

The future of collaborative robotics AI in Industry 5.0: An academic perspective with a practice approach

William Wai Leong Lee

Abstract

Over the last five decades, the industrial robotics market has undergone significant evolution, particularly in its integration within the automotive sector. Technological advancements encompassing software, electronics, electromechanical, biomechanics, and kinematics have profoundly influenced automation trends. The ongoing progress in computing power is poised to facilitate AI technologies, thereby accelerating the development of autonomous robots. These compact, mobile, self-driving systems are anticipated to assume pivotal roles in future industrial and manufacturing landscapes – thru data from sensors, AI compute and machine control.

In contemporary times, this evolution has resulted in the emergence of collaborative robots, commonly referred to as cobots, where humans work alongside these digital counterparts. As these autonomous robots continue to integrate into human work environments, there is a foreseeable trajectory towards potential full replacement of organic roles, unless proactive measures are taken to incorporate human oversight and intervention.

Thus, to facilitate collaborative interaction between humans and robots, it is essential that robots comprehend tasks, procedures, and human actions adeptly. They should demonstrate rapid and robust learning capabilities in near real-time, exhibiting tolerance towards pose variations and the ability to generalize across related tasks. This necessitates a paradigm shift towards a more humanlike cognitive framework for robots, enabling adaptation to human behaviors and work environments. This includes capabilities such as autonomous visual exploration for object recognition and learning new tasks through observation of human demonstrations and instructions.

To enhance collaboration between humans and robots, incorporating additional sensory modalities such as tactile feedback, natural speech, and dialogue interactions beyond perception and planning is crucial. This advanced human-robot collaborative paradigm entails a novel approach in robot design, emphasizing human-like attributes to anticipate and complement the actions of human team members in workplace settings. Furthermore, enabling robots to possess situational awareness enhances their ability to discern task structures during operations.

The capacity to comprehensively summarize and retain operational information enables robots to efficiently retrieve necessary data for task completion in collaboration with humans. Over time, robots can anticipate human task performance, either during brief absences or through enhancements. This approach draws on the Theory of Mind to advance Robotics AI beyond mere collaboration, integrating human-centric autonomy. This ensures that future workplaces, rather than being fully automated for the fear of being replaced, foster robots' ability to emulate human-like tasking within Human-Robot Teams.

Keywords

Collaborative, Robotics, Artificial Intelligence, Human-Robot, Human-centric

10th International Conference on Socio-Technical Perspectives in IS (STPIS'24) August 16-17 2024, Sweden.

 willylee.wleong@gmail.com (William Wai Leong Lee)

 0000-0001-7750-618X.



© 2024 Copyright for this paper by its author. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)

1. Introduction

When collaborating with humans, robots must extend their focus beyond their immediate environment and task execution; they need to develop a meticulous understanding of their human partners' reasoning processes. Effective support in complex tasks compel that robots share pertinent information with their human counterparts. However, indiscriminate sharing of all available information can lead to human annoyance and distraction, as not all information may be pertinent or novel in every situation. Thus, it is crucial for robots to determine the appropriate timing and type of information to communicate, which demands an understanding of the human partner's knowledge and situational awareness.

In prior research, our team introduced the concept of Theory of Mind, which involves selecting information-sharing actions based on an evaluation of relevance and an estimation of human beliefs. This study integrates this concept into a communication assistant designed for a collaborative human-robot setting and evaluates the performance benefits derived from this integration. The human's belief state is estimated based on task progression and gaze positions. Utilizing this belief estimate, the system predicts the most likely subsequent actions and assesses their impact on future rewards. If a simulated belief update from a robot's communication action is projected to lead to a future plan with higher expected rewards, while considering the explicit cost of communication, the system will opt to assist the human with pertinent information.

To evaluate this approach, we have studied a task that is challenging for the human, thereby creating scenarios where additional information communicated by the robot could be beneficial. Naive human participants performed this task alongside the robot, generating data to assess the impact of the human-centric communication concept on various performance measures. Compared to conditions without information exchange, participants who received assistance from the robot were able to recover from unawareness significantly earlier. This approach acknowledges the costs associated with communication and effectively balances providing necessary interruptions with avoiding unnecessary ones, surpassing other methods in this regard.

Belief inference enables an assistance model that empowers humans to make informed decisions independently, rather than patronizing them. The integration of Theory of Mind-based Communication into a collaborative human-robot interaction setting marks a significant advancement in the development of intelligent robotic systems. By accurately estimating human belief states and predicting the impact of potential communication actions, robots can provide timely and relevant information that enhances human performance in complex tasks. This approach not only improves task efficiency but also fosters a more harmonious human-robot partnership by respecting the human's cognitive load and avoiding unnecessary distractions.

