model was trained on as part of its performance assessment framework. CL, particularly online learning variants, provides the mechanisms to first detect these drifts, often through continuous monitoring and alerting tools, and then adapt the model accordingly. This adaptive loop, aligned with the MA-NOLO project's goals, ensures the AI remains robust, reliable, and resilient throughout its entire lifecycle in the CEC.

Current trends in continual learning

To address the dual challenges of plasticity (learning new information) and stability (avoiding catastrophic forgetting), the CL landscape offers a spectrum of solutions ranging from non-parametric context adaptations to sophisticated parametric updates. The lightest end of the spectrum lies In-Context Learning (ICL), where models adapt behavior based on examples provided in the prompt window. While computationally cheap, it suffers from context rot and is strictly limited by the context window size [11]. Retrieval-Augmented Generation (RAG) extends this by retrieving relevant information from an external buffer. While RAGs provide high capacity without altering model weights, they fail to compress knowledge, effectively deferring the learning problem to a storage and retrieval bottleneck that can be prohibitive for latency-sensitive edge applications.

We also have parametric methods, such as replay and regularization, as well as Parameter-Efficient Fine-Tuning (PEFT). While the classical parametric approach is fi-ne-tuning with replay, where the model is updated on new data mixed with a subset of old data, this approach is practical against forgetting. Still, it is fundamentally unscalable for lifelong learning at the edge, as the rehearsal buffer grows indefinitely, consuming scarce storage and compute resources. PEFT techniques, such as LoRA (Low-Rank Adaptation) and Adapters, offer a compelling middle ground [12]. By freezing the vast majority of the model's parameters and training only small, inserted modules, PEFT achieves targeted updates with minimal computational overhead. However, these methods have historically been viewed as having low capacity, struggling to integrate vast amounts of new knowledge over an extended lifecycle without saturation. Finally, a new research trend points towards Sparse Memory Architectures [13] and Mixture of Experts (MoE) [14]. Unlike monolithic models, these architectures route inputs to specific, sparse subsets of parameters (or "experts"). MoEs allow for high capacity by adding new experts over time, though they can still incur significant memory overhead [15]. Sparse Memory

Layers represent a novel “frugal” approach where specific Feed-Forward layers are replaced by massive, sparse key-value memory pools [14].

Efficiency and frugal adaptation

Beyond robustness, the CEC demands efficiency. It is computationally and financially infeasible to retrain massive models from scratch every time new data becomes available. CL offers a path to efficient adaptation. This can be achieved through several approaches. One is continual fine-tuning, where a model undergoes iterative adaptation to new domains or tasks with minimal computational overhead, effectively adapting to new contexts without catastrophic forgetting [16].

Another approach is the use of frugal models that are efficient by design. The MA-NOLO project [6,17], for instance, proposes developing “Trustworthy Efficient AI for Cloud-Edge Computing” by focusing on both model-centric and data-centric methods that can operate within the energy and computational constraints of the edge. Another approach is the use of frugal models that are efficient by design.

A key paradigm for achieving this rapid adaptation is meta-learning [18]. This approach utilizes a meta-learner model, trained on a range of learning tasks, to determine the optimal compression method or efficient adaptation strategy. This enables a base model to be updated for new tasks or domains (e.g., in “Domain Adaptive Few-Shot Continuous Learning”) with minimal additional data. This high-level learner, as illustrated in Figure 3, can orchestrate the update process, deciding how and when to replace or fi-ne-tune models at the edge to maintain peak performance and efficiency.

