INTEND: Human-Like Intelligence for Intent-Based Data Operations in the Cognitive Computing Continuum

Rustem Dautov[1,*], Hui Song[1], Dumitru Roman[1], Erik Johannes Husom[1], Sagar Sen[1], Vilija Balionyte-Merle[1], Donatella Firmani[2], Francesco Leotta[2], Jerin George Mathew[2], Jacopo Rossi[2], Lorenzo Balzotti[2], Andrea Morichetta[3], Schahram Dustdar[3], Thijs Metsch[4], Valerio Frascolla[4], Ahmed Khalid[5], Giada Landi[6], Juan Brenes[6], Ioan Toma[7], Róbert Szabó[8], Christian Schaefer[9], Seonghyun Kim[9], Cosmin Udroiu[10], Alexandre Ulisses[11], Verena Pietsch[12], Sigmund Akselsen[13], Arne Munch-Ellingsen[13], Irena Pavlova[14], Dessisslava Petrova-Antonova[14], Hong-Gee Kim[15], Changsoo Kim[16], Bob Allen[16], Sunwoo Kim[17] and Eberechukwu Paulson[17]

[1]*SINTEF Digital, Norway*
[2]*Sapienza Università di Roma, Italy*
[3]*TU Wien, Austria*
[4]*Intel Deutschland GmbH, Germany*
[5]*Dell Technologies, Ireland*
[6]*Nextworks, Italy*
[7]*Onlim GmbH, Austria*
[8]*Ericsson Research, Hungary*
[9]*Ericsson AB, Sweden*
[10]*CS Group, Romania*
[11]*MOG Technologies, Portugal*
[12]*FILL GmbH, Austria*
[13]*Telenor ASA, Norway*
[14]*GATE Institute, Sofia University, Bulgaria*
[15]*Seoul National University, South Korea*
[16]*AiM Future, South Korea*
[17]*Hanyang University, South Korea*

Abstract

This paper outlines a research roadmap of the INTEND project towards the development of a cognitive computing continuum that leverages human-like intelligence for intent-based data operations. The primary objective of this initiative is to create a system that can adapt dynamically to varying contexts by understanding and acting upon the intents of stakeholders in a decentralised and strategic manner. The project focuses on three main research pillars: resource management through intelligent agents, decentralised decision-making inspired by human cognitive processes, and enhancing human-AI interaction using Generative AI technologies.

Keywords

Cognitive Computing, Intent-Based Data Operation, Machine Learning, Federated Systems, AI Coordination

1. Introduction and Motivation

The increasing complexity of data operations in a computing continuum poses significant challenges due to the need for dynamic adaptation to diverse and fluctuating contexts. This complexity arises from managing vast and heterogeneous data sources, orchestrating resources efficiently, and ensuring optimal performance across distributed environments. Traditional, static approaches are inadequate as they fail to accommodate the rapid changes in data volume, variety, and velocity. Additionally, the integration of various stakeholder requirements and the need for real-time decision-making further complicate the landscape. The decentralised nature of modern computing systems demands sophisticated coordination mechanisms that can reconcile local autonomy with global strategy. Addressing these challenges requires advanced machine learning (ML) algorithms, continual learning pipelines, and seamless human-AI interactions to create a flexible and adaptive system capable of managing the intricate dynamics of data operations effectively [1]. In other words, this requires enhancing existing cloud-edge environments with advanced human-like intelligence and cognition capabilities. Such a cognitive computing continuum would be able to tackle several key challenges. These include learning to utilise and adapt the diverse and complex hardware within the continuum, managing the distribution and dynamic nature of resources, and bridging the cognitive gap between human stakeholders and data operation machines, ensuring mutual understanding and trust despite differing communication media.

To this end, the EU-funded collaborative project INTEND[1] aims to build next-generation cognitive computing continuum systems by leveraging human-like intelligence to interpret and act upon human intents in a decentralised manner [2]. The overall goal of this research initiative underpinned by three research pillars is to create a sophisticated framework that integrates adaptive resource management, decentralised decision-making, and enhanced human-AI interaction to streamline data operations in diverse domains such as digital manufacturing, telecommunications, smart cities, robotics systems, and video streaming. This paper presents a high-level overview of the research roadmap towards creating such a cognitive computing continuum with advanced human-like intelligence to achieve the novel **intent-based data operation**[2] in the continuum.