Finally, our study demonstrates that a robot's ability to understand and anticipate human reasoning significantly enhances cooperative task performance. The Theory of Mind-based Communication framework provides a robust foundation for developing communication assistants that can intelligently decide when and what information to share, leading to more effective and user-friendly human-robot collaborations and interactions. Future research should explore the scalability of this approach in more diverse and dynamic environments, further refining the balance between informativeness and efficiency in human-robot communication.

2. Research and Development Methodology

Theory of Mind (ToM) has the potential to profoundly enhance human-robot collaborations and interactions, particularly within the realms of collaborative learning and assistive activities. This whitepaper approach aims to elucidate the application of ToM in these contexts, outlining both theoretical underpinnings and practical implications.

Collaborative Human-Robot Learning – Understanding Intentions, Adaptive Learning, Feedback and Improvement: Robots endowed with ToM capabilities can infer the intentions and goals of human partners. Such robots can anticipate needs and offer timely assistance, thereby facilitating an enriched learning environment. The ability to understand and predict human intentions is fundamental to seamless collaboration and can significantly elevate the efficacy of the learning process.

This approach allows robots to gauge the mental states and learning progress of their human counterparts. The adaptive capacity enables robots to tailor their instructional approaches based on real-time assessments. For instance, when a robot detects signs of confusion in a learner, it can adjust its teaching strategy by providing further explanations or demonstrations, thus fostering a more conducive learning atmosphere.

Robots when well-enabled, can deliver personalized feedback grounded in their understanding of the learner's mental states. Such targeted feedback addresses specific areas of difficulty, promoting improved learning outcomes. By leveraging this capability, robots can discern refined learner needs and respond with precise and relevant instructional modifications.

Assistive and Augmentative Actions – Context-Aware Assistance, Enhanced Communication, and Proactive Support: ToM equips robots with the ability to provide contextually appropriate assistance. In healthcare settings, for example, a robot could interpret a patient's pain or discomfort and respond in a supportive manner. Context awareness is crucial for delivering meaningful and effective assistance that aligns with the specific needs and circumstances of the human user.

Effective communication is pivotal in human-robot collaborations and interactions. This allows robots to understand the human perspective, thereby facilitating more effective

communication. By determining the optimal moments and manners to share information, robots can avoid cognitive overload and enhance cooperative dynamics. The understanding fosters a more fluid and harmonious interaction.

This enablement permits robots to anticipate future needs and offer proactive support. In the industrial and manufacturing environments, for instance, a robot could foresee when a worker requires a particular tool and ensure its availability in advance. The anticipatory capability enhances operational efficiency and supports human workers in a proactive, rather than reactive, manner.

Practical Implementation – Data Collection, Machine Learning Models, Integration with World Models, and Human-Centric Designed Autonomy: The implementation of ToM calls for robust data collection mechanisms. Utilizing sensors and advanced data analytics, robots can gather extensive information regarding human behavior, expressions, and interactions. This data serves as the foundation for developing accurate cognitive models with the application of various machine learning techniques for inferring mental states from collected data. Supervised learning techniques, which utilize labeled data, are particularly effective in training them. The precision of Collaborative AI models hinges on the quality and comprehensiveness of training data, e.g. especially by learning from human demonstrations.

For a holistic understanding, it must be integrated with world models that encapsulate the environment and tasks. This integration facilitates a comprehensive situational awareness, enabling robots to make informed decisions based on both the inferred mental states and the contextual factors of the environment. The design of Robotics AI systems must prioritize human-centric principles, focusing on usability, comfort, and effectiveness. Ensuring that these systems are intuitive and user-friendly is paramount for achieving successful humanrobot collaboration.

Challenges and Future Directions – Ethical Considerations, Accuracy and Generalization, Interdisciplinary Research and Real-World Applications: The deployment of ToM in human-robot collaborations and interactions enable rigorous ethical considerations. Respecting privacy and ensuring the ethical use of data are critical components of responsible Collaborative AI application. Ethical guidelines must be established and adhered to, ensuring that it enhances human experiences without compromising individual rights. Enhancing the accuracy of these models and ensuring their generalizability across diverse individuals and contexts remains a significant challenge. Continuous advancements in machine learning and interdisciplinary research are essential for developing robust and reliable Robotics AI systems.