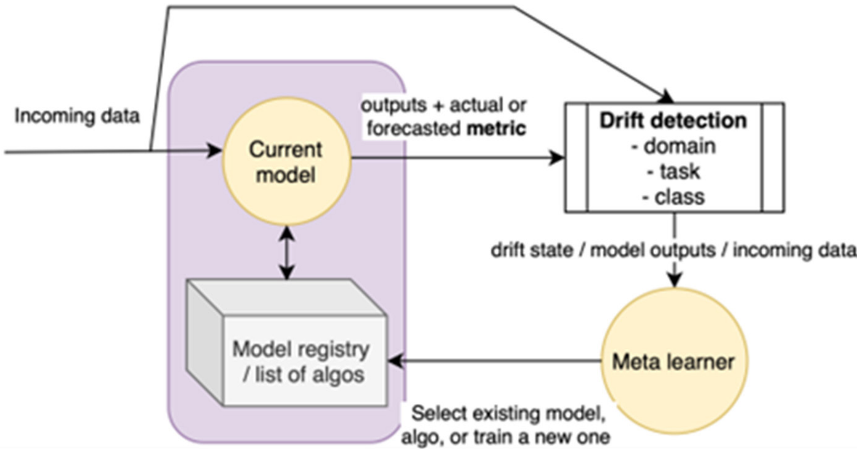

Figure 3: Meta-learning-driven model update and replacement process.

As illustrated in Figure 3, incoming data and model performance metrics are continuously monitored by drift detection mechanisms, whose outputs inform a meta-learner. Based on detected drift and resource constraints, the meta-learner selects an existing model, applies continual fine-tuning, or triggers the training of a new model, enabling efficient and adaptive learning in dynamic Cloud-Edge settings.

On the model-centric side, MANOLO investigates techniques to reduce the computational footprint of models. This includes model compression methods like pruning and quantization. Another technique explored is knowledge distillation, wherein a large, complex “teacher” model transfers its knowledge to a much smaller, computationally cheaper “student” model. This “student” model can then be deployed efficiently at the edge, retaining the performance of the larger model while meeting resource constraints. This focus on adaptation is also reflected in the project’s work on Domain Adaptive Few-Shot Continuous Learning, which provides a direct mechanism for efficient model updates.

On the data-centric side, MANOLO enhances efficiency by improving the quality and relevance of the data used for training. This is crucial for CL, as training on irrelevant or poor-quality data is a primary source of inefficiency. The project employs multiple data inspection techniques, including noise detection algorithms to filter data streams. Furthermore, it leverages drift detection algorithms (such as ADWIN [5] and Page Hinkley [19]) to identify statistical shifts in the incoming data. By detecting such changes, the system can trigger model adaptation intelligently and avoid wasting resources on retraining when it is not necessary, or on data that is anomalous or no longer relevant [15].

Beyond just inspecting and filtering data, data-centric efficiency can be achieved by actively compressing the data itself. Techniques such as dataset distillation [20] focus on synthesizing a small, highly representative dataset from a large, raw data stream. This compact, synthetic dataset embeds the core information of the original, enabling models to be trained or updated with a fraction of the computational cost and memory footprint. These compression and distillation techniques stand alongside other dynamic approaches, such as the use of growing neural networks. For instance, models like the Growing Neural Gas (GNG) are designed to learn online and can grow their network structure (neural network growth) incrementally as they encounter new data patterns [21]. This allows the model to adapt its complexity directly to the data stream, representing an-other highly efficient paradigm for stream-based learning in the CEC.

CL at the CEC in Practice: MetaOS and Orchestration

The principles of CL are not only applicable to application models but are also increasingly critical to the management of the continuum itself. This is a core theme across next-generation MetaOS projects, which are building intelligence directly into the system's orchestration layer to manage the inherent dynamism and complexity of the CEC. For in-stance, projects like ICOS are developing an Intelligent MetaOS that uses a dedicated Intelligence Layer for AI-driven optimization, resource management, and scheduling [1].

The architectural approaches to this orchestration vary significantly, which in turn dictates how CL must be implemented. Many projects, including ICOS, NebulOuS, and NEPHELE, adopt a resource-sharing model based on an "Agent/Controller" design, which is leveraged by MANOLO using a Controller/Node design. This architecture, whether implemented as a distributed model (like ICOS and aerOS) or a Hierarchical one (like NEPHELE and NebulOuS), creates logical points of intelligence where AI-driven decisions are concentrated. In the case of ICOS, this is referred to as the Intelligence Layer [1].

