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.

To this end, the EU-funded collaborative project INTEND[1] aims to build next-generation computing continuum systems by leveraging human-like intelligence to interpret and act upon human intents in a decentralised manner. The overall goal of this research initiative underpinned by the three 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 video streaming, digital manufacturing, telecommunications, smart cities, and robotics systems. This paper presents a research roadmap towards creating such a cognitive computing continuum with advanced himan-like intelligence to achieve the novel *intent-based data operation*[2] in the continuum.

2. INTEND Project in a Nutshell

The main concept and objective of the proposed research roadmap is **intent-based data operation** in the cognitive continuum, which raises the automation of data operation into a new level, by allowing data stakeholders (e.g., data scientists, engineers, owners, consumers, end-users) to work together with the cognitive continuum via *shared intents* [2]. The data operation intent captures what the stakeholders expect (or intend) their data pipelines to perform, from a high abstraction level and a business perspective, in natural language, while a virtual data operator in the continuum overlooks the resource management and dynamic adaptation of the data pipelines, to fulfil the stakeholder's intents. Intent-based data operation allows data stakeholders to have full control of the data pipelines, without directly handling the resources, and therefore, it transfers data operation from resource-oriented to business-oriented. Intent-based data operation requires the cognitive continuum to work in a similar

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 [1] 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.

way as a human representative overlooking the automatic data operation, and therefore requires advanced human-like intelligence.

The envisioned intent-based data operation requires a cognitive computing continuum with advanced intelligence to meet the following key challenges:

- Heterogeneous and complex hardware in the continuum: The cognitive continuum needs to learn by itself how to utilise and adapt the many diverse resources in the continuum for efficient data processing.
- The distribution and dynamicity of the computing continuum means adaptation actions at different places, for different resource types, often interacting with each other, and coordinated for strategic overall goals.
- The cognitive gap between human stakeholders and the data operation machine: The two sides 'think' and 'work' in different ways and speak different languages, but they need to understand and trust each other.

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 [3], powered by a series of improvements, e.g., in the direction of generative AI, neural-symbolic AI, and deep reinforcement learning. Such human-like intelligence has the potential to eventually disrupt how people use the cloud-edge computing continuum. By exploiting these latest AI breakthroughs, it is possible to bring the next-level human-like intelligence to create a cognitive computing continuum, allowing the latter to adapt, think and talk 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**: Chat with various data stakeholders (real humans) in natural language to understand how they intend their data pipelines to perform, and explain to the stakeholders what was done according to the given intents.

2.1. Research Pillars of the INTEND Project

TODO: Some intro text

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 efficiently.

Pillar II: Decentralised and Strategic Decision Making This challenge 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 [4] 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 challenge 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 stakeholders' 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 where stakeholders can express 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.

3. Conceptual Architecture

The described research pillars underpin the conceptual architecture of the envisioned cognitive continuum for intent-based data operations depicted in Fig. 1. The bottom shows data pipelines (bold arrows) running on a continuum of diverse resources, from various resource groups (in hexagons), such as central cloud, 5G infrastructure, on-premises cluster, far-edge devices, etc. We will build a set of tools to be deployed into the continuum, which bring human-like intelligence in these three directions, and together realise intent-based data operation.

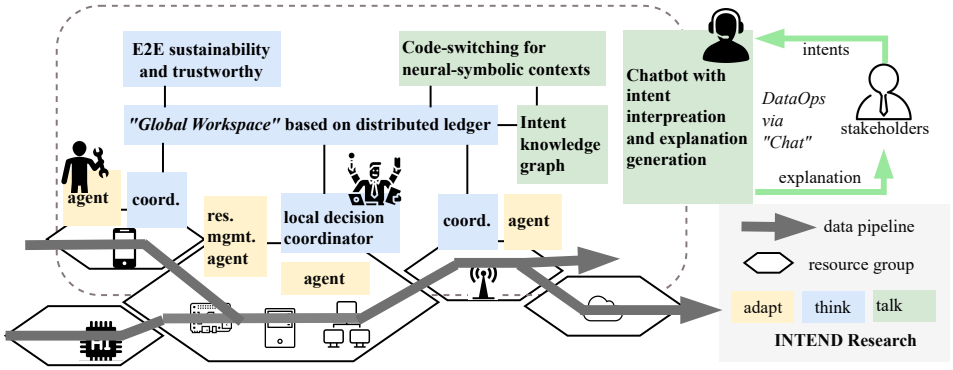

Figure 1: INTEND Conceptual Architecture.

Intelligent resource management (aka the "adapter"): We handle the hardware diversity by continual learning, i.e., to learn autonomously what is the best way to use the resources in the continuum, based

on observing how the data pipelines perform on the current and similar resources. The results will be a set of intelligent resource management agents (yellow blocks in the diagram). An agent overlooks a particular adaptation problem (e.g., data placement, workload orchestration) within a limited resource group. Each resource group may have multiple agents looking at different adaptation problems, based on different AI approaches and models.

Federated and strategic decision making (aka the "thinker"): We handle the distributed and dynamic nature of computing continuum by decentralised and federated decision coordination, to compare and combine the adaptation decisions made by different AI models, into globally optimal adaptation. The coordinator itself will be based on decentralized AI structures like federated learning and swarm learning, guided by the Global Workspace Theory [4] about how human brain make strategic decision in a multi-objective, concurrent and decentralised way. The decision coordination will be based on measures of end-to-end data security and sustainability in a distributed, dynamic and multi-vendor continuum.

Intent-based stakeholder interface (aka the "talker"). We will leverage neuro-symbolic AI, Large Language Model (LLM), and generative AI, to bridge the cognitive gap between AI and humans. The core of this part is a knowledge graph to represent the data operation semantics and the stakeholder intents. With the help of code-switching, the common knowledge graph serves as reference to the various AI-based decision makers and coordinators. For human interaction, we will use LLM to extract stakeholder intents from artefacts and natural language, and use generative AI to explain in natural language how and why the AI-based agents and decision coordinators make certain decision. The extraction and generation will be integrated into a chatbot for data operation.

4. 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. 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] 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.

[2] 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.

[3] 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.

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