Steps towards a trustworthy and reliable human-machine symbiotic collaboration

Giorgos Flouris (Foundation for Research and Technology – Hellas, FORTH, Greece)

Satoshi Nakamura (Nara Institute of Science and Technology, NAIST, Japan)

Wen Gu (Japan Advanced Institute of Science and Technology, JAIST, Japan)

Rafik Hadfi (Kyoto University, Japan)

Loizos Michael (Open University of Cyprus & CYENS Center of Excellence, Cyprus)

Theodore Patkos (Foundation for Research and Technology – Hellas, FORTH, Greece)

Seitaro Shinagawa (Nara Institute of Science and Technology, NAIST, Japan)

Katsuhito Sudoh (Nara Institute of Science and Technology, NAIST, Japan)

Hiroki Tanaka (Nara Institute of Science and Technology, NAIST, Japan)

Koichiro Yoshino (Institute of Physical and Chemical Research, RIKEN, Japan)

Teaser

This article describes the proceedings of the discussions of Group 1 of the 4th JST/ERCIM Joint Workshop 2023. The workshop was a 3-day meeting that took place in Kyoto, Japan, aiming to strengthen the research ties between European and Japanese researchers in the fields of AI, Big Data, Human-Computer Interaction and the Internet of Things, and to explore collaboration opportunities between the two communities. Group 1 dealt with the topic of identifying research steps and challenges necessary to achieve trustworthy and reliable human-machine symbiotic collaboration in increasingly complex real-world scenarios.

Main text

Robotics is experiencing a paradigm shift from heavy-duty industrial robots towards intelligent agents operating in our homes, interacting with people with no technical skills, in environments that are open, complex, and unpredictable. For AI to cope with such environments, a set of diverse AI challenges must be addressed, and a set of new skills must be obtained.

Europe, having a strong capacity in research and academia, as well as a well-established ecosystem of scientists, developers, suppliers and end-users, already set these challenges among its strategic,

short/mid-term objectives, as evidenced for instance by the recently announced Strategic Research Innovation and Deployment Agenda of the ADR Association [L1], or the Statement on the Future of AI in Europe published by the CLAIRE organization [L2]. These objectives are coupled with a clear desire to preserve certain ethical values, such as trustworthiness and transparency, often enforced through regulation frameworks, such as the forthcoming AI Act. Following a similar path, Japan is also a pioneer in the way results in cutting-edge AI and Robotics research are adopted by the industry, boosting innovation and maximizing economic and societal impact.

Research collaboration opportunities between Europe and Japan relevant to human-machine symbiotic collaboration is creating a fruitful ground for bringing forward positive results to both ends, if structured on top of a well-defined plan. This was the objective of Group 1 of the 4th JST/ERCIM Joint Workshop. Human-machine collaboration can be achieved across various domains, including education, healthcare, business, society, and economics, through appropriately implemented and deployed smart assistants. Several examples are illustrated in Figure 1, along different scales and domains.

To promote trustworthiness and reliability in these domains, machines should be able to explain actions and reasons for those actions. Contestability (the ability to propose methods, trial and error, and to collaborate as well as humans do) is also important. Indicators and criteria for quantifying aspects of explainability and contestability are necessary, respecting fairness, bias, and accountability, depending on their purpose.

Socially aware communication and interaction encompass diverse aspects, including cognitive, linguistic, cultural, and social differences that shape human connections. Embodiment and the understanding of verbal and non-verbal communication like intentions and emotions, is crucial.

In other words, such systems require cognitive skills, which highlight the necessity for reasoning abilities aligned with human thought, particularly in exhibiting a “Theory of Mind”, i.e., understanding others’ beliefs, intents, desires, and emotions. These reasoning abilities include the use of common sense and causality to predict human actions in various contexts, especially where implicit knowledge is involved. They also encompass the use of argumentative reasoning for rational discourse and decision-making, along with social cognition to navigate social interactions and norms. Together, these skills are vital for AI systems to integrate effectively into human settings, and are expected to require the adoption of insights from Cognitive Science, Psychology, and the social sciences, despite the anticipated interdisciplinary challenges.

Furthermore, a symbiotic environment highlights the need to integrate contextual information. This environment thrives on the collaboration between humans and artificial agents, augmenting human capabilities and emphasizing the necessity for adaptability. Social and cultural awareness play pivotal roles, necessitating adherence to regulations and social norms to avoid misunderstanding. Establishing common ground through defining, aligning, and using new concepts becomes pivotal, posing shared understanding and facilitating smooth interactions across diverse cultural and social contexts, aligned with human societal values.

