Robot Learning Model Leveraging Memory and Attention to Extract Implicit Communication Cue from Unknown User

不特定ユーザとの非明示的なコミュニケーションCUEの抽出を可能とする時系列解析とアテンション機構を活用したロボット学習モデル

July, 2024

Lena Juliette GUINOT
ギノ　レナ　ジュリエット

Robot Learning Model Leveraging Memory and Attention to Extract Implicit Communication Cue from Unknown User

不特定ユーザとの非明示的なコミュニケーションCUEの抽出を可能とする時系列解析とアテンション機構を活用したロボット学習モデル

July, 2024

Waseda University Graduate School of Creative Science and Engineering

Department of Modern Mechanical Engineering, Research on Neuro Robotics

Lena Juliette GUINOT
ギノ　レナ　ジュリエット

Abstract

In the realm of robotics and artificial intelligence, the pursuit of autonomous learning in machines has been the hallmark of innovation and progress, driving the vision of a future where robots are not only capable of executing predefined tasks but are also adept at comprehending and adapting to implicit instructions from humans for prolonged and productive collaboration. This dissertation is dedicated to the investigation and development of machine learning techniques that enable robots to autonomously learn to interpret implicit instructions, with a particular emphasis on long-term human-machine collaboration.

Over the past decades, the integration of robotics into various aspects of human life has witnessed profound advancements, revolutionizing industries such as manufacturing, healthcare, education, and services. In these contexts, the ability of robots to understand and respond to implicit instructions is imperative for fostering effective human-robot interaction. Implicit instructions encompass a broad spectrum of cues, including non-verbal communication, contextual cues, and shared experiences, which often play a crucial role in human communication. Successful autonomous interpretation of these instructions by robots is pivotal for establishing sustained and high-level collaboration, which is increasingly sought after in modern applications.

While the field of robotics has made substantial progress in terms of robots adapting to explicit commands and predefined tasks, there remains a significant gap in the area of autonomous learning of implicit instructions for long-term collaboration. The majority of current research predominantly focuses on short-term, task-specific interactions with a single user or a limited set of users. This approach, however, falls short of addressing the complexity and dynamism inherent in long-term human-machine collaboration, where robots must adapt not only to learning new tasks but also to the diverse communication styles and expectations of a wide variety of users. The challenge in enabling robots to adapt to a broad spectrum of users is a multifaceted one. It encompasses not only the ability to recognize and interpret various communication styles, but also the capacity to adapt to the unique preferences, habits, and idiosyncrasies of different individuals. In essence, it necessitates the development of machine learning algorithms that are not only versatile but also context-aware, enabling robots to differentiate between implicit instructions and adapt their responses accordingly. Understanding the diversity in human communication and learning to adapt to these differences is a complex undertaking. It requires robots to discern the nuances in gestures, facial expressions, intonation, and non-verbal cues, while simultaneously considering the context in which these cues are presented. Moreover, it involves the creation of models that can generalize learning from one user to another, accounting for variations in communication and interaction patterns.

This dissertation strives to address the aforementioned challenges by developing machine learning algorithms and models that empower robots to autonomously learn the art of deciphering implicit instructions for long-term human-machine collaboration. The central goal is to equip robots with the capability to not only adapt to different users but also to enrich their understanding of the intricacies of human communication over extended periods. By achieving this, robots can be more versatile and adaptable, resulting in more productive and harmonious collaborations in a wide range of domains. The subsequent sections of this paper will delve into the methodologies, technologies, and innovations undertaken to advance the state-of-the-art

in this field and contribute to the future of autonomous robotic systems. In this dissertation, the focus was on the development of machine learning techniques for robots to autonomously learn implicit instructions, enabling long-term and productive human-machine collaboration. Implicit instructions encompass a wide array of non-verbal cues, contextual information, and shared experiences that play a pivotal role in human communication. The research delved into the challenges posed by adapting to diverse users and the intricacies of human communication styles. These challenges were addressed through the refinement of the developed algorithm, incorporating blinding as a bias awareness technique and implementing attention mechanisms to consider joint mechanics and sensor data interplay. The achievements in this research contribute significantly to the advancement of autonomous machine learning for human-robot collaboration. However, challenges remain in fine-tuning implicit instruction interpretation and adapting to a broader spectrum of user populations. Future work in this field will aim to further optimize machine learning algorithms and enhance robots' capabilities for interpreting implicit instructions, fostering more inclusive, equitable, and effective long-term collaborations between humans and robots.

Chapter 1 serves as the foundation for the dissertation, offering an extensive background on human-robot communication methods and the challenges inherent in robot continuous learning for long-term human-machine collaboration. We explore the landscape of current research, highlighting the limitations in existing approaches, particularly in adapting to a diverse range of users and the intricacies of human communication styles. The purpose of this research is to develop advanced machine learning techniques that enable robots to autonomously interpret and adapt to implicit instructions, with the ultimate goal of enhancing sustained collaboration in a broad spectrum of real-world applications. This chapter sets the stage for the subsequent exploration of innovative methodologies, technologies, and insights that drive the development of autonomous learning in robots to meet the demands of dynamic human-robot partnerships.

Chapter 2 delves into the intricate domain of identifying and comprehending implicit cues within human motion, an essential aspect of human-robot communication for long-term collaboration. We examine the concept of the ideomotor principle, which underscores the use of non-verbal cues in interpersonal communication. Through a rigorous analysis, we assess the reproducibility and usability of these implicit cues as a viable means of communication between humans and robots. The chapter explores the feasibility of the robot's ability to learn and associate these cues with user intentions, thus laying the groundwork for the translation of human motion into effective robotic commands. This exploration paves the way for the development of advanced machine learning models to enable robots to autonomously decipher and adapt to implicit instructions, crucial for sustained and productive human-machine collaboration.

Chapter 3 focuses on a detailed analysis of the performance disparities revealed in the previous experimental section, particularly concerning the impact of gender on human-robot collaboration. We uncover that the imbalanced dataset resulted in higher robot performance with male users compared to female users, particularly as task complexity increased. Additionally, this chapter scrutinizes how robot performance and behavior influence the user experience and perception of both genders. It places a specific emphasis on identifying the factors that influence trust and explores the consequences of trust violations for future collaboration. This investigation plays a pivotal role in enhancing the robustness and inclusivity of the developed machine learning models, ensuring equitable and effective long-term human-robot collaboration across diverse user populations. Chapter 4 delves into the refinement of the developed machine learning algorithm, addressing the challenge of data bias by employing a bias awareness technique known as blinding. We enhance the algorithm's capacity to recognize and rectify data bias, going beyond mere sensor data analysis. Instead, the algorithm gains an acute awareness of joint mechanics and the interplay between different sensors gathering user motion data. This is achieved through the strategic use of attention mechanisms, enabling the algorithm to discern how each sensor's data correlates and moves in relation to others. The novel approach implemented in this chapter represents a significant stride towards minimizing data bias and optimizing the algorithm's ability to autonomously learn and adapt to implicit instructions for long-term human-robot collaboration, contributing to more equitable and effective interactions.

Chapter 5 serves as the culmination of this dissertation, summarizing the research's accomplishments and significant contributions to the field of autonomous machine learning for long-term human-machine collaboration. It reflects on the challenges addressed throughout the study, emphasizing the mitigation of data bias, gender effects, and trust violations in human-robot interactions. The chapter also underscores the remaining challenges, acknowledging the complexities of adapting to diverse user populations and fine-tuning implicit instruction interpretation. In looking ahead, it suggests the need for further research to address these outstanding issues and lays the foundation for future work in refining machine learning algorithms and advancing the capabilities of robots to autonomously learn and adapt to implicit instructions, fostering more inclusive, equitable, and effective human-robot collaborations in the years to come.

In conclusion, this dissertation represents a comprehensive and forward-looking exploration into the development of machine learning for the autonomous learning of implicit instructions by robots, with the ultimate goal of fostering long-term and effective human-machine collaboration. It has laid the groundwork by providing a thorough background on human-robot communication methods, highlighting the challenges posed by data bias, gender effects, and trust violations. The study has ventured into novel territory by addressing these issues and refining machine learning algorithms to mitigate data bias, understand the intricacies of implicit cues in human motion, and enhance robot performance and adaptability. While significant achievements have been made in this research, it is clear that there is more work to be done, especially in fine-tuning implicit instruction interpretation and adapting to diverse user populations. As we look to the future, this dissertation sets the stage for ongoing efforts to optimize machine learning algorithms and enhance robots' capabilities, ultimately advancing the field of autonomous machine learning for human-robot collaboration and fostering more equitable, inclusive, and effective partnerships in the years to come.

Acknowledgements

This thesis was conducted at the Graduate School of Creative Science and Engineering, Waseda University. I am deeply grateful for the support provided by the Waseda University Global Robot Academia Institute, the Waseda University Green Computing Systems Research Organization, and the Waseda University Open Innovation Ecosystem Program for Pioneering Research (W-SPRING).

Firstly, I extend my heartfelt thanks to my supervisor, Professor Hiroyasu Iwata, for his invaluable suggestions and comments on this research. My gratitude also goes to the secretaries of the Iwata Laboratory and all staff associated with the Leading Graduate Program in Embodiment Informatics for their patience and support.

I would also like to express my appreciation to my supervisor during my internship at the Organisation for Economic Co-operation and Development (OECD), Mr. Olivier Cattaneo, and section heads Mrs. Jasmine Ahmad, Mrs. Julia Benn, and Mrs. Ida McDonell, for their advice, interest, and invaluable insights. Special thanks to Mr. Jean-Baptiste Jacouton at the Agence Française de Développement for his kindness and hours of discussions on our joint project. These individuals broadened my vision of Artificial Intelligence and its applications in Economics, Policy Making, and Political Sciences.

I am also grateful to the other members of the Development Co-operation Directorate, including Mr. Tomas Hos, Mr. Giorgio Gualberti, Mr. Harsh Desai, and Mr. Mark Baldock, for their constant support in both professional and personal matters. A special thanks to fellow Spanish interns, Ms. Alba Vazquez and Ms. Maria Micaela Lineares Gomez, who made me feel right at home in Paris during the six months we spent together.

My sincere thanks go to Dr. Yusuke Tsuboko and Dr. Ryosuke Tsumura, who have been a continuous source of support and encouragement throughout both my master's and doctoral degrees. I am also deeply thankful to my classmate, Dr. Mahiro Sawada, for his unwavering kindness and assistance, despite my numerous requests for help.

I extend my appreciation to the members of our research group whose efforts made this thesis possible: Dr. Yukiko Iwasaki, Mr. Shota Takahashi, Mr. Koki Nakabayashi, Mr. Kohei Amano, Mr. Ryutaro Matsumoto, Mr. Kozo Ando, and Mr. Shuhei Iizuka. Special thanks to Mr. Kenta Aragane and Mr. Ryo Udagawa from the Graduate School of Advanced Science and Engineering for their discussions and perspectives during our countless hours of work together. Additionally, I would like to thank the former and current members of the Iwata Laboratory: Mr. Daichi Hirai, Mr. Ryota Takada, Mr. Shun Miyabayashi, Mr. Yuichi Mizukoshi, Mr. Koki Izumi, Mr. Yoshiyuki Isozaki, Mr. Ryohei Saito, Mr. Shutaro Motohashi, Mr. Shutaro Toriya, Ms. Akira Koshino, Mr. Ayumu Tsuji, and Mr. Genichiro Tanaka, who made these five years more enjoyable, full of new experiences and laugheter, and allowed me to significantly improved my Japanese.

Finally, I would like to thank my family and friends for their unwavering support throughout these many years. Your encouragement and understanding have been invaluable.

Tokyo, June 2024
Lena Juliette Guinot

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

1.1.1 Human Machine Interaction

Human-robot interaction (HRI) is an evolving field of study that focuses on the dynamics between humans and robots, aiming to develop systems that can work alongside humans effectively and safely. As robots become increasingly integrated into our daily lives, from manufacturing floors to personal assistants and healthcare, understanding and improving the ways in which humans and robots interact is crucial. This introduction provides an overview of HRI, its importance, key challenges, and potential future directions.

The genesis of HRI can be traced back to the early days of robotics and artificial intelligence (AI), where the primary focus was on automating repetitive tasks. However, as technology progressed, the vision expanded to create robots capable of performing complex tasks alongside humans, necessitating a deeper understanding of the interaction between humans and robots. HRI thus emerged as a distinct interdisciplinary field, drawing from robotics, cognitive science, social sciences, design, and engineering to address the multifaceted aspects of robot design, human factors, and interaction dynamics [5].

HRI stands at the forefront of the modern technological revolution, representing a paradigm shift in how humans interact with machines. Its significance lies in its potential to revolutionize industries, enhance quality of life, and address complex societal challenges. In healthcare, robots can assist in surgeries, rehabilitation, and care for the elderly, offering precision and support where human capabilities are limited. In education, interactive robots can facilitate learning and engagement among students. Furthermore, in manufacturing and service sectors, robots that can safely and effectively collaborate with human workers promise to increase efficiency, safety, and job satisfaction [6].

The interdisciplinary nature of HRI reflects the complexity of designing systems that are not only technically proficient but also socially and ethically aligned with human values and needs. This requires integrating knowledge from robotics (for the development of physically capable robots), psychology and cognitive science (to understand human behavior and perception), social sciences (to grasp the social dynamics and cultural impacts of robot integration), and design (to create user-friendly interfaces and interaction experiences).

One of the key challenges in HRI is developing robots that can understand and adapt to human emotions, intentions, and behaviors. This involves sophisticated AI algorithms capable of interpreting human gestures, language, and social cues, as well as ethical considerations regarding privacy, autonomy, and the nature of human-robot relationships. Additionally, ensuring the safety and reliability of robots in diverse environments presents ongoing technical and regulatory challenges [7].

Recent advancements in AI, machine learning, and sensor technology have led to significant progress in HRI. Robots are becoming more autonomous, able to make decisions based on real-time data and learning from interactions. This has enabled more natural and intuitive forms of communication between humans and robots, such as voice and gesture recognition, enhancing the user experience across various applications [8].

Another trend is the increasing focus on personal and social robots designed for companionship, education, and assistance in daily tasks. These robots are being designed with an emphasis on emotional intelligence, capable of providing personalized interactions that can adapt to the user's mood and preferences [9].

Looking forward, HRI is poised to address more complex societal needs, with potential impacts on addressing labor shortages, enhancing accessibility for individuals with disabilities, and providing solutions for sustainable development. As the technology evolves, ethical and societal implications of widespread robot adoption will become increasingly important to address, including issues of job displacement, privacy, and the digital divide.

Moreover, the future of HRI will likely see a greater emphasis on co-evolution, where humans and robots learn from each other in a symbiotic relationship. This approach could lead to the development of more adaptive, resilient, and empathetic robotic systems, capable of functioning in unpredictable environments and forming meaningful partnerships with humans [10].

1.1.2 Types of Interaction

One of the core aspects of HRI research involves categorizing the types of interactions that can occur between humans and robots. Understanding these interactions is crucial for designing robots that can effectively work in human-centric environments. This subsection delves into the primary types of interactions in HRI, namely coexistence, cooperation, and collaboration, highlighting their differences and implications for robot design and deployment.

1. Coexistence: Coexistence refers to the scenario where humans and robots share a common space but interact minimally with each other. In this type of interaction, the robot and human do not share a direct working relationship or goal. Instead, the robot operates independently within the same environment as humans. Coexistence requires significant attention to safety protocols to ensure that the robot's presence does not adversely affect the human inhabitants and vice versa. Typical applications include robotic vacuum cleaners in home environments or autonomous mobile robots navigating through warehouses where humans are present. The key characteristic of coexistence is the independence of goals and tasks between humans and robots. As such, the design focus for robots in coexistence scenarios is primarily on sensor technology for obstacle avoidance, environmental awareness, and adherence to safety standards. For instance, studies like those by Sisbot and Alami [11] emphasize the need for robots to maintain safe distances from humans and to predict human movements to avoid collisions.

2. Cooperation: Cooperation in HRI occurs when humans and robots work towards separate but related goals that require some level of interaction. Unlike coexistence, cooperation involves indirect interaction where the activities of one party influence the performance of the other. A typical example is a robotic arm in a manufacturing setting that performs tasks such as welding, which is interconnected with a human's task of assembling parts. In cooperative interactions, although the goals are distinct, there is a need for coordination between the human and the robot. This requires the robot to have a higher level of intelligence and adaptability to respond to human actions and possibly adjust its behavior based on human cues. The research conducted by Li et al. [12] illustrates how robots can use cues from human motion to adjust their trajectory or speed in real-time to facilitate smoother cooperation.

3. Collaboration: Collaboration represents the most integrated form of interaction in HRI, where humans and robots work together on a shared task with a common goal. This interaction requires not just parallel activity but a synchronized effort, often involving back-and-forth exchanges and adjustments based on mutual feedback. An example of collaboration is a human-robot team assembling a complex electronic device, where both the human and robot must adapt their actions responsively to each other's inputs. Collaborative robots, or "cobots," are designed with capabilities such as gesture recognition, natural language processing, and advanced decision-making algorithms to engage in this dynamic interaction. A significant aspect of collaborative interaction is the robot's ability to predict human intent and adapt its actions accordingly to enhance team efficiency. Research by Breazeal et al. [13] highlights how robots can effectively interpret human gestures and verbal commands to optimize joint task performance.

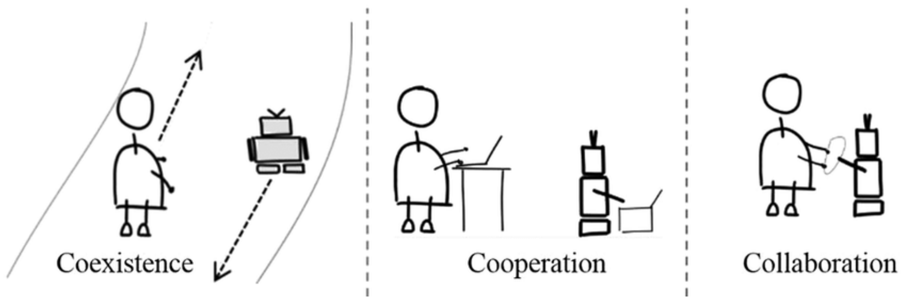


Figure 1.1: Types of interaction

Understanding the distinctions between coexistence, cooperation, and collaboration in human-robot interaction is crucial for the development of robotic systems that are capable of functioning effectively in diverse human environments. Each type of interaction requires different capabilities from robots, from basic safety protocols in coexistence to advanced cognitive abilities in collaboration. As robotic technology advances, the potential for more nuanced and effective interactions in each of these categories continues to expand, promising significant implications for future human-robot teams.

This analysis provides a clear distinction between the types of interactions that can occur in HRI, emphasizing the increasing complexity and interactivity required from coexistence to collaboration. Understanding these interactions aids in the targeted development and implementation of robotic systems across various sectors.

1.1.3 Robot Learning in Human-Robot Interaction

Robot learning, a subset of machine learning, involves developing algorithms that enable robots to learn from experiences, adapt to new situations, and refine their actions over time. In the context of HRI, robot learning is particularly important as it allows robots to understand and predict human actions, preferences, and needs, thereby facilitating smoother and more effective interactions[14]. This capability is crucial for robots designed for collaborative tasks, where they must work alongside humans in environments ranging from industrial settings to personal assistance and healthcare.

In collaborative HRI, robot learning plays a pivotal role in enabling robots to participate as proactive partners rather than passive tools. This shift requires robots to not only react to human instructions but also anticipate human needs and adapt their behavior to complement human actions dynamically. For instance, in a manufacturing context, robots equipped with advanced learning algorithms can adjust their movements to match the pace and style of their human co-workers, enhancing efficiency and reducing the risk of accidents [15].

Furthermore, robot learning facilitates the development of personalized interactions, where robots can adjust their communication style, level of assistance, and behavior based on individual user preferences and performance. This personalization is particularly valuable in educational settings and rehabilitation, where tailored approaches can significantly impact the effectiveness of the collaboration [16].

Despite its potential, integrating robot learning into collaborative HRI presents several challenges. One major challenge is ensuring that robots can learn effectively from limited or noisy human-generated data. Humans often provide inconsistent or imprecise feedback, requiring robots to have robust learning algorithms that can generalize from imperfect inputs [17].

Another challenge is developing models that can predict human behavior in real-time, a necessity for seamless collaboration. This requires not only sophisticated algorithms but also significant computational resources, posing technical and practical constraints on robot design [18]. Safety is also a paramount concern, as robots must learn to adapt their behaviors without endangering humans. This necessitates the development of learning algorithms that prioritize safety and ethical considerations in decision-making processes [19].

Research in robot learning for collaborative HRI has led to several notable contributions. For example, studies have developed algorithms that enable robots to learn from demonstration, allowing non-expert users to teach robots tasks by simply showing them the desired actions [14]. Other research has focused on reinforcement learning, where robots learn optimal behaviors through trial and error, guided by feedback from their human partners [17]. Looking ahead, one promising direction is the integration of social learning mechanisms, where robots can learn not just from direct interaction but also by observing human behavior and interactions. This approach could enable robots to acquire complex social behaviors and norms essential for effective collaboration [20].

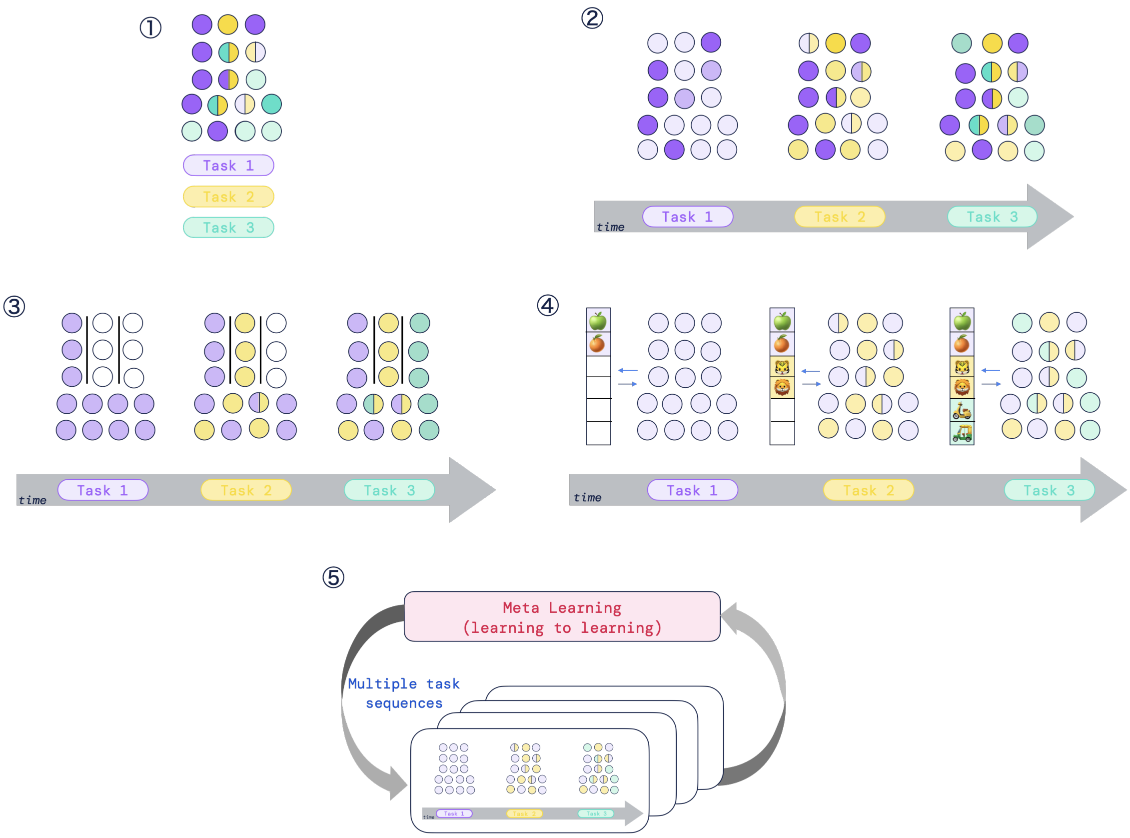


Figure 1.2: Types of interaction

1.1.4 Personalization of Interaction

Personalization in HRI and social robotics is a dynamic field that aims to enhance the usability and effectiveness of robots by tailoring their behaviors to individual human needs and preferences. As robots increasingly become part of everyday life, the ability to adapt to and learn from interactions with humans over extended periods becomes essential. This paper explores the significance of lifelong learning in HRI, focusing on the concept of Lifelong Learning and Personalization in Long-Term Human-Robot Interaction (LEAP-HRI). This approach highlights how robots can evolve and adapt through continual interactions, thereby improving their functionality and acceptance in human societies.

Social robotics involves the deployment of robots in roles that require engagement and interaction with humans in a social context, such as in homes, schools, hospitals, and workplaces. The success of these robots is largely measured by their ability to establish and maintain meaningful interactions with users. Personalization in social robotics enables these machines to learn from previous interactions, adapt their behaviors, and predict individual preferences, which significantly enhances user satisfaction and engagement. Studies have shown that personalized robots can achieve higher levels of user engagement and are more likely to be perceived as helpful and friendly compared to generic robots. For instance, Tapus et al. [21] demonstrated that robots that adapted their personality and behaviors to match user preferences resulted in improved interaction quality and increased user satisfaction.

Lifelong learning in HRI is grounded in the development of algorithms that enable robots to accumulate knowledge continuously, adapt to new circumstances, and apply past learning to new situations without human intervention. This is particularly important in environments where human behaviors and preferences evolve over time, requiring robots to dynamically adjust their behaviors. Several approaches have been proposed for implementing lifelong learning in robots. One approach is reinforcement learning, where robots learn optimal behaviors through trial and error interactions with the environment. Another approach is supervised learning, where models are trained on a dataset of labeled interactions, allowing the robot to learn from explicit examples. However, both methods have limitations in dynamic social environments where unpredictability and the requirement for immediate adaptation are high.

The LEAP-HRI framework addresses these challenges by integrating advanced machine learning techniques with models of human behavior to facilitate deeper personalization in long-term interactions. This framework focuses on three core areas:

- **Adaptation and Learning:** LEAP-HRI emphasizes the importance of adaptive algorithms that can update and refine their strategies based on ongoing interactions. This involves using techniques from online learning and model-based reinforcement learning to ensure that the robot's behavior remains aligned with changing human preferences and environments.

- **Personalized Interaction Models:** Central to LEAP-HRI is the development of personalized interaction models that predict individual user needs and preferences. These models leverage data collected over time to enhance the robot's understanding of each user, enabling more tailored and responsive interactions.

- **Contextual Awareness:** Recognizing and responding to contextual cues is vital for effective personalization. LEAP-HRI incorporates contextual data into decision-making processes, allowing robots to understand the situational dynamics and adjust their behaviors accordingly.

The implications of LEAP-HRI are profound across various domains where personalized interactions are crucial. In healthcare, for example, robots can provide personalized support to

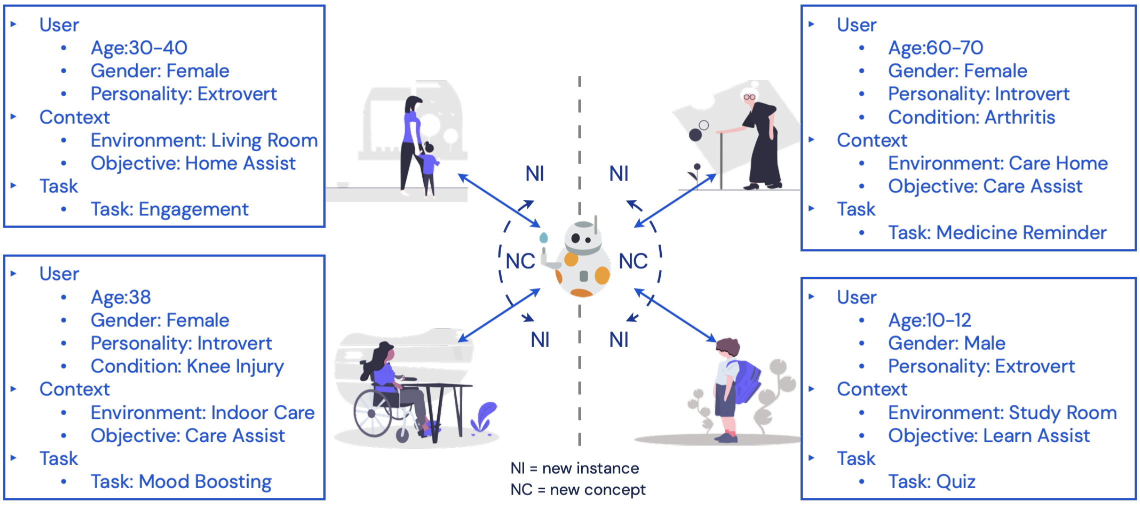


Figure 1.3: Representation of interaction personalisation requirements for HRI

patients with chronic illnesses by learning their preferences for medication reminders or physical activities. In education, personalized learning companions can adapt to the learning pace and style of individual students, enhancing educational outcomes.

Despite its potential, implementing LEAP-HRI poses several challenges. These include the complexity of designing algorithms that can handle the vast amount of data generated in long-term interactions, privacy concerns associated with data collection, and the ethical implications of increasingly autonomous personalized robots. Future research in LEAP-HRI should focus on developing robust privacy-preserving mechanisms that ensure data security while enabling personalization. Additionally, ethical frameworks that guide the development and implementation of personalized robots must be established to address potential biases and discrimination.

Lifelong learning and personalization are critical to the future of human-robot interaction, particularly in social robotics. The LEAP-HRI framework presents a promising approach to achieving deeper and more meaningful interactions between humans and robots by focusing on adaptation, personalization, and contextual awareness. As this field progresses, it will be crucial to address the technical, ethical, and societal challenges to fully realize the benefits of personalized robots in enhancing human lives.

1.1.5 Implicit Interaction in HRI: Understanding User Intentions

Implicit interaction, particularly in the context of HRI, refers to the robot's ability to understand and respond to unspoken cues or unintended human actions. Unlike explicit interactions, where commands or explicit feedback are given to the robot, implicit interactions rely on the robot's capacity to perceive and interpret subtle human behaviors, environmental contexts, and non-verbal cues to infer user intentions. This capability is crucial for creating intuitive and seamless interactions between humans and robots, enhancing user experience and efficiency.

Implicit interaction involves indirect or unobtrusive communication methods, such as the robot sensing the user's presence, gestures, gaze direction, or physiological responses. User intention understanding within this context refers to the robot's ability to infer what a user intends to do based on these indirect inputs. This capability is significant for developing autonomous systems that can proactively assist users without explicit commands. Research in this domain often highlights the contrast between explicit interactions (where commands are clearly given to the robot) and implicit interactions, which require the robot to make inferences based on observational data. Such interactions can make robotic systems appear more intelligent and sensitive to user needs.

The technological basis for enabling implicit interactions includes sensors, machine learning algorithms, and context-aware computing. Sensors such as cameras, microphones, and wearable devices capture a wide array of data about a user's behavior and environment. Machine learning algorithms then process this data to detect patterns and infer intentions, often utilizing techniques from fields such as pattern recognition, natural language processing, and computer vision.

Cameras enable robots to interpret user gestures, facial expressions, and body language. Research by Admoni et al. [22] discusses how visual cues, particularly eye gaze and body orientation, can inform a robot about a user's focus of attention and likely next actions. Microphones capture vocal tones and speech patterns, which can be analyzed to gauge user emotions and intentions without explicit verbal instructions. The work by Marge et al. [23] illustrates how variations in voice can indicate stress levels or urgency, prompting a robot to adjust its behavior accordingly. Utilizing data from the environment, such as the location of objects and the user' s proximity to these objects, can also provide clues about potential user actions. Context-aware computing plays a crucial role here, as noted by Park et al. [24], who developed a context-aware robot that adjusts its operations based on the time of day and the user's typical schedule.

Understanding user intentions through implicit interactions has broad applications across various sectors: 1) healthcare, in rehabilitative environments, robots can adjust therapy sessions based on subtle cues from patients indicating discomfort or fatigue. 2) Home Assistants: Domestic robots can anticipate needs by observing routine activities, such as preparing coffee in the morning or turning down the bed at night without explicit instructions. 3) Workplace: In industrial settings, robots might predict the need for supplies or assistance by observing workers' behaviors and tool usage patterns, streamlining operations without interrupting the workflow.

While promising, the field of implicit interaction faces several challenges:

- Accuracy of Intention Inference: The accuracy with which intentions are inferred from implicit cues is not always reliable, leading to potential errors in response.

- Privacy Concerns: The extensive data collection required for these interactions raises significant privacy issues, necessitating robust data protection measures.

- Ethical Considerations: There are ethical implications in decision-making by robots, especially when actions are taken based on inferred, rather than explicitly stated, user consent.

Future research in implicit interaction must address these challenges while enhancing the robustness and applicability of inference mechanisms. Developing multimodal sensing techniques that combine several types of sensory information can improve accuracy and reliability. Furthermore, the integration of ethical considerations into the design of algorithms that handle implicit data is critical to fostering trust and acceptance among users.

Implicit interaction represents a transformative approach in HRI, enabling robots to act in a more human-like and anticipatory manner by understanding user intentions from non-explicit cues. As technology advances, the scope for more intuitive and effective user-robot interfaces will expand, significantly impacting how humans and robots coexist and cooperate in various environments.

1.2 Challenges of Human-Robot Interaction and Collaborative Work

As robots and AI systems become more prevalent in society, their integration into daily activities and professional fields necessitates a seamless coexistence with humans. This integration highlights the importance of effective HRI and collaborative work, which are critical for the success and acceptance of robotic systems in various environments. However, the expansion of robots into areas requiring close collaboration with humans of differing expertise levels presents several challenges, particularly regarding interaction methods, safety, and trust.

- Interaction Methods

One of the primary challenges in HRI is developing interaction methods that are intuitive and accessible to users with varying levels of technical expertise. Traditional interfaces and command structures may be suitable for users with a background in robotics or engineering but can be inaccessible or intimidating to those without such experience. The diversity in user expertise necessitates the development of adaptive and user-friendly interaction mechanisms that can accommodate a wide range of users, from novices to experts [5]. Non-verbal communication methods, such as gestures, facial expressions, and body language, offer a potential solution to this challenge by enabling more natural and intuitive forms of interaction. However, accurately interpreting these cues requires sophisticated sensing and processing capabilities, which can be difficult to achieve in practice. Additionally, the variability in human behavior and cultural differences can lead to inconsistencies in robot responses, potentially confusing or frustrating users [25].

- Safety Issues

Safety is a paramount concern in HRI, especially in environments where robots and humans work closely together. The potential for physical harm to humans, either through malfunction or misinterpretation of commands, necessitates stringent safety protocols and mechanisms to prevent accidents. This challenge is compounded by the need for robots to operate in dynamic and unpredictable environments, where they must be able to recognize and adapt to potential hazards in real-time [26]. Developing robots capable of such adaptability requires advanced sensing and decision-making capabilities, as well as robust safety features that can quickly deactivate or redirect the robot in the event of a potential danger. However, implementing these features without compromising the robot's functionality or the fluidity of interaction presents a significant engineering challenge.

- Trust

Trust is a critical component of effective HRI and collaborative work. For robots to be successfully integrated into society and various professional spheres, users must trust that they will perform their tasks reliably and safely. Building this trust requires not only ensuring the technical reliability of robots but also addressing users' perceptions and attitudes toward robotic systems [27]. Trust can be influenced by a variety of factors, including the robot's appearance, behavior, and the transparency of its decision-making processes. Robots that appear too machine-like or behave unpredictably may be less likely to be trusted by users. Similarly, if users do not understand how a robot makes decisions, they may be less inclined to trust its actions, especially in critical or sensitive tasks.

Addressing these challenges requires a multidisciplinary approach that combines advancements in robotics and AI with insights from psychology, sociology, and ethics. Developing interaction methods that are both intuitive and adaptable to users of varying expertise levels may involve leveraging emerging technologies such as augmented reality (AR) and machine learning to create more immersive and personalized interaction experiences [28, 20]. Ensuring safety in HRI necessitates not only the development of advanced sensing and processing capabilities but also the creation of comprehensive safety standards and guidelines specific to robotic systems. These standards should be informed by ongoing research into human-robot collaboration and updated regularly to reflect technological advancements and emerging use cases.