2. INTEND Project in a Nutshell

The cornerstone of the proposed research initiative is **intent-based data operation**, which brings the computing continuum automation to a whole a new level by allowing data stakeholders (e.g., data scientists, engineers, owners, consumers, end-users) to collaborate and interact with the cognitive continuum via **shared intents** [4]. In simple terms, intents capture the stakeholders' expectations in

RuleML+RR'24: Companion Proceedings of the 8th International Joint Conference on Rules and Reasoning, September 16–22, 2024, Bucharest, Romania

*Corresponding author.
rustem.dautov@sintef.no (R. Dautov); hui.song@sintef.no (H. Song); dumitru.roman@sintef.no (D. Roman); erik.johannes.husom@sintef.no (E. J. Husom); sagar.sen@sintef.no (S. Sen); vilija.balionyte-merle@sintef.no (V. Balionyte-Merle); donatella.firmani@uniroma1.it (D. Firmani); leotta@diag.uniroma1.it (F. Leotta); mathew@diag.uniroma1.it (J. G. Mathew); j.rossi@diag.uniroma1.it (J. Rossi); lorenzo.balzotti@uniroma1.it (L. Balzotti); andrea.morichetta@tuwien.ac.at (A. Morichetta); dustdar@dsg.tuwien.ac.at (S. Dustdar); thijs.metsch@intel.com (T. Metsch); valerio.frascolla@intel.com (V. Frascolla); ahmed.khalid@dell.com (A. Khalid); g.landi@nextworks.it (G. Landi); j.brenes@nextworks.it (J. Brenes); ioan.toma@onlim.com (I. Toma); robert.szabo@ericsson.com (R. Szabó); christian.schaefer@ericsson.com (C. Schaefer); kim.seonghyun@ericsson.com (S. Kim); cosmin@c-s.ro (C. Udroiu); alexandre.ulisses@mog-technologies.com (A. Ulisses); verena.pietsch@fill.co.at (V. Pietsch); sigmund.akselsen@telenor.com (S. Akselsen); arne.munch-ellingsen@telenor.com (A. Munch-Ellingsen); irena.pavlova@gate-ai.eu (I. Pavlova); dessislava.petrova@gate-ai.eu (D. Petrova-Antonova); hgkim@snu.ac.kr (H. Kim); changsoo.kim@aimfuture.ai (C. Kim); bob.allen@aimfuture.ai (B. Allen); remero@hanyang.ac.kr (S. Kim); enpaulson2@hanyang.ac.kr (E. Paulson)

[1] https://intendproject.eu/

[2] The concept of **intent** in an autonomous management framework is inspired by Intent-Based Networking [3] and refers to a declarative goal that describes the properties of a satisfactory outcome. This gives the framework the flexibility to explore various options within the solution space and find the optimal one.

terms of how their data pipelines are intended to perform; they are expressed using natural language, typically from a business perspective at a rather abstract level. These intents are then translated into actionable instructions by the computing continuum, which performs continuous resource management and dynamic adaptation of the data pipelines. Such translation from a natural language requires the continuum to become **cognitive** by mimicking the human brain operation, so that it can function in a way similar to a human administrator who possesses the domain knowledge required to perform efficient and effective data operation in the continuum. This way, such intent-based data operation allows the stakeholders to have full and scalable control of their data pipelines without directly handling the underlying computing and networking resources [5, 6]. Thus, the task of data operation evolves from resource-oriented to business-oriented.

This described vision requires the envisioned cognitive continuum to demonstrate advanced human-like intelligence, which is far from trivial to implement. The recent breakthroughs in AI research have shown unprecedented human-like intelligence to handle creative tasks such as drawing pictures, composing music, and writing articles [7], powered by a series of improvements, e.g., in the direction of Generative AI, neural-symbolic AI, and deep reinforcement learning. Similarly, such human-like intelligence has the potential to eventually disrupt the way people use the cloud-edge computing continuum. More specifically, by exploiting these AI advances, it is possible to create a cognitive computing continuum capable of **adapting**, **thinking** and **talking** like humans:

- **Adapt like humans**: continually learn how to use heterogeneous and unconventional resources in an effective way, to keep adapting data pipelines accordingly.
- **Think like humans**: make strategic decisions that coordinate and harmonise adaptation actions at different places in the continuum for different purposes, like the human brain takes decisions in a multi-objective, concurrent, and decentralised way.
- **Talk like humans**: communicate with various human stakeholders in natural language to understand how they intend their data pipelines to perform, and report back explaining what has been done according to the given intents.

2.1. Research Pillars of the INTEND Project

The described three human-inspired features of the envisioned cognitive system underpin the research pillars of the INTEND project.