In contrast, projects like FLUIDOS and NEMO are architected with a decentralized orchestration, enabling a P2P resource-sharing model. This fundamental design choice avoids a central controller in favor of distributed, autonomous agents. This architectural split has profound implications for adaptation: the centralized/hierarchical models concentrate the need for CL for MetaOS support in their controller (although CL may still need to occur at local nodes at the application level), while the decentralized models must dis-tribute the MetaOS-support-related continual learning process itself. For example, NEMO explicitly explores parallel learning mechanisms, such as gossip learning [13], where agents continually learn and share updates with their peers without a central coordinator.

The orchestration layer (at the controller in distributed projects) must be adaptive [22]. The ICOS Intelligence Layer, for example, uses AI models to predict resource utilization (e.g., CPU) to make proactive scheduling decisions. However, in a dynamic environment characterized by fluctuating workloads, network states, and resource availability, these predictive models are highly susceptible to concept drift [3,22]. A model trained on yesterday's network traffic patterns will likely fail today, resulting in suboptimal and inefficient resource allocation. Therefore, the orchestrator's own AI models must be continual learners, constantly updating themselves from the stream of system telemetry.

This is reflected in the goals of several MetaOS projects [2]. FLUIDOS explicitly plans to consolidate online learning capabilities and leverage feedback from telemetry services to refine orchestration, which is a textbook implementation of a CL loop for system management. Similarly, ICOS is working to extend its AI-driven scheduling and Federated Learning (FL) integration. FL, in this context, provides a powerful mechanism for the orchestrator to continually learn from distributed nodes without centralizing sensitive telemetry data, enhancing both efficiency and privacy [9] (Figure 4).

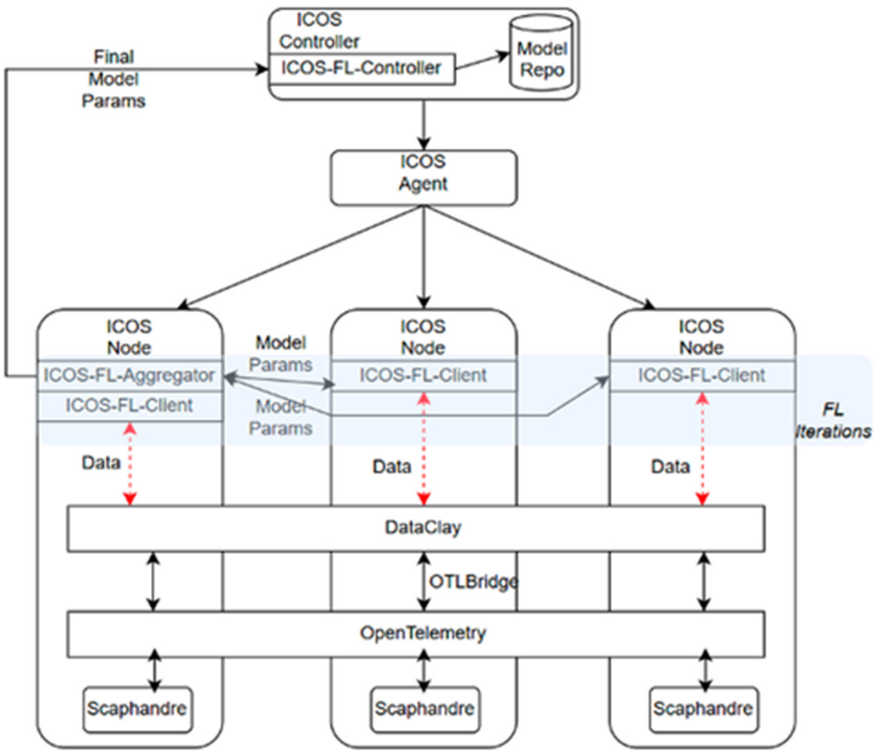

Figure 4: Federated Learning architecture in ICOS.
Source: [1].

A prime example of this adaptive orchestration is in intelligent task placement [22]. This is a complex problem that must adapt to real-time network delays and node failures. Research in this area is already applying adaptive learning techniques, such as Distributed Reinforcement Learning (DRL), to create policies that learn online to minimize delay and optimize task offloading [23]. This approach is being actively researched in projects like NEMO, which aims to use advanced RL algorithms for workload orchestration. This DRL-based orchestrator is, in effect, a continual learning agent that constantly fine-tunes its strategy based on new data about network performance.

