Theme 1: Trustworthy and Reliable Human-Machine Symbiotic Collaboration

Positions and Proposals of the 5th Joint JST/ERCIM Workshop

Contributors:

- **Giorgos Flouris**, FORTH, Greece (fgeo@ics.forth.gr)
- **Theodore Patkos**, FORTH, Greece (patkos@ics.forth.gr)
- **Rafik Hadfi**, Kyoto University, Japan (rafik.hadfi@i.kyoto-u.ac.jp)
- **Antonis Argyros**, FORTH, Greece (argyros@ics.forth.gr)
- **Davide Ceolin**, CWI, The Netherlands (davide.ceolin@cwi.nl)
- **Richárd Farkas**, University of Szeged, Hungary (rfarkas@inf.u-szeged.hu)
- **Kristian Fenech**, Eötvös Loránd University, Hungary (fenech@inf.elte.hu)
- **Bence Gercuj**, SZTAKI, Hungary (gercuj.bence@sztaki.hun-ren.hu)
- **Wen Gu**, Japan Advanced Institute of Science and Technology, Japan (wgu@jaist.ac.jp)
- **Michael Loizos**, Open University of Cyprus & CYENS Center of Excellence, Cyprus (loizos@ouc.ac.cy)
- **András Lőrincz**, Eötvös Loránd University, Hungary (lorincz@inf.elte.hu)
- **Satoshi Nakamura**, Nara Institute of Science and Technology, Japan (s-nakamura@is.naist.jp)
  Kaoru Ota, Muroran Institute of Technology, Japan (ota@muroran-it.ac.jp)
- **Seitaro Shinagawa**, Nara Institute of Science and Technology, Japan (sei.shinagawa@is.naist.jp)
- **Katsuhito Sudoh**, Nara Institute of Science and Technology, Japan (sudoh@is.naist.jp)
- **Hiroki Tanaka**, Nara Institute of Science and Technology, Japan (hiroki-tan@is.naist.jp)
- **Zsolt Viharos**, SZTAKI, Hungary (viharos@sztaki.hun-ren.hu)
- **Kei Wakabayashi**, University of Tsukuba, Japan (kwakaba@slis.tsukuba.ac.jp)
- **Edgar Weippl**, TU Wien, Austria (EWeippl@sba-research.org)
- **Satoshi Yagi**, Kyoto University (yagi.satoshi.5c@kyoto-u.ac.jp)
- **Koichiro Yoshino**, RIKEN, Japan (koichiro.yoshino@riken.jp)

Endorsers: [List of endorsers (individuals/organizations/associations): to be decided]

Purpose of this document

The purpose of this document is to raise awareness and call for coordinated actions towards achieving future Trustworthy and Reliable Human-Machine Symbiotic Collaboration, as proposed by senior members of the European Research Consortium for Informatics and

Mathematics (ERCIM) and the Japan Science and Technology Agency (JST) who participated in a series of workshops over the last years. The document is endorsed by key individuals, organizations and associations in the field. In alignment with the European and Japanese research agenda pertaining short- and mid-term objectives and strategic plans, the document describes challenges and recommendations for progress in three critical areas of application, and underscores the need for Europe and Japan to join forces in research and innovation, in order to strengthen the world-leading role of both parties in this area.

Entering the Age of Human-Machine Symbiosis

The tremendous progress of AI technologies over the last years is already transforming almost every aspect of the lives of citizens, especially in developed countries, and there is consensus that it will continue to impact society at an increasing pace. Non-expert humans are not only using intelligent autonomous machines, but are progressively becoming dependent on them to perform even simple everyday tasks. Almost without noticing, intelligent systems are obtaining new roles as our assistants, our companions, and soon our long-living -virtual or even physical- friends. The challenges from the symbiosis of humans with intelligent autonomous systems are too important to ignore, introducing societal, behavioural, cognitive, economical, and of course technological aspects to consider. The latest wave of progress in AI foreshadows the magnitude of achievements we anticipate to experience by the end of the decade.