- **Pillar I: Adaptive Resource Management**: Dynamically allocating resources across data pipelines while balancing various runtime contexts such as data orchestration, resource availability, performance, and energy consumption, is a pressing challenge. Traditional static methods are inadequate due to the rapid changes and complexities in data operations. Developing effective ML agents that can continually learn and adapt to new scenarios is crucial. These agents must make real-time decisions to optimise resource use without complete system visibility, ensuring local and global performance alignment and accommodating diverse stakeholder requirements and unpredictable environmental changes. To address this, INTEND will implement ML agents using both supervised learning from historical data and reinforcement learning from current behaviours. These agents will dynamically reconfigure resource allocations, ensuring optimal performance under changing conditions. A detailed understanding of possible adaptation actions, such as data block movement and bandwidth adjustments, will be developed. The appropriate ML algorithms will be selected, and continual learning pipelines will be established to enable agents to adapt to new data and scenarios.
- **Pillar II: Decentralised and Strategic Decision Making**: This involves creating a system where local resource management agents operate autonomously while coordinating globally. The agents are required to make independent decisions and share them for collective evaluation, ensuring strategic coherence. The difficulty lies in balancing local autonomy with global strategy, as agents must continuously learn and adapt without complete information. This coordination must effectively integrate diverse perspectives to form optimal strategies, addressing the complexities of distributed environments and ensuring robust, adaptive decision-making across the entire computing continuum. To address this, a novel coordination framework inspired by the Global Neuronal Workspace

Theory [8] will be created, where multiple AI models act as both resource management agents and decision coordinators. This decentralised system allows for partial information operation, with agents continuously learning and adapting without needing a complete system view. The federated coordinator will evaluate, revise, and merge local decisions to form globally optimal strategies, ensuring that local autonomy is maintained while achieving strategic coherence.

- **Pillar III: Enhanced Human-AI Interaction**: This involves creating seamless integration between human intuition and AI's computational power. It is required that the intelligent data operation tools built in INTEND can effectively interpret and act on human intentions, which requires sophisticated natural language processing, context awareness, and adaptability. The goal is to enable intuitive interactions where humans can express their goals and the AI can autonomously manage data operations to meet those objectives. This demands balancing complexity, accuracy, and user experience to foster effective collaboration and achieve meaningful outcomes in dynamic, intent-driven environments. The project aims to leverage Generative AI technologies to facilitate natural language interactions between stakeholders and self-adaptation agents. This technology will facilitate the expression of intents directly in natural language, simplifying the interaction process by removing the need for intermediate translation and predefined communication protocols, enabling more intuitive and flexible interactions. By addressing interoperability challenges through Generative AI's common-sense knowledge, the project aims to create a seamless interface for the stakeholders to express their intents directly in natural language, which the AI can comprehend and act upon. To address the limitations of Generative AI, domain-specific knowledge representations such as knowledge graphs will be integrated to improve decision-making in complex scenarios requiring detailed quantitative analysis.

2.2. Conceptual Architecture

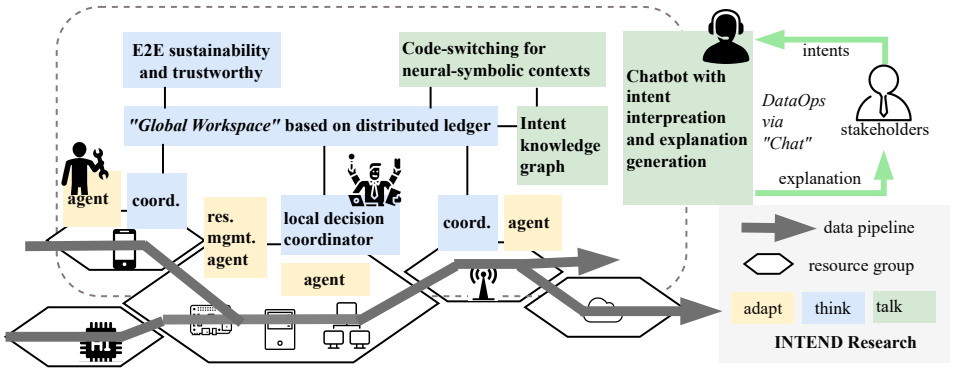

Figure 1: Conceptual architecture of the cognitive computing continuum envisioned by INTEND.