Ultimately, the principles of trustworthy AI, as outlined in the MANOLO project [6], also apply in this context. Just as MANOLO calls for continuous monitoring and alerting for application models, a trustworthy MetaOS must do the same for its own internal models. By applying CL to the orchestration layer itself, the system becomes resilient, self-adapting, and capable of maintaining efficiency and trustworthiness over its entire lifecycle [17].

Challenges and Trends to Solve Them in the Cloud-Edge Continuum

Fundamental conflicts between the static nature of traditional AI development and the dynamic reality of edge operations obstruct the realization of a truly intelligent CEC. This section summaries the primary challenges identified across the previous sections, along with the specific solutions already mentioned proposed to address them.

Non-stationary data and drifts

In the CEC, data is generated as an infinite, high-velocity stream rather than being stored in a static repository. Environmental conditions, user behaviors, and network states are in constant flux. Traditional models, trained offline on historical datasets, operate on the assumption that future data will resemble past data. This assumption is frequently violated in the continuum, leading to concept drift, a phenomenon in which the statistical properties of the target variable change over time, resulting in a silent but catastrophic degradation in model accuracy [5]. To counter this, systems must transition from static deployment to online learning, with integrated drift detection, and trigger immediate adaptation mechanisms, as discussed earlier. This ensures that the model's knowledge re-mains aligned with the current state of the environment, maintaining robustness without requiring human intervention.

Resource constraints and training inefficiency

Retraining massive deep learning models from scratch is computationally prohibitive and energy-intensive, rendering it infeasible for resource-constrained edge devices. Furthermore, training on raw data streams often involves processing vast amounts of redundant, noisy, or irrelevant information, which creates significant processing bottlenecks and wastes energy. The solution lies in Frugal AI and Data-Centric Efficiency.

a. Firstly, the reviewed techniques, such as continual fine-tuning and meta-learning, enable models to adapt to new tasks with only a few examples, thereby drastically reducing the computational requirements for updates. Additionally, growing neural networks offer a structural solution, where the model architecture itself expands or contracts based on the complexity of the incoming data stream.

b. In the latter, efficiency may be further enhanced by data distillation and noise filtering. By algorithmically selecting only the most representative and high-quality data points for training, the system reduces the computational burden of the learning process itself.

Orchestration complexity and heterogeneity

The CEC is not a monolithic entity; it is composed of diverse architectures ranging from centralized clouds to decentralized swarms. Orchestrating workloads across this heterogeneity requires predicting resource availability (CPU, bandwidth) in real-time. However, static orchestration rules fail when network dynamics shift, and centralized control mechanisms struggle to scale to the sheer number of edge nodes found in P2P environments, such as those explored by FLUIDOS and NEMO. Thus, a MetaOS should adopt Adaptive Intelligence (e.g., AI-driven Adaptive Orchestration).

This is visible in projects such as ICOS and NEPHELE, where an Intelligence Layer continually updates workload prediction models, ensuring accurate task placement de-spite changing network conditions [1]. Here, FL allows nodes to collaboratively train a global model without sharing raw data, thereby preserving privacy.

Key Findings

The review of the state-of-the-art in the Cloud-Edge Continuum (CEC), alongside the analysis of current Meta-Operating System (MetaOS) projects like ICOS, FLUIDOS, and MANOLO, reveals a clear consensus: static intelligence is obsolete in this dynamic environment (See a summary in Figure 5). The following key findings summarize the critical necessity of CL and its supporting technologies for enabling a robust, efficient, and trust-worthy CEC.