The convergence of insights from psychology, neuroscience, and Artificial Intelligence is crucial for the advancement of these models. Interdisciplinary research can provide a more delicate understanding of mental state inference, facilitating the development of more sophisticated and effective systems. Expanding the application of Collaborative AI across various domains, such as education, healthcare, industrial and manufacturing automation,

presents promising avenues for future research. The practical benefits of Robotics AI in enhancing human-robot collaboration are vast, and exploring these applications can lead to significant societal advancements.

In Summary: Integrating ToM into human-robot collaboration holds immense potential for creating more intuitive, effective, and supportive systems. By understanding and responding to human mental states, robots can significantly enhance both collaborative learning and assistive actions. As R&D progresses, the ethical, accurate, and interdisciplinary development of Collaborative Robotics AI will be pivotal in realizing its full potential across various real-world applications.

3. The Practice Approach for Industrialization

Operational Sensing in Human-Machine Collaborative with Real-Time Monitoring of Operations – Operational sensing, a crucial component of human-machine collaborative interaction, involves the real-time monitoring of ongoing processes through the deployment of sensors and data analytics. This practice is fundamental for comprehending the current state and performance of operations. By harnessing these technologies, organizations can derive actionable insights into their processes, ensuring they operate at optimal efficiency.

Analyzing Task Structures – Identifying the structure of tasks within operations constitutes another vital element of operational sensing. Process mining techniques, which scrutinize event logs, are utilized to unveil the sequence of activities. The application of machine learning algorithms further refines this analysis by discerning patterns and structures within these tasks, thereby providing a comprehensive understanding of operational workflows.

Summarization and Storage of Operational Data – The summarization of operations is facilitated by natural language processing (NLP) techniques. Advanced models, such as GPT-4, a large language model (LLM), can generate concise summaries of operations based on key events and actions. These summaries distil complex operational data into more digestible formats, enhancing quick comprehension and decision-making.

Structured Storage Solutions – To ensure the efficient retrieval of summarized data, implementing a knowledge management system is essential. These systems store summaries in structured formats, such as knowledge graphs. This structured approach enables easy access to historical data, supporting future queries and analyses.

Human Cognitive Sensing in Operational Environments – Understanding the human mind is a pivotal aspect of human cognitive sensing. Cognitive models and sensors, including EEG (electroencephalogram) and eye-tracking devices, are employed to infer mental states and intentions. Machine learning algorithms play a significant role in interpreting these signals, providing deeper insights into human cognitive processes.

Summarizing Visual Experiences – Combining large language models (LLMs) with gaze-tracking data facilitates the generation of summaries that reflect what a human operator sees and does. This integration assists in evaluating task proficiency and provides timely assistance, thereby enhancing overall operational efficiency.

Sensing Affective States and Cognitive Load - The detection of affective states involves using sensors to capture physiological signals, such as heart rate and skin conductance, as well as facial expressions. These signals are indicative of emotions and understanding them allows systems to tailor interactions and support to the user's emotional state for effective collaboration.

Impact on Memory and Performance - Affective states and cognitive load significantly influence memory retention and recall. Systems can adapt their assistance based on the detected cognitive load, optimizing learning and performance. By doing so, they can better support users in high-stress or cognitively demanding environments.

Multi-modal Interaction and Grounding – Human intentions can be inferred using multi-modal inputs, including speech, gestures, and gaze. Contextual understanding and situational awareness are essential for resolving ambiguities and accurately interpreting human intentions. This multi-modal approach ensures a more holistic understanding of user needs and actions.

Providing Contextual Assistance – Based on the inferred intent and situational context, agents can offer relevant help, whether it is information, guidance, or task automation. This tailored assistance enhances the user's experience and efficiency in completing tasks.

Resource Efficiency and Multi-modal Fusion – Efficiency in operational sensing involves optimizing algorithms and processes to run effectively on devices such as the HoloLens. This optimization provides balancing computational load and responsiveness, ensuring that the system performs efficiently without compromising speed or accuracy.

Integrating Multi-modal Data – The fusion of data from various sensors – visual, auditory, and tactile – creates a comprehensive understanding of the environment and the human's actions. This multi-modal integration is critical for developing systems that can perceive and respond to complex, dynamic scenarios in real-time.

Effective Communication Strategies – Effective communication in operational environments entails the use of clear, concise language and visual aids. It is important to consider the user's knowledge level and context to ensure that the information conveyed is easily understood. This clarity aids in better decision-making and operational efficiency.

Establishing Common Ground – Building a shared understanding, or common ground, between the agent and the human is crucial. Aligning the agent's responses with the human's knowledge and expectations helps in establishing trust and improving

collaboration. This common ground ensures that interactions are more productive and mutually beneficial.