The described research pillars underpin the conceptual architecture of the envisioned cognitive continuum for intent-based data operations depicted in Fig. 1. The bold arrows at the bottom depict data pipelines deployed and running on a heterogeneous continuum composed of various resource groups (depicted as hexagons), such as central clouds, 5G infrastructures, on-premise data centres, edge gateways, etc. Human stakeholders interact with the cognitive continuum using a chatbot interface, both to communicate their intents and receive AI-generated explanations. The actual cognitive continuum is underpinned by the threefold human-like intelligence, i.e., adapt, think, and talk:

- **Adaptive Resource Management**: This is performed by intelligent resource management agents (depicted as yellow boxes). Each agent overlooks a particular adaptation problem (e.g., data placement, workload orchestration) within its resource group. Each resource group may have multiple agents looking at different adaptation problems, based on different AI approaches and models.
- **Decentralised and Strategic Decision Making**: This is handled by decentralised and federated decision coordination (depicted as blue boxes), which will compare and combine multiple adaptation suggestions proposed by different AI models aiming for a globally optimal decision. The global

workspace will be built using the distributed ledger technology. Apart from available hardware resources, the decision coordination will also consider such aspects as end-to-end data security, trustworthiness and sustainability in a distributed, dynamic and multi-vendor continuum.

- **Enhanced Human-AI Interaction**: This bridges the cognitive gap between the computing continuum and humans and is performed by a combinations of technologies (depicted as green boxes). The core element is a knowledge graph capturing data operation semantics and stakeholder intents. With the help of code-switching for neuro-symbolic AI, the common knowledge graph serves as a reference for the various AI-based decision makers and coordinators. Human-computer interaction via chatbots uses LLMs to translate stakeholder intents from natural language, and Generative AI to explain back how and why the AI-based agents and decision coordinators reached certain decisions.

3. Conclusion

Taken together, the described objectives, the research pillars and the conceptual architecture of the INTEND project outline a comprehensive roadmap for developing a cognitive computing continuum that enhances intent-based data operations with human-like intelligence. By integrating adaptive resource management, decentralised decision-making inspired by cognitive theories, and Generative AI for improved human-AI interaction, the INTEND project aims to create a robust and flexible framework for managing data pipelines in distributed computing environments. Supported by a collaborative EU-funded project, this initiative will result in open-source tools and a prototype platform, demonstrating the approach's effectiveness across several application domains.

Acknowledgments

This work has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101135576 (INTEND).

References

[1] A. Morichetta, V. C. Pujol, S. Dustdar, A roadmap on learning and reasoning for distributed computing continuum ecosystems, in: 2021 IEEE International Conference on Edge Computing (EDGE), IEEE, 2021, pp. 25–31.

[2] D. Firmani, F. Leotta, J. G. Mathew, J. Rossi, L. Balzotti, H. Song, D. Roman, R. Dautov, E. J. Husom, S. Sen, V. Balionyte-Merle, A. Morichetta, S. Dustdar, T. Metsch, V. Frascolla, A. Khalid, G. Landi, J. Brenes, I. Toma, R. Szabó, C. Schaefer, C. Udroiu, A. Ulisses, V. Pietsch, S. Akselsen, A. Munch-Ellingsen, I. Pavlova, H.-G. Kim, C. Kim, B. Allen, S. Kim, E. Paulson, INTEND: Intent-Based Data Operation in the Computing Continuum, in: CEUR Workshop Proceedings, volume 3692, 2024, pp. 43–50.

[3] P. H. Gomes, M. Buhrgard, J. Harmatos, S. K. Mohalik, D. Roeland, J. Niemöller, Intent-driven closed loops for autonomous networks, Journal of ICT Standardization 9 (2021) 257–290.

[4] M. Tomasello, M. Carpenter, J. Call, T. Behne, H. Moll, Understanding and sharing intentions: The origins of cultural cognition, Behavioral and brain sciences 28 (2005) 675–691.

[5] B. Oliveira, N. Ferry, H. Song, R. Dautov, A. Barišić, A. R. Da Rocha, Function-as-a-service for the cloud-to-thing continuum: a systematic mapping study, in: 8th International Conference on Internet of Things, Big Data and Security-IoTBDS, 2023, pp. 82–93.

[6] H. Song, R. Dautov, N. Ferry, A. Solberg, F. Fleurey, Model-based fleet deployment in the iot–edge–cloud continuum, Software and Systems Modeling 21 (2022) 1931–1956.

[7] J. Song, D. Yip, Exploring the Intersection of AI Art and Film: A Case Study of Giant, in: 2023 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), IEEE, 2023, pp. 347–352.

[8] G. A. Mashour, P. Roelfsema, J.-P. Changeux, S. Dehaene, Conscious processing and the global neuronal workspace hypothesis, Neuron 105 (2020) 776–798.