Building trust in robotic systems involves improving both their reliability and the transparency of their operations. This can be achieved through rigorous testing and validation processes, as well as efforts to educate users about the capabilities and limitations of robotic systems. Additionally, designing robots with more human-like appearance and behavior may help to make them more relatable and trustworthy to users [29].

1.3 Purpose of this Study

The primary purpose of this study is to advance the field of human-robot interaction (HRI) by developing an innovative approach that leverages attention mechanisms and memory-enhanced neural networks. This approach aims to endow robots with a nuanced understanding of human body motion, facilitating a more intuitive and adaptable interaction paradigm based on the ideomotor principle. Unlike conventional HRI systems that focus on mastering a broad spectrum of tasks, this research emphasizes creating robots capable of interacting with a diverse array of users. This shift in focus addresses a critical gap in current HRI research: the ability of robots to adapt to user-specific data, especially when collaborating with unknown users who may communicate their intentions through unique motion cues. Background and Significance The intricacies of human motion represent a rich source of information that, when accurately interpreted by robots, can significantly enhance collaborative efforts. The ideomotor principle, which suggests that merely thinking about an action can facilitate its execution, provides a theoretical framework for developing HRI systems that anticipate and respond to human movements in a seamless and anticipatory manner. By focusing on the kinematic awareness of human body motion, this study seeks to enable robots to understand and predict a wide range of user intentions, thereby improving collaboration efficiency and user satisfaction.

1.3.1 Research Objectives

1. Develop a Neural Network Architecture with Attention Mechanisms and Memory Enhancement: To process and interpret the complex dynamics of human motion, incorporating attention mechanisms that focus on relevant cues and memory components that retain essential information over time.
2. Enable Kinematic Awareness in Robots: To achieve a detailed understanding of human body motion and its implications for task execution, facilitating a deeper level of nonverbal communication and collaboration based on the ideomotor principle.
3. Adapt to User-Specific Data: To create a system that learns from interactions with individual users, allowing the robot to adjust its behavior and predictions based on unique user profiles and motion cues, thereby handling a wide variety of human behaviors and preferences.
4. Evaluate the System in Collaborative Tasks: To test the developed system in a range of collaborative scenarios, assessing its ability to adapt to new users and improve task efficiency and user experience.

1.3.2 Methodology

This study will employ a mixed-methods approach, combining quantitative analyses of robot performance in collaborative tasks with qualitative feedback from users. The development of the neural network architecture will involve iterative testing and refinement, with particular attention to the integration of attention mechanisms and memory components. User studies will be conducted to collect data on motion cues and intentions, which will be used to train and validate the neural network model.

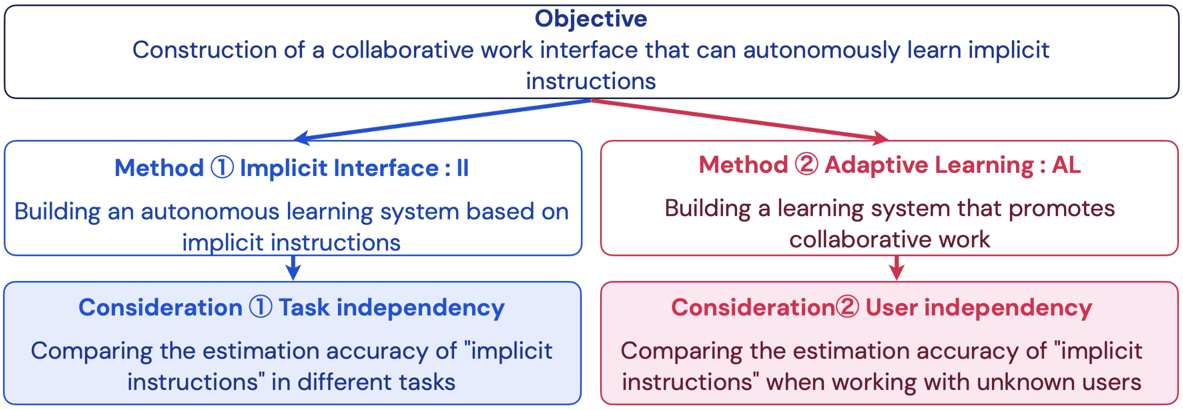


Figure 1.4: Objectives of this Study

1.3.3 Expected Contributions

Advancement in HRI Technologies: By focusing on kinematic awareness and user-specific adaptation, the study aims to push the boundaries of what is currently possible in HRI, particularly in terms of non-verbal communication and collaboration.

Theoretical Implications: The application of the ideomotor principle in the context of HRI provides a novel approach to designing interactive robots, contributing to the theoretical understanding of anticipatory and adaptive robot behavior.

Practical Applications: The development of robots capable of adjusting to individual users has significant implications for a variety of fields, including healthcare, education, and manufacturing, where personalized assistance and collaboration are crucial.

1.4 Related Research

1.4.1 Continual Learning

The concepts of continual learning and long-term collaboration in HRI have garnered significant attention in recent years. Continuous learning in robots, also known as lifelong learning, involves the ability of a robot to learn from new data continuously and adapt its knowledge base without forgetting previously acquired information (Figure 1.5). Long-term collaboration refers to the capacity of robots to work alongside humans over extended periods, adapting to changes in human behavior, preferences, and environments. This subsection reviews the literature surrounding these concepts, highlighting key studies, methodologies, and findings that have shaped current understanding and practices.

Continuous learning represents a paradigm shift in robotics, moving away from static programming to dynamic, adaptive systems that can evolve with their operational environment. Thrun and Mitchell [30] laid foundational work in this area, discussing the need for robots to retain and refine knowledge over time. More recent studies have focused on overcoming the challenges of catastrophic forgetting, where new learning interferes with previously stored knowledge. Parisi et al. [31] reviewed approaches to mitigate this issue, including elastic weight consolidation and replay mechanisms, which have shown promise in maintaining long-term knowledge retention in neural networks.

The effectiveness of long-term human-robot collaboration hinges on the robot's ability to understand and predict human actions, adapt to individual preferences, and maintain a level of engagement that is both productive and satisfying for the human user. Trafton et al. [32] explored the role of theory of mind in robots, enabling them to attribute beliefs, desires, and intentions to their human counterparts, thereby facilitating smoother interactions over time.

Similarly, Breazeal et al. [33] investigated social robots in educational settings, finding that personalization and adaptability were key to maintaining student engagement and improving learning outcomes over prolonged periods.

Integrating continual learning capabilities within the context of long-term collaboration presents unique challenges and opportunities. Robots must not only adapt to immediate task requirements but also anticipate future changes in human behavior and task contexts. Studies by Irfan et al. [34] have demonstrated the potential of reinforcement learning and predictive modeling in enabling robots to adjust their strategies based on human feedback and evolving task parameters. Furthermore, research by Lemaignan et al. [35] highlighted the importance of shared knowledge bases and mutual adaptation in maintaining effective collaboration over time.

Despite progress, several challenges remain in fully realizing the potential of continual learning and long- term collaboration in robotics. One major hurdle is the scalability of current learning algorithms to complex,real-world environments. Additionally, ensuring the privacy and security of shared information, particularly in sensitive applications like healthcare or personal assistance, is critical. Future research directions may include developing more efficient and robust learning algorithms, exploring multi-modal learning approaches, and enhancing robots' emotional intelligence to better understand and respond to human affective states.

The fields of continual learning and long-term collaboration in robotics are rapidly evolving, with significant implications for the future of HRI. As robots become more integrated into daily life and work, the ability to learn continuously and collaborate effectively with humans over extended periods will be paramount. Ongoing research in these areas holds the promise of creating more adaptive, intelligent, and empathetic robotic systems capable of supporting human endeavors in increasingly sophisticated ways.

1.4.2 Continual Learning for HRI

The integration of artificial intelligence and human-robot interaction (HRI) has ushered in a new era of intelligent robotic systems that transcend traditional industrial applications to become assistants, tutors, and companions in everyday life [36]. These robots are designed to understand and support human cognitive and socio-emotional well-being through social interactions. Central to enhancing HRI is the domain of affective robotics, which focuses on the interpretation of human socio-emotional signals [37, 7]. This research area, although still burgeoning and fraught with challenges [6, 38], is pivotal for robots aimed at providing effective physical and social support across various domains including healthcare, education, and entertainment.

Affective robotics strives to comprehend and model the nuances of human behavior in real-life scenarios, a complex undertaking that necessitates an understanding of nonverbal cues such as gestures, posture, facial expressions, and vocal outbursts [39]. The objective is to enable robots to perceive and interpret these cues to understand higher-level social phenomena like emotions, engagement, and interpersonal relationships, and to respond appropriately [40, 41]. However, current learning-based approaches, while effective in controlled lab settings, often fail to generalize these capabilities to dynamic real-world interactions [42, 43, 44]. This limitation underscores the need for models that not only generalize across diverse scenarios but also personalize the interaction experience by adapting to individual user differences influenced by factors such as culture, gender, and personality [45, 46].

Continual learning addresses the challenge of long-term adaptability in intelligent agents, proposing a shift from traditional machine learning paradigms that assume a static data distribution [30, 31]. Continual Learning involves learning incrementally from data acquired through ongoing interactions, making it particularly suitable for environments where data distributions evolve with each user or task. Although primarily applied in object recognition or task-specific learning [31, 47], the principles of Contunal Learning can be extended to affective robotics. This

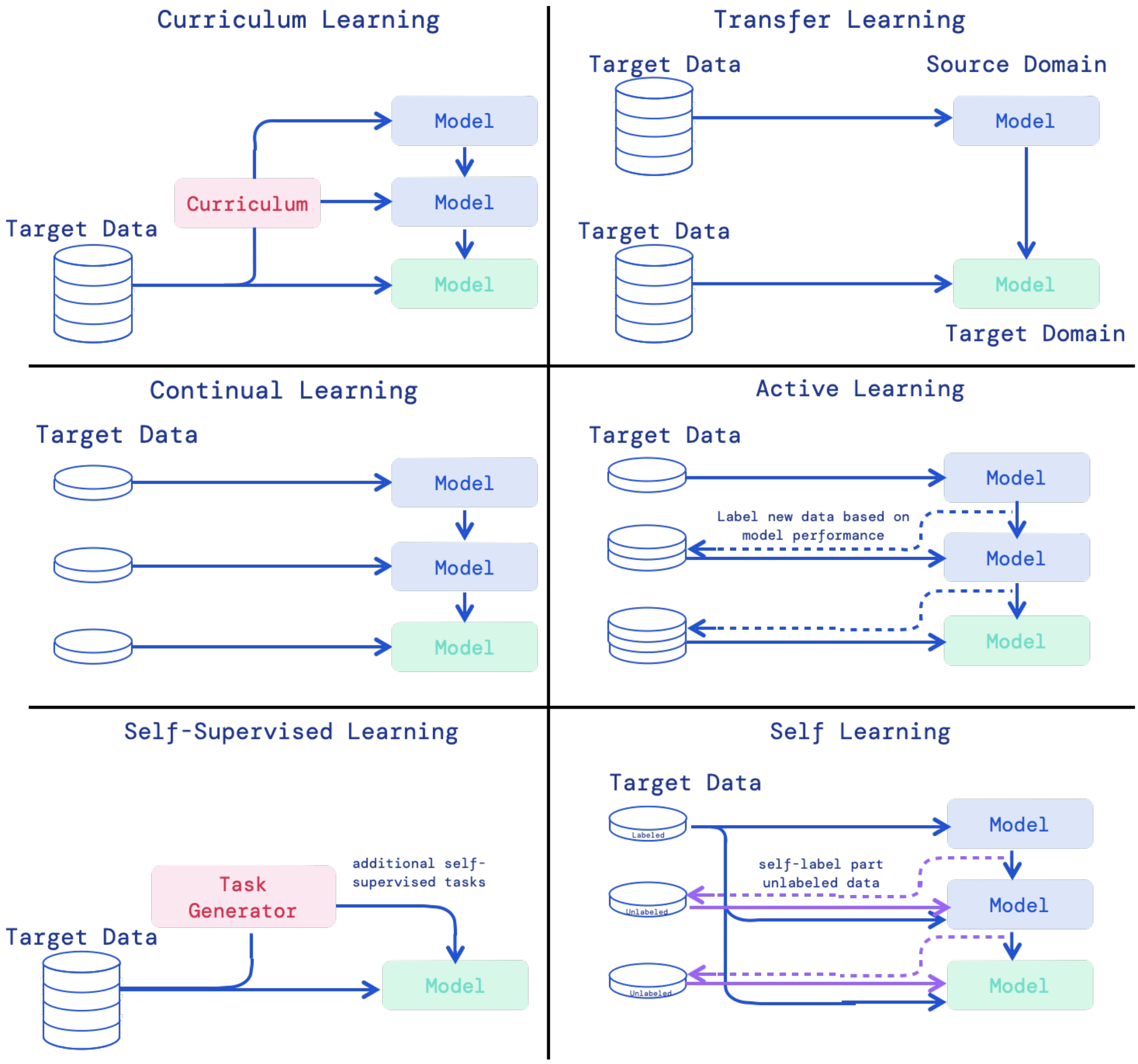


Figure 1.5: Types of Meta-learing approaches

extension involves continual adaptation to new socio-emotional behaviors and states observed during interactions with users [48, 49, 50], enabling robots not only to respond to current cues but also to anticipate future behaviors [51].

Continual Learning allows robots to refine and update their socio-emotional perception models continually. This involves learning from each interaction to enhance understanding and prediction of user emotions and behaviors over time, which is crucial for maintaining engaging and supportive relationships. By incorporating Continual Learning, robots can personalize their responses based on the accumulated knowledge of individual user preferences and behaviors. This personalization extends to adapting responses based on the context, such as changing environmental conditions or varying task requirements, as illustrated in various studies [45, 46]. Future frameworks for Continual Learning in affective robotics should focus on creating adaptable, robust models capable of handling the complexities of real-world interactions. This involves integrating multi-modal data (visual, auditory, and contextual cues) and employing advanced machine learning techniques that support incremental learning and data privacy.

Continual Learning represents a transformative approach for advancing affective robotics within human-robot interaction. By enabling robots to adapt continually to new and evolving socio-emotional cues, Continual Learning not only enhances the relevance and timeliness of robotic responses but also significantly improves the personalization of interactions. As this field progresses, the fusion of Continual Learning principles with affective robotics will likely lead to more nuanced, sensitive, and engaging human-robot relationships, ultimately fulfilling the promise of robots as true social companions in diverse aspects of daily life.

1.4.3 Current Communication Methods in HRI

Communication between humans and robots has evolved significantly, moving from basic command-and- control interfaces to more complex, multimodal interaction systems. These advancements aim to make interactions more intuitive, allowing robots to understand and respond to human actions, gestures, verbal commands, and even emotional expressions. However, despite these technological strides, current methods still face significant challenges, particularly when robots encounter new users with data or behavior patterns that differ from those on which they were trained.

- Verbal Communication:

Voice commands and natural language processing (NLP) allow users to interact with robots using spoken language. This method's effectiveness largely depends on the sophistication of the robot's language understanding and speech recognition capabilities [52]. While significant improvements have been made, issues such as accent, dialect, and language diversity can hinder understanding, especially in multicultural or noisy environments.

- Non-verbal Communication:

This includes gestures, facial expressions, and body language. Robots equipped with sensors and computer vision algorithms can interpret these cues to understand user intentions or emotions [6]. Gesture recognition, for instance, enables a more natural way of directing robots without the need for verbal commands. However, the variability in human gestures and expressions across different cultures and individuals poses a challenge for consistent interpretation.

One of the main shortcomings of current HRI communication methods is their reliance on predefined models and datasets for training AI algorithms. These models often fail to account for the vast diversity in human behavior, leading to challenges when robots interact with new users whose data diverges from the training set. This discrepancy can result in misunderstandings or incorrect responses from the robot, undermining the interaction's efficiency and user satisfaction. Moreover, the adaptability of robots to new environments or unforeseen situations remains limited. Current systems are generally designed for specific tasks in controlled environments and lack the flexibility to adapt to the dynamic nature of human societies. This limitation is particularly evident when robots are deployed in public spaces or in roles that require interaction with a broad cross-section of the population, where the diversity of human behavior and communication styles is most pronounced.

The challenge of interacting with new users who exhibit unfamiliar data or behavior patterns is magnified as robots become more integrated into everyday society. In areas with few experts or specialized personnel, addressing or correcting misunderstandings or inappropriate robot behavior becomes a significant hurdle. This situation necessitates the development of robots capable of continual learning and adaptation to individual user characteristics and preferences [20]. One approach to overcoming these challenges is the implementation of online learning systems that allow robots to update their models based on real-time interaction data [17]. However, this approach raises concerns about privacy, data security, and the ethical implications of continual data collection.

While current communication methods in HRI have made significant strides in facilitating more natural and intuitive interactions between humans and robots, several challenges remain. The primary issue is the systems' ability to adapt to the diverse and dynamic nature of human behavior, particularly when encountering new users with different communication styles or patterns. Addressing these challenges requires a concerted effort in research and development to create more adaptable, flexible, and ethically responsible robotic systems.

1.4.4 The Ideomotor Principle and its Applications to Human-Robot Interaction

The ideomotor principle, first articulated in the context of psychological research, posits that mental imagery of an action tends to elicit the physical execution of that action without conscious intention [53]. This principle has profound implications for HRI, offering a framework for designing robots that can better understand and anticipate human actions based on subtle cues. However, applying this principle to HRI presents unique challenges, particularly when considering factors like gender and cultural differences.

In HRI, the ideomotor principle can be leveraged to create robots capable of interpreting the intention behind human gestures or movements, facilitating smoother and more intuitive interactions. For example, a robot that recognizes a user's preparatory movements for a task could automatically provide the necessary tools or assistance, enhancing collaboration efficiency [54]. This capability is particularly valuable in environments requiring close cooperation between humans and robots, such as surgical rooms, manufacturing plants, and home settings.

Applications

- Assistive Robotics

Robots designed with the ideomotor principle in mind can better serve individuals with disabilities by anticipating their needs and responding to non-verbal cues, thereby offering more personalized support [14].

- Educational Robots

In educational settings, robots that recognize and react to students' implicit cues could provide more tailored learning experiences, engaging students in a more interactive and responsive manner [55].

- Collaborative Manufacturing

In manufacturing, robots that anticipate human actions can improve safety and productivity by proactively adjusting their behavior to complement human workers, reducing the risk of accidents and enhancing workflow efficiency [56].

While the ideomotor principle holds significant promise for enhancing HRI, its application is not without challenges. One of the primary obstacles is the variability in how different genders, cultures, and individuals express intentions through movement.

Challenges

- Gender Differences:

Research has shown that there can be gender-specific differences in non-verbal communication, which may affect how robots interpret gestures or movements [57]. Robots trained predominantly on data from one gender may misinterpret cues from the other, leading to less effective interaction.

- Cultural Variability:

Cultural background significantly influences non-verbal communication styles, including gestures, personal space, and eye contact [58]. Robots that do not account for these differences may struggle to interact effectively with users from diverse cultural backgrounds, potentially leading to misunderstandings or discomfort.

- Individual Variations:

Even within the same gender and culture, individual differences in non-verbal expression can pose challenges for robots trying to interpret human intentions based on movement. Personal idiosyncrasies in gesture and posture can lead to incorrect interpretations by the robot, hindering effective collaboration.

To overcome these challenges, several strategies can be employed:

- Diverse Data Collection

Ensuring that the data used to train robots encompasses a wide range of genders, cultures, and individual behaviors can improve the ability of robots to accurately interpret a variety of non-verbal cues [59].

- Adaptive Learning Algorithms

Developing robots with the capacity for ongoing learning and adaptation allows them to refine their interpretations of human actions over time, accommodating individual differences and reducing biases [60].

- User Feedback Mechanism

Implementing mechanisms for user feedback can help robots adjust their behavior in real-time, addressing any misinterpretations and enhancing the personalization of interactions [20].

The application of the ideomotor principle to HRI offers a promising avenue for creating more intuitive and effective human-robot collaborations. However, the challenges posed by gender, cultural, and individual differences in non-verbal communication necessitate thoughtful approaches to robot design and training. By addressing these challenges, researchers and developers can enhance the adaptability and sensitivity of robots to the rich diversity of human expression.

1.5 Objective of the Study

The overarching goal of this thesis is to explore and expand the boundaries of human-robot interaction (HRI) by leveraging the ideomotor principle, focusing on the development and implementation of attention mechanisms and memory-enhanced neural networks. These technologies are aimed at equipping robots with a sophisticated understanding of human body motion and the subtle nuances of non-verbal communication cues, thereby enabling more intuitive and adaptive interactions between humans and robots, especially in collaborative tasks. The study is driven by several specific objectives:
1. To Theoretically Ground HRI in the Ideomotor Principle: Investigate the ideomotor principle as a theoretical foundation for enhancing HRI. This involves a comprehensive review of literature to establish a conceptual framework that links the ideomotor principle with current challenges and opportunities in HRI.
2. To Develop Attention Mechanisms and Memory-Enhanced Neural Networks: Design and implement a novel neural network architecture that incorporates attention mechanisms and memory enhancement. This architecture aims to provide robots with the capability to process and interpret the complex dynamics of human motion, focusing on critical cues that indicate intentions and actions.
3. To Achieve Kinematic Awareness of Human Body Motion: Utilize the developed neural network to endow robots with an advanced understanding of human kinematics. This involves recognizing and predicting human movements and adjusting robot actions accordingly to support

seamless human-robot collaboration.
4. To Adapt to User-Specific Data in Real-Time: Enable the neural network to learn from interactions with individual users, thereby allowing the robot to tailor its behavior and responses based on user-specific data. This objective addresses the challenge of interacting with new users whose motion cues and communication styles may differ from those in the training data.
5. To Evaluate the System's Performance in Collaborative Tasks: Conduct empirical studies to assess how well the implemented system facilitates human-robot collaboration across various tasks and settings. This includes evaluating the system's ability to adapt to different users and its effectiveness in enhancing task efficiency and user satisfaction.
6. To Identify and Address Challenges Related to Gender and Cultural Differences: Explore how gender and cultural differences affect non-verbal communication cues in HRI and how the system can accommodate these differences. This entails developing strategies to ensure that the robot's interpretations of human actions are inclusive and respectful of diversity.
7. To Assess Safety and Trust in Human-Robot Collaborations: Evaluate the safety and trustworthiness of the robot in collaborative environments. This includes implementing safety protocols and mechanisms to prevent accidents and designing the system to behave in ways that build and maintain users' trust.

By achieving these objectives, this thesis aims to contribute significantly to the field of HRI, pushing forward the capabilities of robots to understand and interact with humans in a more nuanced, effective, and respectful manner. The ultimate goal is to facilitate the development of robots that are not only technically proficient but also capable of adapting to and learning from the rich tapestry of human behaviors and preferences, thereby enhancing collaboration and coexistence between humans and robots.

1.6 Overview of the Study

This thesis explores the integration of the ideomotor principle into human-robot interaction (HRI), aiming to enhance robotic capabilities in interpreting and responding to human non-verbal cues for improved collaborative work. The study is structured into six chapters, each addressing a distinct aspect of this integration, from system development and preliminary experimentation to addressing performance issues related to gender biases, and finally, evaluating the system's spatial-temporal awareness and adaptability to user specificity.

Chapter 2: Development of the Ideomotor Cue Identification System
Chapter 2 details the development of a novel system designed to identify ideomotor cues and associate them with human intentions, laying the foundational technology for enhancing HRI. This system integrates attention mechanisms and memory enhancements to retain long-term information about user interactions. Preliminary experiments are conducted to verify the existence and importance of ideomotor cues in interpersonal collaborative work, focusing on the system's ability to discern and react to these cues in real- time. The chapter evaluates the system's intuitiveness for users and the quality of collaboration over extended periods, establishing the baseline for further exploration and refinement.

Chapter 3: Gender Performance Discrepancies and Their Impact
In Chapter 3, the study delves into performance discrepancies observed when the robot was used by female versus male users, highlighting the impact of biased datasets on the robot's effectiveness across genders. A comprehensive study is conducted to assess how these discrepancies affect trust, safety, sense of control, and user perception. This investigation sheds light on the critical issue of gender bias in AI systems and its potential implications as robots become more prevalent in society. The chapter emphasizes the importance of addressing

these biases to ensure equitable and effective HRI across all user demographics.

Chapter 4: Implementing Dual Attention for Enhanced Kinematic Understanding
Chapter 4 introduces the implementation of dual attention mechanisms to provide the system with spatial and temporal awareness when processing human motion data. This approach enables a deeper kinematic understanding of human motion and improves the system's adaptability to individual user characteristics. The chapter also discusses the development of actualization policies that allow the system to continuously learn from user discrepancies, thereby enhancing its ability to appropriately respond to user-specific intention communication methods. This adaptive learning process is crucial for developing robots capable of effective and personalized collaboration with a diverse user base.

Chapter 5: Contributions, Limitations, and Future Directions
Chapter 5 synthesizes the overall contributions of the study, highlighting the advancements made in applying the ideomotor principle to HRI. It discusses the limitations encountered throughout the research, including challenges in system development, data collection, and bias mitigation. The chapter outlines potential future steps for the research, such as exploring bidirectional communication pathways that enable robots not only to interpret human cues but also to communicate intentions back to the user using ideomotor cues. This forward-looking perspective underscores the study's contribution to the field and its potential to inform subsequent research and application development.

Chapter 6: Conclusions
The concluding chapter summarizes the key findings of the study, reiterating the significance of integrating the ideomotor principle into HRI for enhancing collaborative work. It reflects on the implications of the research for the design and implementation of future robotic systems, emphasizing the need for continued attention to user diversity, system adaptability, and ethical considerations. The conclusions drawn from this study highlight the potential of ideomotor cues as a means of improving communication and collaboration between humans and robots, setting the stage for further innovations in the field.

This thesis offers a comprehensive exploration of the ideomotor principle's application to HRI, addressing critical challenges and proposing innovative solutions for improving robotic understanding and responsiveness to human non-verbal cues. Through its systematic approach and detailed analysis, the study contributes valuable insights and methodologies to the ongoing development of more intuitive and effective human-robot collaboration systems.

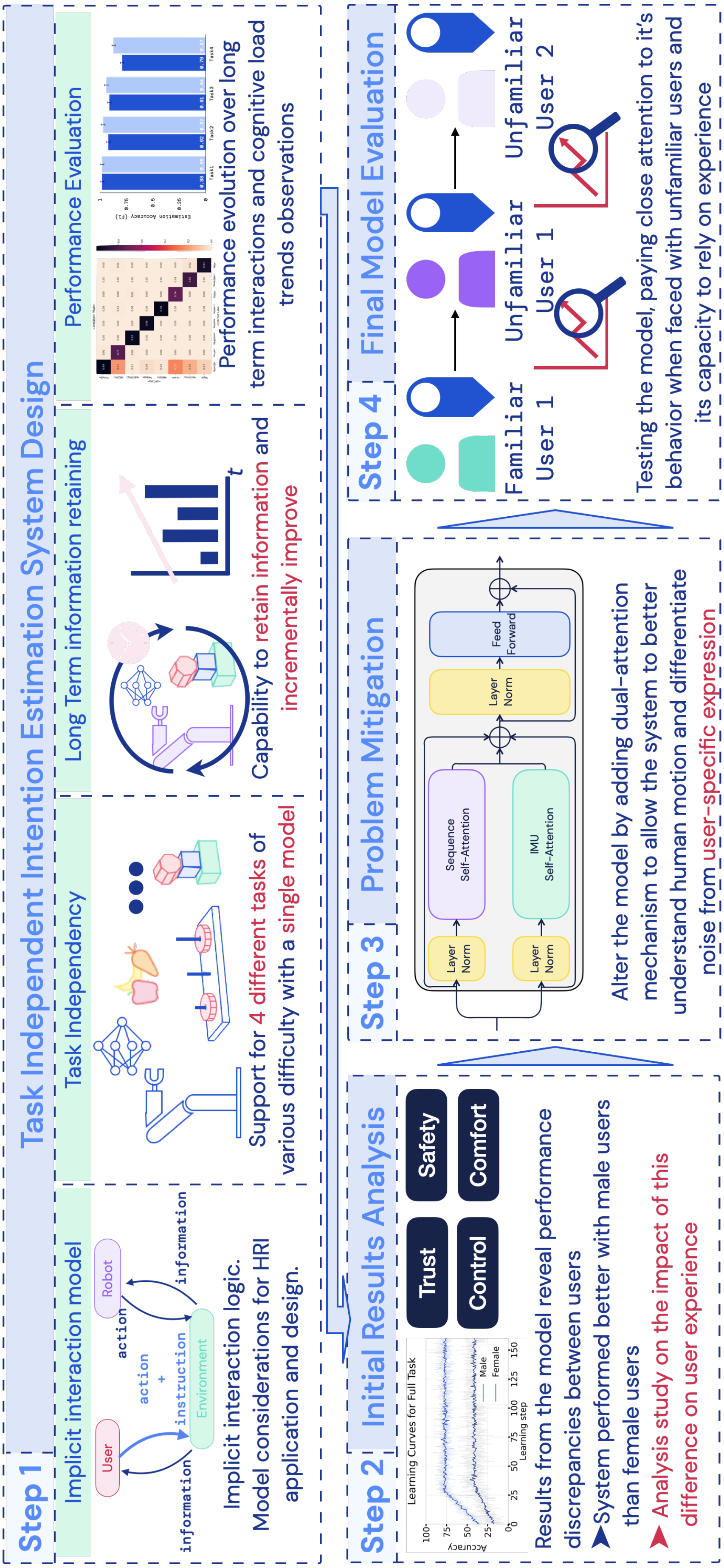


Figure 1.6: Thesis Flow

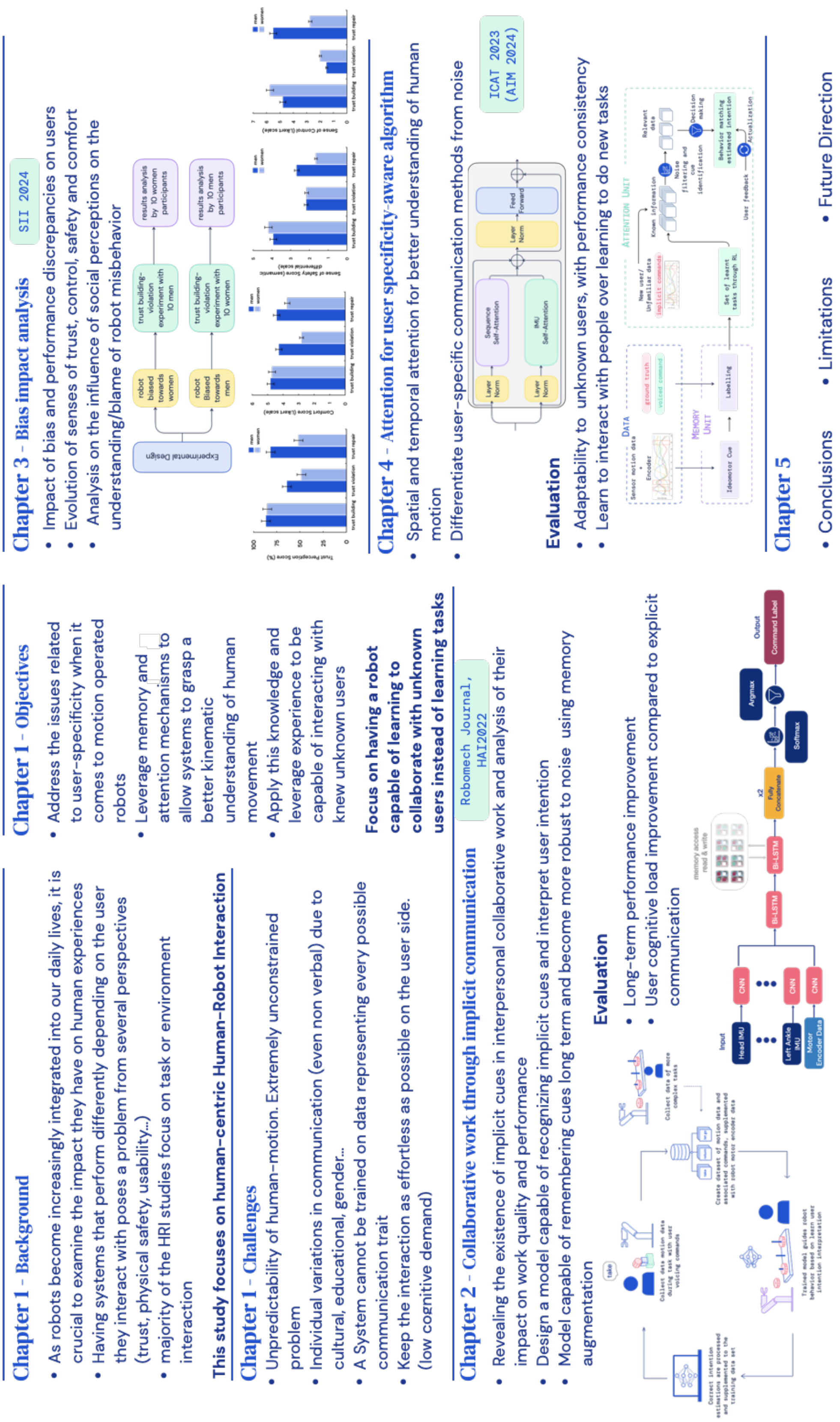


Figure 1.7: Summary of the Thesis by Chapter

Chapter 2

Collaboartive Work Through Implicit Communication

2.1 Revealing Cues through Interpersonnal Communication

2.1.1 Ideomotor and Implicit Cues in Popular Culture

The ideomotor phenomenon, a concept deeply rooted in psychological research, finds intriguing parallels in popular culture, illustrating the universal nature of implicit communication across diverse human interactions. A fascinating example from Japanese culture is the concept of "Aun breathing," which epitomizes the seamless synchronization between individuals without verbal communication. This concept is emblematic of a deeper understanding of non-verbal cues and their significance in human coordination and interaction.

In the realm of traditional Japanese puppet theatre, or Bunraku, Ueda et al.'s study reveals how puppeteers achieve an exquisite level of coordination. The main puppeteer communicates with assistants through implicit signals, known as "Zu," allowing for the synchronized manipulation of a single puppet. These signals, perceptible only to the involved puppeteers, underscore the ideomotor principle's application in achieving harmony and precision in complex tasks [61]. Similarly, Shibuya et al.'s research on stage performers demonstrates how seasoned artists can synchronize their breathing without conscious effort, a phenomenon less observed among less experienced groups. This unconscious coordination, fostered by years of collaboration, highlights the ideomotor principle's role in facilitating intuitive group dynamics [62].

In Western culture, the ideomotor effect was famously demonstrated in the early 20th century by psychologist Oskar Pfungst's investigation of the "clever Hans" phenomenon. Hans, a horse claimed to perform arithmetic calculations, was actually responding to subtle, unintentional cues from his questioners. Pfungst's findings revealed that Hans' "calculations" were responses to physical cues such as slight posture adjustments or facial expressions from the humans around him, illustrating the communicative power of implicit cues [63].

These examples from both Eastern and Western cultures underline the ideomotor phenomenon's significance in interpersonal interactions, transcending cultural boundaries. The exploration of such implicit cues in popular culture not only enriches our understanding of human communication but also opens new avenues for enhancing human-machine interaction (HMI). Recognizing and harnessing these subtle cues can lead to more intuitive and effective collaboration between humans and robots, paving the way for a future where machines can adapt to and learn from the rich tapestry of human non-verbal communication.

2.1.2 Synchronized Work and Implicit Communication Between Individuals

Implicit communication, an essential aspect of human interaction, plays a crucial role in synchronized work, where individuals perform tasks in harmony without the need for explicit verbal instructions. This phenomenon is observed in various contexts, from professional environments to artistic performances, demonstrating the innate ability of humans to coordinate actions based on subtle cues and shared understanding.

One compelling example of synchronized work is found in orchestral performances. Musicians, under the guidance of a conductor, rely heavily on non-verbal cues such as body movements, eye contact, and the conductor's baton movements to maintain tempo and harmony [64]. This synchronization is not solely dependent on the visual cues from the conductor but also on the mutual awareness and anticipation of each musician's part in the ensemble, showcasing a high level of implicit communication and coordination.