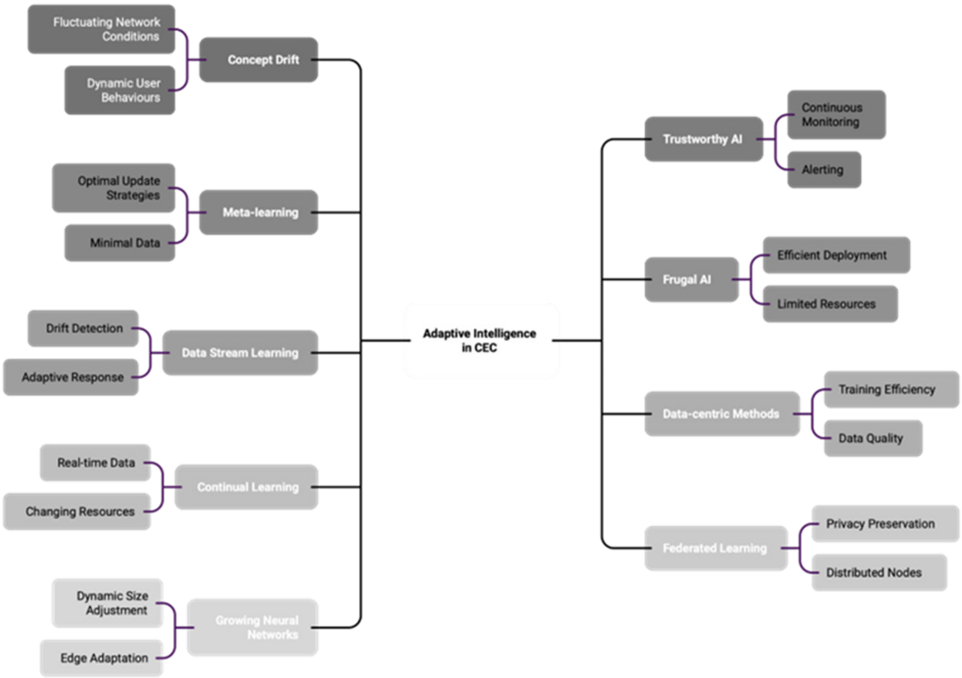

Figure 5: Adaptive Intelligence in the CEC. Challenges (intermediate nodes) and solutions (leaf nodes) leading to findings.

a. Traditional static ML models are insufficient for the CEC due to their inability to adapt to the continuous, non-stationary data stream, which leads to concept drift and degrades performance. Concept drift is an inherent and pervasive challenge in the CEC, driven by fluctuating network conditions, dynamic user behaviours, evolving sensor environments, and unpredictable workloads.

b. Trustworthy AI in CEC requires continuous monitoring, alerting, and adaptation loops, extending beyond models to the MetaOS itself to ensure long-term robustness and resilience.

c. Meta-learning offers an efficient mechanism for rapid adaptation, allowing models to autonomously select optimal update strategies for new tasks with minimal data.

d. Frugal AI techniques are critical for deploying efficient, lightweight AI at the edge, where resources are limited.

e. Data stream learning and drift detection form the backbone of adaptive intelligence in the CEC, enabling models to identify and respond to changes in the data quickly.

f. Data-centric methods significantly improve training efficiency, minimizing wasteful computation and reducing the burden of irrelevant or low-quality data.

g. CL is required for application-level models and MetaOS-level orchestration, which must adapt to real-time data and changing resource conditions.

h. An AI-driven orchestration layer (such as ICOS intelligent layer) relies on CL to maintain accuracy in predictive tasks, such as workload forecasting and task scheduling, which is highly vulnerable to drift.

i. FL can enhance continual adaptation in a privacy-preserving manner, enabling learning from distributed nodes without centralizing sensitive information.

j. Different orchestration designs create different CL needs, and decentralized systems may leverage other distributed learning methods, such as gossip learning.

k. Growing neural networks (such as Growing Neural Gas) can adjust their size over time, which allows models at the edge to adapt their structure as data changes continually.

l. CL is essential for maintaining the robustness and reliability of the AI system in the CEC, which supports the Trustworthy AI principle by enabling real-time adaptation.

These findings have profound implications for the design of future continuum plat-forms. They suggest that the current separation between system orchestration and AI ap-plication lifecycle must be bridged; the MetaOS of the future must be an inherent learning system, not just a static manager.

Future Research Directions

While the necessity of CL in the CEC is evident, several research avenues remain open to realize the vision of an autonomous, self-adaptive MetaOS fully. Future lines of work should focus on standardizing interfaces for drift-aware orchestration, allowing ap-plications to signal MetaOS when they require additional resources for retraining. Additionally, significant research is needed into decentralized continual learning protocols that can operate robustly in P2P architectures without converging to suboptimal states. The integration of Neuro-symbolic AI could also be explored to provide explainability alongside the adaptability of CL, further solidifying the trustworthiness of autonomous decision-making in the continuum. This is summarized in Figure 6 and covered in the following subsections.