For the above work suggests that operational sensing, which encompasses real-time monitoring, task structure analysis, and the integration of advanced technologies, plays a vital role in enhancing the efficiency and effectiveness of human-machine interactions for collaboration. The summarization and storage of operational data, understanding human cognitive and affective states, and employing multi-modal interaction techniques are key components of this process.

By optimizing resource efficiency and ensuring effective communication, these systems can significantly improve performance and decision-making in complex operational environments from a practical approach, with pragmatism for industrialization and early deployment of Collaborative Robotics AI FutureNow.

4. Conclusion and Recommendations

This whitepaper consolidates foundational concepts in Artificial Intelligence (AI) crucial for the development of sophisticated Collaborative Robotics AI systems capable of deep understanding, decision-making, and human interaction. It explores several key areas including multimodal sensing, world modelling, Theory of Mind (ToM), multi-agent interaction, and human-centric response generation, elucidating their theoretical underpinnings, practical applications, and future research trajectories.

Multimodal sensing involves integrating diverse sensory inputs (e.g., cameras, microphones, tactile sensors) to capture comprehensive environmental and human interaction data. This approach enables Robotics AI systems to construct holistic representations of their surroundings, facilitating robust interpretation and interaction in real-world scenarios. Sensor fusion techniques amalgamate data from these sensors using advanced algorithms, thereby enhancing the synthesis of unified situational models.

The concept of world modelling entails creating detailed representations of the environment, encompassing objects, attributes, and their interrelationships. Such models equip Robotics AI systems with foundational knowledge of current conditions, essential for predicting future states, informed decision-making, and adaptive responses in dynamic environments.

For ToM in Collaborative AI refers to the capability of attributing mental states – beliefs, intents, desires – to oneself and others, enabling the inference and anticipation of human behavior. Rooted in psychological theories, this approach enhances Robotics AI systems' ability to engage intuitively and responsively in social contexts, thereby improving interaction quality in applications such as human-robot interaction and social robotics.

While effective multi-agent interaction among multiple agents (both human and AI) relies on communication, coordination, and collaboration towards shared goals. Algorithms and

protocols designed for multi-agent systems enable information exchange, negotiation, and cooperative problem-solving, crucial for domains such as autonomous vehicles and collaborative robotics.

Moreover, human-centric response generation involves tailoring Collaborative AI responses based on contextual cues, historical interactions, and inferred user preferences and emotional states. By integrating insights from ToM with Robotics AI systems deliver personalized and contextually relevant assistance across various domains, enhancing user satisfaction and engagement.

The applications and implications of these advanced Collaborative AI concepts span diverse domains including education, healthcare, manufacturing, and customer service. Personalized learning platforms adapt educational content based on student progress, while healthcare systems monitor patient conditions and deliver customized medical interventions. These applications highlight the potential to enhance human quality of life and operational efficiency across various sectors.

However, implementing Collaborative Robotics AI presents challenges, including ethical concerns regarding privacy and data usage, necessitating robust safeguards. Achieving accurate mental state inference requires sophisticated algorithms capable of generalizing across diverse contexts and individuals, underscoring the need for continual improvement and ethical considerations in Collaborative AI development. Lastly, advancing the capabilities through multimodal sensing, world modelling, multi-agent interaction, and human-centric response generation holds transformative potential across various applications. Ongoing R&D in these areas promises to enhance Robotics AI systems' adaptability, responsiveness, and ethical framework, thereby fostering more effective human-machine collaboration and improved societal outcomes.

Future research directions aim to further refine Collaborative AI models for enhanced personalization and applicability across diverse domains. Therefore, having interdisciplinary efforts integrating insights from psychology to neuroscience, are destined to refine its capabilities and broaden for the real-world impact.

Their implementation in autonomous systems, virtual assistants, and interactive technologies represent evolving frontiers where Theory of Mind-equipped Collaborative AI can facilitate more intuitive, adaptive, and socially adept interactions. The developments signify a transformative trajectory towards human-centred solutions capable of navigating complex socio-technical landscapes.

These concepts enable the systems to perceive, interpret, and interact with the world akin to human cognitive processes and social dynamics; by harnessing the capabilities and technologies of certainty to redefine human-machine collaboration, personalized assistance, and societal engagement across diverse sectors. Future endeavours will continue to refine and expand these capabilities, driving innovation towards more empathetic, context-aware, and effective Robotics AI.