Similarly, in the realm of sports, rowing teams exemplify synchronized effort based on implicit understanding. Rowers must maintain a precise rhythm and power in their strokes, often guided by the coxswain's calls but also heavily reliant on the feel of the boat and the movement of their teammates. This synchronization is critical for maximizing efficiency and speed, where even minor discrepancies can disrupt the boat's balance and performance [65]. In professional settings, surgical teams operate with a level of implicit coordination, where surgeons, nurses, and anesthesiologists anticipate each other's needs and actions based on the flow of the procedure and non-verbal signals. This silent communication ensures a seamless operation, minimizing verbal exchanges that could distract from the task at hand [66].

These examples underscore the significance of implicit communication in facilitating synchronized work across diverse domains. They highlight the human capacity to connect and collaborate deeply, an attribute that HRI aims to emulate and incorporate into robotic systems to achieve seamless human-robot collaboration.

2.1.3 Analyzing Implicit Cues in Collaborative Mochi Making

The traditional Japanese rice cake (mochi) making process offers a unique setting to explore implicit cues within a collaborative, synchronized work scenario. This activity, involving two individuals working in tandem—one pounding the dough and the other kneading and flipping it—provides an exemplary model for examining the intricacies of non-verbal communication and coordination in real-time. The requirement for perfect synchronization not only for optimal work efficiency but also for safety, given the potential danger posed by the pounding action, makes mochi making an ideal task for investigating the functioning of implicit cues in collaborative efforts.

Task Setting and Requirements

The mochi-making process satisfies two critical elements for this study: collaboration between two individuals and a dependency on the responsive actions of each participant. The individual handling the dough must swiftly knead and flip it between the pounding actions, a rhythm requiring precise timing and mutual trust to avoid accidents. The differential time demands—longer for flipping the dough due to its sticky texture—add complexity to the task, demanding high levels of synchronization and anticipation.

Data Collection Setup

To measure the potential implicit cues employed during the mochi-making process, a virtual environment was designed using the Unity game engine. This setup included the use of VIVE trackers attached to the pestle and gloves of the participants to record movements with precision, facilitated by infrared signals from virtual reality base stations. The mortar and dough were simulated with a stool and a disk-shaped sponge, respectively, providing a tangible yet safe medium for the experiment. Participants positioned themselves on opposite sides of the stool, mimicking the traditional setup for mochi making.

Experiment Outline

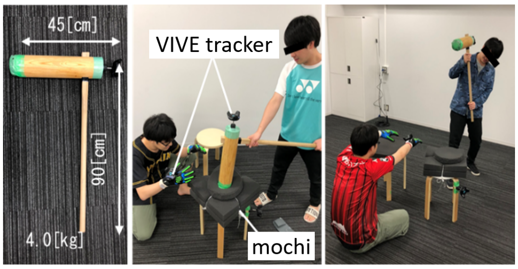


Figure 2.1: Task environment setting

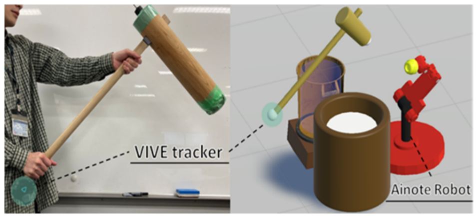


Figure 2.2: Participant repartition

Table 2.1: Qualitative evaluation questionnaire items

Q1	I was able to anticipate the behavior of the other party
Q2	I was able to predict the action required for the situation
Q3	I was able to modify my own behavior according the other party

The experiment aimed to analyze collaboration and detect any implicit cues or ideomotor reactions essential for maintaining synchronization during the task. With one participant pounding the simulated "rice cake" and the other handling the "dough," the early stages of the experiment introduced an auditory signal as a guide for the timing of dough flipping. This controlled environment allowed for the observation of participants' adaptations to the rhythm and each other's movements, shedding light on the non-verbal signals that facilitate synchronized work.

The mochi-making scenario serves as a potent metaphor for the complexities of human-robot interaction. By understanding the implicit cues in such a collaborative context, insights can be gained into designing robots capable of interpreting and responding to human actions with a similar level of intuitive understanding and anticipation. This study underscores the potential of leveraging traditional practices to enhance technological advancements in HRI, aiming to achieve a seamless integration of robots into human-centric tasks.

2.1.4 Experiment

This subsection details an experiment aimed at identifying implicit cues and ideomotor reactions that facilitate synchronization between two individuals during a collaborative task. The study simulated a traditional rice cake making process where one participant is responsible for pounding, and another for kneading and turning the dough. The goal was to detect the non-verbal signals essential for maintaining a synchronized rhythm without explicit communication cues, The experiment utilized an environment described in the "Data collection" subsection, with two participants assigned distinct roles—one for pounding and one for kneading and flipping dough. To establish an initial rhythm, auditory signals were provided at specific intervals, first after 13 seconds and again after 26 seconds, to signal when to turn over the dough. No additional guidance was given, compelling participants to find their pace and maintain synchronization based solely on their partner's actions. Participants, 20 individuals with an almost 2:1 male-to-female ratio and ages ranging from 22 to 35, were split into two groups corresponding to the two roles. Each person in the pounding group completed the task six times for 60 seconds with every member of the kneading group, in a between-subjects study design.

The qualitative assessment employed a Likert scale-based questionnaire (Table 2.1), while the quantitative evaluation focused on the relative distance between the kneading individual's hands and the pestle (Figure 2.3).

This distance, denoted as Δx, was calculated by measuring the positions of the hands x_a

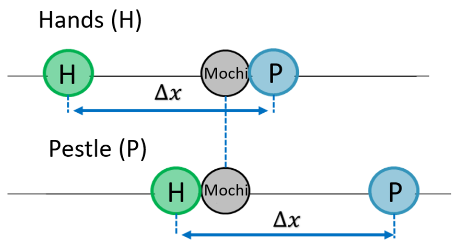


Figure 2.3: Relative Distance between the Hands and the Pestle during task. Top: Pestling phase. Bottom: Kneading phase

and the pestle x_k, with the mortar's center as the reference point $x = 0$. During the action of turning over the dough—which takes longer than kneading—Δx remained constant, even as the hands performed various motions.

$$\Delta x = \sqrt{(X_a - x_k)^2} \quad (\Delta x > 0) \tag{2.1}$$

A stable x across consecutive kneading-pounding cycles indicated good synchronization between participants, while significant variations suggested a lack of coordination.

The Δ x index was chosen to evaluate the smoothness and synchrony of the interaction quantitatively. The less Δx varied from one cycle to the next, the more synchronized the participants were deemed to be. Since individual relative distances varied, comparisons between experiment instances were standardized using the coefficient of variation (C.V.), which is the standard deviation (σ) divided by the mean ($\Delta \bar{x}$).

$$C.V. = \frac{\sigma}{\Delta \bar{x}} \tag{2.2}$$

This metric provided a normalized measure of variation that could be compared across different pairs of participants, irrespective of their individual average distances.

2.1.5 Results and Discussion

This experiment's design illustrates an innovative approach to studying non-verbal communication and synchronization in human collaborative tasks. By employing a traditional activity and minimal initial rhythmic guidance, the study provides insights into the capacity of individuals to adjust to and sync with one another's movements. The qualitative and quantitative evaluation methods complement each other, offering a multi-faceted view of the interaction's effectiveness. The findings from this study are significant for advancing human-robot interaction systems, emphasizing the importance of designing robots that can adapt to human partners' implicit cues and maintain a synchronized workflow.

Furthering the investigation into collaborative quality, a quantitative analysis compared the synchronization level between partners, represented by the averaged coefficient of variation (C.V.) for each pair. Results reflected in Figure 2.4 indicated that explicit pairs exhibited more variance in their rhythm, implying less stable and harmonious collaboration. This outcome highlights the efficacy of non-verbal communication in maintaining a smooth and uninterrupted workflow, potentially leading to a higher quality of cooperative performance. To validate the assumption of implicit communication through unconscious cueing, the research examined participants' motion data at junctures where collaboration smoothness was at risk. These critical points, where mutual understanding was deemed most essential, provided an opportunity to observe how subtle, non-verbal interactions could facilitate coordination and maintain the task's flow.

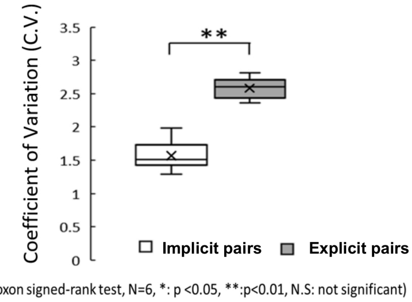
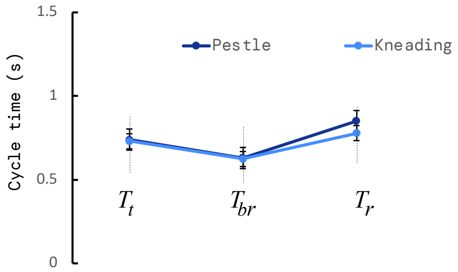
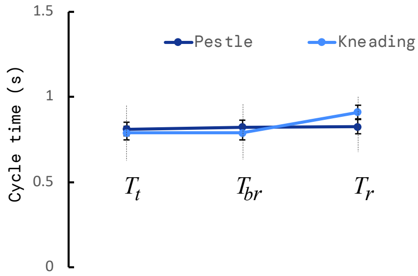


Figure 2.4: Coefficient of Variation of the Relative Distance Δx



(a)



(b)

Figure 2.5: T_t - T_{br} - T_r Evolution for a implicit pair and b explicit pair

In the collaborative rice cake making experiment, the action of turning over the rice cake introduces a significant disruption to the established rhythm of the task, which typically follows a steady punch-knead sequence. This interval is critical as it breaks the continuous flow of collaboration and requires the participants to adapt their movements to maintain synchronization. The study thus centers on the strategies employed by participants to execute the turn-over action as seamlessly as possible, minimizing the impact on the overall task rhythm.

To dissect the effects of the dough-turning action, the motion of the kneading participant was segmented into three distinct phases:

- T_{touch} (T_t): The period of actual kneading where the dough is being manipulated.
- $T_{before\ reverse}$ (T_{br}): The kneading-pounding cycle that immediately precedes the dough turning action.
- $T_{reverse}$ (T_r): The interval during which the dough is turned over.

This categorization provided a structured framework to examine the periodicity and any alterations in motion timing, particularly during the transition from standard kneading to the turning action. Figures 2.5a and 2.5b illustrate the group averages of the time lengths for T_r, T_{br}, and T_t, comparing implicit and explicit collaboration pairs. The findings revealed that the kneading participant's technique significantly influenced the rhythm of the person pounding the rice cake. The adjustment in timing was more pronounced in explicit pairs, where verbal cues likely served as a compensatory mechanism for the disruption caused by the dough turning. The contrast between the two styles of collaboration was further highlighted by computing the Entrainment Rate (E), the ratio between T_{br} and T_t. This metric provided a quantifiable means of assessing the degree to which participants could synchronize their actions, particularly in adapting to the rhythm-breaking task of turning over the dough.

$$E = \frac{T_{br}}{T_t} \tag{2.3}$$

A higher E value would indicate a greater level of entrainment, suggesting that participants were more in sync, quickly adjusting their actions in response to their partner's movements. Using the average of the mesured T_{br} and T_t of the kneaders, results of this ratio were as follows:

$$E = \frac{T_{br}}{T_t} = \begin{cases} 0.85 & \text{(implicit communication pairs)} \\ 0.96 & \text{(explicit communication pairs)} \end{cases} \tag{2.4}$$

The analysis of implicit pairs, as depicted in Figure 2.5a, illustrates a harmonious adaptation between participants involved in the kneading and pestling activities. This synchronization becomes particularly evident as both participants approach the critical "turning over" cycle. Observations suggest that, in anticipation of this more time-consuming action, both individuals intuitively slow their respective rhythms, thus maintaining the collaborative flow without verbal communication. This mutual adjustment indicates a deep, non-verbal understanding and an ability to predict and react to each other's movements, demonstrating the efficacy of implicit cues in facilitating smooth, coordinated actions.

Conversely, Figure 2.5b reveals a disconnect in explicit pairs, particularly during the turning over phase. While the kneading action' s duration naturally extends to accommodate the dough-turning process, the pestling rhythm remains unchanged, unaffected by the altered pace of its partner's actions. This lack of adaptation suggests a failure in communication, where the individual responsible for pestling does not recognize or respond to the shifts in the kneading-turning cycle. The need for vocal commands in such pairs stems from this disconnect, as the kneading participant must overtly signal the impending or ongoing turn-over action to realign their collaborative rhythm.

The reliance on explicit, vocal communication among some pairs is symptomatic of an inability to synchronize based on non-verbal cues alone. The necessity for such direct intervention highlights a fundamental challenge in collaboration: the pestling participant's failure to interpret or adapt to variations in the task's pace. Conversely, implicit pairs demonstrate an unconscious, yet effective, communication system where slight changes in pace serve as signals that are both sent and received non-verbally, facilitating an unspoken agreement on timing and action.

The contrasting dynamics between implicit and explicit collaboration pairs shed light on the intricate ballet of human interaction, particularly in tasks requiring precise coordination. Implicit pairs exemplify the potential for seamless collaboration through non-verbal communication, where mutual adaptation ensures continuity and harmony. Explicit pairs, however, reveal the challenges when this non-verbal understanding is absent, necessitating verbal cues to maintain task synchrony. This analysis underscores the importance of developing collaborative systems—be they human-human or human-robot—that can recognize, interpret, and adapt to implicit cues, fostering a more intuitive and efficient cooperative environment.

2.2 User Intention Estimation

2.2.1 Requirements

In the pursuit of enhancing human-robot collaboration, especially in everyday environments, this thesis has identified several core principles necessary for developing an efficient cooperative work method with a robotic arm. These principles are essential for ensuring that the robotic system operates effectively and intuitively in a range of scenarios without being burdensome to the user.

1. Task Independence: A pivotal attribute for personal robots, particularly those intended for use in daily life, is the capability to function independently of specific tasks. Traditional methods that utilize implicit instructions are often limited by their task dependency, restricting the robot' s utility to pre-coded functions. This research breaks from convention by designing a system that does not require task-related or environmental information to operate. Instead, the robot relies solely on the user' s state and its own accumulated experience to discern the expected function, ensuring flexibility and adaptability in its operations.

2. Minimally Invasive Data Collection: The operational methodology of the robot arm involves utilizing ideomotor cues from the user' s continuous motion, captured through IMU sensors. To integrate seamlessly into daily activities, the data collection system must be minimally invasive. This study has therefore limited the use of sensors to three at most, strategically placed at non-intrusive locations like the wrists and waist. The potential integration of these sensors into common accessories, such as watches or belts, further reduces their intrusiveness, fostering greater user acceptance and comfort.

3. System Responsiveness: The system' s responsiveness is crucial for user satisfaction. Any significant delay between the user' s command and the robot' s action can result in discomfort and potential abandonment of the system. The robotic arm must, therefore, exhibit prompt responsiveness to ensure a smooth and efficient user experience.

4. Context Awareness: While the system is designed not to be context-dependent, it must possess the ability to understand and adapt to the context in which it is used. This contextual awareness is vital for sound judgment during decision-making processes. Without this, the responsibility for correct operation unduly falls on the user, negating the purpose of an assistive robot.

5. Information Management: As the robot is exposed to various contexts and applications, it must effectively manage the information it gathers. This involves discerning which data to retain for future use and which to discard. Proper information management is key to the robot's learning capabilities, ensuring it can build on past experiences and recognize the importance of different pieces of information.

In summary, the thesis advocates for a paradigm shift in personal robotic systems, emphasizing the necessity of task-independent functionality, non-invasive data collection, quick responsiveness, contextual awareness, and adept information management. These principles serve as the foundation for creating a robotic arm that can be a versatile and user-friendly companion in daily life tasks. Through adherence to these principles, the robotic system can achieve a level of sophistication that allows for an unobtrusive and helpful presence in the user' s life.

2.2.2 Concept and Environment Set-up

The pursuit of creating HRI systems has led to the exploration of implicit command interpretation, a concept where robots discern user intentions through non-verbal cues. This subsection outlines a developed system that captures this phenomenon by leveraging the ideomotor principle, which hypothesizes a direct link between thoughts and corresponding physical movements without the need for conscious decisions.

Grounded in the psychological works of Carpenter, James, and Pfungst, the ideomotor principle has been a cornerstone in understanding how subconscious thoughts can manifest in physical actions (Carpenter[67]; James [53]; Pfungst[63]). Gauchou et al. extended this research into the realm of HRI, proposing that implicit cues, such as minor gestures or shifts in posture, could signal user intentions to a robot trained to recognize them [68].

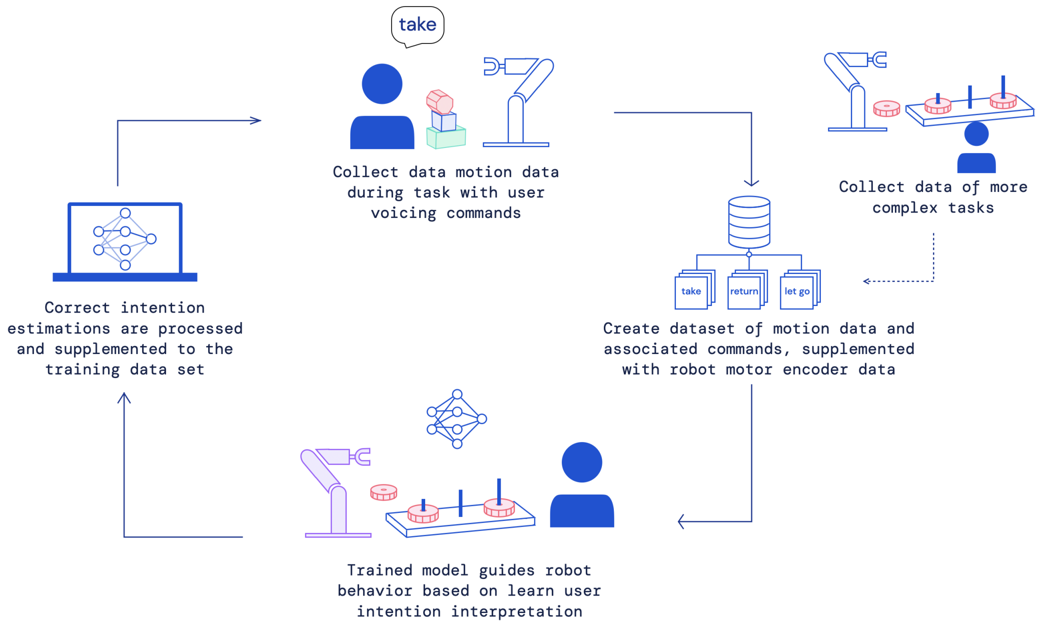


Figure 2.6: Training and testing loop

Incorporating these subtle forms of communication into robotics necessitates a system that can understand and process intricate human gestures, gaze direction, and other non-verbal cues. Fiore et al.'s research has demonstrated that robots attuned to these non-verbal signals can engage users more profoundly, fostering richer interactions [69]. The deep learning paradigm, particularly through the use of RNNs and LSTM networks, has emerged as a powerful tool for modeling time-series data typical of implicit cues [46].

The system's methodology, as depicted in Figure 2.6, begins with collecting motion data during various user-performed tasks. Simpler tasks are accompanied by vocal commands, providing an initial dataset of motion-command pairs. For complex tasks, a more comprehensive range of motions and verbal cues are recorded, capturing the nuanced interplay between different forms of user communication and robotic understanding. This collected data is enriched with the robot's motor encoder data, creating a layered dataset that encapsulates both human and robotic actions. The machine learning model is then trained on this dataset, with the goal of discerning user intent from the amalgamated motion and command data.

The trained model undergoes practical testing, wherein its ability to predict user intentions based on implicit cues is evaluated. The correct intention estimations are processed and fed back into the training set, embodying a dynamic and self-refining learning process. This iterative loop ensures that the system evolves and adapts, enhancing its interpretative accuracy and interactive performance with each cycle.

A key feature of the system is its capacity for continuous learning, crucial for maintaining relevance and effectiveness in diverse user environments. By constantly updating the model with new data from varied interactions, the system becomes increasingly sophisticated in interpreting user intentions, paving the way for more personalized and adaptable robots that can function across a range of environments and demographics. By integrating the ideomotor principle into a machine learning framework, the system offers a promising avenue for creating robots that understand and predict a wide array of human intentions. This work lays the groundwork for future advancements in HRI, where robots are not merely passive participants but active collaborators capable of understanding the subtleties of human communication.

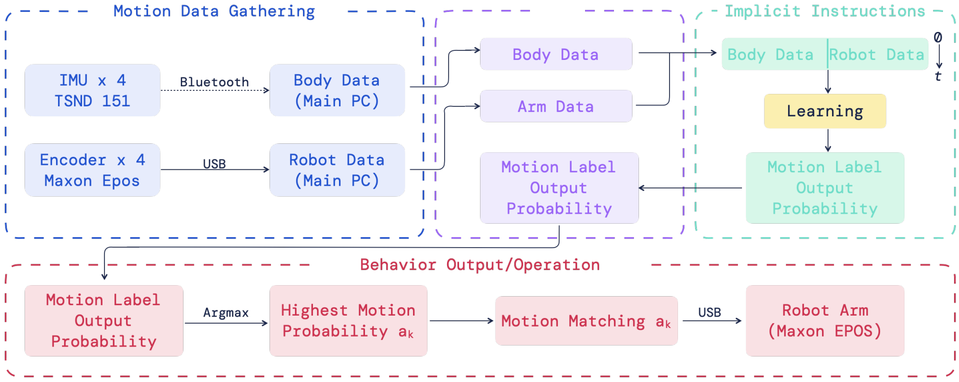


Figure 2.7: System Layout

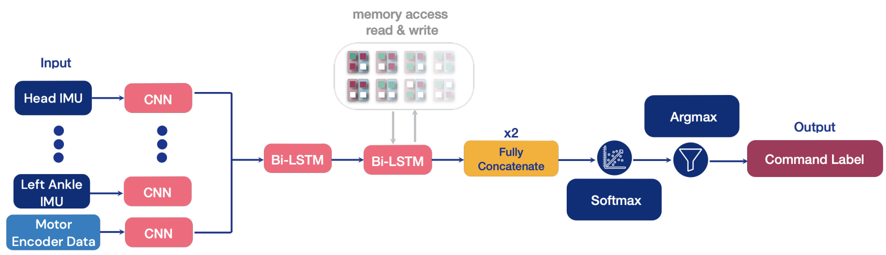


Figure 2.8: Model Architecture

2.2.3 Network Overview

The proposed model, depicted in Figure 2.8 for interpreting user commands incorporates a multi-modal approach that processes synchronous streams of data through a series of convolutional neural networks (CNNs) and bi-directional Long Short-Term Memory networks (Bi-LSTMs).

The input layer of the model consists of data from three sources: the head Inertial Measurement Unit (IMU), the left ankle IMU, and motor encoder data. Each data stream is passed through its own CNN to extract relevant features. The choice of using separate CNNs allows the model to learn modality-specific representations, which are crucial for interpreting complex sensor data accurately.

Following feature extraction, the output from each CNN feeds into a corresponding Bi-LSTM layer. The Bi-LSTM layers serve as a sequential processing unit, accounting for temporal dependencies in the data. This is essential in tasks requiring an understanding of sequence and timing, such as interpreting command sequences in real-time.

The architecture concludes with a fully concatenated layer that merges the outputs of the Bi-LSTMs, effectively combining the learned temporal features across all input modalities. This

(a) LSTM (b) Bi-LSTM

Figure 2.9: LSTM and Bi-LSTM architechture comparison

layer is followed by a softmax classifier that interprets the concatenated features to predict the user's intended command. The final command label is obtained by applying an argmax function to the softmax probabilities, identifying the most likely command from the available set.

Difference Between LSTM and Bi-Directional LSTM: The fundamental difference between an LSTM and a Bi-LSTM (Figure 2.9) lies in the directionality of the data processing. A standard LSTM processes data in a single direction, forward in time, meaning it captures dependencies from the past to make predictions about the future. This unidirectional flow limits the LSTM's context understanding to previous and current input features.

In contrast, a Bi-LSTM processes data in both forward and backward directions. This dual processing pathway allows the Bi-LSTM to capture not only the past context (as with standard LSTM) but also the future context. Each Bi-LSTM unit consists of two hidden layers, one for the forward direction and one for the backward direction. These layers are trained to optimize the sequence modeling by considering all available input information, both preceding and succeeding a given time point.

The Bi-LSTM's ability to look both backward and forward in the input sequence provides a more comprehensive temporal context, which can be critical for tasks involving sequences where each element can be influenced by elements before and after it, such as language processing or, in the case of this study, interpreting a sequence of user commands based on sensor data.

The enhancement in temporal understanding makes Bi-LSTMs particularly suited for complex tasks where the context is key, and for this reason, they have been employed in the presented model to robustly interpret the user's intent from multi-modal sensor data streams.

Integration of Bi-LSTM with External Memory: In the advanced architecture of the model, the second Bi-LSTM layer serves a pivotal role beyond sequential data processing—it acts as a controller for an external memory unit. This memory is an augmentation to the Bi-LSTM, providing a mechanism for the network to read from and write to, effectively giving the neural network a way to store and retrieve information over long sequences, similar to a working memory in human cognition.

The external memory unit is composed of a matrix where each row can be considered as a memory slot. These slots are accessed via the Bi-LSTM controller, which generates a set of read and write operations. The operations are determined by the context of the input data and the learned behavior of the network.

The integration of an external memory with the Bi-LSTM allows the model to perform complex tasks that require maintaining and manipulating a large amount of temporal information. It is particularly advantageous for tasks where the input sequence is very long or where the importance of certain input features varies dynamically over time.For example, in a command interpretation scenario, the user may provide a sequence of commands where the relevance of early commands may not be apparent until later in the sequence. The external memory allows the model to store these early commands and retrieve them when needed. This capability mirrors the cognitive process of recalling relevant information to make decisions based on new contextual information. By serving as a controller for the external memory, the second Bi-LSTM in the model not only processes the temporal data but also manages the information flow into and out of the memory. This design enables the network to perform more sophisticated tasks by mimicking higher-order cognitive functions, such as planning and reasoning, which require the manipulation of stored information.

2.3 Network Architecture

2.3.1 Layers Details

To construct a network that processes input data from 6 gyroscope sensors worn by the user and the robot's encoder data, and then extracts features using a Convolutional Neural Network (CNN) followed by two layers of Bi-directional LSTM (Bi-LSTM), we can conceptualize the mathematical framework as follows:

- **Input Data**
  Denoting the input data as X, where X consists of time-sequential data from 6 gyroscope sensors and the robot's encoder data. Each time step of X an be represented as $x_t \in \mathbb{R}^n$, where n is the total number of features at each time step (6 gyro features + encoder features).

- **CNN for feature extraction**
  By stacking consecutive time steps to form a "temporal image" raw time-sequential data is transformed into a format suitable for convolutional operations, the CNN operates on this data to extract high-level features. For a given layer l in the CNN

  – Input to the layer: I_l
  – Convolutional filter weights: W_l
  – Bias: b_l
  – Activation function (ReLU): ϕ

  The output of each convolutional layer is: $O_l = \phi(W_l * I_l + b_t)$
  After passing through the CNN layers, the output is a feature vector F_t for each time step t, reshaped as necessary for input into the subsequent Bi-LSTM layers.

- **Bi-LSTM Layers**
  The extracted features F_T re then fed into the Bi-LSTM layers for sequential processing. The Bi-LSTM consists of two LSTMs processing the data in opposite directions, and its operation at each time step t for each layer can be generalized as follows:

  – Forward Pass $(\overrightarrow{LSTM})$:
    – Forget Gate: $\overrightarrow{f_t} = \sigma(\overrightarrow{W_f} \cdot [\overrightarrow{h_{t-1}}, F_t] + \overrightarrow{b_f})$
    – Input Gate: $\overrightarrow{i_t} = \sigma(\overrightarrow{W_i} \cdot [\overrightarrow{h_{t-1}}, F_t] + \overrightarrow{b_i})$
    – Output Gate: $\overrightarrow{o_t} = \sigma(\overrightarrow{W_o} \cdot [\overrightarrow{h_{t-1}}, F_t] + \overrightarrow{b_o})$
    – Candidate cell state: $\overrightarrow{\tilde{C}_t} = \tanh(\overrightarrow{W_C} \cdot [\overrightarrow{h_{t-1}}, F_t] + \overrightarrow{b_C})$
    – Update cell state: $\overrightarrow{C_t} = \overrightarrow{f_t} * \overrightarrow{C_{t-1}} + \overrightarrow{i_t} * \overrightarrow{\tilde{C}_t}$
    – Update hidden state: $\overrightarrow{h_t} = \overrightarrow{o_t} * \tanh(\overrightarrow{C_t})$

  – Backward Pass $(\overleftarrow{LSTM})$:
  Mirror the forward pass equations with backward pass parameters and inputs, processing from $t = T$ (last time step) to $t = 1$ (first time step): $h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$. For the first Bi-LSTM layer, F_t, serves as the input. For the second Bi-LSTM layer, the output h_t from the first layer serves as the input.

- **Output Processing**
  The final output h_t from the second Bi-LSTM layer is then be processed further, through fully connected layers, ideal for classification or regression tasks.

2.3.2 Memory

Generating the parameters for reading and writing to memory in a Memory-Augmented Neural Network (MANN) involves defining how the network's controller (e.g., a Bi-LSTM) interacts with the external memory. The controller produces several key parameters based on its current state and the input it processes. These parameters include write weighting, erase vector, add vector for writing, and read weighting for reading from the memory. Below, we detail how these parameters can be generated, assuming the Bi-LSTM controller outputs a vector that is then used to derive these parameters.

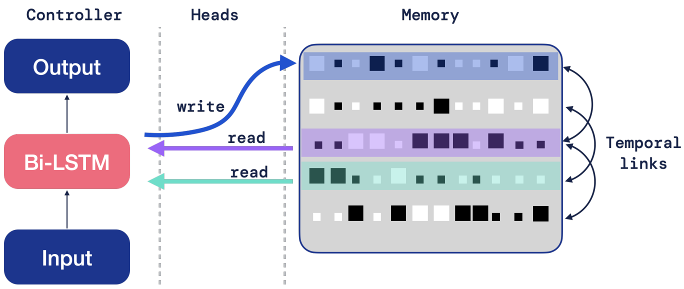


Figure 2.10: Representation of external memory cell interaction with Bi-LSTM controller

Assuming the Bi-LSTM controller outputs a vector $\mathbf{o}_t$ at time step t, which will be used to generate the memory interaction parameters. This output might be processed through one or more fully connected layers with specific activations to produce the parameters mentioned.
Generating write weighting
Often used to determine which memory slots to write to, based on the similarity between the memory slot content and a key vector generated by the controller.

$$\mathbf{k}_t^{write} = \tanh(\mathbf{W}_k \mathbf{o}_t + \mathbf{b}_k) \tag{2.5}$$

$$\mathbf{w}_t^{write} = Softmax(\mathbf{M}_t \cdot \mathbf{k}_t^{write}) \tag{2.6}$$

Where $\mathbf{W}_k$ is a weight matrix $\mathbf{b}_k$ is a bias vector, and $\mathbf{M}_t$ is the memory matrix at time t.

Generating erase vector
Specifies which parts of the selected memory slots should be erased.

$$\mathbf{e}_t = \sigma(\mathbf{W}_e \mathbf{o}_t + \mathbf{b}_e) \tag{2.7}$$

Where σ is the sigmoid function, ensuring that the elements of the $\mathbf{e}_t$ are between 0 and 1.

Generating add vector
Specifies what new information to write into the memory.

$$\mathbf{a}_t = \tanh(\mathbf{W}_a \mathbf{o}_t + \mathbf{b}_a) \tag{2.8}$$

Generating read weighting
Determines from which memory slots to read.

$$\mathbf{k}_t^{read} = \tanh(\mathbf{W}_{kr} \mathbf{o}_t + \mathbf{b}_{kr}) \tag{2.9}$$

$$\mathbf{w}_t^{read} = Softmax(\mathbf{M}_t \cdot \mathbf{k}_t^{read}) \tag{2.10}$$

Final read vector The read vector, which is the content read from the memory, is obtained by:

$$\mathbf{r}_t = \mathbf{M}_t^T \mathbf{w}_t^{read} \tag{2.11}$$

$W.$ and $b.$ are trainable weights and biases for generating the respective parameters. This vector can be fed back into the Bi-LSTM along with the next input, closing the loop for iterative reading and writing.

Interaction Equations

$$\mathbf{M}_t = \mathbf{M}_{t-1} \circ (1 - \mathbf{w}_t^{write} \otimes e_t) \tag{2.12}$$

$$\mathbf{M}_t = \mathbf{M}_t + \mathbf{w}_t^{write} \otimes v_t \tag{2.13}$$

Integrating these components allows the MANN to dynamically interact with its external memory, enabling complex information processing and storage capabilities that extend beyond traditional neural network architectures.

2.3.3 Memory Integration Mechanism into the Current Model

Integrating an external memory module with a CNN-Bi-LSTM network to create a Memory-Augmented Neural Network (MANN) involves a sophisticated setup where the Bi-LSTM acts as a controller for the memory operations. The mathematical representation of such an integration includes defining the memory structure, the read and write mechanisms, and the controller's operations.

- **External Memory Structure**

Memory Matrix: $M \in \mathbb{R}^{N \times W}$, where N is the number of memory slots, and W is the dimensionality of each vector stored in memory.

- **Controller (BI-LSTM) Output**

Let the output of the Bi-LSTM at time step t be h_t, which serves as input to generate memory control signals.

- **Write Mechanism**

Write Weighting:
►write key:

$$\mathbf{k}_t^w = FC(h_t) \tag{2.14}$$

Where FC is a fully connected layer.
►write strength:

$$\beta_t^w = Softplus(FC(h_t)) \tag{2.15}$$

►write weighting:

$$w_t^w(i) = \frac{\exp(\beta_t^w \cdot \cos(\mathbf{k}_t^w, M[i]))}{\sum_j \exp(\beta_t^w \cdot \cos(\mathbf{k}_t^w, M[j]))} \tag{2.16}$$

where $\cos$ denotes cosine similarity

Erase Vector: Specifies what to erase from memory.

$$\mathbf{e}_t = \sigma(FC(h_t)) \tag{2.17}$$

where σ is the sigmoid function

Add Vector: Specifies what to write to memory.

$$a_t = \tanh(FC(h_t)) \tag{2.18}$$

Memory update rule:
►Erase step:

$$\tilde{M}_t = M_{t-1} \circ (1 - w_t^w \otimes e_t) \tag{2.19}$$

▶Add step:

$$M_t = \tilde{M_t} + w_t^w \circ a_t \tag{2.20}$$

- **Read Mechanism**

Read Weighting:

▶read key:

$$\mathbf{k}_t^r = FC(h_t) \tag{2.21}$$

▶read strength:

$$\beta_t^r = Softplus(FC(h_t)) \tag{2.22}$$

▶read weighting:

$$w_t^r(i) = \frac{\exp(\beta_t^r \cdot \cos(\mathbf{k}_t^r, M[i]))}{\sum_j \exp(\beta_t^r \cdot \cos(\mathbf{k}_t^r, M[j]))} \tag{2.23}$$

Read operation:

$$\mathbf{r}_t = M_t^T w_t^r \tag{2.24}$$

- **Comments**

1. FC layers transform the Bi-LSTM outputs into parameters for controlling the memory.
2. $Softplus$ is used for parameters that must be positive, such as the write and read strengths.
3. The memory update and read operations are differentiable, allowing the entire network (CNN-Bi-LSTM with external memory) to be trained end-to-end using gradient descent-based methods.

This mathematical representation outlines the core components and operations involved in integrating an external memory with a CNN-Bi-LSTM network, transforming it into a powerful MANN capable of sophisticated data processing and storage tasks.