In this context, ERCIM and JST worked together to explore how best to join forces, in order to promote research excellence towards achieving future human-machine symbiotic environments ***built on top of responsible technologies and ethical values***. Europe and Japan have steadily developed over the past decades a strategic partnership that adheres to common key principles and objectives while facing common societal challenges. Following the will of both sides to intensify their cooperation on key global challenges, this white paper wishes to drive the attention of relevant stakeholders and supporting bodies on the opportunities and the risks that future human-machine symbiotic environments introduce and on the need to support a European-Japanese collaboration on these aspects. Our intention is to underscore the need to take action urgently, in the form of targeted investments in research and innovation, well-incentivised multi-stakeholder participation and clear regulatory roadmapping, as we expect that by 2030 such environments will no longer be considered uncommon. Without proper consideration, their implementation will not meet the desirable standards. Multi-disciplinary research is both essential and timely, and if structured on top of appropriate planning and supporting instruments, will create a fruitful ground for yielding positive results for both regions, significantly helping them to promote research excellence and increase growth and industrial competitiveness.

Major Use Cases

Human-machine symbiosis will affect holistically a broad range of human activity; to understand the extent, we consider three characteristic domains and the challenges they pose. It is noteworthy that these domains constitute top priorities in both the European and Japanese research agenda; at the same time, their diversity, sketched also in Table 1,

justifies the large-scale impact that future developments in human-machine collaboration are going to have.

The first is the *industrial* use case, where the worker (user of the robot) needs to collaborate with the robotic machinery (typically a robotic arm or some mobile platform) in order to execute a task in an optimal and safe manner. The main characteristics of this scenario compared to the others is the relatively high predictability of the setting, and the need for just-in-time action and cooperation, often in safety-critical situations. The robot should be able to augment and/or complement the user's sensory capabilities and physical limitations, predict human behaviour, and ensure seamless interoperability while taking into account the human partner's physical limitations. Natural forms of interaction will be crucial, as well as the ability of the robot to enhance its skills during the collaboration (e.g., through learning by demonstration and adaptive/transfer learning). In the industrial setting, energy consumption considerations also become important.

The second is the *home* use case, in which a home-dweller (who can be an elder, child or person with disabilities) has to interact with a social robot, typically embodied, with the aim to improve the person's well-being (e.g., for therapeutic reasons, companionship, or support in home tasks). An important characteristic of this use case is that the person is typically a non-expert, and thus the robot should be able to interact using more natural forms of communication. Moreover, in most applications in this use case, the robot should be able to exhibit emotional depth and empathy in order to fulfil its goals. Importantly, it should also be able to forecast human behaviour employing theory of mind techniques, learn to apply social norms and ethics, and be compliant with existing laws and regulations. Note that these are culture- and country-specific, which raises the need for online learning, (co-)adaptation and personalisation capabilities. While symbiotic intelligent systems are convenient and helpful, their constant employment can shift cognitive and memory-based tasks, alter cognitive strategies and capabilities, influence behaviour, and foster dependencies; mitigating these effects for the human users is another important challenge associated with this use case. Explainability and predictability of the robot's behaviour is also a necessary feature. Importantly, the highly interactive nature of this use case will often raise the need for negotiation, argumentation and persuasion (e.g., to persuade a reluctant human to take their medication).

The third is the *disaster management* use case, which focuses on disaster management and particularly the *mitigation*, *prevention*, *response*, and *recovery* aspects. In this case, adequate human-machine symbiosis can enhance the resilience of societies, particularly in disaster-prone regions like Japan and Europe. The human users range from first responders and local government officials to citizens affected by the disasters. Robots and intelligent systems, such as drones, autonomous vehicles, and mobile robotic systems, are deployed in various forms to address the unpredictable and often hazardous environments associated with natural and human-made disasters (e.g., nuclear fuel leaks, etc.). Some critical aspects must be considered. In disaster management, aspects such as adaptability, communication, and contextual awareness are paramount. Robots must operate autonomously while sharing information in real-time with human responders. Given disaster settings' unpredictability and high stakes, the symbiosis should support dynamic decision-making, allowing robots to adjust their objectives as new local information arrives.