5. Figure

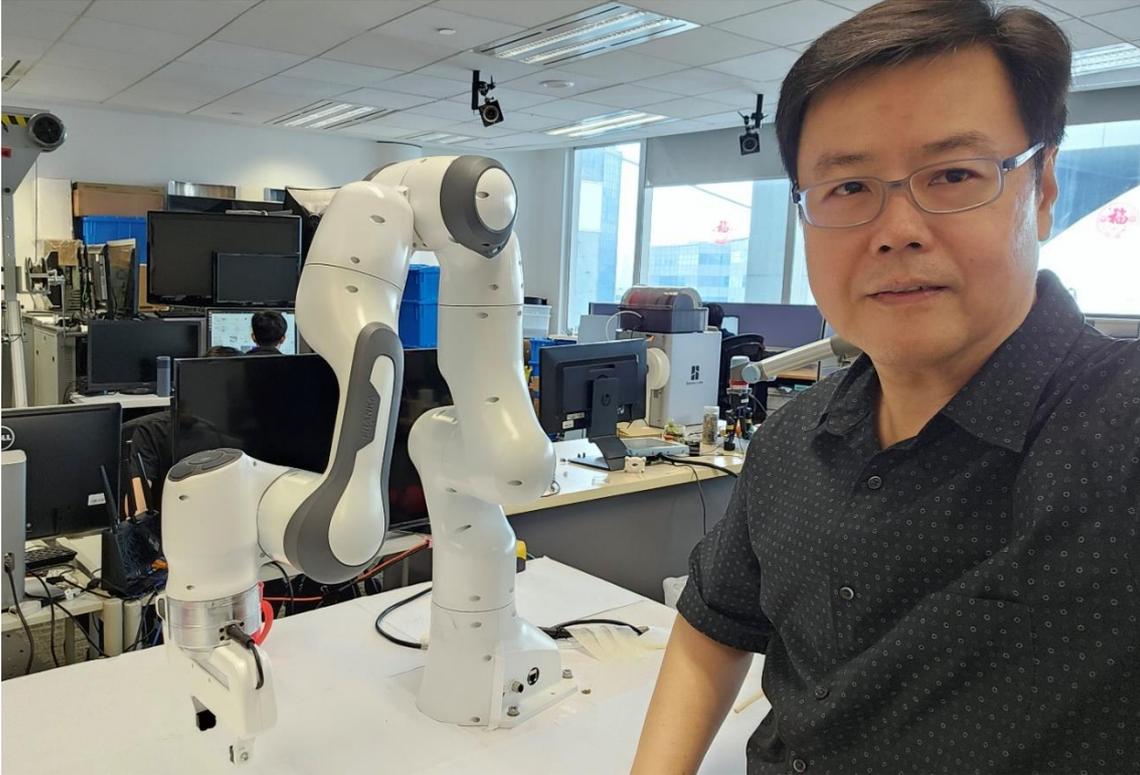


Figure 1: Collab-AI Research at A*STAR, Agency for Science, Technology and Research at Institute for Infocom Research (i2r) Lab, Author: William Wai Leong Lee¹

6. Further Discussions

The text discusses advanced concepts within cognitive architecture and process understanding. It outlines several key approaches: establishing common ground truth understanding, extracting task structure of processes, integrating commonsense reasoning, implementing episodic memory within cognitive architecture, and utilizing knowledge graphs of processes. These approaches aim to enhance system capabilities such as communication, task automation, decision-making, learning from past interactions and demonstrations, and optimizing workflows:

1. Common Ground Truth Understanding: This approach focuses on achieving consensus among different systems or agents regarding factual information or concepts. Effective communication and collaboration hinge upon this shared understanding
2. Extracting Task Structure of Process: This involves identifying and delineating the sequential steps or components inherent in specific tasks or processes. Such structuring facilitates efficient organization and automation of tasks

3. Commonsense with Reasoning: Integrating basic human-like understanding and logical reasoning capabilities into systems enables them to navigate everyday situations and make intuitive decisions akin to human cognition

4. Episodic Memory for Cognitive Architecture: Episodic memory enables systems to recall and learn from past experiences or interactions. In the context of cognitive architecture, this capability supports continuous improvement in performance based on learned knowledge

5. Knowledge Graphs of Process: Knowledge graphs provide a structured representation of information, illustrating relationships between various entities. Applied to processes, these graphs map out procedural steps and their interconnections, thereby aiding in comprehensive understanding and optimization of workflows

The above mentioned are advanced concepts in cognitive architecture and process understanding serve to augment system functionalities crucial for enhanced performance across various domains, including communication, automation, decision-making, learning from demonstrations, and process optimization towards Industry 5.0 as Machine Intelligence of Collaborative Human-Robot and Human-Centric Autonomy.

References

- [1] Theory of Mind Based Assistive Communication in Complex Human Robot Cooperation
<https://arxiv.org/pdf/2109.01355.pdf>