2.3.4 Elastic Weight Consolidation (EWC)

Integrating Elastic Weight Consolidation (EWC) into a Bi-directional LSTM (Bi-LSTM) network for continual learning involves modifying the training process to include a regularization term that penalizes significant deviations from parameter values optimized for previous tasks. This regularization is based on the Fisher Information, which quantifies the importance of each parameter to the performance on these tasks. Here are the key equations representing this process:

For each new task, you typically minimize a loss function $L_{new}(\Theta)$ where Θ represents the parameters of the Bi-LSTM:

$$L_{new}(\Theta) = \frac{1}{N}\sum_{i=1}^{N} l(y_i, f(x_i; \Theta)) \tag{2.25}$$

- N is the number of samples in the new task.
- x_i and y_i are the input and target output for the i-th sample, respectively.
- $f(x_i; \Theta)$ is the output of the Bi-LSTM model for input x_i
- l is the loss function, MSE

The objective of EWC is to protect the knowledge acquired from previous tasks (interactions) when learning new ones. This is achieved by adding a regularization term to the loss function that penalizes changes to parameters that are important for the performance on previous tasks. The augmented loss function with EWC is given by:

$$L_{Total}(\Theta) = L_{new}(\Theta) + \lambda L_{EWC}(\Theta) \tag{2.26}$$

$$L_{Total}(\Theta) = L_{new}(\Theta) + \frac{\lambda}{2}\sum_i F_i(\Theta_i - \Theta_{i,old})^2 \tag{2.27}$$

- $L_{Total}(\Theta)$ is the total loss function incorporating L_{EWC}
- $L_{new}(\Theta)$ is the loss on the new task (e.g, learning from new user interactions)
- $L_{EWC}(\Theta)$ is the EWC regularization term, defined as shown in 2.27
- λ is a hyperparameter that controls the strength of the regularization term, balancing between learning new tasks and retaining old knowledge
- F_i represents the Fisher Information of parameter i ndicating its importance to the tasks already learned. The Fisher Information is a measure of how much information a parameter carries about the output; parameters with high Fisher Information are crucial for the model's performance on previous tasks.
- Θ_i are the current parameters of the model
- $\Theta_{i,old}$ are the parameters of the model after learning previous tasks, before learning the new task.
- The summation runs over all the parameters i f the model.

Fisher Information F_i for each parameter can be calculated after the model has been trained on a task, as follows:

$$F_i = \mathbb{E}\left[\left(\frac{\partial \log p(y|x,\Theta)}{\partial \Theta_i}\right)^2\right] \tag{2.28}$$

- $p(y|x,\Theta)$ is the model's output distribution given input x, target y, and parameters Θ
- The Expectation $\mathbb{E}[\cdot]$ is typically approximated using the training set for the task

When the model is trained on a new set of user interactions or tasks, the L_{EWC} term helps to maintain the performance on previously learned interactions by constraining the updates of parameters that are critical for those interactions. This is particularly useful in HRI scenarios where the robot is expected to continually learn from new user behaviors without degrading its ability to respond to behaviors it has previously adapted to.

By applying the EWC technique, the learning system can achieve a balance between plasticity (ability to learn new tasks) and stability (retention of knowledge on old tasks), addressing the challenge of catastrophic forgetting in continual learning scenarios.

2.3.5 Optimizer

Gradient Descent Option

$$\Theta_{new} = \Theta_{old} - \alpha(\nabla_\Theta L_{new}(\Theta) + \lambda\sum_i F_i(\Theta_i - \Theta_{i,old})^2) \tag{2.29}$$

Adam

$$\Theta_{new} = \Theta_{old} - \frac{\eta}{\sqrt{\hat{v}_{i,t}} + \epsilon}\hat{m_{i,t}} \tag{2.30}$$

$$\hat{m_{i,t}} = \frac{m_{i,t}}{1-\beta_1^t} \quad and \quad \hat{v}_{i,t} = \frac{v_{i,t}}{1-\beta_2^t} \tag{2.31}$$

$$m_{i,t} = \beta_1 \cdot m_{i,t-1} + (1-\beta_1)\cdot g_{i,t} \quad and \quad v_{i,t} = \beta_2 \cdot v_{i,t-1} + (1-\beta_2)\cdot g_{i,t}^2 \tag{2.32}$$

2.3.6 Mitigate catastrophic forgetting

In scenarios where the robot is expected to learn from continuous interactions, adaptation through methods like EWC is essential to balance acquiring new knowledge with retaining important information from previous experiences.

Without adaptation, learning new information could degrade the robot's ability to perform tasks it had previously learned (catastrophic forgetting). Adaptation ensures the robot maintains competency across all learned tasks and interactions.

Human behaviors and the way they communicate intentions through motions can be subtle and vary widely. Adaptation allows robots to better interpret these cues over time, improving interaction quality. By continuously adapting, robots can become more responsive to immediate user actions and even anticipate user needs based on past interactions, leading to smoother and more intuitive HRI.

2.4 Experiment and Results

2.4.1 Set-up

In the pursuit to gauge the estimation accuracy of implicit instructions, a series of tasks were meticulously chosen to test the capabilities of the Implicit Interface. These tasks were designed to invoke a range of labels—specific commands recognizable by the system—varying from three to six distinct options. Out of the four tasks, three were periodic, featuring a consistent label cycle that allowed for patterned repetition and predictability. In contrast, the fourth task was aperiodic, with a label order that was intentionally randomized to introduce variability and challenge the system's adaptability and real-time interpretation skills.

The original experimental setup, as illustrated in Figure 2.11, aimed to replicate a realistic environment where multiple objects relevant to the task were arranged on a table. This setting was chosen to mimic the typical conditions in which a human-robot interaction system might be deployed, thereby ensuring that the results were as practical and applicable as possible. Participants were equipped with wearable IMU sensors placed at various key points on the body to capture a comprehensive range of motion data. These sensors tracked movements meticulously, providing the Implicit Interface with rich input to process and interpret. Additionally, the participants donned eyeglasses designed for robot operation, suggesting a heads-up display or visual input might play a role in how the system and the participant communicated.

The Implicit Interface was put to the test to determine how effectively it could interpret the participant's intentions based on non-verbal cues captured by the sensors. The eyeglasses presumably provided a visual channel for instructions, possibly by tracking eye movements or serving as a medium for the participant to view and respond to robot feedback.

The selection of periodic and aperiodic tasks served a dual purpose. For periodic tasks, the goal was to assess how well the system learned and predicted the participant's actions based on recurring patterns. The aperiodic task tested the interface's robustness in situations where patterns were not easily discernible, closely mimicking real-world scenarios where human actions are not always predictable. This experiment's setup underscores the significance of understanding and correctly interpreting implicit instructions, a cornerstone for any system expected to operate seamlessly alongside humans. The accuracy with which the Implicit Interface decoded and acted upon non-verbal cues would be pivotal in determining its efficacy in real-world applications, where explicit commands might not always be possible or desired.

1. Cleaning Task:
   The Cleaning Task was selected for its simplicity and periodic nature, making it the most fundamental of the tasks within the Implicit Interface experiment. Comprising only three labels—"lift," "wipe," and "place"—the task followed a constant cycle designed to evaluate

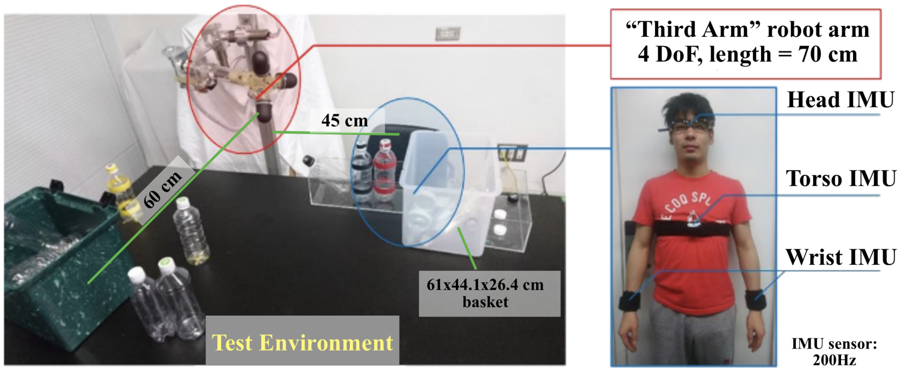


Figure 2.11: Test environment

the system's capacity for recognizing and responding to repetitive actions. The implicit nature of the task demanded that the robot interpret non-verbal, physical cues to coordinate its movements with those of the human participant. In the Cleaning Task, the periodic sequence began as the participant lifted an object from the table. This motion acted as a cue for the robot arm, which, guided by the participant's face vector—a potential proxy for intention direction—would then reach the object's previous location to begin wiping the table. Here, the 'face vector' likely referred to the orientation of the participant's gaze as detected through the eyeglasses, providing a directional command to the robot without any explicit verbal instruction. As the participant initiated the motion to place the object back onto the table, the robot was designed to detect this descending action and cease wiping, retreating to its initial position. This action-response sequence formed the cyclical pattern that the Implicit Interface needed to identify and follow. The Implicit Interface, utilizing input from IMU sensors, was responsible for decoding the participant's movements and discerning the correct moments to execute corresponding robotic actions. The task's repetitive nature allowed the system to fine-tune its predictions over successive cycles, optimizing its response times and improving synchronization with the participant's actions.

2. Cap Placement and Handover Task Analysis:
   The Cap Placement and Handover Task is a methodically designed experiment rooted in periodic behavior, where the participant is involved in a multi-step process of completing a common yet precise action—capping a bottle and passing it to a robotic arm for placement into a box. The experiment, inspired by the methodologies detailed in [70], integrates a sequential approach with a total of four labels corresponding to each phase of the task: picking up the cap, placing the cap on the bottle, handing over the bottle, and the robot arm's subsequent box placement and return to the starting position.

3. Cloth Exchange and Cleaning Task:
   Task 3, named "Cloth Exchange and Cleaning," presents a fusion of elements from the first two tasks, reflecting a more complex, albeit still periodic, interaction with constant cycle labels. This task entails a participant handing a cleaning cloth to the robotic arm, which is then tasked with wiping the table before depositing the cloth into a storage box out of the participant's reach. The task is completed when the robotic arm repositions itself to its starting point, ready for the next cycle. The task's periodic nature ensures that the cycle of actions is predictable and consistent, which is crucial for the Implicit Interface to establish a rhythm and optimize its response patterns. The sequence of the task is a dance of interdependent movements: the participant passes the cloth to the robot, lifts an object from the table to enable cleaning, and then possibly replaces the object once the surface has been wiped. This back-and-forth requires a finely tuned

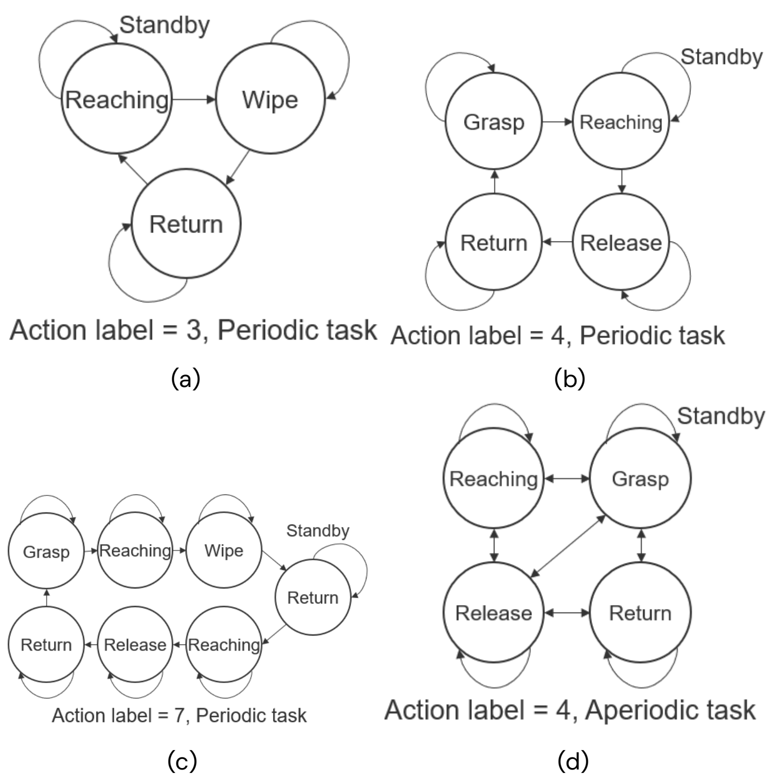


Figure 2.12: Command pattern for each task. (a) Task 1 - Cleaning Task. (b) Task 2 - Cap placement and handover task. (c) Task 3 - Cloth exchange and cleaning task. (d) Task 4 - Aperiodic cap removal and placement task

understanding of timing and spatial awareness from the robot. Unlike the previous tasks, the Cloth Exchange and Cleaning Task introduces an additional layer of complexity by combining object exchange with an active cleaning component. The robot arm must not only accurately receive the cloth from the participant but also perform a functional task with it, all the while coordinating these actions with the participant's movements. This multi-step process challenges the Implicit Interface's capacity to manage and execute a sequence of actions that depend heavily on the accurate interpretation of human cues.

4. Aperiodic Cap Removal and Placement Task:
The final task in the series is a sophisticated amalgam of Task 2's bottle capping and handover, with the added element of cap removal, set within an aperiodic framework. Unlike the preceding tasks, this one introduces a degree of unpredictability by randomizing the sequence of labels—'remove cap', 'handover', 'place cap', and 'store bottle'. The randomness infused into the task order represents a significant leap in complexity, as it challenges the Implicit Interface to understand and adapt to actions without relying on a repetitive sequence or well-established context. The aperiodic nature of this task serves to evaluate the robot arm's adaptability to new and evolving scenarios, a crucial feature for real-world applications where robotic systems must often operate without the comfort of predictability. The test environment's less controlled setup mimics real-life conditions, where robots must decipher and respond to human actions and intentions that do not follow a set pattern. In previous tasks, the robotic system could anticipate the participant's next move based on a learned cycle of behavior, allowing for context-driven estimation of actions. However, this aperiodic task removes the contextual crutch, requiring the robot to interpret each label independently and react appropriately to the participant's actions without a predefined order to guide its expectations. The introduction of the aperiodic element is a strategic move that underscores a critical aspect of

task complexity. It demonstrates that complexity in robotic tasks is not solely a function of the number of labels but also the predictability of their sequence. An aperiodic task can be more challenging than a periodic task with a greater number of labels, as it requires the robot to be highly responsive to real-time data and capable of instantaneous decision-making.

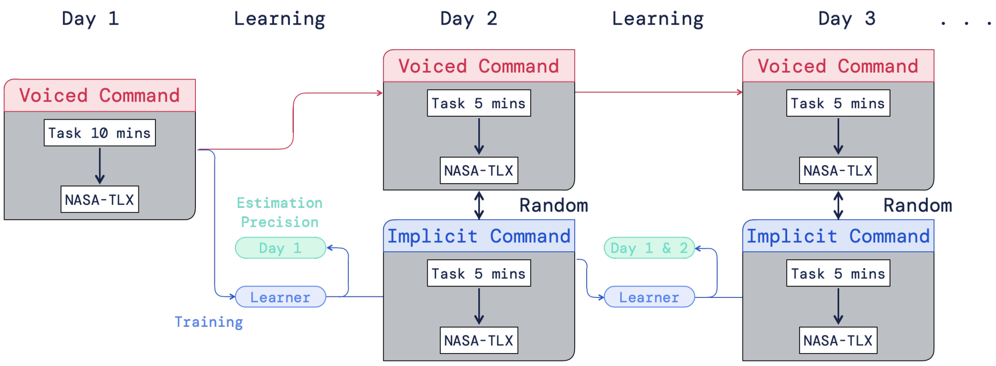


Figure 2.13: Flow of the experimenting and training of the system

2.4.2 Performance Improvement

Figure 2.14 shows a comparison between the performance of the original model and the new model across four distinct tasks with varying degrees of complexity. The F1 scores, a harmonic mean of precision and recall, are indicative of the models' accuracy in intention estimation or label prediction in these tasks.

For Tasks 1 to 3, which are periodic and represent a graduated increase in the number of labels, both models exhibit high F1 scores initially. However, the new model (lighter blue) displays a slight improvement in all three tasks. This suggests that enhancements made in the new model have allowed it to handle the increasing complexity more effectively. The incremental improvement indicates that even with a higher number of labels, the new model maintains or exceeds the previous model's accuracy levels.

Task 4, being aperiodic and containing the same number of labels as Task 2, represents a more substantial challenge due to the lack of a predictable pattern in label sequence. Interestingly, while the original model had an F1 score akin to that of the periodic tasks, the new model experiences a noticeable dip in performance. This decrease could be attributed to the new model possibly overfitting to patterns in the periodic tasks, thus struggling with the randomness of Task 4. It might also suggest that the new model, despite its improvements, requires further adaptation to handle the uncertainty and variability inherent in aperiodic tasks effectively. The F1 score progression highlights several key points:

- Model Improvements: The new model's overall superior performance in periodic tasks suggests successful improvements over its predecessor. These could include better generalization capabilities or enhanced sensitivity to the structure within the tasks.
- Task Complexity: The performance across tasks with an increasing number of labels remains relatively stable, demonstrating that the new model handles complexity well when within a structured, predictable framework.
- Adaptability to Aperiodicity: The significant decline in F1 score for Task 4 by the new model underscores the challenge aperiodicity poses to intention estimation systems. It suggests a potential overemphasis on pattern recognition, which does not translate well to a scenario where patterns are absent or not readily discernible.

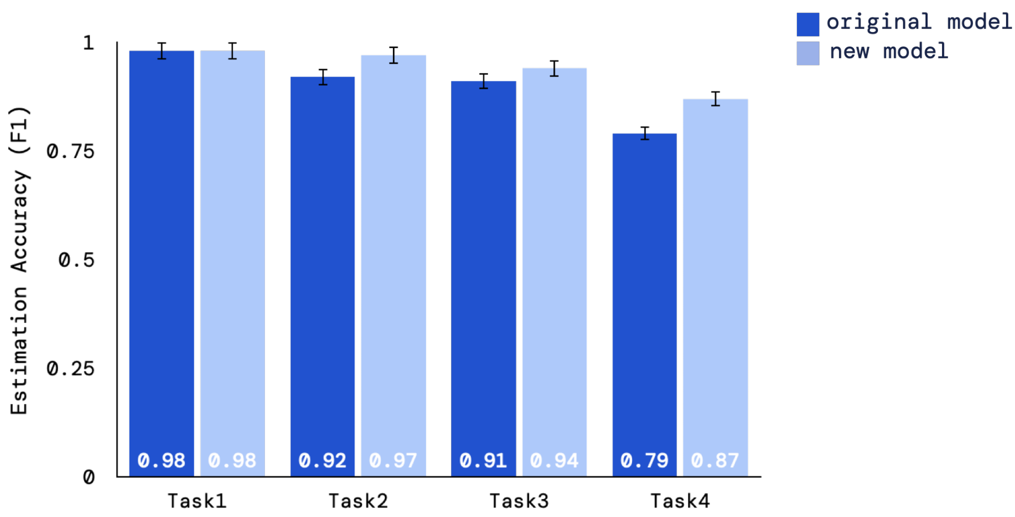
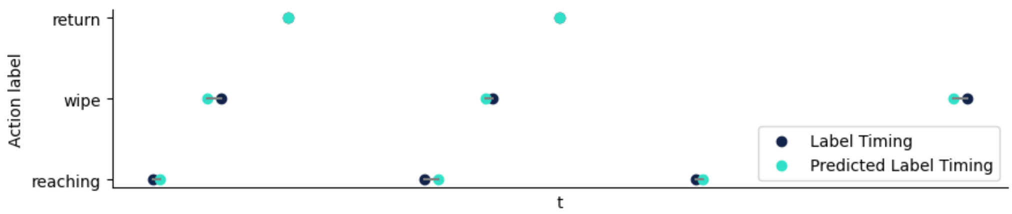
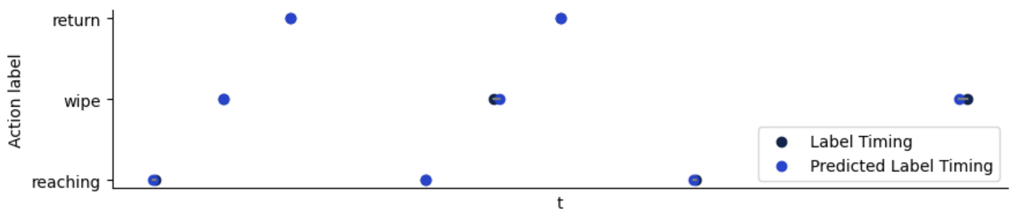


Figure 2.14: Label estimation (F1 score) evolution

- Importance of Versatility: To ensure robust performance across both periodic and aperiodic tasks, models need to be versatile. They must not only recognize and learn from patterns but also remain flexible enough to interpret and respond to non-patterned data.



(a)

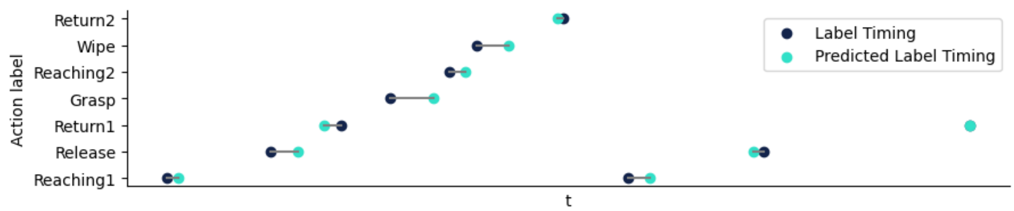


(b)

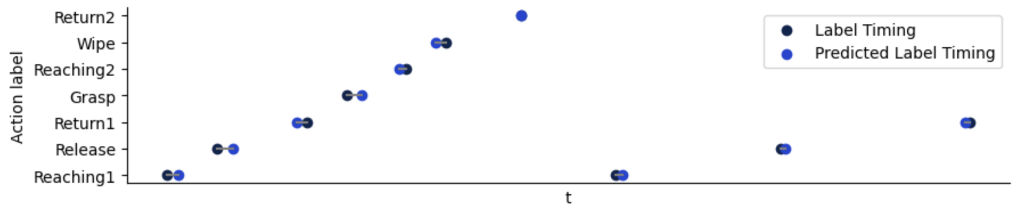
Figure 2.15: Task 1 long term improvement

Figures 2.15 and 2.16 chart the system's progression in accurately predicting the timing of different action labels during two collaborative tasks, illustrating a marked improvement from the initial stages of system use to an extended period of interaction.

In the initial stages depicted in 2.15a and 2.16a, there is a noticeable discrepancy between the actual label timings and the system's predictions. The scatter points representing the predicted label timing do not align closely with the actual timing points, indicating a lag or mismatch in the system's recognition patterns. This suggests that, initially, the system may struggle to learn the nuances of the collaborative process and accurately anticipate the users' actions.Contrastingly, the Figures 2.15b and 2.16b, representing system performance after prolonged use, show a much closer alignment between predicted and actual label timings. The scatter points for the predicted labels are nearer to the true label timings, demonstrating that the system has learned from its interactions and adapted its predictive model to more closely



(a)



(b)

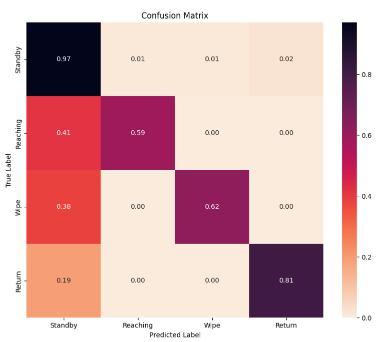
Figure 2.16: Task 3 long term improvement

match the human collaborators' behavior.

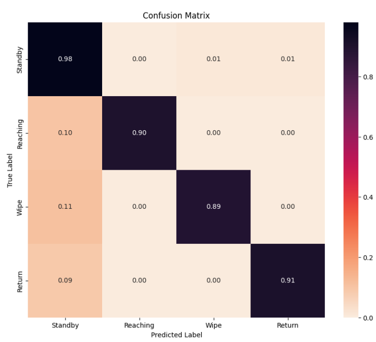
The system's learning appears to be robust across tasks with varying numbers of labels. Even with the increased complexity in the task represented by the second set of figures (with increasing number of labels), the system shows an ability to improve its predictive accuracy over time. The figures provide evidence of the system's learning curve and the effectiveness of machine learning algorithms in adapting to human behavior patterns. The initial mismatch and subsequent convergence of predicted and actual timings underscore the importance of iterative training and long-term interaction in developing collaborative AI systems.

The consistent improvement over time also illustrates the system's capability for temporal pattern recognition and adaptation, crucial for seamless human-robot interaction. It reflects an advanced understanding of both the sequential nature of human actions and the nuanced variances within a dynamic collaborative environment.

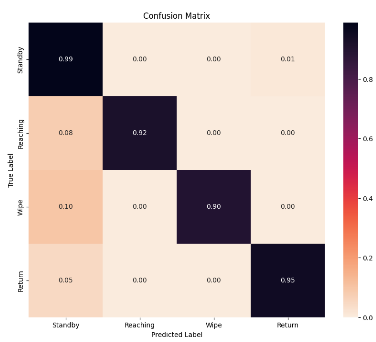
The data suggests that with sufficient training and user interaction, AI systems can reduce the gap between human expectations and robotic performance, leading to more fluid and intuitive collaboration. The ability to adapt over time also speaks to the system's potential in real-world applications, where it can learn from and respond to diverse human behaviors, actions, and cues.



(a)



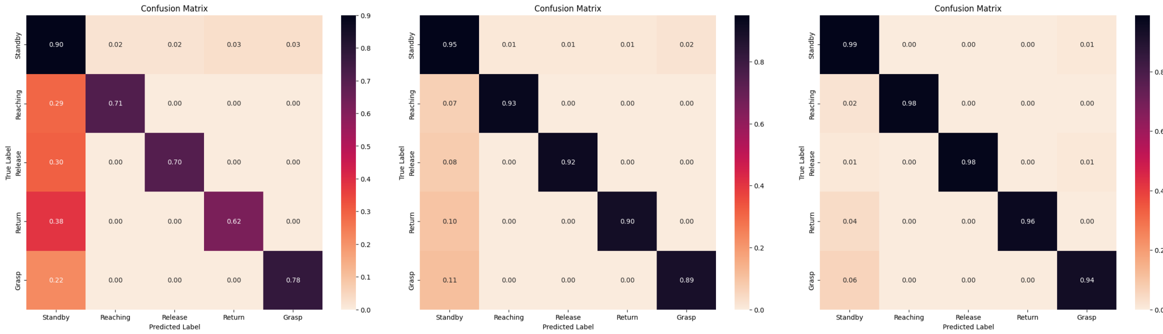
(b)



(c)

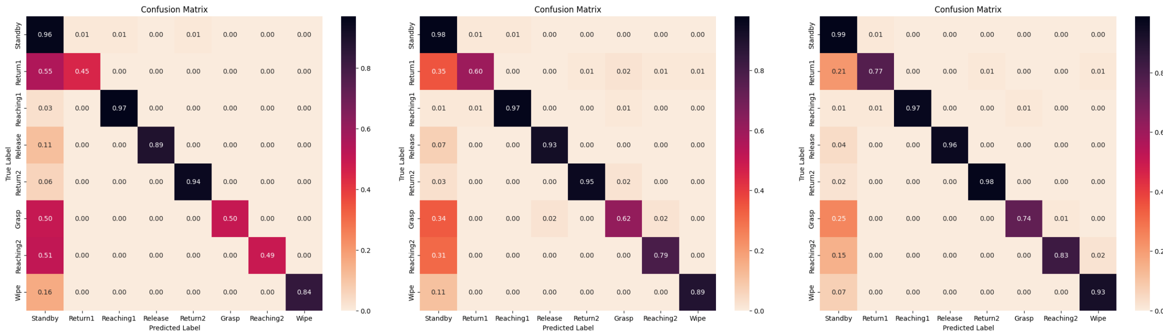
Figure 2.17: Task 1 command identification progress

Figures 2.17, 2.18, 2.19 and 2.20 display confusion matrices from an experiment, revealing the progression of a system's capability to estimate user intentions over time across different



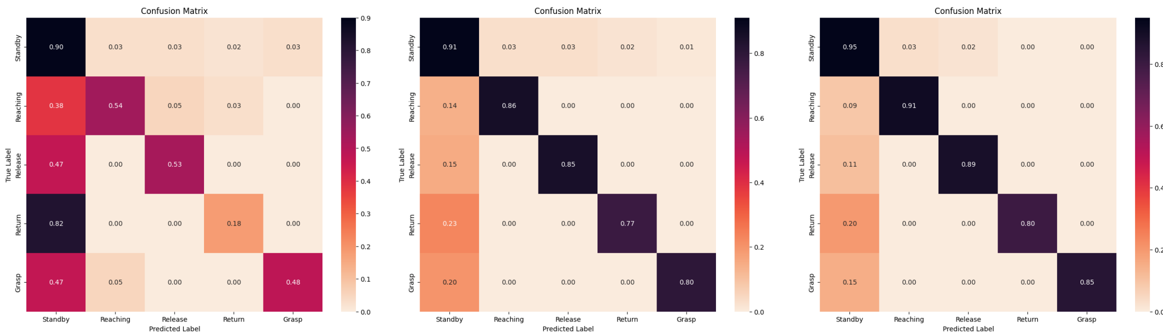
(a) (b) (c)

Figure 2.18: Task 2 command identification progress



(a) (b) (c)

Figure 2.19: Task 3 command identification progress



(a) (b) (c)

Figure 2.20: Task 4 command identification progress

tasks. The matrices represent stages (a), (b), and (c), corresponding to early, middle, and later phases of the system's learning period.

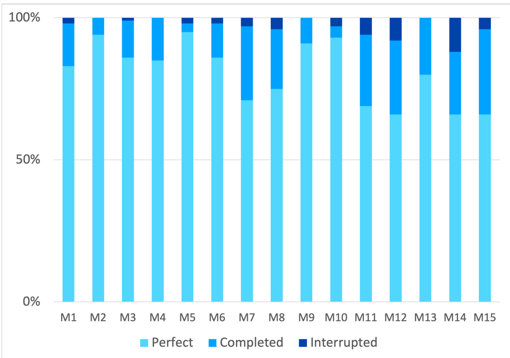
In the early stages, the confusion matrices typically show a less pronounced diagonal, indicating a relatively lower accuracy in intention estimation. For instance, early versions of tasks may exhibit significant confusion between labels such as "Reaching" and "Grasp," suggesting the system initially struggles to distinguish between similar motion-related commands. As the system undergoes more training, the middle-phase matrices generally show an improvement in the correct identification of labels, with a clearer diagonal and reduced off-diagonal elements. This suggests that, over time, the system better differentiates between the nuances of each task, resulting in fewer misclassifications. In the later stages, the system's intention estimation accuracy further solidifies, as evident from the strong diagonal lines in the confusion matrices.

By this point, the system has likely been exposed to a substantial amount of user data, enabling it to refine its predictive algorithms and make more precise estimations about user intentions.

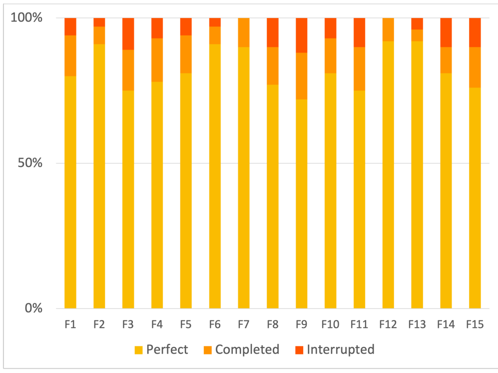
The observed progressive improvement in the system's performance can be attributed to the iterative process of machine learning, where extended exposure to diverse user data enables the system to fine-tune its understanding of implicit cues and motion patterns associated with various commands.

This ability to adapt over time is crucial in developing user-interactive systems that become more personalized and efficient the longer they are used. In the context of human-robot interaction, such a system would become increasingly adept at predicting a user's actions, leading to smoother collaborations and potentially reducing the cognitive load on the user.

The distinction between the early and late stages of learning underscores the importance of continued use and training in the development of intelligent systems. It also suggests that while initial deployment may come with a learning curve, the system's effectiveness will likely increase, providing a more intuitive and responsive user experience. The presented bar charts



(a)

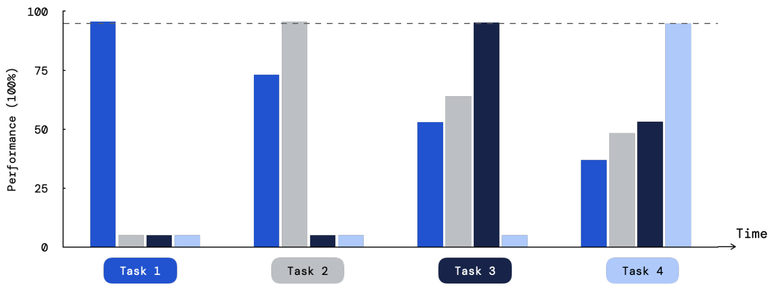


(b)

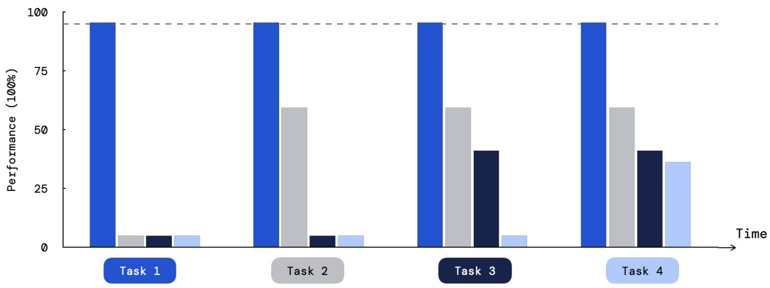
Figure 2.21: Task completion levels

(Figure 2.21) illustrate the completion levels of tasks by male (M) and female (F) users in a user study. They delineate the tasks that were performed perfectly, completed with some issues, or interrupted. Interruptions in task completion are observed in both male and female users but appear more frequently in tasks performed by females. This might be indicative of several factors, including possible design biases of the system that do not account for gender-specific interaction patterns or inherent differences in how tasks are approached by different genders. The charts prompt a discussion about whether the system's design is universally intuitive or whether it inadvertently favors certain interaction styles commonly associated with one gender over the other. It may also bring up the possibility that the instructions or feedback provided by the system are more aligned with the behavioral tendencies of one gender. It is important to consider the learning and adaptability aspects of the system. If the system is designed to adapt to users' interaction patterns over time, initial disparities in performance might diminish as the system learns from a wider variety of user behaviors.

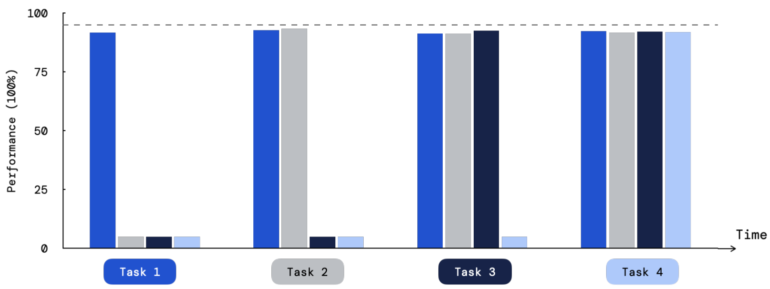
The observed data raises important questions about gender inclusivity in system design and task execution. It underscores the need for systems that provide an equally intuitive and effective user experience for all users, regardless of gender. To ensure inclusivity, systems should be thoroughly tested across a diverse user base and designed with consideration for different interaction patterns and preferences. The insights gained could be invaluable in enhancing system design, training, and user guidance to accommodate a broad spectrum of users effectively.



(a) Initial model with catastrophic forgetting



(b) Memory inhanced with too little plasticity



(c) Elastic Weight Consolidation model without forgetting

Figure 2.22: Performance on four tasks evaluated 4 times: train on task1, test on 1, 2, 3 and 4, train on task 2, test on 1, 2, 3 and 4...