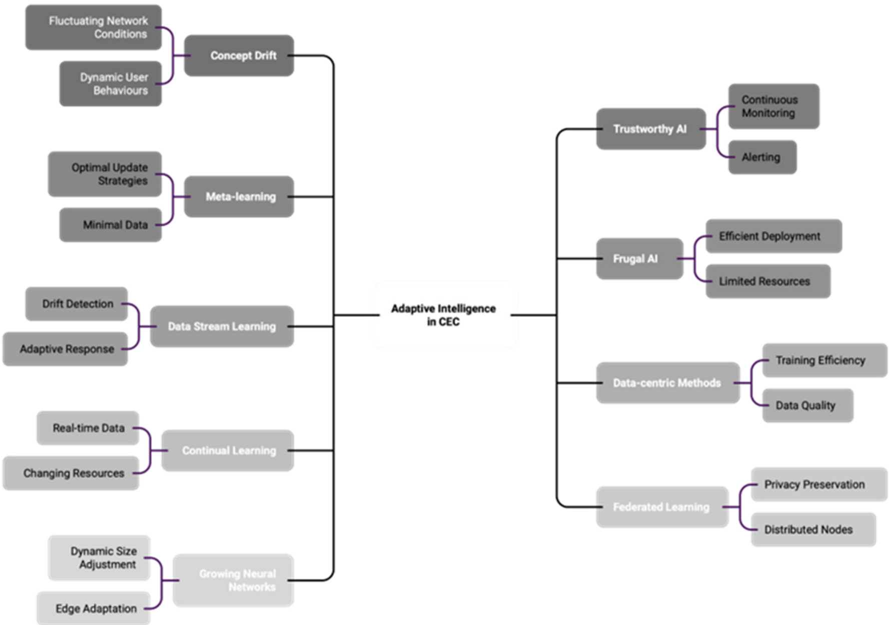

Figure 6: Future research directions.

Standardization of drift-aware interfaces

Current orchestration mechanisms operate primarily on resource metrics. There is a critical need to standardize "Drift-Aware Interfaces" that allow applications to signal their learning state to the MetaOS. Future work should define protocols that enable an application to report a concept drift event or a confidence drop as a standard telemetry metric. This would allow an Intelligence Layer or orchestrators to distinguish between a simple resource spike and a fundamental need for model retraining, triggering specific workflows rather than generic scaling.

From FL to fully decentralized learning

While FL is a step forward, it often retains a dependency on a central aggregator, which contradicts the fully decentralized P2P nature of architectures like FLUIDOS and NEMO. Future research should focus on alternative learning mechanisms and swarm intelligence protocols. These approaches enable edge nodes to exchange model updates asynchronously with their neighbors, eliminating the need for a central coordinator. Key challenges here include ensuring model convergence in unstable networks and preventing catastrophic forgetting when learning from independent and identically distributed data across heterogeneous peers.

Neuro-symbolic AI for explainable adaptation

In CEC environments, where an orchestrator transfers a critical safety task from the Edge to the Cloud, operators need to understand the rationale behind this decision. Future work should explore neuro-symbolic AI, combining the adaptive power of neural net-works with the transparency of symbolic logic rules. This aligns with the Trustworthy AI principles emphasized in MANOLO [6], ensuring that the system's adaptive decisions are not only efficient but also verifiable and explainable to human operators.

Environmental-aware carbon-aware continual learning efficiency

While CL avoids full retraining, iterative updates still consume energy. Future re-search must develop carbon-aware CL metrics that weigh the accuracy gain of a model update against its energy cost. The orchestration layer should be capable of delaying a learning update if the energy grid is currently dirty or if the expected performance boost does not justify the power consumption, effectively implementing a "Green Intelligence" policy across the continuum.