Dimension	Industrial setting	Home setting	Disaster management
Predictability	High	Medium	Low
Scale of coordination	Low	Medium	High
Scale of interaction	One-to-one	One-to-few	Many-to-many
Emotional depth, empathy	Low	High	Medium
Current level of development	Medium/low	Medium/low	Low
Information availability	High	Medium	Low

Risks, Challenges and Opportunities

The type of symbiotic relationship that is being established between humans and intelligent machines has no match in the history of humankind. The challenges to consider are many. From the technical standpoint, developing intelligent entities that interact for long periods of time with non-expert human users in open and unpredictable environments goes beyond the current capabilities of existing AI methods and Robotics systems. For future intelligent machines to exhibit a behaviour that is not only effective, but also closer to human intuition and intellect, they need to narrow the chasm between artificial and natural intelligence and make progress on skills that humans excel in. These include among others:

- the ability to provide real-time responses to complex interactions and decision-making;
- a general understanding of how the world works;
- the exploitation of commonsense knowledge that is hidden, yet pervasive, in the majority of human-to-human interactions;
- the competence to engage with multi-modal forms of communication in synergetic task execution with humans, exhibiting varying degrees of coordination demands;
- advanced skills in object manipulation and navigation in unstructured environments.

The scale of excellence needed for such skills differs in each of the three scenarios described above; yet, certain aspects are equally important across any scenario, where humans and machines will be called to work together. These include the requirements for intelligent systems to operate based on medium- or low-quality information, to ensure safe and trusted interactions, to explain with grounded justifications their decisions, to adopt a principled approach in respecting the AI ACT, the Japanese Guidelines for Business, and similar legal and ethical regulations.

As it is evident, the technical requirements cannot and should not be dealt with ignoring non-technical ones. Human-machine symbiosis is on the cross-section of interest between diverse disciplines. The societal issues that emerge present novel areas of investigation pertaining to both the intelligent system and the humans themselves. On the one hand, the soft skills needed by artificial entities must comply with social norms and values, properly configurable to the characteristics of different population groups, including also cultural aspects. On the other hand, the abundance of intelligence systems and the flood of

information already give rise to behavioural changes in humans. For human-machine interactions to be helpful, we need to explore early enough and understand the cognitive footprint that they leave on humans, and the risks involved. Concerns have been raised related to the erosion of memory and attention, to over-reliance and unjustified trust on our artificial companions in critical situations, to an incremental loss of cognitive skills and reduction of problem-solving abilities, etc.

EU - Japan Collaboration

Europe and Japan face significant challenges, including an aging population, climate crises, economic stagnation, and energy demands, that require innovative and collaborative solutions. At the same time, they adhere to common principles and a strong desire to align research progress with responsible practices, especially concerning AI development. By leveraging their complementary strengths, a partnership between Japan and Europe offers immense opportunities in research and technology. Europe's focus on foundational and neuro-symbolic AI aligns well with Japan's excellence in manufacturing and robotics engineering, creating a robust framework for developing advanced, human-machine symbiotic technologies. Moreover, Japan's cultural acceptability of robots provides a unique advantage, positioning Japanese society as an ideal large-scale testbed for evaluating AI integration in daily life. At the same time, Europe can contribute throughout its mature regulatory framework to establishing comprehensive guidelines for trustworthy AI. These collaborations would amplify the shared European and Japanese values of industrial competitiveness and geopolitical positioning in the face of growing competition from the United States and China.

Conclusions

ERCIM and JST, recognizing the global impact of future human-machine symbiotic collaboration, along with the risks that lie ahead and the great potential for offering responsible and trustworthy AI, ask the participation of individuals, organizations, and policy makers to join forces in planning the appropriate next steps towards a better future.To achieve the goal of a beneficial human-machine symbiosis that avoids symptoms of parasitism, the current race of progress in AI should not be the force that will determine the characteristics of future collaborative environments, but the other way around: our desirable standards for an AI-enabled human-centred future should help drive research towards even more astonishing achievements in AI and Robotics.