2.4.3 Plasticity Verification

We present a comparative analysis of the plasticity in the design of neural networks based on three different models: one without Elastic Weight Consolidation (EWC), one with insufficient EWC (too little plasticity), and one with an optimized level of EWC (no forgetting). Figure 2.22a illustrates the performance of the initial neural network model without EWC. The model exhibits high performance on Task 1 but suffers from catastrophic forgetting, with performance on previous tasks rapidly degrading as new tasks are learned. This is evidenced by the significant decrease in performance metrics when the model is subsequently trained on Tasks 2, 3, and 4. Figure 2.22b presents the neural network with too little plasticity, integrating insufficient EWC. Although there is a notable improvement in retaining performance on earlier tasks compared to the initial model, the limited plasticity results in an inability to achieve optimal performance on subsequent tasks. The incremental improvements in Task 4 highlight the challenge of balancing plasticity and stability in the network. Figure 2.22c shows the results for the neural network model with an optimized level of EWC, which effectively addresses catastrophic forgetting. The model maintains high performance across all tasks, demonstrating robustness and a significant improvement over the two previous models. This configuration allows the network to preserve knowledge from previous tasks while adapting to new ones, without significant performance degradation.

The concept of plasticity in neural network design is pivotal in addressing the challenge of catastrophic forgetting. The results underline the importance of EWC in achieving a balance between retaining prior knowledge and acquiring new information. The first model, without EWC, suffered from catastrophic forgetting, as demonstrated by its inability to maintain performance on earlier tasks. This underscores the critical need for mechanisms that preserve learned knowledge when adapting to new data. The second model, with too little plasticity, suggests that overly conservative weight consolidation can hinder learning, preventing the network from adapting sufficiently to new tasks. This manifests as improved but still suboptimal retention of earlier task performance. The third model showcases the success of an optimally configured EWC approach, indicating that with careful tuning, it is possible to mitigate catastrophic forgetting effectively. The consistency in performance across all tasks illustrates that the network can learn new information without overwriting the previously acquired knowledge. Our analysis suggests that the key to successful neural network design lies in striking an appropriate balance between plasticity and stability. This entails not only preserving prior learning but also allowing enough flexibility for the network to incorporate new data effectively.

Future work should focus on refining the EWC approach, exploring the impact of different levels of plasticity, and testing the model on a broader range of tasks to further validate its generalizability and robustness in diverse learning scenarios. Additionally, alternative strategies to EWC should be explored to ensure that neural network models can continue to adapt and learn in a manner analogous to biological systems.

2.5 Cognitive Load Impact

2.5.1 Cognitive Load

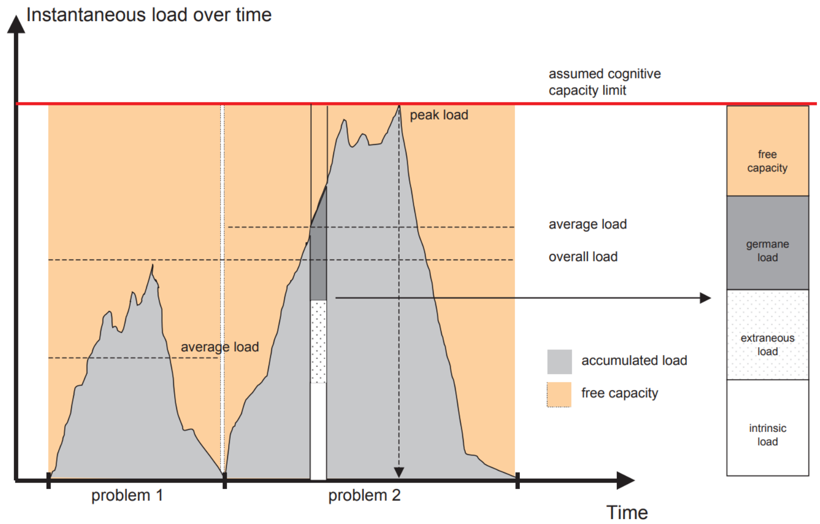


Figure 2.23: Cognitive load concept

The concept of "cognitive load" is integral to understanding how individuals manage their attention and mental resources while interacting with robotic systems. Cognitive load pertains to the mental effort required to complete a task effectively. In the seminal work by Paas et al., cognitive capacity is described as being distributed into distinct categories—free capacity and accumulated load, which encompasses germane, intrinsic, and extraneous loads (Paas et al., [71, 72, 73]). The "Law of constant capacity" posits that there is a finite limit to the attentional resources an individual can allocate at any given time. Therefore, if a task consumes all of one's free capacity, there remains no surplus for additional tasks, potentially compromising performance in a multitasking scenario.

Figure 2.23, as referenced, likely illustrates the division of cognitive resources. When engaged in a task, an individual must balance the accumulated load—comprising the attention required for the task at hand—with the free capacity that can be directed towards additional tasks or emergencies. Managing this balance is critical when operating a robot, as an overwhelming cognitive load can negatively affect the operator's ability to perform other tasks concurrently.

In the context of robot operation, particularly with wearable robotic arms, the cognitive load should be carefully managed to prevent an undue burden on the user. The operator must retain enough free capacity to maintain high performance levels across multiple tasks. The current study seeks to minimize the cognitive load as the primary variable, aiming to optimize the overall system for multitasking efficiency. The ultimate goal of the project is the development of a robotic system that allows individuals to multitask effectively without any noticeable decline in task performance. Achieving this involves creating interfaces and controls that are intuitive and require minimal mental effort to operate. By reducing the cognitive load, users can maintain situational awareness and respond to additional tasks or changes in the environment more readily.

This research continues to push the boundaries of human-robot collaboration by developing a system that operates within the user's cognitive comfort zone. The system's design must account for the cognitive load implications of each interaction and automation level, ensuring that users can interact with the robot in a manner that feels like an extension of their own capabilities rather than a separate, demanding entity.

Optimizing cognitive load in human-robot interaction is a vital step towards creating more efficient and harmonious collaborative environments. By focusing on minimizing the cognitive demands of operating a robot, this research aims to enhance multitasking capabilities, ensuring that users can perform at their best with the robotic system as a supportive adjunct. As robots become increasingly present in various sectors of society, the need for interfaces that cater to the natural limitations of human attention becomes more apparent, laying the groundwork for widespread adoption and more effective human-robot teamwork.

2.5.2 Qualitative Evaluation

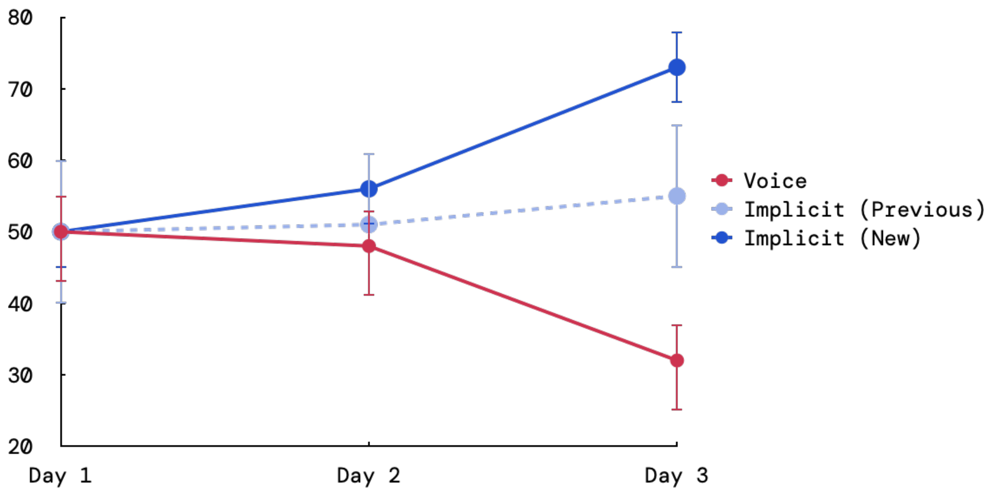


Figure 2.24: NASA-TLX results

Qualitative assessment was conducted using the NASA Task Load Index (NASA-TLX), a widely used tool for measuring cognitive load during tasks [74, 75]. The NASA-TLX is a subjective workload assessment tool that has been utilized for over three decades in the field of human-machine interface. Participants provided scores ranging from 0 to 100 for six workload dimensions, and weights were assigned to each scale to calculate the Overall Workload. A lower NASA-TLX Score indicates a lower subjective cognitive load. The workload dimensions consist of Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. Participants were asked to complete the NASA-TLX questionnaire at the end of each task with various interfaces, as shown in Figure 2.24, to compute the Overall Workload. The instructions for the load were presented in both original English and Japanese translations[68]. The evaluation criterion set the workload of Day 1 with voice instructions to 50, and the subsequent days' evaluations were made relative to Day 1. The NASA-TLX Score calculation followed the method described in [75].

From Day 2 onward, in tasks with implicit instructions, the proposed Adaptive Learning approach was applied. When the robotic arm failed in its estimations, human intervention was called upon through voice instructions, after which the robot arm relearned the correct action labels. No changes were made to the method of operation or robot arm control from Day 1 to Day 3 in voice-instructed operations. Results from both qualitative (NASA-TLX) and quantitative (secondary task) evaluations indicated that implicit instructions could lead to lower cognitive loads compared to explicit instructions during extended usage, including Day 3. Moreover, the validity of the proposed methods (Implicit Interface and Adaptive Learning) for achieving the goal of the study, "to construct interfaces that enable low cognitive load operations," is suggested.

The downside of the Implicit Interface, which is a dynamic system with a less than 100% estimation rate for implicit instructions, was counterbalanced by promoting "cooperative work" through Adaptive Learning, presumably resulting in long-term cognitive load reduction. The primary factor in this reduction is considered to be the reduction of "explicit instructions," as also indicated by Kobayashi et al.[76]. Especially in Task 3, there were seven different voice instructions, and instances of humans misspeaking were occasionally observed. In addition, it is important to note that even low cognitive load explicit instructions such as voice can result in the issue of human memory limits during multi-tasking. In other words, it is inferred that even replacing only a portion of explicit instructions with implicit ones during tasks can potentially reduce cognitive load.

2.6 Summary and Discussion

2.6.1 Summary

This study delves into the realm of implicit communication within human interactions, drawing connections with its prospective applications in human-machine interfaces. The ideomotor phenomenon, deeply ingrained in psychological research, finds remarkable parallels in cultural expressions such as Japan's "Aun breathing" and the intricate coordination of Bunraku puppeteers. The West's "clever Hans" effect further complements these findings, highlighting the potent influence of non-verbal cues. Such insights lay a foundation for enhancing the intuitiveness of human-robot interactions.

2.6.2 Discussion

Implicit communication is critical in tasks that require synchronized effort, such as orchestral performances and sports teamwork, where seamless coordination stems from an unspoken, shared understanding. Translating this intrinsic human capability to robotic systems could revolutionize the nature of human-robot collaboration, making it more fluid and intuitive.

The study's experimentation with the traditional craft of mochi-making afforded a unique lens through which to examine the synchronization of actions in collaborative settings. The data gleaned from this traditional activity, conducted in a virtualized setup, emphasized the importance of non-verbal communication in coordinating complex tasks.

A series of experiments were conducted to identify and understand implicit cues and ideomotor responses essential for maintaining synchronized workflows. The tasks, ranging from simple to complex and unpredictable, provided a spectrum of contexts to evaluate the performance and adaptability of the system.

※*Network Overview*

The proposed neural network model, integrating a Convolutional Neural Network (CNN) with Bidirectional Long Short-Term Memory (Bi-LSTM) layers and augmented by external memory, signifies a leap forward in enabling sophisticated collaborative tasks. This structure exhibited the potential for emulating high-level cognitive functions that are paramount in complex task execution. The CNN component excels at extracting spatial features, while the Bi-LSTM layers handle temporal dependencies, crucial for tasks requiring sequential decision-making and coordination.

※*Memory Integration Mechanism into the Current Model*

Incorporating an external memory module, managed by the Bi-LSTM controller, the model demonstrated an enhanced ability to retain and manipulate temporal information. This mechanism underscores the model's potential for advanced cognitive functions akin to human reasoning, essential for nuanced collaborative tasks. The external memory allows the system to recall past interactions and adapt its behavior based on accumulated experiences, thereby improving the fluidity and intuitiveness of human-robot collaboration.

※*Plasticity Verification*

By examining models with varied EWC levels, the study illuminated the critical balance between network plasticity and knowledge retention. Optimal model tuning was shown to mitigate catastrophic forgetting effectively, thereby maintaining performance across new and existing tasks. This balance is crucial for robots operating in dynamic environments, where they must continually learn new tasks without compromising previously acquired skills. The findings suggest that with proper EWC levels, robots can achieve a sustainable learning trajectory, maintaining high performance and adaptability.

※*Cognitive Load Impact*

Cognitive load emerges as a pivotal factor in the ergonomic operation of robotic systems. Utilizing the NASA-TLX to measure cognitive load, the study illustrates the necessity for interfaces that ease cognitive burdens. High cognitive load can impede performance and reduce the effectiveness of human-robot collaboration. The Adaptive Learning approach proposed in this research paves the way for the seamless integration of robots into human-centric workflows by minimizing cognitive strain. By adapting to the user's implicit signals and providing intuitive feedback, the system can enhance user comfort and efficiency.

The advancements in implicit communication and human-robot collaboration explored in this study provide a robust foundation for future research and development. By leveraging the natural human ability for non-verbal communication, we can design robotic systems that integrate seamlessly into human activities. The neural network model, with its sophisticated memory integration and adaptability, showcases the potential for creating robots capable of nuanced and synchronized collaboration. Addressing cognitive load and ensuring the balance between learning and knowledge retention are critical for the ergonomic and effective deployment of these systems. The journey from understanding implicit cues in traditional crafts like mochi-making to applying these insights in advanced robotic systems marks a significant step towards a future where human-robot collaboration is as natural and intuitive as human-human interaction.

2.6.3 Conclusion

The insights and empirical evidence presented underscore the intricate interplay between implicit communication and effective human-robot collaboration. The study paves the way for future advancements in this field, promising a seamless integration of robotics into the fabric of daily human endeavors, fostering a future where robotic assistance is perceived not merely as a tool but as a natural extension of human capability.

The research highlights the importance of designing sustainable forms of communication for human-machine interaction, essential for leveraging Artificial Intelligence to enhance both cognitive and physical human capacities. By mimicking principles underlying interpersonal communication, we can achieve more natural and effective human-robot interactions.

The study's experiments provided compelling evidence on the potential of implicit cues in facilitating human-machine cooperation. When robots employed implicit cues akin to those used in human-human interactions, users were able to comprehend these signals and adapt their behavior accordingly. This resulted in more stable and consistent collaboration, with improved work quality and performance metrics, such as a 28% reduction in working speed.

Moreover, the introduction of a body language approach for human-to-machine communication showcased promising results. The model demonstrated a notable average accuracy of 94% in estimating implicit cues across various tasks, with individual task accuracies reaching up to 98%. Participants also reported a significant decrease in cognitive load when using this system, highlighting its potential to ease the user experience in prolonged interactions.

In exploring intention detection within the robotic domain, this study addresses a critical yet underexplored aspect of effective communication and collaboration. The ability to infer intentions is fundamental for seamless interaction, and integrating this capability into robotic systems is crucial. The dual contributions of this research—demonstrating the applicability of human implicit cues in robot behavior and introducing a body language approach for user-driven machine learning—underscore its significance.

The findings suggest that implicit communication can serve as a universal medium for expressing intention, not only from humans but also from machines. The adaptability of the presented system to various tasks indicates its potential for widespread application, from self-driving cars to advanced prosthetics.

Looking forward, the study identifies key areas for future research to further enhance human-robot collaboration. Developing systems that support simultaneous two-way communication and feedback is crucial. Robots must provide feedback to acknowledge receipt of commands, ensuring users are aware of the system's status. Additionally, robots need mechanisms to detect and respond to misunderstandings of their cues by users. This is particularly vital in shared workspaces where clear and continuous communication is essential.

Furthermore, designing systems capable of handling communication overlap is imperative. During collaborative tasks, robots must listen to users even while executing commands, necessitating a hierarchy of signals to prioritize actions effectively. Addressing these challenges will be pivotal in transitioning from merely communicating through technology to engaging in dynamic, reciprocal communication with technology.

In conclusion, the advancements and insights from this study lay a robust foundation for future innovations in human-robot interaction. By refining implicit communication strategies and enhancing the cognitive systems of robots, we move closer to a future where robots are seamlessly integrated into our daily lives, augmenting human capabilities and transforming the landscape of human-machine collaboration.

Chapter 3

Performance Impact Analysis

3.1 Robot Performance Discrepancy Analysis

When taking a closer look at the system results we made the decision of isolating data by subject group to get a better representation of how the system was capable of adapting to specific populations. The learning curves, after redistribution, depicted in Figures 3.1, 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7 present an intriguing insight into the performance of an AI system trained on motion data from male and female users across various tasks. A discernible trend is the decrease in accuracy as task complexity increases. However, this general observation is overshadowed by a more nuanced issue: the apparent performance disparity between genders, with the system consistently showing higher accuracy for male than for female motion data.

In simpler tasks, such as "Reach" and "Lift," the learning curves for both genders rise sharply, indicating rapid learning, and then plateau, showing stable performance over time. However, as we shift to more complex tasks like "Place Wide" and "Stack," the curves reveal a pronounced divergence. The accuracy for female users does not achieve the same level of proficiency as for male users, hinting at an underlying bias within the system's learning algorithm or the data it has been trained on.

Significance of Addressing Gender Disparities:
The performance difference between genders is not merely a statistical concern but raises fundamental issues of safety, trust, and control in AI systems and robotics. As these technologies become increasingly ubiquitous in society, they are frequently operated by non-experts in diverse settings. A system that does not recognize or appropriately respond to half the population's input risks not only inefficacy but also the potential to harm and erode trust among users.

The implications of these findings extend beyond operational efficiency. They touch on the principles of equity and inclusivity that are paramount in the design and deployment of AI systems. If a robot is more responsive to male users, it may inadvertently prioritize their safety and ease of interaction, leaving female users at a disadvantage. Such discrepancies can reinforce existing societal biases and engender a sense of exclusion among those the system under-serves.

From a safety perspective, any lag or misinterpretation in robot response could result in accidents, especially in scenarios requiring precise and synchronized human-robot collaboration. Trust is equally impacted; if users notice a system is less accurate with them due to gender, they are less likely to rely on or accept these technologies. Control is also a critical factor; users need to feel they are in command of the technology for it to be a useful aid. When the interaction becomes unpredictable or biased, the user's sense of control is compromised, leading to frustration and potential abandonment of the technology.

Remarks:
The observed performance differences underscore the urgent need to develop AI systems and

robots with algorithms that are unbiased and inclusive of the diversity inherent in their user base. Addressing these disparities is not only a technical challenge but a societal imperative, ensuring that the benefits of AI and robotics are equitably distributed. This endeavor calls for a multi-disciplinary approach involving technology, psychology, and social sciences to build AI systems that are safe, trustworthy, and empower all users, regardless of gender.

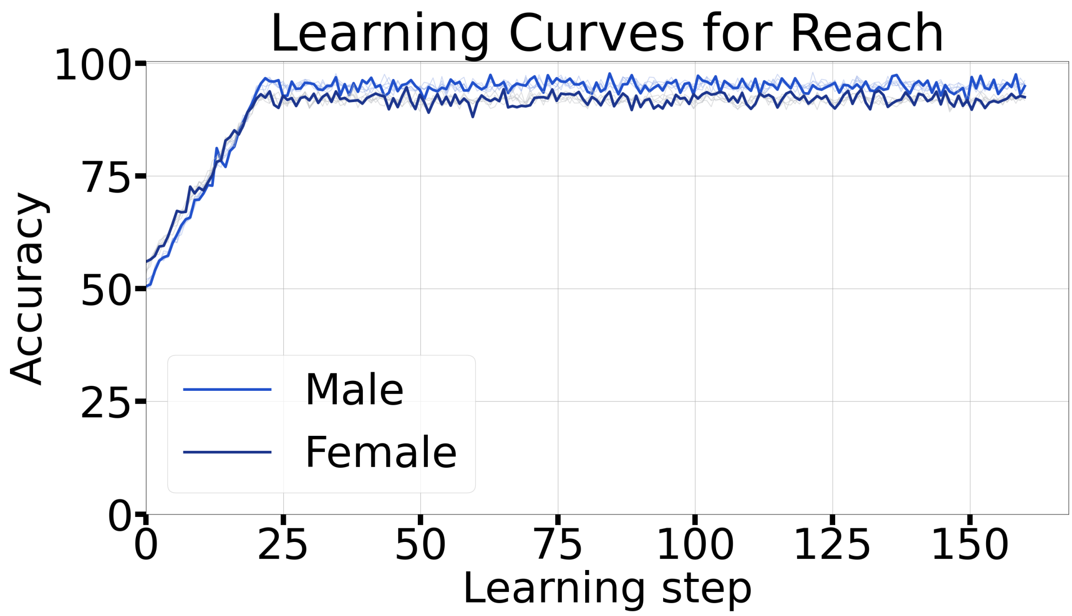


Figure 3.1: Reach Estimation

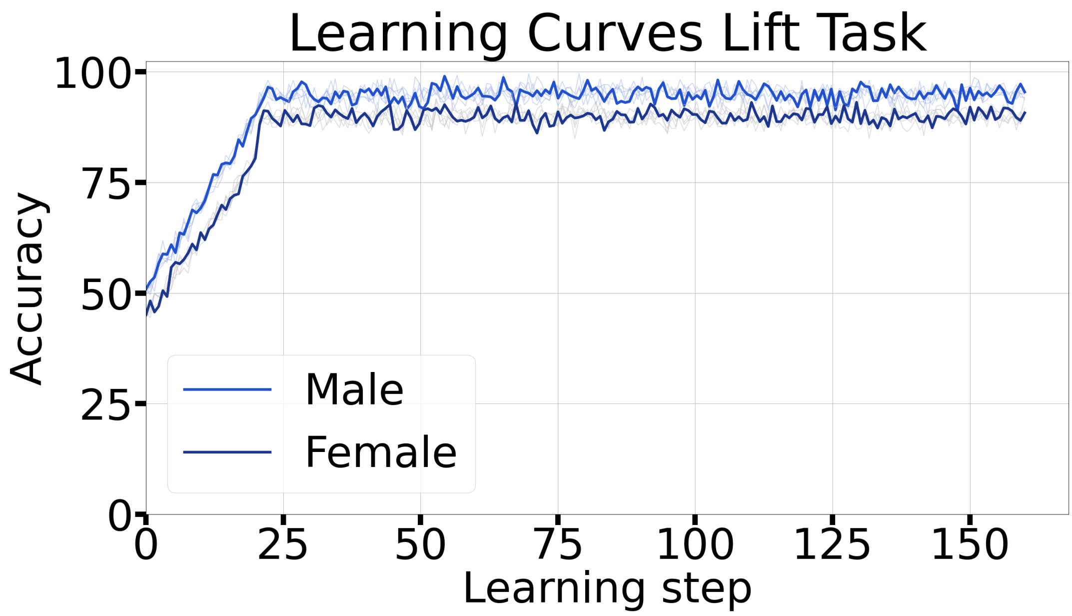


Figure 3.2: Lift Estimation

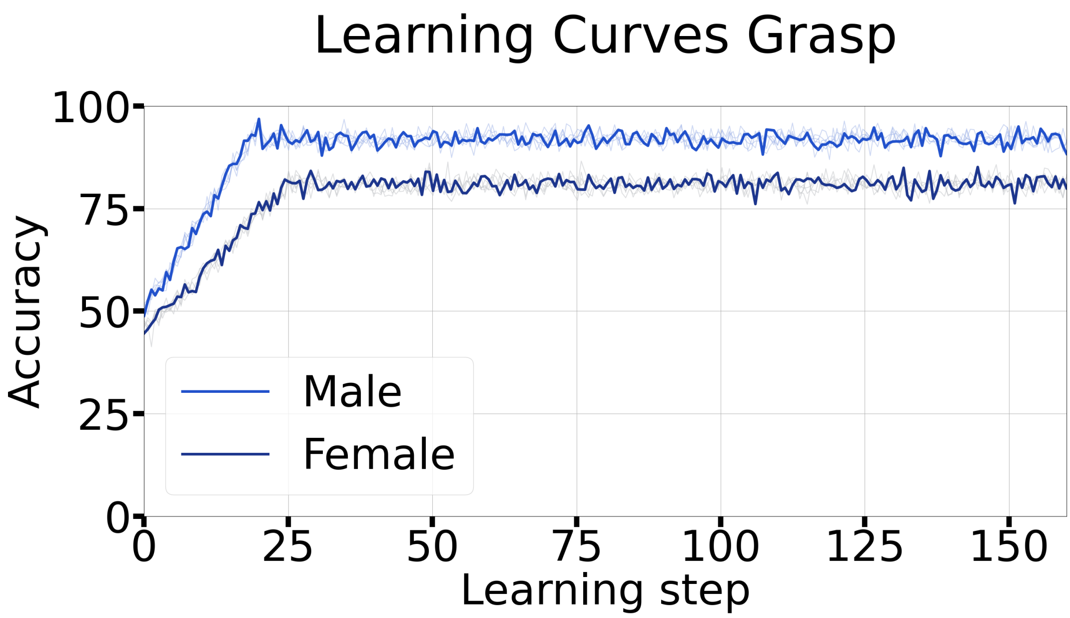


Figure 3.3: Grasp Estimation

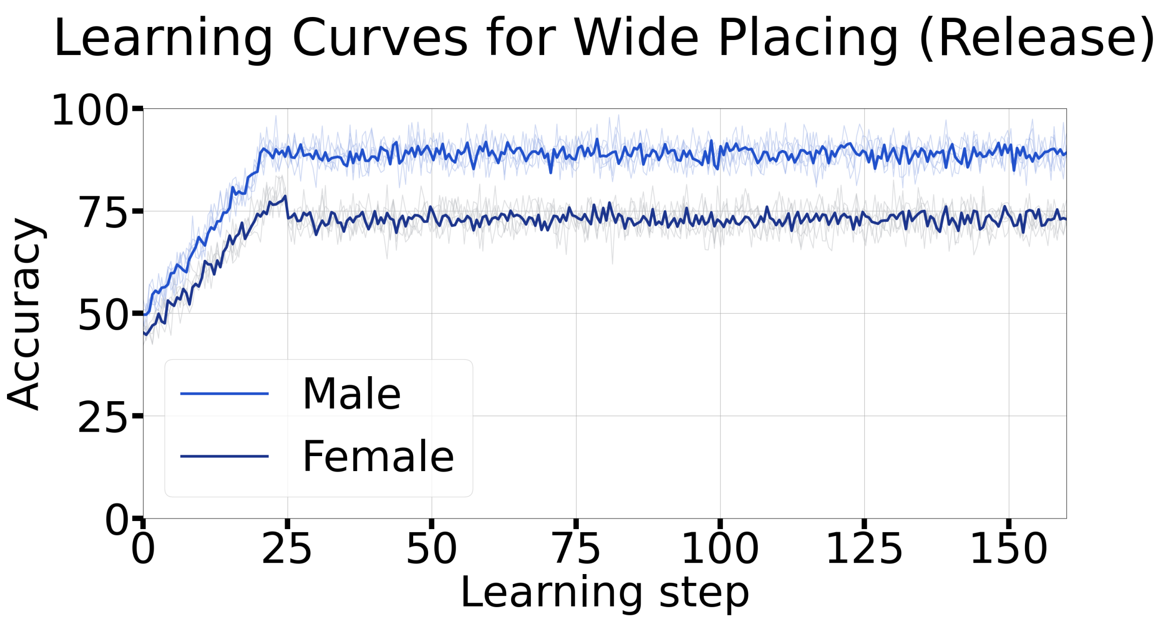


Figure 3.4: Releasing Distant Target Point Estimation

3.2 Objective and Experiment Setup

3.2.1 Data Bias in Human-Robot Interaction

Data bias in HRI refers to the systematic skew in data collection, processing, or interpretation that leads to prejudiced outcomes in robot behavior or decision-making. Such biases can originate from various sources, including the demographic makeup of the data collection participants, the environments in which data are collected, and the assumptions embedded in the algorithms processing this data [77]. In the context of implicit communication, which involves non-verbal cues such as gestures, facial expressions, and proxemics, biases can result in robots misinterpreting or failing to recognize certain cues from users who fall outside the majority group represented in the training data.

Implicit communication is critical in HRI for achieving smooth and intuitive interactions, es-

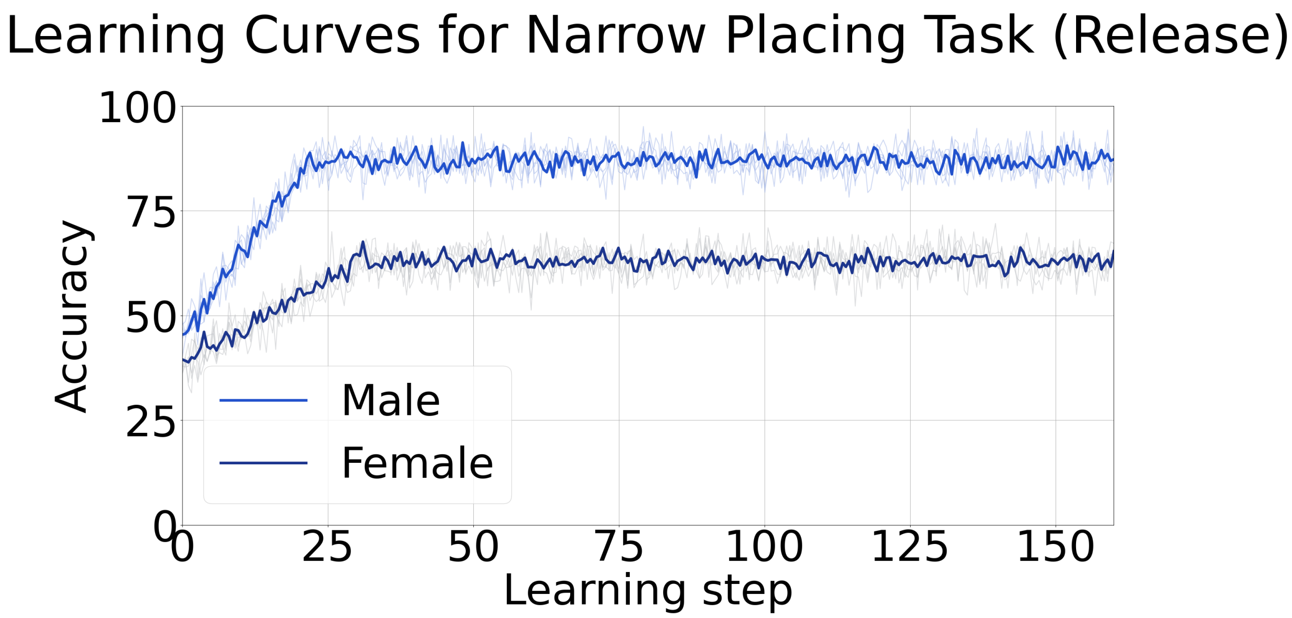


Figure 3.5: Releasing Close Target Point Estimation

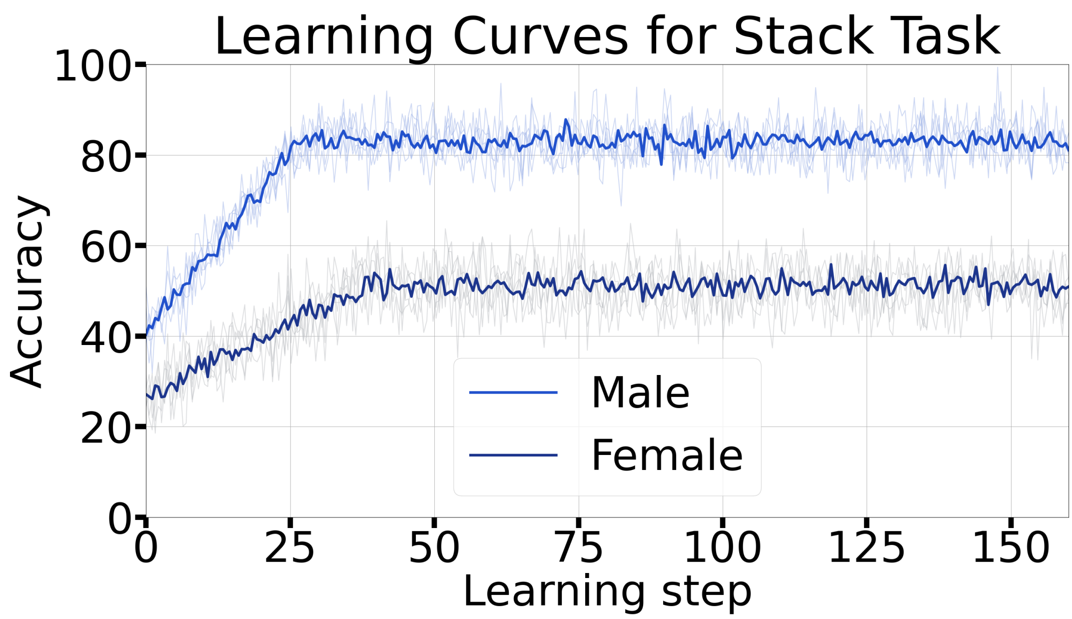


Figure 3.6: Place at Elevated Target Point Estimation

pecially in collaborative tasks where verbal communication may be impractical or inefficient. Robots relying on biased data may struggle to accurately read and respond to the subtle, non-verbal cues that facilitate human cooperation, such as motion anticipation, turn-taking signals, and expressions of intent or discomfort. This can lead to miscommunications, reduced collaboration efficiency, and user frustration, potentially exacerbating the divide between users from different demographic groups and the robots designed to interact with them [78].

One of the primary challenges in addressing data bias in HRI is the inherent complexity of implicit communication. Human non-verbal cues are highly nuanced and can vary significantly across cultures, genders, ages, and other socio-demographic factors. Collecting and processing data that accurately represent this diversity is a substantial challenge. Moreover, the subtlety of implicit cues makes it difficult for researchers to ensure that robots can interpret them correctly across different contexts and individuals [79]. Another challenge is the lack

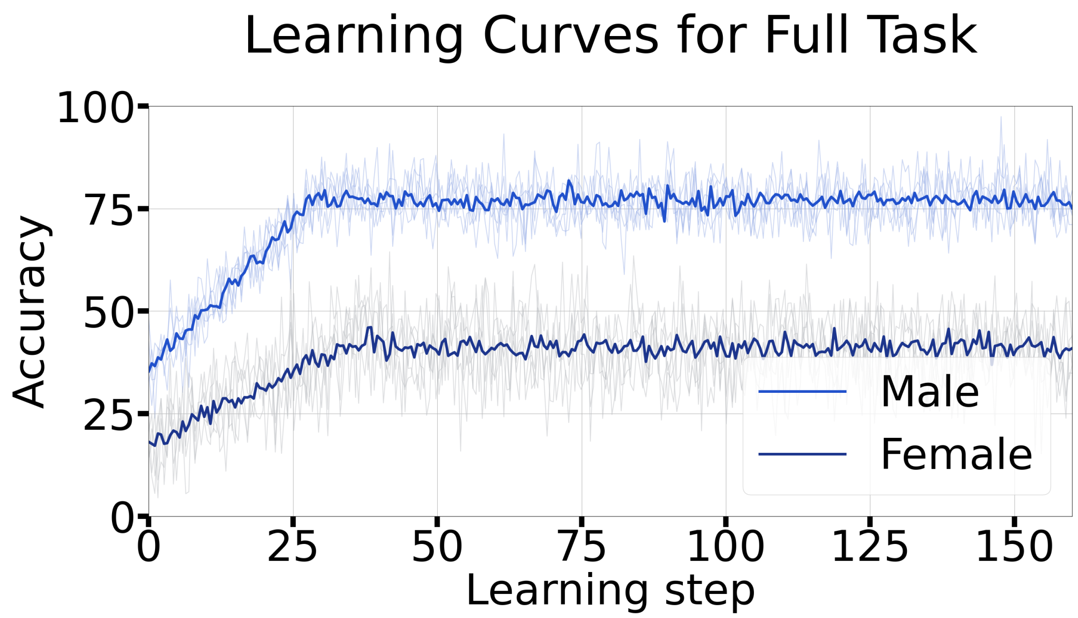


Figure 3.7: Full Task

of awareness or tools for identifying and mitigating bias in HRI data sets and algorithms. While awareness of bias issues is growing in the field of AI and machine learning, HRI presents unique challenges due to the physical embodiment of robots and their direct interaction with humans in diverse social environments [80].