Cross-MetaOS knowledge transfer

As different MetaOS projects mature, the continuum will likely become a cluster of clusters. A major future challenge is enabling knowledge transfer between these heterogeneous systems. A model trained on network anomalies in a NEPHELE-managed smart factory should ideally be transferable to an ICOS-managed logistics hub. Research into standardized model representations (such as ONNX for CL) and teacher-student bridges between different MetaOS architectures will be essential to prevent knowledge silos [24].

Conclusion

The CEC is a dynamic, stream-based environment that invalidates the assumptions of static, batch-trained artificial intelligence. The non-stationary data streams and changing environments inherent to the CEC mean that static models may likely fail. This failure manifests as processing bottlenecks, performance degradation due to concept drift, and a critical erosion of Trustworthy AI principles, particularly robustness and reliability. To prevent this, a paradigm shift is necessary.

CL, in its various forms, is not merely an academic exercise but a vital operational necessity. This ranges from data-centric approaches, such as the online learning and concept drift detection explored in the MANOLO project, to model-centric frugal AI techniques. These include efficient continual fine-tuning [16], model compression, knowledge distillation, and frugal, growing models [21] that can adapt their complexity on the fly.

Furthermore, these principles are fundamental to the intelligent MetaOS platforms being developed to manage the complexity of the continuum. As platforms like ICOS mature with their Intelligence Layer, or as FLUIDOS and NEMO consolidate online learning capabilities for orchestration, their internal AI-driven components must also be continual learners. They explore this through incremental, online learning and reinforcement learning paradigms to avoid becoming obsolete. Ultimately, CL is the key to delivering on the promise of the CEC and the stated goals of projects like MANOLO and ICOS: a system that is simultaneously efficient, autonomous, and trustworthy.

The key findings of this study reinforce that continual, lightweight, and distributed adaptation is fundamental to sustaining performance across both edge-level models and MetaOS orchestration mechanisms. Looking ahead, future research should focus on developing more interoperable drift-aware mechanisms, progressing toward fully decentralized learning architectures, enabling more interpretable adaptive models, improving energy-efficient learning strategies, and supporting knowledge exchange across heterogeneous MetaOS ecosystems. Collectively, these directions point toward the next generation of scalable, transparent, and sustainable continual learning frameworks for the Cloud-Edge Continuum.

Author Contributions

Writing-original draft preparation, Methodology, Supervision-A.L.S.-C..; Writing-Data curation-R.R.; Data curation, Investigation-M.A.; Investigation, Visualization-J.S.; Review & Editing-C.B.; Supervision, Project administration, Funding acquisition-R.S.C. All authors have read and agreed to the published version of the manuscript.

Funding

This work was partially supported by the project "Trustworthy Efficient AI for Cloud-Edge Computing (MANOLO)", funded by the European Commission under Project Code/Grant Number 101135782 through the HORIZON EU Program.

Miguel Aspis and Andrés L. Suárez-Cetrulo were additionally supported by the project "Open Cloud Edge IoT Platform Uptake in Large Scale Cross-Domain Pilots (O-CEI)", funded by the European Commission under Project Code/Grant Number 101189589 through the HORIZON EU Program.

Data Availability Statement

This article is a review and does not involve the generation or analysis of new datasets. All data discussed in this work are derived from previously published studies, which are cited appropriately within the manuscript. Therefore, no new data are available.

Acknowledgment

The authors extend their sincere gratitude to their partners in the MANOLO and ICOS projects for their valuable insights and support, which helped inspire and shape the motivation behind this research.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Garcia J, Masip-Bruin X, Giannopoulos A, Trakadas P, Cajas Ordoñez SA, et al. (2025) ICOS: An intelligent MetaOS for the continuum. Proceedings of the 2nd International Workshop on MetaOS for the Cloud-Edge-IoT Continuum, Rotterdam, Netherlands, pp. 53-59.

2. Segou O, Skias DS, Velivassaki TH, Zahariadis T, Pages E, et al. (2024) NExt generation Meta Operating systems (NEMO) and Data Space: Envisioning the future. Proceedings of the 4th Eclipse Security, AI, Architecture and Modelling Conference on Data Space, Mainz, Germany, pp. 41-49.

3. Cajas Ordóñez SA, Samanta J, Suárez-Cetrulo AL, Simón-Carbajo R (2025) Intelligent edge computing and machine learning: A survey of optimization and applications. Future Internet 17(9): 417.