Human-centric adaptive learning via automatic feature selection with proper learning strategies is also a promising direction to consider. As desired features in the learning strategies, cognitive development, ability of learning online and ability of making actionable prediction should be considered.

Foundation Models and Large Language Models (LLMs) significantly enhance human-machine interaction, but face limitations in logical reasoning. Integrating neuro-symbolic methods can address these limitations by enriching symbolic reasoning capabilities within LLMs.

Multi-agent-based simulations, coupled with agent communication will fine-tune perceptions and enhance cognitive abilities, facilitating real-world error correction and knowledge emergence.

Overall, the synergy among these technologies is paving the way for more general AI. Each offers unique strengths, pushing towards more versatile and efficient systems. This collaborative evolution highlights a trend towards AI that is more context-aware and autonomous, progressively integrating into a unified AI ecosystem.

To promote collaboration between Europe and Japan on the above topics, we have planned activities that take into account the cultural differences and backgrounds, as well as the characteristics of the two research ecosystems, starting with a SWOT analysis (Strengths, Weaknesses, Opportunities, Threats). One of the aspects of complementarity could be neuro-symbolic AI in the EU and deep learning AI and humanoid robots in Japan. More applications to EU-Japan collaboration calls will be encouraged, as well as tools to improve networking, such as postdoctoral fellowships, professorial sabbaticals, researcher exchanges, or student exchanges/internships. Joint academic ventures such as joint workshops, joint editorship of journal special issue(s), joint online degree (open university style) would also be helpful in promoting collaboration.

References:

[1] https://adr-association.eu/
[2] https://claire-ai.org/

Please contact:

Giorgos Flouris

Foundation for Research and Technology – Hellas, FORTH, Greece

fgeo@ics.forth.gr

Satoshi Nakamura

Nara Institute of Science and Technology, NAIST, Japan

s-nakamura@is.naist.jp

Figures:

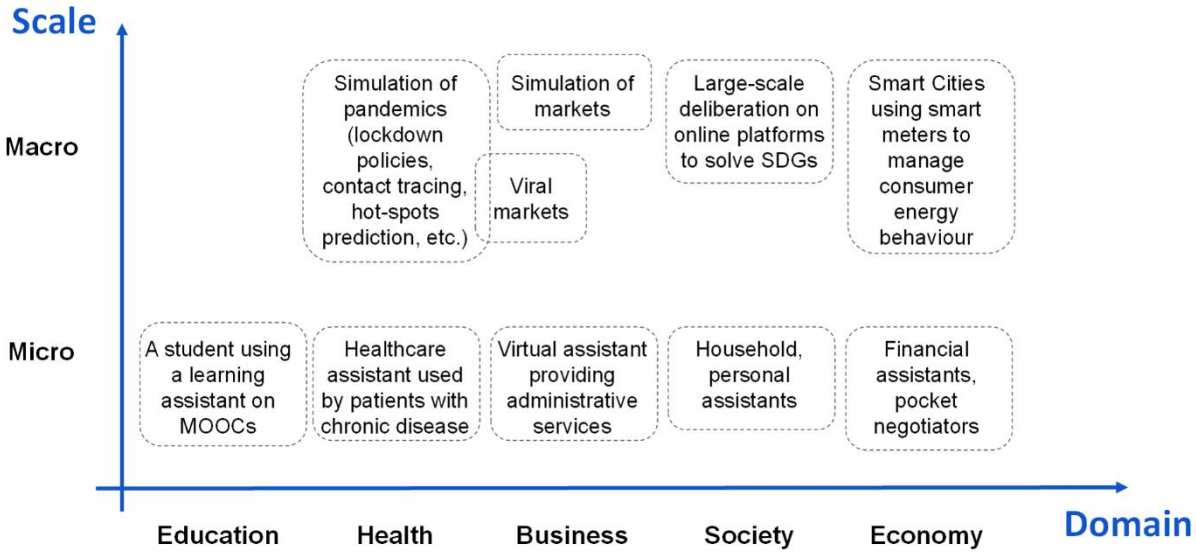

Figure 1. An organisation of the relevant use cases along 5 different domains and for different scale (micro or macro)