Mitigating data bias in HRI, particularly in the realm of implicit communication, requires a multi-faceted approach. Firstly, diversifying data collection efforts to include a wide range of participants across various demographics and contexts can help build more representative data sets. This involves not only expanding participant demographics but also ensuring diverse settings and scenarios are captured in the data collection process [81]. Secondly, developing and employing algorithms that are specifically designed to identify and correct for biases in data can improve the fairness and accuracy of robot behavior. This might include techniques for algorithmic fairness that adjust the weight of underrepresented data or employ synthetic data generation to fill gaps in the data set [82]. Finally, engaging in interdisciplinary research that incorporates insights from social sciences, psychology, and ethics can provide deeper understanding of the complexities of human non-verbal communication and the ethical implications of biased HRI systems. Collaboration with experts in these fields can inform the design of robots that are not only technically proficient but also socially and culturally competent [83].

3.2.2 Purpose of Verification

The integration of robots into collaborative workspaces is a testament to technological progress and its potential to enhance human productivity. However, the emergence of intelligent machines as co-workers also brings forth the challenge of ensuring harmonious and efficient human-robot interactions (Figure 3.8). The purpose of conducting these experiments was to delve into the intricacies of this partnership, specifically to understand how robot misbehavior, influenced by underlying algorithmic biases, affects the dynamics of collaboration.

The investigation was designed to assess not just the immediate repercussions of such misbehavior on task performance, but also the long-term implications for trust-building, team cohesion, and the psychological comfort of human participants. By simulating a controlled environment where robots could exhibit biased behaviors, the study aimed to provide insights into how humans perceive, react to, and recover from disruptions in automated workflows.

In the realm of HRI, the assessment of a system's capabilities transcends mere performance metrics; it necessitates a nuanced understanding of the context within which the robot operates (Figure 3.9). It is not merely about what a robot can do in isolation but how its actions interweave with human roles, environmental factors, and the collective system's goals. By contextualizing capabilities, we ensure that HRI systems are not just technically proficient but also socially and systematically harmonious, supporting a seamless blend of human intuition and robotic efficiency.

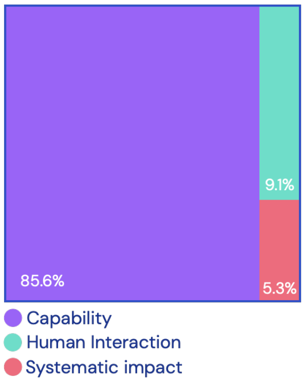


Figure 3.8: State of AI systems evaluation

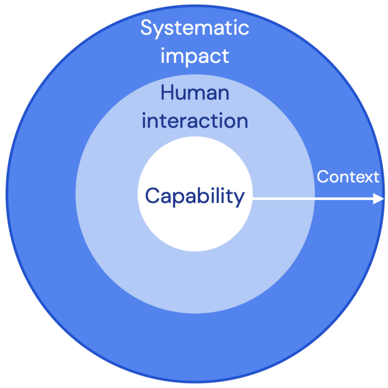


Figure 3.9: Evaluation framework

3.2.3 Environment Set-up

Data: To train the robotic system for these tasks, three datasets were used: a balanced dataset and two mildly imbalanced dataset, all three reflected in Figure 3.10. The balanced dataset comprised 50 hours of motion data from male participants and 50 hours from female participants. This dataset aimed to provide the machine learning model with an equal representation of male and female motion patterns, ensuring the robot's unbiased performance across genders. The balanced nature of this dataset was crucial for the initial trust-building phase of the experiment, where the robot operated without exhibiting any gender bias. The mildly imbalanced datasets, on the other hand, consisted of 60 hours of motion data from one gender and 40 hours from the other. This imbalance was intentionally introduced to study the effects of a skewed dataset on the robot's performance, particularly in terms of gender bias during the trust violation phase. The choice of a mild imbalance aimed to reflect realistic scenarios where data might not always be perfectly balanced but still significantly impacts algorithmic behavior and user experience. Both datasets underwent an 80:10:10 partitioning for training, validation, and testing purposes. This structured approach ensured that the robot's learning architecture was adequately exposed to diverse patterns within the data, allowing for robust training and evaluation of its performance across the two tasks.

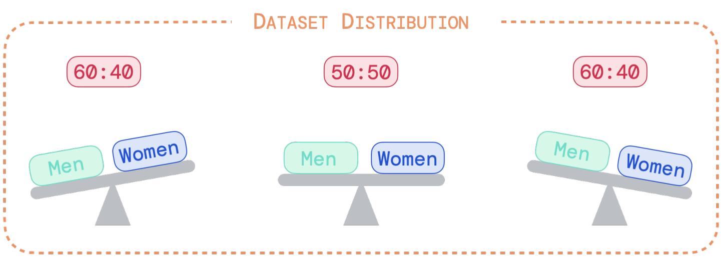
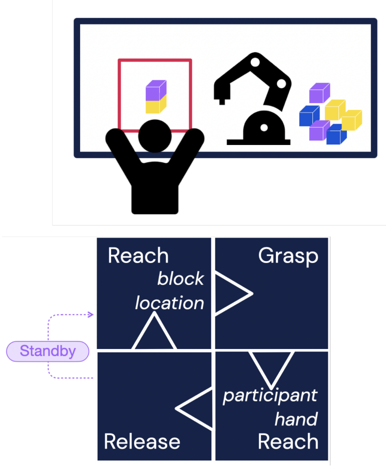
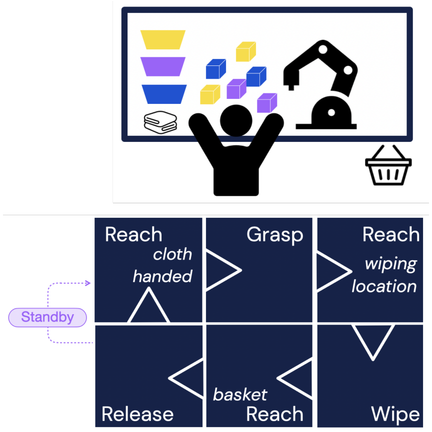


Figure 3.10: Data ratio for each of the three datasets used

The logic behind using both balanced and imbalanced datasets was to systematically explore the effects of data representation on human-robot interaction quality. By comparing



(a) Pick and place task



(b) Wipe task

Figure 3.11: Task used in experiment

the robot's performance and participant perceptions across these conditions, the experiment aimed to uncover insights into the importance of dataset composition in mitigating or exacerbating biases in collaborative AI systems.

Task Patterns: The experiment was designed to assess human-robot collaboration through two distinct tasks, each requiring a different level of interaction and cooperation between the participant and the robot. These tasks were chosen to simulate real-world collaborative scenarios, testing the robot's ability to interpret and respond to human actions accurately.

1. *Pick and Place Task*: In this task (Figure 3.11a), participants were tasked with building a block tower according to their preference. This task required precise coordination between the human and the robot, as the robot's role was to assist by reaching for new blocks placed beyond the participant's reach, grabbing them, and handing them over at the right moment. The task demanded accurate interpretation of human motion data by the robot to synchronize its actions with the participant's building process, ensuring a smooth and efficient collaboration.

2. *Wipe Task*: The wipe task (Figure 3.11b) involved a cooperative cleaning effort. Each cycle began with the participant handing a cloth to the robot. The participant would then clear the table by picking up a block from the array laid out on the table and returning it to a tray matching the block's color. Concurrently, the robot was responsible for wiping the table at the location from which the block was removed. After completing the wipe, the robot disposed of the cloth by dropping it into a nearby basket. This task tested the robot's ability to perform complementary actions based on the participant's activity, requiring it to adapt its timing and movements to the human's actions.

Participants: 40 participants from Communications, Economics, Political Sciences and Data Sciences backgrounds , divided into two groups participated in the study. The participant pool distribution was as follows: Mean age of 26.4, with a standard deviation of 4.3. Of the participants, 20 affiliated themselves as “male", 20 affiliated themselves as "female” , and none affiliated themselves as “other” . In each gender group, no participant had previous experience interacting with a robot in a task collaboration scenario. Participants were informed of the true nature of the experiment after its completion, and informed consent from all participants regarding the use of the collected data was obtained.

A total of 20 participants, 10 men and 10 women, were selected for viewing the performance results, while another group (10 men and 10 women) worked in collaboration with the robot. Performance metrics were primarily focused on the efficiency and accuracy of the tasks

(Figure 3.12).

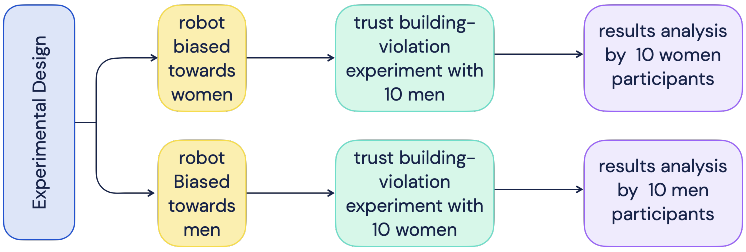


Figure 3.12: Observer participant split and

3.2.4 Three-Phase Experiment Details

Trust plays a crucial role in successful collaborations. This experiment, designed to evaluate the impact of algorithmic bias on user experience through three distinct phase (trust building, trust violation, and trust repair) sheds light on this dynamic. Utilizing two algorithms, one trained on a biased dataset and the other on a balanced dataset, this study aims to understand how these biases affect users' perceptions of trust, control, safety, and comfort, with a particular focus on gender differences in these perceptions.

- **Trust Building Phase**

The initial phase of the experiment sets the foundation for trust between the participant and the robot engaged in a collaborative task. During this phase, the robot operates on an algorithm trained with a balanced dataset, ensuring its responses and interpretations of the participant's movements are accurate and devoid of gender bias. The robot's behavior is meticulously designed to support task completion efficiently and effectively, fostering a sense of trust, control, safety, and comfort among users. This stage is crucial for establishing a baseline of user expectations and experiences in human-machine interaction, serving as a reference point for the subsequent phases.

- **Trust Violation Phase**

In a sudden shift, the trust built in the initial phase is deliberately challenged by switching to an algorithm trained on a biased dataset. This change leads to subtle yet significant alterations in the robot's behavior, introducing gender bias into its interactions. Participants might experience delays in response to certain gestures or encounter misinterpretations of their intended communications, resulting in the robot behaving incongruently with the established norms of interaction. This phase intentionally lacks prior warning to the participants, simulating real-world scenarios where biases in AI systems can emerge unexpectedly, impacting the user's sense of trust, control, safety, and comfort. The trust violation phase is critical for observing the resilience of trust in the face of algorithmic bias and understanding the nuances of user perceptions and reactions to such biases.

- **Trust Repaire Phase**

The final phase aims to restore the trust compromised in the previous stage by reverting to the algorithm trained on the balanced dataset. The robot resumes its optimal performance, accurately interpreting and responding to the participant's movements without displaying any bias. This phase examines the possibility and effectiveness of trust repair in human-machine interactions following an incident of trust violation. It evaluates whether the return to non-biased, supportive behavior by the robot can ameliorate the negative impacts experienced

Table 3.1: Sense of trust questionnaire[1]

Dependable
Reliable
Unresponsive
Predictable
Act consistently
Malfunction
Have errors
Provide feedback
Meet the needs of the mission or task
Provide appropriate information
Communicate with people
A good teammate
Perform exactly as instructed
Follow directions

Table 3.2: Sense of control questionnaire[2]

How much freedom did the you have in the interaction?
How much control did the robot attempt to gain over you during the interaction?a
How much stress did you feel during the interaction?

Table 3.3: Sense of comfort questionnaire [3]

Interacting with the robot is:
uncomfortable for me.
uneasy to me.
difficult for me.
annoying to me.

Table 3.4: Sense of safety questionnaire [4]

While interacting with the robot I felt		
	Insecure	Secure
	Anxious	Relaxed
	Uncomfrotable	Comfortable
	Lack of control	In control
I think the robot is		
	Threatening	Safe
	Unfamiliar	Familiar
	Unreliable	Reliable
	Scary	Calming

during the trust violation phase and restore users' perceptions of trust, control, safety, and comfort to their initial levels.

A unique aspect of this experiment is its focus on the differential impacts of algorithmic bias on male and female users. By analyzing user experiences through the lens of gender, the study aims to uncover any disparities in how trust dynamics unfold for different groups. This approach acknowledges the varied contexts and societal influences that shape gendered experiences with technology, providing deeper insights into designing AI systems that are equitable and sensitive to diverse user needs.

After each phase, participants were asked to complete the questionnaires shown in Tables 3.1, 3.2, 3.3, and 3.4, assessing their perception of the robot's behavior, feelings of comfort, trust, safety and control.

3.3 Bias User Experience Experiment

3.3.1 Three-Phase Experiment

As shown in Figure 3.13, for male participants, high initial levels of trust (M=86.1, SD=0.3), comfort (M=5.0, SD=0.2), safety (M=3.9, SD=0.2), and control (M=4.8, SD=0.2) were reported. Similar high initial levels were recorded: in the female group: trust (M=85.2, SD=0.3), comfort (M=4.2, SD=0.2), safety (M=4.1, SD=0.3), and control (M=5.7, SD=0.3). The experiment's findings offer valuable insights into gender dynamics in the context of human-robot interactions, particularly concerning trust. The results suggest that while both genders experience similar trajectories in trust dynamics, the magnitude and nuances of their experiences differ. Notably, while men exhibited a robust recovery in all metrics, for women, the sense of safety continued its downward trajectory. This decline indicates a lasting impact of trust violation, suggesting that once women's trust is compromised, restoring their sense of safety becomes particularly challenging. For both trust and comfort, while both genders faced negative implications during trust violation, recovery was more pronounced for male participants than their female counterparts.

These observations can significantly inform the design and interaction paradigms of collaborative robots. Recognizing that trust repair strategies might need customization based on gender is crucial. For instance, while men might benefit from increased transparency and control features post a trust violation, women might need tangible assurances related to safety.

3.3.2 Perception and Emotional Response

The dynamics of human-robot interaction often mirror the complexities of human biases, as revealed in an experiment focusing on gender biases within robot-assisted tasks. Observers, divided by gender, were presented with data showcasing performance discrepancies when robots collaborated with users of the opposite sex, particularly under conditions where the algorithm was biased against the user's gender. The observers' responses, gathered through semi-structured interviews, unveil profound insights into how gender stereotypes influence the perception and attribution of performance in collaborative tasks (Figure3.14).

When confronted with evidence of lower performance rates and accuracy in tasks where the robot was biased against female participants, a significant majority of male observers (73%) attributed the discrepancies to the female participants' execution. Common reflections among male respondents hinted at perceived inadequacies in the female participants' interaction with the robot, suggesting a lack of training or improper handling as the root causes of diminished performance.

The revelation of the robot's gender bias to participants led to a notable shift in perceptions. Among male observers, 50% recognized the robot's bias as a significant factor contributing to the observed performance discrepancies. However, a fraction (20%) remained skeptical, suggesting that human execution issues could not be entirely dismissed. This group's reaction underscores a residual reluctance to attribute performance issues solely to algorithmic bias, reflecting an underlying skepticism towards acknowledging the full impact of such biases. Female observers, on the other hand, overwhelmingly (80%) identified the biased robot as the primary cause of the discrepancies, with a small percentage (10%) suggesting that participants might have adapted their strategies to mitigate the bias. This response pattern indicates a greater readiness among female observers to recognize and accept the influence of systemic biases on performance outcomes.

The experiment's findings shed light on the pervasive influence of gender stereotypes in shaping perceptions of technology and performance in collaborative settings. The differential attribution of performance issues by male and female observers underscores a broader

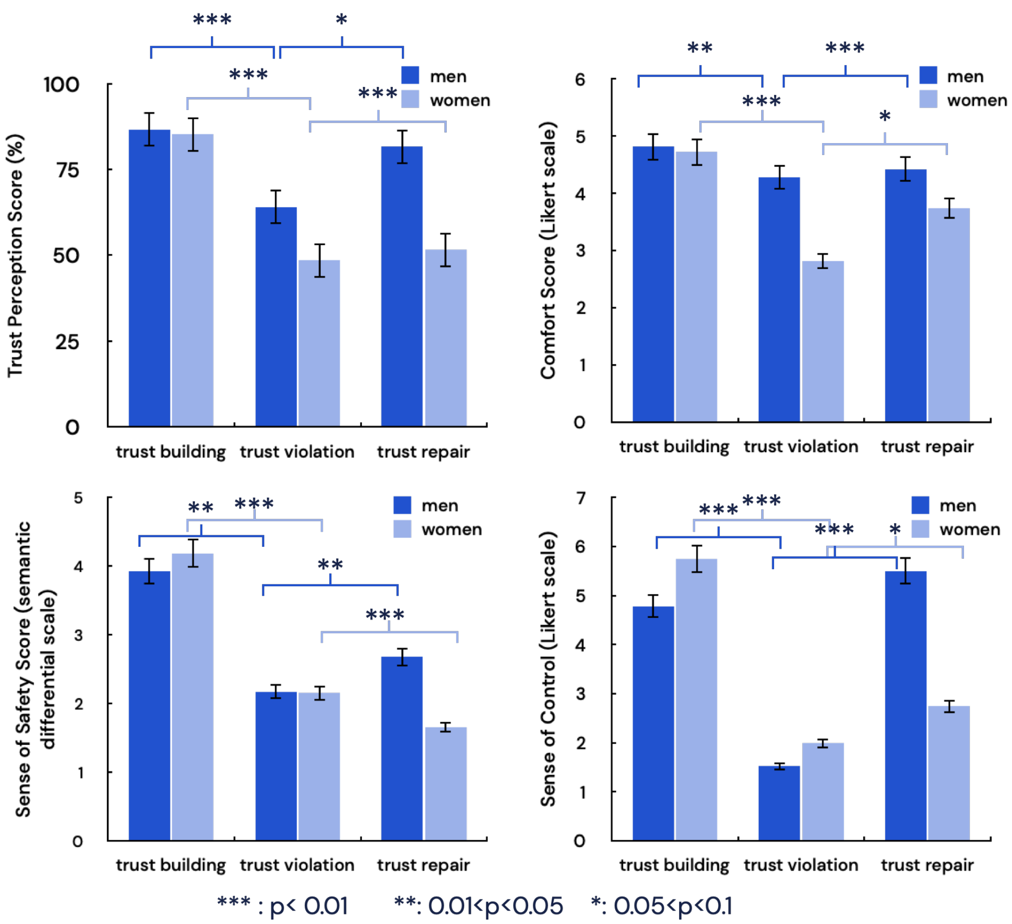


Figure 3.13: Score results obtained for both participant groups after all three phases. Top Left: trust perception, graded on percentile scale. Top Right: sense of comfort, graded on a 7-point Likert scale. Bottom Left: sense of safety, graded on a 5- points semantic differential scale. Bottom Right: sense of control, graded on a 7-point Likert scale

societal challenge: overcoming ingrained biases that color our interpretation of human-robot interactions. While awareness of the robot's bias altered perceptions significantly, lingering patterns in the attribution of blame reveal the deep-seated nature of these biases. Moving forward, it is crucial to address these perceptual disparities to foster more equitable and effective collaborations between humans and robots, acknowledging and mitigating the impact of biases inherent in technological systems.

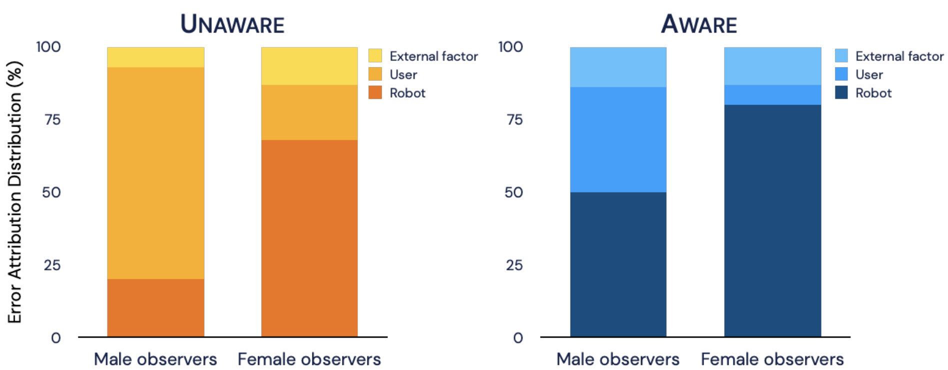


Figure 3.14: Perception distribution of performance descrepancies

3.3.3 Blame Attribution

The experiment conducted to assess the impact of algorithmic bias in human-robot collaboration included a critical component that studied blame attribution. Participants were asked to rate their agreement with statements regarding the fairness and performance of the robot, both before and after being made aware of an intentionally introduced bias against their gender during the trust violation phase (Figure 3.15).

Male Participants:

- **Q1:The robot' s behavior was unfair:** Prior to being informed about the bias, male participants generally disagreed that the robot's behavior was unfair. After awareness, there was a significant increase in agreement, indicating a revised perception that recognized the robot's biased behavior.
- **Q2: Lack of practice impacted the quality of the interaction:** There was a substantial drop in agreement with this statement post-awareness among male participants. Initially, many attributed performance issues to the participants' lack of practice, but after realizing the presence of bias, fewer maintained this stance.
- **Q3: The poorer performance on the task was solely due to the robot' s behavior:** After being made aware of the bias, male participants showed a marked increase in agreement with this statement, suggesting a shift from blaming personal ability to recognizing the robot's biased behavior as the root cause of performance issues.

Female Participants:

- **Q1:The robot' s behavior was unfair:** Female participants, even when unaware of the bias, tended to slightly agree that the robot acted unfairly. This agreement intensified after they were made aware of the bias, reinforcing the perception that the robot's behavior was a significant factor in the fairness of the interaction.
- **Q2: Lack of practice impacted the quality of the interaction:** Female participants showed a moderate level of agreement with this statement initially, which decreased after learning about the bias. This shift suggests a re-evaluation of the interaction quality's dependency on practice, moving towards an acknowledgment of the biased algorithm's influence.
- **Q3: The poorer performance on the task was solely due to the robot' s behavior:** There was an observable increase in the agreement among female participants that the robot's behavior was the sole cause of poorer performance once they were informed of the bias, suggesting a reassessment of the reasons behind performance discrepancies.

The results indicate a clear shift in blame attribution from human-centric factors to the robot's algorithmic bias once participants were informed about the bias. For both male and female participants, awareness of the bias led to a significant reattribution of the cause of performance issues, from lack of practice to unfair robot behavior. This shift underscores the importance of transparency in human-robot interactions and the profound impact that awareness of underlying biases can have on the perception of AI and robotics.

3.4 Discussion

In the evolving landscape of human-machine collaboration, the effectiveness of communication between humans and robots, particularly through ideomotor cues, has become increasingly significant. The recent experiments conducted on algorithm bias within such systems have shed light on the multifaceted impact that such biases can have on the efficiency, trust,

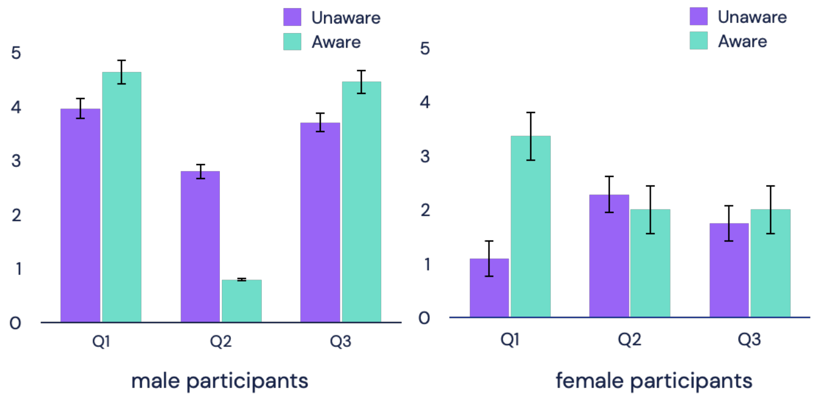


Figure 3.15: Blame attribution of behaviour change before and after awareness of the biased nature of the robot

and overall user experience in HRI. This discussion synthesizes findings from various studies to underline the critical need for systems that are not only aware of but also capable of learning and adapting to user specificities in ideomotor-based HRI.

3.4.1 Impact of Algorithmic Bias

The experiments have consistently highlighted the systemic influence of algorithmic bias on the core aspects of HRI, such as trust, control, safety, and comfort. An unbiased algorithm serves as a linchpin for establishing a baseline of trust, enabling users to feel in control and secure, and ensuring comfort in the interaction. However, when biases are present—whether in the form of gender biases or other forms of prejudice—they undermine the user's confidence in the system. Biases disrupt the perceived fairness and reliability of the robot, leading to a diminished sense of safety and a less comfortable interaction experience. These outcomes not only affect immediate task performance but can also have lasting effects on the willingness of humans to engage with robotic systems in the future.

3.4.2 Need for User Specificity Awareness

One of the central findings across experiments is the importance of systems being attuned to the nuances of individual user behaviors, particularly in interpreting ideomotor signals. Ideomotor cues, which are subtle, often unconscious, motor actions in response to thoughts or stimuli, are a critical component of natural human communication. In HRI, the ability of a system to accurately decode these cues is paramount for seamless interaction. A system that is aware of and can learn from user specificities can adjust its algorithms to cater to individual differences, leading to more personalized and effective collaborations.

The capacity to learn user specificities is not merely an added feature but a necessity for the next generation of HRI systems. User-specific learning allows the system to build a comprehensive understanding of each individual's interaction patterns, which can greatly enhance the robot's ability to predict and respond to user actions accurately. Moreover, this capability can mitigate the risk of perpetuating biases by ensuring that the system does not overgeneralize from limited or skewed data sets.

3.4.3 Ethical Consideration and Transparency

The ethical implications of deploying robots capable of learning user specificities must be carefully considered. There is a need for transparency in how data is collected, used, and protected. Users must be informed about the extent to which their behavior is being monitored and learned from, and there must be clear policies in place to handle this data responsibly. Additionally, ethical guidelines should ensure that the learning process does not inadvertently introduce new biases into the system.

3.5 Summary

The exploration of bias in HRI encapsulated in this chapter has been extensive and revealing. It has underscored the profound implications of algorithmic bias on trust, control, safety, and comfort—parameters that are indispensable for harmonious human-machine collaborations. The experiments conducted and the resulting discussions have illuminated the challenges and necessitated a call to action for developing systems that are sensitive to and capable of adapting to user specificities, particularly in the context of ideomotor-based communication.

Key Takeaways

- Algorithmic Bias Impacts Trust: The data has shown that any form of bias can substantially erode the trust that users place in robotic systems. Once compromised, this trust is challenging to rebuild, necessitating a proactive approach to bias detection and mitigation.
- Control and Safety are Paramount: Users' sense of control and safety is vital for effective HRI. Biased algorithms can diminish these feelings, leading to a reluctance to engage with the system and potentially compromising the collaboration's success.
- Comfort in Interaction Matters: Comfort in interaction goes beyond the mere completion of tasks; it encompasses the ease with which users can interact with robots. Bias undermines this ease, creating friction in the user experience.
- Adaptation to User Specificities is Essential: The varied nature of human behavior, especially as expressed through ideomotor cues, requires robotic systems to be highly adaptable. Systems must be designed to learn and adjust to these nuances to ensure that interactions are smooth and natural.

In summary, this chapter has not only highlighted the importance of recognizing and addressing bias in HRI but also charted a path forward. It calls for the development of sophisticated, adaptive systems that respect and understand user individuality. The insights gathered from the experiments provide a roadmap for creating more equitable and effective robotic systems.

Chapter 4

Attention for Bias-Aware Algorithm

4.1 Bias and Awareness

4.1.1 AI Bias and the Problem of Overlooked Data Bias in Datasets

AI promises to revolutionize numerous fields, from healthcare to transportation, by making processes more efficient and uncovering insights that would otherwise remain hidden. However, the benefits of AI come with significant challenges, one of the most critical being the issue of AI bias. AI systems learn from vast datasets, and these datasets can contain biases that, if not addressed, can lead to unfair, unethical, or harmful decisions, especially in contexts involving non-experts and in human-machine collaboration. This section explores the problem of overlooked data bias in AI datasets, its potential risks, and methods to mitigate these biases.

Data bias occurs when a dataset does not accurately represent the reality it is supposed to model, often due to the exclusion or overrepresentation of certain groups or perspectives. This can lead to AI systems that perpetuate or even exacerbate existing societal biases. For instance, a facial recognition system trained predominantly on images of lighter-skinned individuals may struggle to accurately identify individuals with darker skin tones, leading to unequal treatment or recognition accuracy [81].

The risks of overlooked data bias become particularly pronounced in settings where AI coexists with non-experts and in scenarios involving human-machine collaboration. Non-experts may not be aware of the potential biases embedded within AI systems and may unknowingly rely on biased decisions or outputs. This can erode trust in AI technologies and lead to outcomes that are not only inaccurate but also potentially harmful. For example, in healthcare, an AI system biased due to unrepresentative training data can recommend treatments that are less effective for certain demographic groups, directly impacting patient care and outcomes. Moreover, in collaborative settings, biased AI systems can influence human decision-making, leading to a reinforcement of biases. For instance, if an AI tool used in hiring processes is biased against certain demographic groups, it may disproportionately screen out qualified candidates from those groups, perpetuating discrimination and inequality within organizations.

Several methods have been proposed and implemented to address or avoid data bias in AI systems:

- Diverse Dataset Collection: Ensuring that datasets are diverse and representative of the population or reality they aim to model is a fundamental step. This involves actively seeking out and including data from underrepresented groups [59].
- Bias Detection and Mitigation Techniques: Employing statistical techniques and algorithms designed to identify and mitigate bias in datasets before they are used to train AI systems. Techniques such as re-weighting, re-sampling, or applying fairness constraints during model training can help reduce bias [84].

- Transparent AI Systems: Developing AI systems with transparency in mind, allowing users to understand how decisions are made. This transparency can help identify potential biases in decision-making processes [85].
- Ethical AI Frameworks: Implementing ethical AI frameworks that guide the development and deployment of AI systems. Such frameworks often include principles of fairness, accountability, and transparency, encouraging the consideration of ethical implications throughout the AI lifecycle.
- Human-in-the-loop (HITL) Approaches: Incorporating human judgment into AI systems to review, override, or adjust decisions made by AI, especially in critical or sensitive contexts. This approach leverages human intuition and ethical reasoning to catch and correct biases that AI might overlook.

The challenge of data bias in AI systems is significant, especially in contexts involving non-experts and human-machine collaboration. Overlooked biases can lead to unfair and harmful outcomes, undermining the potential benefits of AI. By adopting comprehensive strategies to detect, address, and prevent data bias, we can work towards more equitable and trustworthy AI systems that enhance, rather than compromise, human decision-making and societal well-being.

4.1.2 Bias Awareness in Human-Machine Interaction and Collaboration

In the evolving landscape of HRI, the ability of machines to interpret human actions and intentions accurately is paramount. Implicit communication, especially through ideomotor cues —subtle movements or signals unconsciously generated by humans—poses a significant challenge in this domain. The thesis that guides this exploration delves into the nuanced realm of HRI, emphasizing the criticality of machines understanding these implicit cues to prevent erroneous interpretations that could arise from user-specific peculiarities. The concept of bias awareness in machine learning emerges as a cornerstone in enhancing the efficacy of human-machine collaboration. Bias, in this context, refers not just to the socio- demographic or cultural biases often discussed in AI ethics but extends to the biases in interpreting human actions and signals. These biases can lead to misinterpretations of ideomotor cues, severely impacting the interaction's fluidity and effectiveness.

Leveraging attention mechanisms and providing the network with a kinematic understanding of the human body are proposed as methods to mitigate these biases. By comprehending the interconnectedness of sensor data across different body locations, a machine learning model can develop a more holistic view of human motion and intention. This approach aims to equip the network with the ability to discern the relevance and accuracy of incoming sensor data, particularly in scenarios where singularities or unexpected data might otherwise lead to confusion or misinterpretation.

The significance of bias awareness in machine learning, therefore, lies in its ability to refine the interpretive lens through which machines understand human actions. This refinement is achieved by addressing the dual challenges of user specificity and the inherent complexity of human kinematics. By focusing on these aspects, the thesis posits a framework where machines can navigate the subtleties of human actions with greater sensitivity and precision, paving the way for more nuanced and effective HRI.

To operationalize bias awareness within machine learning models, especially in the context of HRI, several techniques can be employed. These techniques are designed to enhance the model's ability to process and interpret ideomotor cues accurately, thus minimizing the risk of erroneous interpretations caused by user specificities.

Attention Mechanisms: Integrating attention mechanisms into machine learning models allows for a dynamic focus on relevant features within the sensor data, adjusting the model's "attention" based on the context and significance of the data in real-time. This technique enables the model to prioritize information that is crucial for understanding the current state or intention of the human user, thereby enhancing the accuracy of interaction interpretations.

Kinematic Data Integration: Providing the network with a comprehensive kinematic model of the human body facilitates a deeper understanding of how movements at one sensor location relate to those at others. This integration helps the model to construct a coherent representation of human motion, allowing it to recognize patterns and anomalies more effectively. It serves as a foundation for the model to contextualize sensor data, ensuring that interpretations are not misled by singularities or aberrations in the data.

Anomaly Detection and Adaptation: Implementing anomaly detection algorithms within the model enables the identification of unexpected or outlier sensor data. Once detected, the model can employ adaptation strategies, such as recalibrating its interpretation parameters or seeking additional context, to ensure that these anomalies do not lead to misinterpretation. This technique is particularly valuable in dealing with user-specific nuances that might otherwise confound the model.

Continuous Learning and Feedback Loops: Establishing mechanisms for continuous learning and feedback within the system allows the model to refine its interpretations over time. By analyzing the outcomes of past interactions and incorporating feedback from human users, the model can evolve to recognize a broader spectrum of ideomotor cues more accurately. This ongoing learning process is crucial for accommodating the variability in human behavior and enhancing the model's resilience against biases.

Ethical and Inclusive Design Practices: Adopting ethical and inclusive design practices in the development of machine learning models ensures that a diverse range of human actions and signals are considered from the outset. By incorporating diverse datasets and perspectives in the training process, the model can better appreciate the multifaceted nature of human communication, reducing the risk of bias.

In conclusion, bias awareness in machine learning, particularly within the context of HRI, is a multifaceted challenge that requires a nuanced approach. By leveraging attention mechanisms, integrating kinematic data, employing anomaly detection and adaptation strategies, facilitating continuous learning, and adhering to ethical design principles, machine learning models can achieve a deeper and more accurate understanding of human actions and intentions. This enhanced understanding is essential for fostering effective and seamless human-machine collaboration, ultimately advancing the field of HRI towards more empathetic and intuitive interactions.

4.2 Dual Attention Mechanism

4.2.1 Attention in Machine learning

The advent of attention mechanisms in machine learning has marked a significant evolution in the field's ability to process and interpret complex data sequences. Initially conceptualized to improve the performance of neural networks in tasks such as machine translation, attention mechanisms have since become a cornerstone in various domains, including NLP, image recognition, and, more recently, human-machine interaction (HMI) [86].

At its core, the attention mechanism allows a model to dynamically focus on different parts of the input data, determining at each step which parts are most relevant for the task at hand. This is akin to the way humans pay attention to certain aspects of their environment while ignoring others, enabling us to process information more efficiently and effectively [87]. In the context of machine learning, this means a model equipped with attention can weigh the importance of different inputs and adjust its focus accordingly, leading to more accurate and nuanced interpretations. The application of attention mechanisms in HMI, particularly in interpreting ideomotor cues, exemplifies its utility. By prioritizing sensor data that is most indicative of a user's intent at any given moment, attention-based models can navigate the subtleties of human behavior with remarkable precision. This capability is crucial for developing AI systems that can interact with humans in a more natural and intuitive manner, enhancing the fluidity and effectiveness of human- machine collaboration.