4. Lourenço A, Rodrigo J, Gama J, Marreiros G (2025) On-device edge learning for IoT data streams: A survey. arXiv Preprint.

5. Suárez-Cetrulo AL, Quintana D, Cervantes A (2023) A survey on machine learning for recurring concept drifting data streams. Expert Systems with Applications 213(Part A): 118934.

6. Karanxha G (2025) Evaluating transparency in artificial intelligence systems: Adapting the Z-Inspection® framework for the MANOLO Project.

7. Hickman E, Petrin M (2021) Trustworthy AI and corporate governance: The EU's ethics guidelines for trustworthy artificial intelligence from a company law perspective. European Business Organization Law Review 22(4): 593-625.

8. Wickramasinghe B, Saha G, Roy K (2023) Continual learning: A review of techniques, challenges, and future directions. IEEE Transactions on Artificial Intelligence 5(6): 2526-2546.

9. Wang Z, Wu F, Yu F, Zhou Y, Hu J, Min G (2024) Federated continual learning for edge-AI: A comprehensive survey. arXiv Preprint.

10. Rakholia R, Suárez-Cetrulo AL, Singh M, Simón-Carbajo R (2024) Advancing manufacturing through artificial intelligence: Current landscape, perspectives, best practices, challenges and future direction. IEEE Access 12: 131621-131637.

11. Hong K, Troynikov A, Huber J (2025) Context rot: How increasing input

tokens impacts LLM performance.

12. He XO (2024) Mixture of a million experts. arXiv Preprint.
13. Hegedűs I, Danner G, Jelasity M (2021) Decentralized learning works: An empirical comparison of gossip learning and federated learning. Journal of Parallel and Distributed Computing 148: 109-124.
14. Lin J, Zettlemoyer L, Ghosh G, Yih WT, Markosyan A, et al. (2025) Continual learning via sparse memory finetuning. arXiv Preprint.
15. Aspis M, Cajas Ordóñez SA, Suárez-Cetrulo AL, Simón-Carbajo R (2025) DriftMoE: A mixture of experts approach to handle concept drifts. arXiv Preprint.
16. Gunasekara N, Pfahringer B, Gomes HM, Bifet A (2023) Survey on online streaming continual learning. Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, Macao, SAR, pp. 6628-6637.
17. Arazo E, Stoev H, Bosch C, Suárez-Cetrulo AL, Simón-Carbajo R (2024) Xpression: A unifying metric to optimize compression and explainability robustness of AI models. World Conference on Explainable Artificial Intelligence, Springer Nature, Switzerland, pp. 370-382.
18. Anderson R, Koh YS, Dobbie G, Bifet A (2019) Recurring concept meta-learning for evolving data streams. Expert Systems with Applications 138: 112832.
19. Sebastião R, Fernandes JM (2017) Supporting the page-Hinkley test with empirical mode decomposition for change detection. International Symposium on Methodologies for Intelligent Systems, Springer Nature, Switzerland, pp. 492-498.
20. Yu R, Liu S, Wang X (2023) Dataset distillation: A comprehensive review. IEEE Transactions on Pattern Analysis and Machine Intelligence 46(1): 150-170.
21. Suárez-Cetrulo AL, Cervantes A (2017) An online classification algorithm for large scale data streams: iGNGSVM. Neurocomputing 262: 67-76.
22. Cajas Ordóñez SA, Samanta J, Suárez-Cetrulo AL, Simón-Carbajo R (2024) Adaptive machine learning for resource-constrained environments. International Workshop on Discovering Drift Phenomena in Evolving Landscapes, Springer Nature, Switzerland, pp. 3-19.
23. Giannopoulos A, Suárez-Cetrulo AL, Masip-Bruin X, D’Andria F, Trakadas P (2024) Placing computational tasks within edge-cloud continuum: A DRL delay minimization scheme. European Conference on Parallel Processing, Springer Nature, Switzerland, pp. 36-45
24. Biderman D, Portes J, Ortiz JJG, Paul M, Greengard P, et al. (2024) Lora learns less and forgets less. arXiv Preprint.