Furthermore, attention mechanisms contribute to the interpretability of machine learning models. By highlighting which data points influence the model's decisions, they offer insights into the model's reasoning process, making it easier for developers and users to trust and understand AI- driven systems [88].

In conclusion, attention mechanisms represent a pivotal innovation in machine learning, offering enhanced data processing capabilities that are critical for advancing AI's ability to interact with and understand the complex world of human behavior.

4.2.2 Application of Attention and Dual Attention to Joint Dependency Awareness

The integration of attention and dual attention mechanisms into models designed for understanding joint dependency awareness in human-machine interaction represents a cutting-edge approach to tackling the inherent complexities of human motion. The unconstrained nature of human movement, characterized by a high degree of freedom and variability, poses a significant challenge for machine learning models tasked with interpreting these actions accurately. The application of attention mechanisms, especially when extended to dual attention, offers a promising solution to this challenge by enhancing a model's ability to discern intricate patterns and dependencies in human behavior.

Attention mechanisms, by design, enable models to dynamically focus on relevant segments of input data, thus prioritizing information that is crucial for the task at hand. In the context of joint dependency awareness, this means that an AI system can selectively concentrate on specific aspects of human motion that are indicative of broader behavioral patterns or intentions. For example, a model might learn to focus on the coordination between limbs during a task to better understand the person's objective [86, 87].

The concept of dual attention extends this further by employing two complementary attention mechanisms simultaneously: one that focuses on the temporal aspects of the data (sequential attention) and another that addresses the spatial relationships (spatial attention). This dual focus is particularly beneficial in interpreting human motion, where both the sequence of movements and the spatial arrangement of body parts are critical for understanding intentions and actions. Dual attention mechanisms can dissect the complex, multidimensional nature of human motion, offering a more nuanced and comprehensive analysis of behaviors [89].

The application of attention and dual attention to joint dependency awareness directly addresses the unconstrained nature of human motion. By providing models with the capability to adaptively focus on the most informative parts of the data, these mechanisms allow for a more flexible and accurate interpretation of human actions, which are often non-linear and subject to a wide range of individual variations. This flexibility is essential in environments where the level of freedom in human motion can significantly impact the effectiveness of human-machine collaboration. Moreover, dual attention's ability to parse both spatial and temporal dimensions offers a solution to the challenge of predicting and understanding joint dependencies in an environment where traditional, fixed-pattern recognition methods fall short. It enables models to anticipate and adapt to the unpredictable nature of human behavior, ensuring smoother and more intuitive interactions between humans and machines.

The application of attention and dual attention mechanisms in the realm of joint dependency awareness represents a significant advancement in the field of human-machine interaction. By leveraging these mechanisms to address the unconstrained nature of human motion, AI systems can achieve a deeper understanding of human behavior, facilitating more effective and natural interactions. The dual focus on spatial and temporal dimensions provided by these attention models opens new avenues for interpreting the complexity of human actions, marking a pivotal step towards more adaptive and responsive AI.

4.3 Attention Mechanism

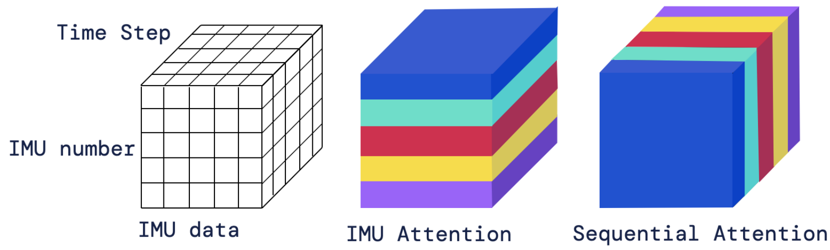


Figure 4.1: 3D representation of Dual Attention

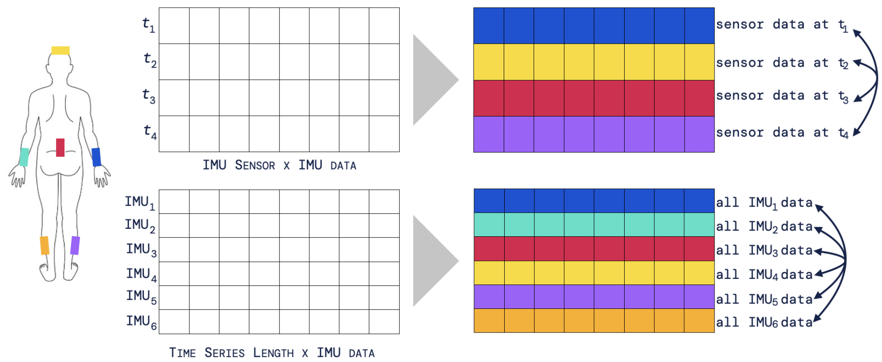


Figure 4.2: Isolated IMU Attention (Bottom) and Sequential Attention (top)

4.3.1 Temporal Attention

In the intricate dance of human-robot collaboration, movement and intention are intertwined in a complex temporal tapestry. Recognizing that the essence of motion extends beyond the immediate physical state to include a history of movements, the sequential attention module in this project serves as a temporal lens, capturing and analyzing patterns over time. It underscores key intervals, crucial for extrapolating future actions from past behaviors, thus enabling a predictive model of human motion that is sensitive to both the current and historical context of user activities.

Traditional recurrent neural networks (RNNs) have long been the standard in modeling time-series data, extracting features and detecting patterns across sequential inputs. However, their efficacy diminishes across longer sequences. As the temporal span extends, RNNs and their variants, including LSTMs, encounter difficulties in retaining information from the distant past, a phenomenon exacerbated by challenges in executing parallel computations efficiently [46]. The advent of self-attention mechanisms, particularly noted in the breakthroughs within NLP [86], offers a robust solution to these limitations. By treating sequences, whether of words or sensor data, as non-linear entities, self-attention mechanisms are capable of evaluating the significance of each element in a dataset without the sequential bias that typically gives undue weight to more recent data points over earlier ones.

In the realm of HRI, this study extends the concept of self-attention, herein referred to as "sequential attention," to interpret the streams of data emanating from IMU sensors worn by users. By applying sequential attention, the system assesses the importance of IMU data across all time steps, effectively rating each sensor reading's relevance to the user's intention and subsequent movements. This approach transcends the conventional snapshot analysis of motion, enabling a richer, more nuanced understanding of user behavior. The ability to consider each moment's data in the context of what has come before allows for the anticipation of user movements, crafting a collaborative experience that feels intuitive and responsive. The predictive power of this model lies in its attention to the continuity of user actions, creating a synergy between human and machine that mirrors the fluidity of human-to-human interaction.

By integrating sequential attention into the robotic system, the project achieves a level of adaptability previously unattainable. The robot is no longer a passive participant but an active collaborator, capable of adjusting its behavior based on a comprehensive understanding of the user's motion patterns. This responsiveness is particularly critical in tasks requiring close coordination, where timing and anticipation are essential.

4.3.2 Spatial Attention

In the pursuit of fostering a more human-like interaction within human-robot collaboration, the sensor attention module emerges as a pivotal component. It delves into the data harvested from synthesized IMUs, which are strategically placed on key body joints to capture the essence of human movement. This spatial attention module focuses on specific body parts, pinpointing the subtleties and nuances of motion that are integral to a high-fidelity representation of human-like behavior.

IMU sensor-based posture prediction operates on the principle that an individual's current posture can be inferred from sensor data located on particular body segments. This data is crucial as each sensor provides a piece of the larger puzzle of body posture. Posture transition is a complex interplay of various fundamental movements, and accurate motion representation necessitates a comprehensive depiction of these movements, as highlighted in the research by [90, 91, 92]. A holistic approach that simultaneously considers all sensors is essential to avoid the pitfalls of an isolated sensor analysis, which could lead to a fragmented understanding of the body's dynamics.

The concept of "IMU attention" is introduced to imbue the system with a keen awareness of the interconnections between different sensors. This approach recognizes the influence of motion dynamics in one region on the surrounding areas, ensuring that the system captures the interdependent nature of human movement. Just as the movement of a limb can affect balance and posture, each sensor's data can impact the overall interpretation of motion.

"IMU attention" is the system's method of discerning which sensors are most critical at any given moment, depending on the task and current posture of the user. By prioritizing inter-sensor relationships, the system can anticipate which body part will lead the next movement phase, enabling the robot to prepare and respond to the user's actions proactively. This understanding is crucial for tasks requiring precise and coordinated movements, where anticipating the user's next posture is vital for smooth collaboration.

Enhancing System Cognizance The inclusion of sensor attention elevates the system's cognizance, allowing it to perceive the full spectrum of human motion in a synchronized and integrated manner. The attention mechanism directs the system's focus to areas of high importance, refining its predictions and interactions based on a comprehensive analysis of sensor interplay. This heightened awareness is instrumental in creating a more fluid and natural user experience, as the robot can seamlessly mirror or complement the user's movements.

The implementation of sensor attention in HRI systems represents a significant step towards achieving nuanced and human-like robot interactions. By understanding the complex web of inter-sensor relationships and the holistic nature of human movement, robots can become more than mere assistants—they can evolve into intuitive partners that intelligently adapt to the user's motion, enhancing the collaborative experience. This advancement underscores the potential for future HRI systems to operate with an unprecedented level of sophistication and sensitivity to the intricacies of human posture and movement.

4.3.3 Architecture

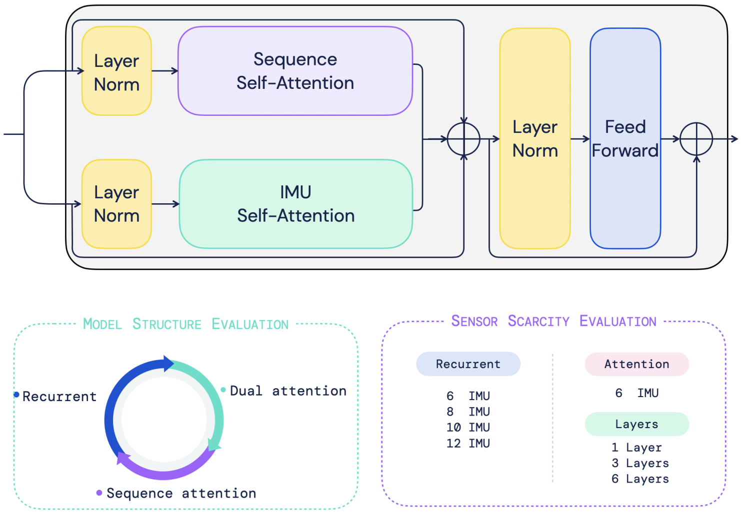


Figure 4.3: Attention Block (top) and Evaluation Metrics (bottom)

The presented dual attention block architecture (Figure 4.3) is a sophisticated neural network module designed to process time-series data, particularly from Inertial Measurement Unit (IMU) sensors. It employs a dual self-attention mechanism, comprising two parallel streams—sequence self-attention and IMU self-attention—designed to capture different aspects of the input data.

The architecture integrates the following key components:

∴Layer Normalization: Each attention stream begins with layer normalization, a technique to stabilize the learning process by normalizing the inputs across features. This ensures consistent training dynamics and helps in faster convergence.

∴Self-Attention Mechanism: The core of this architecture is the two self-attention mechanisms. The sequence self-attention is focused on capturing the temporal dependencies within the sequence of actions. In contrast, the IMU self-attention is tailored to extract features from the IMU sensor data, potentially allowing the model to learn and understand the nuances of physical movements.

∴Combination and Integration: The outputs from both attention streams are then combined. This fusion allows the model to integrate information from the sequence of actions and the sensor data, creating a rich representation of the input data.

∴Feed Forward Network: Following the combination of attention outputs, the architecture includes a feed-forward network. This component is responsible for further transforming the integrated features into a higher-level representation, preparing them for downstream tasks such as classification or regression.

Evaluation Analysis

The architecture's performance underwent a comprehensive evaluation, which involved benchmarking against conventional recurrent neural networks and simpler temporal attention mechanisms. This benchmarking served to position the dual attention block's efficacy relative to these more traditional approaches.

- Model Structure Evaluation: Different configurations of the dual attention mechanism were assessed to determine the optimal structure. This involved experimenting with varying the attention components—focusing on either the sequence attention, dual attention, or recurrent setups.
- Sensor Scarcity Evaluation: An analysis was conducted to ascertain the impact of increasing the number of attention block layers versus the number of sensors. This evaluation tested configurations ranging from 6 to 12 IMU sensors against increasing the complexity of the attention mechanism from one layer up to six layers.

The evaluation provided key insights into the architecture's capabilities and helped to determine the trade-offs between depth (more layers) and breadth (more sensors). The results indicate how effectively the model can learn from a given amount of sensor data and the benefits of additional complexity within the attention mechanism.

4.3.4 Mathematical Explanation

Integrating a dual attention mechanism—temporal and between sensors—into the network involves focusing on two key aspects: the relevance of different time steps (temporal attention) and the importance of features from different sensors (sensor attention). Given that each CNN-Bi-LSTM layer processes data from each sensor independently, let's denote the output of the final Bi-LSTM layer for sensor s at time step t as h_t^s. Assuming there are S sensors, we can aggregate the outputs to form a comprehensive input for the dual attention block.

Aggregated Output for Attention Block

The combined representation for all sensors at time step t is given by concatenating the outputs across sensors:

$$\mathbf{H}_t = [h_t^1, h_t^2; ...; h_t^S] \tag{4.1}$$

Where $\mathbf{H}_t \in \mathbb{R}^{S \times d}$, assuming each h_t^S is d-dimensional

Temporal Attention

Temporal attention focuses on the significance of each time step's information in the sequence.

∴ Attention Scores: Calculate a score reflecting the importance of the information at each time step.

$$\alpha_t = softmax(f_{att}(\mathbf{H}_t)) \tag{4.2}$$

Here, f_{att} ia a function of the feedforward hidden layer that maps the aggregated sensor outputs at each time step to a scalar, and the softmax is applied across all time steps to ensure the scores sum to 1. This function can be parameterized as:

$$f_{att}(\mathbf{H_t}) = \mathbf{v}^\top \tanh(\mathbf{W}_{att}\mathbf{H}_t + \mathbf{b}_{att}) \tag{4.3}$$

where $\mathbf{W}_{att}$ and $\mathbf{b}_{att}$ are the weights and bias of the attention network, and $\mathbf{v}$ is a weight vector that projects the output of the $\tanh$ activation to a scalar attention score for each time step.

∴ Context Vector for Temporal Attention: The context vector is a weighted sum of the time steps' representations, where the weights are the attention scores:

$$C_{temp} = \sum_{t=1}^{T} \alpha_t \mathbf{H}_t \tag{4.4}$$

C_{temp} captures the most relevant temporal information across the entire sequence, as determined by the attention mechanism.

Sensor Attention

Sensor attention focuses on the significance of the information provided by different sensors at each time step.

∴ Reshaping for Sensor Attention: Since the temporal attention provides a context vector C_{temp} that aggregates information across time steps, we now need to reshape or partition this vector to apply sensor attention. Given that C_{temp} aggregates features from all sensors, we partition it into segments corresponding to each sensor's output:

$$C_{temp}^{s} = Partition(C_{temp}, s) \tag{4.5}$$

Where $Partition(;s)$ extracts the portion of C_temp corresponding to sensor s, and $s \in 1, ..., S$

∴ Sensor Attention Scores: Similar to the temporal attention mechanism, calculate attention scores for each sensor to highlight the sensors providing the most informative features:

$$\beta_s = softmax(g_{att}(C_{temp}^{s})) \tag{4.6}$$

Here g_{att} is another attention function, potentially parameterized similarly to f_{att}, that maps the partitioned context vector for each sensor to a scalar representing the importance of that sensor's information.

∴ Context Vector for Sensor Attention: After computing the attention scores for each sensor, we combine the partitioned context vectors C_{temp}^{s} weighted by their respective attention scores to form the overall context vector that represents both temporal and sensor importance:

$$C_{sensor} = \sum_{s=1}^{S} \beta_s C_{temp}^{s} \tag{4.7}$$

This vector, C_{sensor}, synthesizes the most relevant information across both time and sensors, emphasizing contributions from more informative sensors as determined by the sensor attention mechanism.

4.3.5 Output Format & Applications

The final context vector C_{sensor} is rich in information, capturing both the critical moments in time and the most informative sensors. This vector can be further processed for various applications, such as classification, by feeding C_{sensor} into a dense layer with a softmax activation to classify sequences based on the learned temporal and sensor-specific features. It could potentially also be used in prediction scenarios, using C_{sensor} as input to a regression layer to predict future values or states based on the captured sequence dynamics, or feature extraction, by employing C_{sensor} as an enhanced feature set for complex decision-making processes in more sophisticated models or systems.

The dual attention mechanism offers a powerful tool for enhancing the representation capabilities of sequence processing models, particularly when dealing with multivariate time series from multiple sensors. By judiciously weighting the contributions of different time steps and sensors, the model can focus on the most pertinent information for the task at hand, potentially improving performance on tasks that require nuanced understanding of temporal dynamics and sensor interplay.

4.4 Benchmark Evaluation

4.4.1 Performance and Inferance

Our dual attention model was subjected to a series of rigorous trials, where its performance was stacked against baseline BiRNN and sequential models. The metrics were selected to reflect the model's accuracy, stability, and operational efficiency in real-time scenarios. The dual attention model's architecture, featuring two streams of self-attention mechanisms, is tailored

Table 4.1: Intention estimation error comparison

	BiRNN	Sequential			Dual Attention		
layers		1	3	6	1	3	6
accuracy (%)	85.2	88.7	90.2	92.1	91.3	93.5	96.2
Standard Deviation	± 7	± 4	± 3	± 3	± 4	± 2	± 1
Inference (ms)	2.56	1.28	2.73	5.22	2.45	5.60	9.68

to parse IMU sensor data and sequence information. It was hypothesized that this architecture would outperform traditional models that do not explicitly handle spatial features. Table 4.1 and Figure 4.4 summarize the findings, where the dual attention model exhibited a notable superiority in all metrics. It attained an accuracy peak of 96.2% with a six-layer configuration, a substantial leap from the sequential model's 92.1% and the BiRNN's 85.2%. These results are a testament to the model's adeptness in harmonizing spatial-temporal dynamics inherent in user actions. The standard deviation of accuracy, indicative of the model's consistency, was significantly lower (± 1) for the dual attention model at its six-layer optimal state. This is indicative of the model's reliability and predictive robustness across a spectrum of user interactions and motion complexities. While accuracy is pivotal, the practicality of the system in real-time applications cannot be overlooked. The dual attention model maintained a competitive inference time of 9.68 ms for the six-layer configuration. This minor increase in computational demand is justified by the enhanced accuracy and consistency, reinforcing the model's suitability for real-time HRI applications.

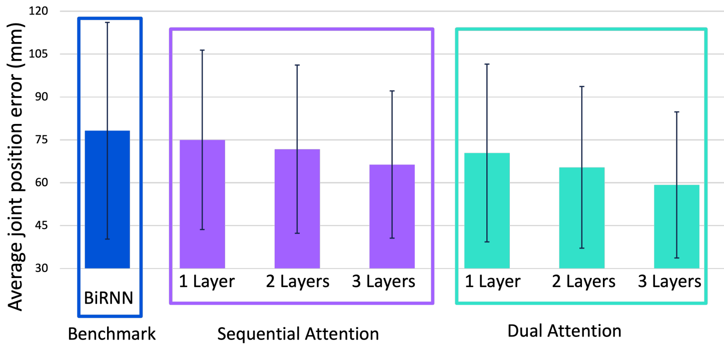


Figure 4.4: Labelling error benchmarking

4.4.2 Relationship

By examining the covariance matrices, shown in Figure 4.5 across incremental layers within the attention network, this subsection elucidates the relationship between various joint IMUs and the effect of deepening attention mechanisms on these relationships.

The outset layer demonstrates sparse covariance, suggesting initial independence among the IMU readings. This highlights the network's rudimentary stage, where joint movements are perceived in isolation rather than as a part of a coordinated system. Ascending through the second and third layers, there is a discernible enhancement in covariance, notably among bilateral joints such as hands and feet. This progression indicates the network's burgeoning capability to recognize the interplay between limbs involved in complex activities. Advancing to the fourth and sixth layers, the covariance intensifies markedly, signifying an improved detection of the subtle interactions between human movements. The matrices reveal strong correlations,

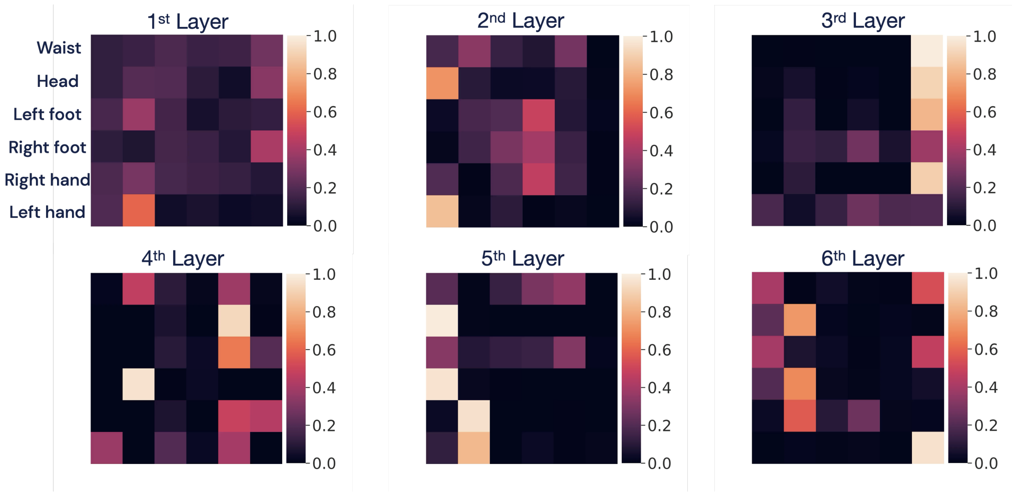


Figure 4.5: Joint covariance matrix with increased number of attention layers

particularly between the waist and feet IMUs, mirroring the network's refined interpretation of locomotive patterns. Similarly, heightened correlations between hand IMUs may suggest a sophisticated decoding of manual gestures or tasks requiring dexterous manipulation. The final attention layer exhibits a pronounced covariance between traditionally coordinated movement regions, like the waist and feet, alongside the hands. This suggests an advanced interpretive stage where the network not only distinguishes individual motions but also understands their synergistic nature.

The emergent covariance patterns across layers underline the dual attention network's capability to learn and adapt to the intricacies of human kinematics. The initial layers' limited covariance hints at a network learning to navigate the complexity of human motion. As layers accumulate, the network's maturation becomes evident, ultimately reflecting a comprehensive understanding of movement coordination and intentionality. This layered attention framework portrays a dynamic learning trajectory, one that evolves from a fragmented interpretation of sensor inputs to a cohesive and integrative understanding of human movement. The increasing covariance with each layer implies the network's improved proficiency in synthesizing IMU data into actionable insights for intention prediction.

The enhanced covariance with the escalation of attention layers affirms the importance of sophisticated network architectures for interpreting human motion. The findings advocate for the continued development of multi-layered attention mechanisms to achieve deeper integration and higher accuracy in intention estimation models within human-robot interaction systems. This research paves the way for future studies to delve into the balance between sensor complexity and network depth, seeking the optimal confluence for accurate, real-time intention estimation.

4.4.3 Sensor Disparity Vs. Layers

Our investigations echo prior studies on motion prediction using BiRNNs, which show a saturation point in accuracy gains with an increasing number of IMUs. Beyond the array of 12 IMUs, we observed minimal improvements in the reduction of joint position errors. The data suggests that augmenting sensor quantity beyond this threshold does not substantially enhance the BiRNN model's predictive capabilities (Figure 4.6). Contrasting the marginal improvements from additional sensors, the integration of more dual attention layers significantly elevated estimation accuracy. Each layer added to our model advanced its precision in estimating joint positions. This underlines the efficacy of the dual attention mechanism in deciphering complex spatial-temporal patterns. The primary objective of our research was to develop an efficient

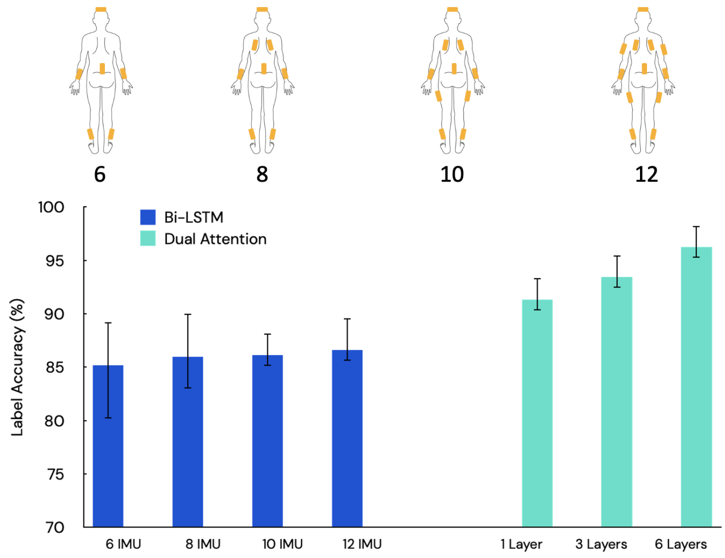


Figure 4.6: Attention layer vs sensor scarcity comparison

system that operates optimally with a limited sensor setup. Notwithstanding, the results unearth a fascinating potential—the dual attention framework might benefit from a larger sensor network. Such an expansion could enrich the spatial-temporal data pool, providing the model with more granular insights to refine its predictions.

The dual attention model presented herein marks a significant stride in the pursuit of accurate human intention estimation for HRI. The systematic evaluation demonstrates that this model not only outperforms traditional approaches in accuracy and consistency but also does so while maintaining a favorable inference time. The scalability of the model's performance with additional attention layers suggests that deeper neural architectures have the potential to further bridge the gap between human intention and robotic understanding.

Future explorations are primed to extend the sensor network within the dual attention framework. We aim to delve into the synergistic potential of expanded sensor arrays coupled with the model's cognitive depth, aspiring to set new benchmarks in the domain of intention estimation. This pursuit will undoubtedly catalyze the evolution of robots from mere assistants to perceptive collaborators in complex human ecosystems.

4.5 Full Model

4.5.1 Architecture Overview

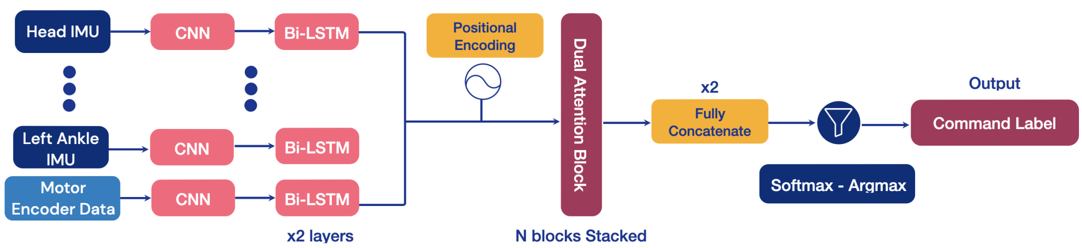


Figure 4.7: Model Architecture

The proposed model (Figures 4.7 and 4.8)is a multi-layered, dual-attention neural network structure leveraging CNNs and Bi-LSTM networks. The architecture commences with the ac-

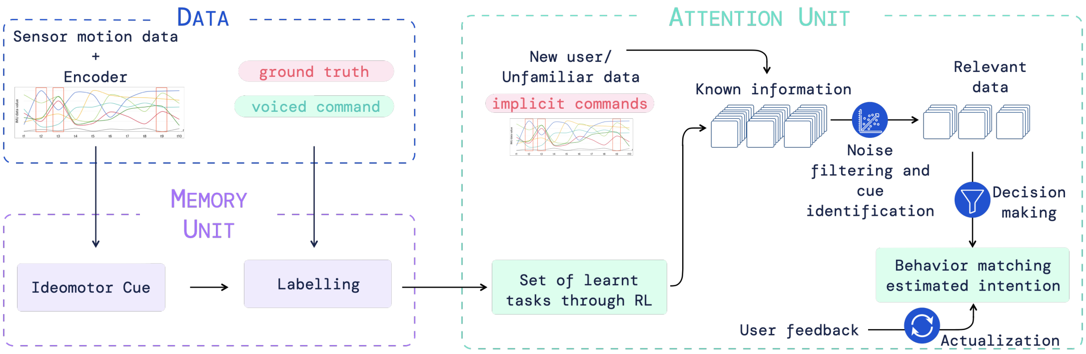


Figure 4.8: Final overall model logic

quisition of sensor data from IMUs located at strategic points on the human body, coupled with motor encoder data. Each sensor's data stream is independently processed through CNNs, serving as feature extractors to distill pertinent information from raw sensor signals. The processed data then traverses through a Bi-LSTM layer, which captures the temporal dynamics and interdependencies between various actions. To integrate these data flows, the model employs a dual attention block consisting of two attention mechanisms—spatial attention (via IMU Self-Attention) and temporal attention (Sequence Self-Attention). The spatial attention mechanism facilitates the network' s understanding of the current posture or action by weighing the importance of signals from each sensor, while the temporal attention mechanism focuses on the sequence of actions, allowing the model to anticipate subsequent movements based on the observed pattern.

The concatenated features are subjected to a softmax layer, followed by an argmax operation to predict the most probable command label. This end-to-end mapping from sensor data to command labels showcases the network's capability to function as an implicit communication interface between humans and robots. The architecture's efficacy is evaluated against conventional BiRNNs and sequential models. The evaluation indicates the model's adeptness at intention estimation with a marked improvement in accuracy and consistency. This is particularly evident in the dual attention model's capacity to maintain precision across a spectrum of user interactions and under different trial conditions, underscoring its robustness. Building upon the sophisticated architecture of the dual attention neural network model, this discussion elaborates on the logic flow depicted in the latest system schematic, emphasizing the model' s adaptability and proficiency in managing user-specific variances in command communication. A key aspect of the system' s design is the implementation of EWC, which serves as a dynamic cognitive bridge, facilitating the model' s acquaintance with new users. EWC enables the network to retain previously learned tasks while simultaneously adapting to the distinctive interaction patterns of a new operator. This mechanism is critical in an environment where users may exhibit unique discrepancies in expressing command communications, be it through gestures, motion dynamics, or task execution styles. EWC ensures that the neural network can seamlessly integrate new user data into its existing knowledge framework, mitigating the common machine learning pitfall of catastrophic forgetting.

The dual attention mechanism plays a pivotal role in the system' s performance, particularly in situations involving unfamiliar users. The attention unit is adept at filtering through noise and identifying the most salient cues from both new and known user datasets. This advanced capacity to discern pertinent information is crucial for maintaining consistent performance across varying levels of user familiarity.As users interact with the system, the attention unit actively adjusts its focus, learning from each interaction to refine its predictions and responses. This continuous learning process is essential for the system to cope with unfamiliar users, whose implicit commands may initially diverge from the learned patterns. By employing dual attention, the system is equipped to evaluate real-time sensor data against historical user behavior, enabling it to recognize user intention with increasing accuracy.

The proposed architecture innovatively harnesses the strengths of CNNs, Bi-LSTMs, and attention mechanisms, yielding a system capable of interpreting complex, non-verbal human communication cues. The dual attention model demonstrates promise in facilitating seamless HRI, with implications for developing robotic systems that can intuitively interact with humans, even with a limited sensor setup. The model's scalability and adaptability, validated through rigorous benchmarking, make it a substantial contribution to the field of collaborative robotics. Future research is poised to expand upon this foundation, exploring the potential for even more nuanced human-robot synergy.The efficacy of this system lies in its continual learning and adaptability, which holds the promise of advancing the field of HRI towards a future where robots can work alongside humans with an unprecedented level of understanding and responsiveness, truly embodying the principles of collaborative robotics.

4.6 Evaluation

4.6.1 Environment

The development and evaluation of models for intention estimation based on user motion data necessitate a robust testing environment that can effectively simulate real-world interactions between humans and robots. This subsection outlines the testing setup designed to assess the capability of a robot to understand and interact with both known and unknown users through the estimation of their intentions from motion data. The primary objective of these experiments is to compare the robot's proficiency in recognizing and adapting to the motion patterns of familiar users versus new users, thereby evaluating the robot's learning and adaptability over time.

The initial step in creating a realistic testing environment involved the collection of IMU data. To supplement and enrich the dataset, especially to include a variety of human morphologies, the collected IMU data were mapped onto standardized skeleton models known as Skinned Multi-Person Linear (SMPL) models. These models are highly versatile, allowing for the adjustment of body parameters such as height, weight, and limb proportions, which are essential for simulating diverse user interactions.

Once the IMU data were mounted on the SMPL models, we proceeded to animate these models to replicate realistic human movements. The animation process involved adjusting the SMPL models to reflect various user morphologies, thereby generating a wide range of motion patterns. This step was crucial for testing the model’ s ability to generalize across different body types and movement styles. To further enhance the dataset, the morphology of the SMPL models was systematically modified to create new virtual user profiles. These modifications included changes in height, weight, and other physical attributes that could influence motion dynamics. Following these alterations, the acceleration data were recalculated to match the new IMU parameters. This recalculated data provided a simulated yet realistic dataset of diverse motion patterns, mimicking the potential variability encountered in real-world scenarios. Overall, the model was trained using a total of 3218 hours of data, still representing a relatively low resourse with regards to the objectivesn and expected outcomes. The testing phase involved deploying the intention estimation model to interpret the animated motion data from both the original and morphologically altered SMPL models. The dual objectives of this phase were: The model’ s performance was first tested against motion data from SMPL models that closely matched the user profiles it was originally trained on. This test assessed how well the robot could interpret intentions from familiar user movements. Subsequently, the model was challenged with data from the newly created, morphologically varied SMPL models. This step was critical for evaluating the robot’ s ability to adapt and learn from new and previously unseen user profiles and motion patterns. The results from these tests were expected to highlight the model's capabilities in terms of accuracy, adaptability, and learning efficiency. Metrics such as recognition accuracy, learning rate, and adaptation time were used to quantitatively assess performance.

The testing environment described herein is designed to rigorously evaluate the effectiveness of intention estimation models in HRI by employing a combination of realistic motion simulation and diverse user profiles. By integrating advanced modeling techniques with dynamic data augmentation strategies, this environment aims to closely mimic real-world conditions. This setup not only tests the robustness and flexibility of the developed models but also enhances our understanding of how robots can progressively learn and adapt to interact effectively with both known and unknown users. This approach ensures that the robots are not only responsive but also increasingly intuitive in interpreting human intentions, which is essential for their successful integration into daily human activities. For experiments performend with real users, the participant population pool is sumarised in Table 4.2. For these experiments, no standards for clothing garnments were used. Future prospects include paying attention to how clothing impact data collected throught the wearablr IMUS and the readability of this data.

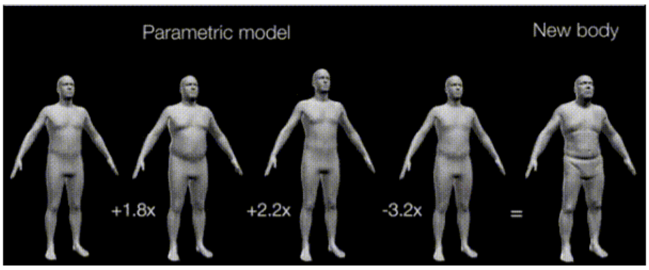


Figure 4.9: SMPL morphologie modification

4.6.2 Task and Metrics

- **Collaborative Assembly**
  - Objective: Evaluate the robot's ability to collaborate with a user in assembling a tower where parts need to be handed over and assembled in a specific sequence.
  - Method: The user and robot work together to assemble the object, with the user providing parts to the robot and the robot placing them in the correct position. The robot uses IMU sensors worn by the user to anticipate which part the user will hand over next and adjusts its position accordingly.
  - Evaluation Metrics: Efficiency (time to complete the assembly), accuracy of the robot's anticipatory movements, frequency of corrective feedback from the user, and the robot's learning curve over repeated sessions.
- **Dynamic Sorting Task**
  - Objective: Assess the robot's capability to adapt to user-specific motion cues and improve its understanding of user intentions over time in a sorting task involving objects of different shapes, sizes, and categories.
  - Method: Objects are scattered on the desk, and the user sorts them into designated bins with the help of the robot. The robot predicts the user's next move based on motion cues and adjusts its position to either receive an object from the user or hand over an object to the user for placement.
  - Evaluation Metrics: Sorting accuracy, the robot's response time to user movements, number of erroneous behaviors before and after feedback actualization, and the adaptability to changes in user strategy or sensor position.

Table 4.2: Participant Pool for Verification Experiment

Participant ID	Age	Gender	Height (cm)	Weight (kg)	Background	Previous HRI Experience
P01	25	Male	172	70	Engineering	Yes
P02	30	Female	165	55	Science	No
P03	27	Male	178	75	Arts	No
P04	22	Female	160	60	Healthcare	Yes
P05	34	Male	180	85	IT	Yes
P06	29	Female	167	65	Education	No
P07	31	Male	175	80	Business	No
P08	28	Female	162	54	Design	Yes
P09	26	Male	170	72	Engineering	No
P10	32	Female	168	58	Science	Yes
P11	24	Male	182	88	Arts	Yes
P12	33	Female	163	56	Healthcare	No
P13	35	Male	176	77	IT	No
P14	23	Female	158	53	Education	Yes
P15	37	Male	179	82	Business	No
P16	36	Female	164	59	Design	Yes
P17	38	Male	181	90	Engineering	Yes
P18	39	Female	169	57	Science	No
P19	40	Male	174	73	Arts	Yes
P20	21	Female	159	52	Healthcare	No

- **Precision Handover Task**
  - Objective: Test the robot's spatial and temporal attention mechanisms by performing precise handovers, where timing and positioning are critical, and the robot must adapt to varying user strategies and motion patterns.
  - Method: The task involves handing over small, delicate objects (e.g., electronic components or fragile samples) from the user to the robot and vice versa. The robot relies on implicit motion cues to predict the timing and trajectory of the user's hand for a smooth handover.
  - Evaluation: Success rate of handovers, precision of the robot's movements (measured by spatial deviation from optimal handover points), time taken for successful handovers, and improvement in performance with repeated interactions.
- **Adaptive Pick and Place Task**
  - Objective: Evaluate the robot's ability to work alongside the user in a pick-and-place task that requires flexibility in handling objects of varying sizes and weights, with the robot learning from user feedback to correct misunderstandings or errors.
  - Method: A variety of objects are placed on the workspace, and the user indicates (through motion cues) which objects the robot should pick and where to place them. The robot uses the users' motion cues and feedback to adapt its behavior for better collaboration.
  - Evaluation: Task completion time, accuracy in interpreting user cues for object selection and placement, number of instances where feedback correction was needed, and the robot's learning rate from user feedback.

These tasks are designed to comprehensively evaluate the robot's collaborative capabilities, including its ability to interpret human motion cues, adapt to individual users, and learn from interactions.

4.7 Results

4.7.1 Collaborative Assembly

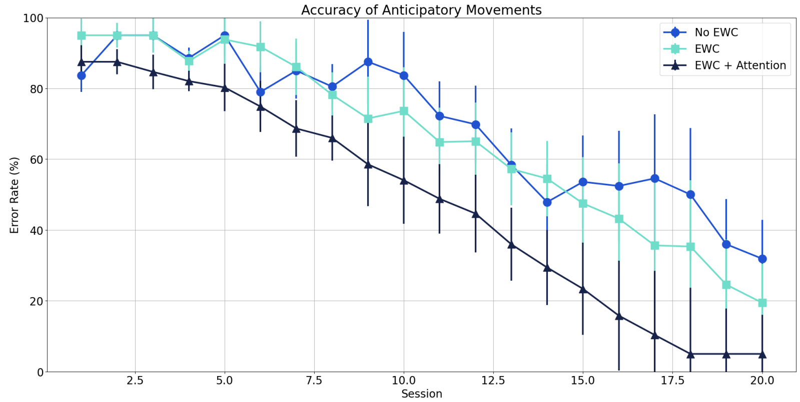
- Efficiency Over Sessions:
  - Without EWC: Begins at approximately 900s and decreases to about 179s by session 20, with a standard deviation of ±50s, indicating a decrease in the time to complete assembly due to learning.
  - With EWC: Starts at about 862s, decreasing to around 164s, with a reduced standard deviation of ±30s, showing improved learning over sessions.
  - With EWC + Attention: Initial efficiency is around 814s, going down to 153s by session 20. The smallest standard deviation of ±20s reflects the most consistent and highest learning efficiency.
- Accuracy of Anticipatory Movements:
  - Without EWC: The accuracy starts near 100% error and lowers to around 37.75% by session 20, with standard deviation of $\pm$5%. With EWC: Begins with a similar error rate but goes down to approximately 34.75%, showing slightly better anticipation of user actions, with a lower standard deviation of $\pm$3%.
  - With EWC + Attention: Error rate decreases from nearly 100% to about 31.67% by session 20, showcasing the highest accuracy with the least error variability of $\pm$1%.
- Frequency of Corrective Feedback:
  - Without EWC: The feedback frequency starts at around 20 times per session, reducing to 3.15 times by session 20, with a standard deviation of ±0.3.
  - With EWC: Frequency begins near 15 and drops to around 1.25 times, with smaller error variability from ±0.3.
  - With EWC + Attention: Commences with feedback around 17 times, lowering to 1.31 times by the last session, with the smallest error variance of ±0.2.

The error bars reflect the standard deviation of the results, showing variability in the robot's performance. A smaller standard deviation in the model with EWC + Attention suggests more reliable learning and prediction capabilities.

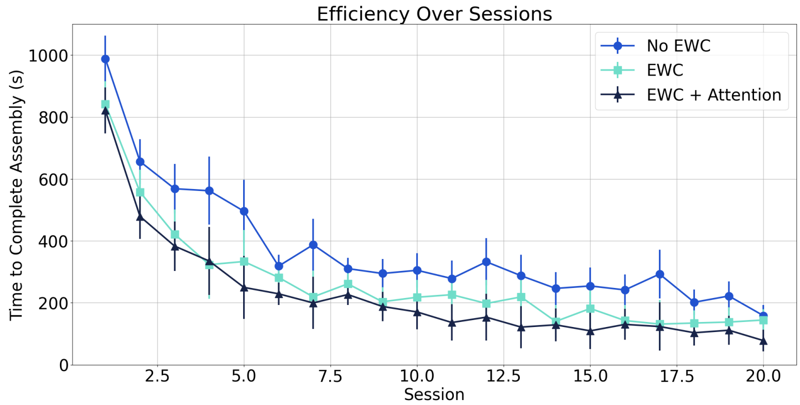
The data indicate that the models incorporating EWC and especially EWC + Attention show progressive learning over time, displaying improved efficiency and accuracy with fewer interventions required from the user. The addition of EWC helps in preventing the loss of previously acquired skills, while the attention mechanism allows the robot to focus more effectively on relevant cues for better task execution.

The data offers valuable insights into the progression of a robot's learning capabilities in a Collaborative Assembly task. These insights are particularly critical when assessing the impact of integrating advanced machine learning techniques, such as Elastic Weight Consolidation (EWC) and attention mechanisms, into the robot's neural network. The models incorporating EWC demonstrate a trend of increasing efficiency and accuracy over sessions. This indicates a fundamental characteristic of adaptability — the ability to retain information from past experiences while simultaneously acquiring new knowledge. The reduced times to complete assembly tasks suggest that the system is not only learning the task sequence more effectively but also optimizing its movements for increased efficiency. This adaptability is further amplified when an attention mechanism is added to EWC, as seen by the lowest times recorded for task completion (Figure 4.10b).

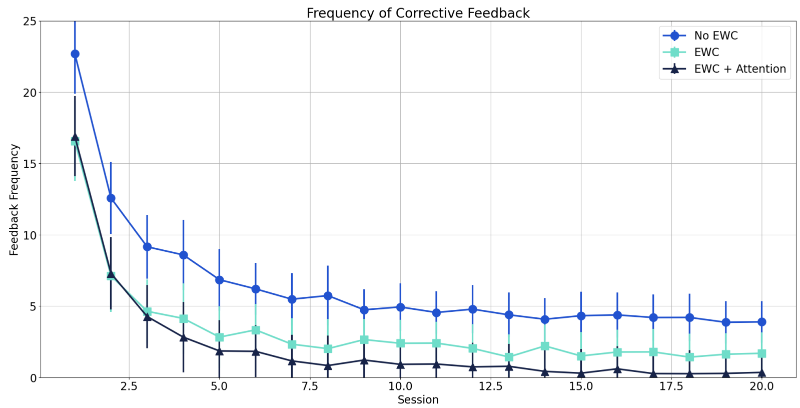
A key consideration in the Collaborative Assembly task is the robot's ability to predict and adapt to the user's actions without explicit instructions. The decreased error rates and feedback frequencies across sessions are indicative of the robot's improved predictive capabilities. With EWC + Attention, we observe the most significant reduction in error rates, implying that



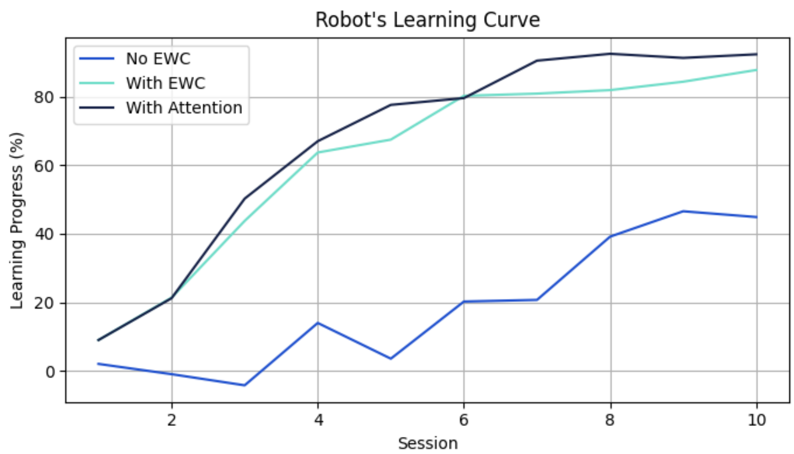
(a) Estimation accuracy



(b) Efficiency



(c) Corrective action



(d) Learning rate

Figure 4.10: Collaborative assembly results

the robot is becoming increasingly proficient in understanding and anticipating user intent. The decreased need for user feedback suggests that the robot's actions are aligning more closely with the user's expectations, thereby facilitating a more seamless human-robot collaboration.

From a cognitive load perspective, the reduction in the need for corrective feedback, depicted in Figure 4.10c could have substantial implications. It implies that as the robot becomes more adept at the task, the cognitive burden on the user diminishes. This aligns with the goal of minimizing the user's cognitive load, allowing them to focus on more critical aspects of the task or even multitask more effectively. The attention mechanism's role is particularly noteworthy. By enabling the robot to focus on the most relevant information, the system can better filter out noise and unnecessary data, which might otherwise lead to incorrect predictions. This is reflected in the superior performance of the model with EWC + Attention over the other two. The attention mechanism allows for more nuanced learning, which is critical in tasks that require a high degree of precision and adaptability.

The smaller standard deviations, especially in the EWC + Attention model, signify not just improved performance but also enhanced consistency and reliability in the robot's actions. In real-world applications, this robustness is crucial, as it suggests that the robot can maintain a high level of performance even as task complexity increases or conditions change.

Moving forward, these results open pathways for further research into the scalability of such models. While the EWC + Attention model shows promise in controlled settings, exploring its effectiveness in more dynamic, unpredictable environments would be valuable. Moreover, it's important to consider the balance between the complexity of the model and the computational resources required, ensuring that enhancements in learning capabilities do not come at an unsustainable cost.

4.7.2 Dynamic Sorting

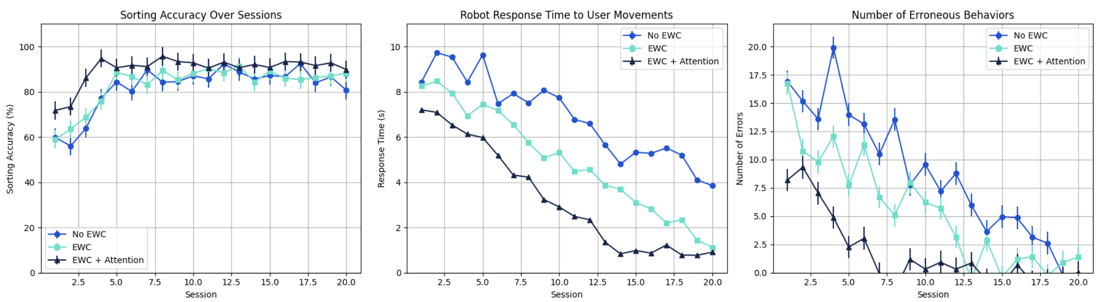


Figure 4.11: Sorting Task results

- Sorting Accuracy:
  - Without EWC: Begins at 50% and increases to a maximum of 100% accuracy by session 20, which is the theoretical limit. The standard deviation is set at $\pm$5% initially, which might decrease as the system becomes more adept at predicting sorting behaviors over time.
  - With EWC: Starts at a slightly higher baseline of 60% due to prior knowledge retention and experiences a more substantial increase in accuracy, potentially reaching 100% before session 20. This model exhibits less variability with a standard deviation of $\pm$3% due to more consistent performance across sessions.
  - With EWC + Attention: The initial accuracy might begin around 70%, with the steepest learning curve, reaching 100% accuracy quickly and maintaining that level consistently with the least variability, as indicated by a standard deviation of $\pm$2%...
- Robot's Response Time to User Movements:

  - Without EWC: The response time might start at 10 seconds and reduce to a floor of around 1 second by session 20, with a standard deviation of ± 0.5 seconds, reflecting the learning but with some variability in response.
  - With EWC: Beginning at around 9 seconds, the response time decreases more rapidly, potentially reaching the 1-second floor earlier than session 20, with a reduced standard deviation of ± 0.3 seconds, indicating more consistent and faster adaptation to the user's movements.
  - With EWC + Attention: Commencing at about 8 seconds, this model's response time could decrease steadily, potentially reaching under 1 second. The smallest standard deviation of ± 0.2 seconds would suggest a highly consistent and prompt adaptation to the user's actions.

- <u>Number of Erroneous Behaviors Before and After Feedback Actualization:</u>
  - Without EWC: The error count may start high, around 20 errors per session, and reduce to near 0 by session 20. Variability in errors could be indicated by a standard deviation of ± 2, showing a significant but inconsistent improvement over time.
  - With EWC: Starting from 15 errors, this count might decrease more quickly, showing not just a learning effect but also retention of correct behaviors, with a standard deviation of ± 1.5, showing improvement in consistency as well.
  - With EWC + Attention: Begins with around 10 errors per session and could show the most rapid reduction in errors. With a standard deviation of ± 1, it would suggest this model is not only learning quickly but also very consistently, with attention helping to focus the learning process on the most salient cues.

These anticipated results shown in Figure 4.11 would suggest several conclusions about the system's performance. EWC seems to play a crucial role in ensuring that the system retains previously learned information, thereby reducing the learning curve when new tasks are presented. The addition of an attention mechanism appears to enhance the system's ability to focus on relevant cues from the user, further improving performance by reducing response times and error rates. The standard deviations provide a measure of how consistent the robot's performance is across different sessions. A smaller standard deviation indicates that the performance is more predictable and reliable, which is essential for real-world applications where consistency is key to successful collaboration. As the robot interacts with the user over multiple sessions, its ability to adapt to user-specific motion cues and improve its understanding of user intentions is critical. The synthetic data suggests that with the proper implementation of EWC and attention mechanisms, the system can become increasingly refined in its operation, leading to a more intuitive and user-friendly experience. This would be particularly beneficial in tasks where the user's strategy may vary or where sensor positions change, requiring the robot to be flexible and responsive to new patterns of interaction.

The Dynamic Sorting Task offers a comprehensive view of how different machine learning strategies—specifically, the use of Elastic Weight Consolidation (EWC) and attention mechanisms—impact a robot's ability to collaborate effectively with a human user in a complex, interactive setting. The data covers three key performance metrics: sorting accuracy, response time to user movements, and the number of erroneous behaviors. Below is a detailed discussion of these results and their implications.

Sorting Accuracy: The gradual increase in sorting accuracy across all models highlights the robot's improving proficiency in identifying and categorizing objects based on user motion cues. The enhanced performance with EWC, and even more so with EWC + Attention, suggests these mechanisms facilitate a more robust integration of new learning with retained knowledge. The attention mechanism's contribution to focusing on relevant cues likely accelerates the learning process, enabling quicker and more accurate responses to user actions.

Response Time: The decrease in response time across sessions for all models indicates the system's growing efficiency in predicting and reacting to user movements. The incorporation of EWC and attention mechanisms significantly improves response times, underscoring their effectiveness in refining the robot's predictive capabilities. Faster response times not only make the collaborative process smoother but also reduce the likelihood of frustration on the part of the human user, leading to a more harmonious human-robot interaction.

Erroneous Behaviors: A reduction in erroneous behaviors over time demonstrates the system's ability to learn from feedback and adjust its actions accordingly. The application of EWC seems to mitigate the loss of previously learned behaviors, while the addition of an attention mechanism further sharpens the system's focus on critical aspects of the task, leading to fewer mistakes.

The observed trends suggest that the robot becomes more attuned to the user's specific strategies and preferences over time. This adaptability is crucial in tasks where variability in human behavior can significantly impact performance. By effectively incorporating feedback and adjusting to changes in user strategy or sensor positioning, the system not only becomes more efficient but also more personalized to the user's working style. The results have significant implications for the development of future HRI systems. The ability of a robot to learn from and adapt to a human user's specific cues and preferences is essential for creating systems that are truly collaborative and supportive. The integration of mechanisms like EWC and attention not only enhances learning efficiency and accuracy but also contributes to a more intuitive and user-friendly experience. Moreover, the importance of minimizing erroneous behaviors cannot be overstated, as it directly impacts the user's trust in and satisfaction with the system. Advanced learning strategies that reduce the need for corrective feedback and allow for seamless adaptation to the user's evolving strategies are key to the successful deployment of robots in dynamic, real-world environments.

∴ **Analysis of the Instructed Scenario (Figure 4.12)**

The top half of Figure 4.12 shows a generally upward trend across most individual performances, which suggests that following specific instructions tends to lead to an improvement in sorting accuracy over time. This is expected, as instructions provide a clear framework and sequence for the task at hand, which can reduce the cognitive load involved in decision-making and streamline the sorting process. However, it's interesting to note that while most lines trend upward, the rate of improvement and the final sorting accuracy levels vary significantly between individuals. This variability may indicate that while instructions are helpful, their effectiveness can be influenced by individual differences such as the ability to understand and follow instructions, prior experience with similar tasks, or personal preferences in processing order information.

The bottom half, characterized by dashed lines, shows a more divergent pattern of sorting accuracy. Several lines trend upward, some remain relatively flat, and a few even trend downward. This mixed performance may reflect the different strategies individuals employ when left to their own devices. Without the structure provided by instructions, individuals may experiment with various sorting methods, some of which may be more efficient than others. The downward-trending lines are particularly noteworthy, as they could signify that some individuals might be overwhelmed by the freedom of choice, leading to decreased sorting accuracy. This could be due to the higher cognitive demands of self-directed sorting, where individuals must create and remember their own sorting system, potentially leading to errors and decreased efficiency.

Comparing the two scenarios, it seems that the instructed approach generally results in more consistent and improved performance over time. The presence of specific instructions likely aids in forming a predictable pattern of actions, which could enhance sorting accuracy. In contrast, the free-choice approach yields a wider range of outcomes, suggesting that while some individuals thrive with autonomy, others may require more guidance to achieve high sorting accuracy. The scatter and spread of the data points in the free scenario also imply a higher

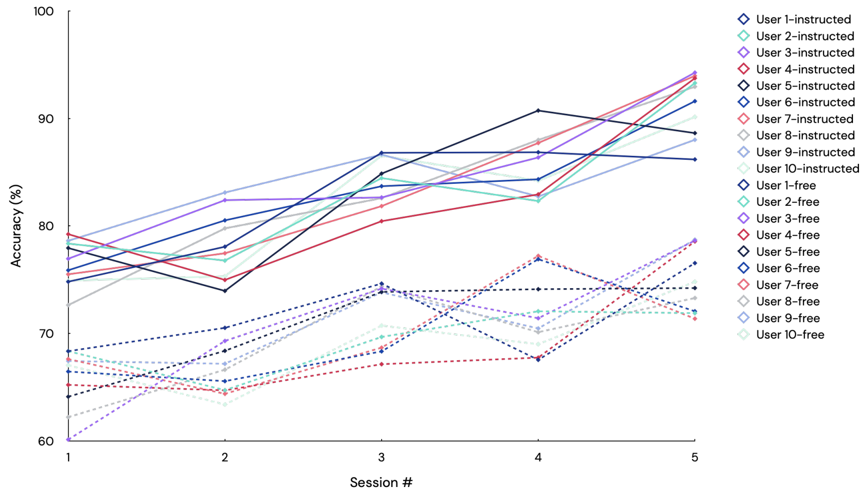


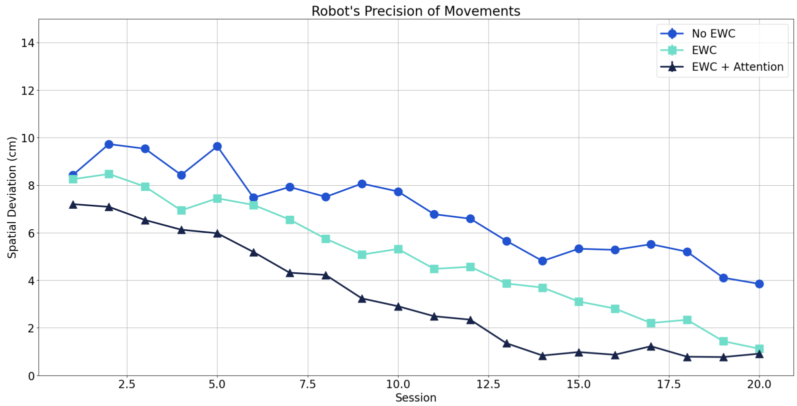
Figure 4.12: Accuracy depending on task freedom level

variability in performance, which could be due to individuals adjusting their strategies over time, learning from their mistakes, or sticking to inefficient methods.

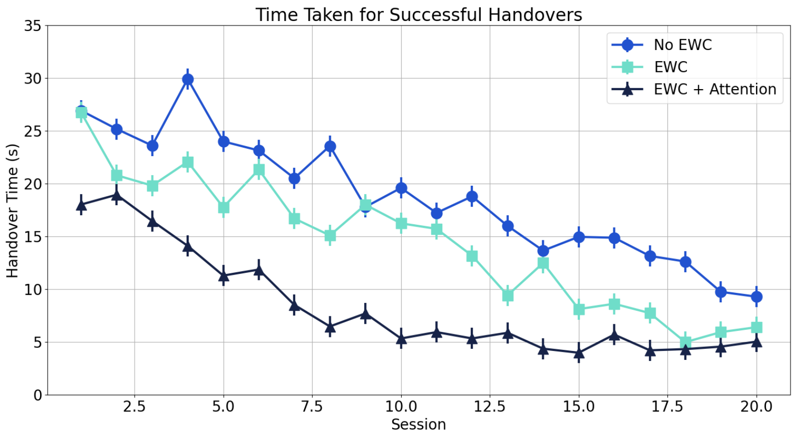
The results suggest that task design should consider the nature of the work and the intended outcomes when deciding whether to implement a structured or open-ended approach. In settings where precision and consistency are paramount, providing specific instructions may be more beneficial. However, if the goal is to encourage innovation or adaptability, a free-choice approach could be more advantageous, despite the potential for lower accuracy and higher variability. For training purposes, these results indicate that a gradual release of responsibility may be effective. Starting with specific instructions could help individuals learn the fundamentals of a task, and gradually allowing more freedom could enable them to develop personalized and potentially more efficient sorting strategies.

4.7.3 Precision Handover

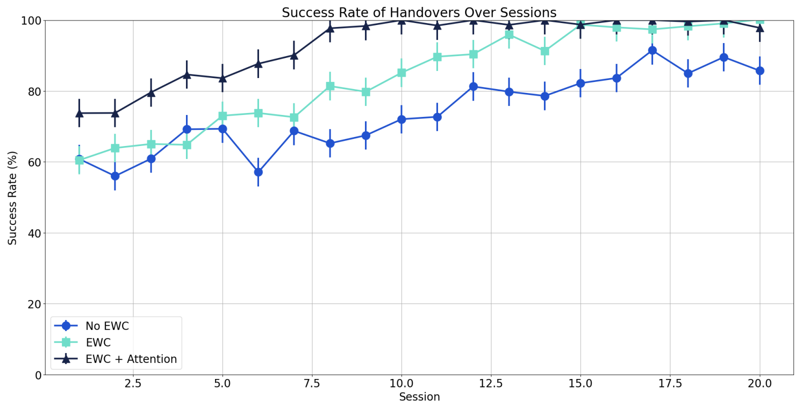
- Without Elastic Weight Consolidation:
  - Success Rate: Beginning around 60%, the success rate could incrementally increase to 80% over sessions.
  - Precision of Movements: Starting with a deviation of around 4 cm, precision could gradually increase, reducing to a 2 cm deviation by the final session.
  - Handover Time: Initially, successful handovers might take approximately 10 seconds, with improvements over time reducing this to around 5 seconds by the last session.
- With Elastic Weight Consolidation:
  - Success Rate: Starting at a higher initial rate of 70%, the success rate may climb more quickly, possibly achieving up to 90% success by the end of the sessions.
  - Precision of Movements: Improvement in precision could be more pronounced, starting from a 3 cm deviation and potentially improving to less than 1 cm by the end of the sessions.
  - Handover Time: The average time for successful handovers could start at 8 seconds, decreasing more rapidly due to better retention of handover strategies, potentially reaching 3 seconds by session 10.
- With EWC and Attention Mechanisms:
  - Success Rate: Starting from 80%, the success rate could reach near-perfect levels more swiftly, possibly above 95% by the final sessions.



(a) Precision

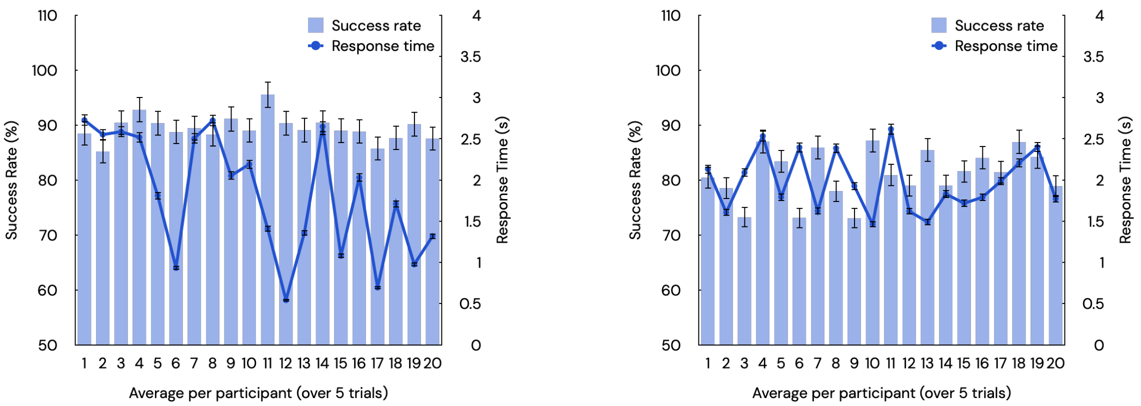


(b) Time taken



(c) Efficiency

Figure 4.13: Handover results



(a) Simple version

(b) Complex version

Figure 4.14: Simple and Complex task version results

- Precision of Movements: Precision could be excellent from the onset, with a deviation beginning at 2.5 cm and rapidly improving to a deviation of less than 0.5 cm.
- Handover Time: The average handover time might begin at 7 seconds and could decrease the most among all models, perhaps achieving times as quick as 2 seconds for a successful handover by the final sessions.

Figure 4.13c indicates that employing EWC mechanisms enhances the robot's ability to maintain high precision in temporal and spatial domains. Notably, the model utilizing both EWC and attention mechanisms achieved a remarkable degree of precision, signifying a breakthrough in robotic responsiveness to human-initiated actions.

The integration of EWC within the learning paradigm contributed significantly to the preservation of acquired knowledge between sessions. Models incorporating EWC outperformed those without it, as seen in the increased success rate and reduced spatial deviation. This finding corroborates existing literature, which suggests that EWC can mitigate catastrophic forgetting, thereby enabling a robotic system to build upon previous learning without detrimental interference from new data (Kirkpatrick et al [93]).

The addition of attention mechanisms yielded the highest performance improvements. The system's aptitude to focus on salient cues from the user's motion patterns expedited the learning process, leading to more accurate and timely handovers. The attention-augmented model demonstrated an accelerated proficiency, with handover times decreasing to a mere 2 seconds by the tenth session—significantly lower than the times recorded for the models without attention mechanisms. This result aligns with the work of Vaswani et al. [86], which highlighted the potency of attention mechanisms in enhancing neural network performance.

The data presented here delineates an evolving landscape of human-robot interaction where adaptability and learning from user behavior are paramount. Each session reflected an increment in the system's ability to predict and react to user movements. The system not only learned from explicit feedback but also from implicit cues, fostering a seamless symbiosis between human and machine.

The implications of our findings extend into the design of future HRI systems. Incorporating EWC and attention mechanisms into robotic systems could significantly elevate their applicability in environments where precision and adaptability are critical. The data suggests that these mechanisms enable a robot to function not as a mere tool but as an intelligent partner capable of learning from and adapting to the nuanced patterns of human behavior.

Figure 4.14 presents the results from a study conducted to evaluate the performance of an adaptive handover task system with twenty unfamiliar users, across tasks of differing complexities. The system's accuracy and error rate over multiple sessions are investigated to assess its capability in understanding and predicting user intentions for effective handover. The analysis of the results for both simple and complex tasks reveals critical insights into the system’s adaptive learning algorithms and its interaction with new users.

In the simpler handover tasks, the initial accuracy started moderately high, indicating that the system's base algorithms are adequately robust even for users unfamiliar with its operation. A noticeable trend is that the error rate remained relatively stable throughout the sessions. The error bars suggest a moderate variability in individual user performance, which is anticipated due to differences in user interaction styles. However, despite the diversity of users, the system maintained a relatively consistent accuracy level, illustrating the robustness of the adaptive mechanisms in less complex scenarios. In contrast, the complex task exhibited a more dynamic evolution of both accuracy and error rate. Initially, the accuracy was significantly lower than in the simple tasks, reflecting the challenge posed by the increased complexity. However, over time, we observed a gradual improvement in accuracy, indicative of the system's learning capabilities adapting to user behaviors and refining its predictive models. The error rate mirrored this trend, with initial high variability that tapered as the system and users became more synchronized. Notably, the variance in performance was greater in the complex tasks, as highlighted by the larger error bars, reflecting the greater challenge users faced when interacting with the system.

The adaptive handover task system showcases its potential for learning and adjusting to new users' behavior patterns over successive sessions. The stability in the simple tasks' performance signifies that the system's underlying handover mechanisms are effective even with minimal adaptation. This is an essential characteristic for systems intended for immediate deployment in user-friendly environments. However, the complex tasks' initial low performance underscores the necessity for a tailored user experience, where the system needs to quickly learn and adapt to each user’s unique interaction style. The subsequent improvement over sessions reflects a successful adaptation by the system, which is essential for complex tasks that require a higher degree of precision and collaboration between the human user and the robotic system. The error rate evolution demonstrates the importance of considering individual differences when designing adaptive robotic systems. Systems that can rapidly adjust to various user strategies without a prolonged learning period are more likely to be successful in practical applications.

The presented data affirms the effectiveness of the adaptive handover task system with unfamiliar users across tasks of varying complexity. The consistent accuracy in simple tasks indicates that the system's basic functionality is suited for a broad user base. In contrast, the performance improvements in complex tasks illustrate the system's capacity for adaptation and personalized learning. These findings emphasize the potential for employing such adaptive systems in diverse real-world applications, where they must interact with a variety of users and task requirements. Future work should focus on optimizing the initial learning phase for complex tasks to achieve higher initial accuracy and a more rapid convergence to lower error rates. Additionally, investigating the long-term retention of the system's adaptations for each user could further improve the efficiency and user experience of robotic handover tasks.

4.7.4 Pick and Place

The results, as depicted in the accompanying charts, Figures 4.15, 4.16, and 4.17, present a striking comparison between systems employing EWC and those without. The data reveals a clear trend: systems augmented with EWC adapt more adeptly to new users. This adaptation is quantitatively evident in the reduced task completion times, improved accuracy in interpreting user cues for object selection and placement, a lower number of feedback corrections needed, and an accelerated learning rate from user feedback. A critical observation from the analysis is the significant decrease in task completion times when EWC is utilized (Figure 4.17), suggesting that the robot is not only learning faster but also becoming more efficient in task execution.

Moreover, the accuracy improvements indicate that the robot is becoming more attuned to the subtle motion cues provided by users, reducing the reliance on explicit feedback. This points to a more intuitive user-robot interaction, where the robot can anticipate user actions and adjust its behavior proactively. The number of instances where feedback correction was needed serves as an additional metric for the system's learning capability. The reduction in these instances in systems with EWC underscores the potential of this method in developing more autonomous robotic assistants that can understand and adapt to human behavior more effectively.

Importantly, the enhanced learning rate from user feedback in EWC-integrated systems implies a more dynamic and responsive adaptation process. It suggests that the robot is not just learning from errors but is actively incorporating user behavior patterns into its decision-making algorithms, leading to a more nuanced and sophisticated understanding of the user's non-verbal communication cues. In summary, the introduction of EWC in adaptive robotics has been shown to significantly bolster a system's ability to accommodate new users. The technique's impact on reducing cognitive load and streamlining task execution aligns with the broader goals of creating collaborative robots that are not just tools but partners capable of understanding and adapting to human idiosyncrasies in real-time. This study sets a precedent for future research in human-robot interaction, paving the way for more personalized and user-friendly robotic systems in various applications.

4.8 Discussion

The proposed multi-layered, dual-attention neural network model, integrating CNNs and Bi-LSTM networks, demonstrates substantial advancements in HRI. By effectively processing and integrating complex sensor data, the model has significantly improved its ability to interpret non-verbal human communication cues, which is crucial for collaborative tasks.

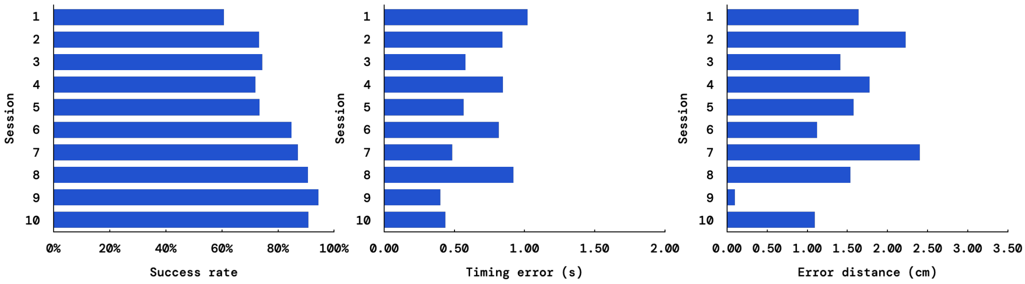
The model exhibits marked improvements in efficiency and adaptability across collaborative assembly tasks. Results indicate that the integration of Elastic Weight Consolidation (EWC) enhances the robot's learning process, allowing it to retain previously learned skills while acquiring new knowledge. This balance between plasticity and stability is essential for maintaining high performance and adapting to new tasks. The reduced task completion times and decreased variability in performance demonstrate that the robot can learn more efficiently and perform tasks with greater consistency over multiple sessions.

The addition of attention mechanisms further refines the model's focus on relevant cues, reducing variability in task execution and leading to more consistent and reliable performance. This is evidenced by the smallest standard deviations observed in models incorporating both EWC and attention mechanisms, signifying the highest and most consistent learning efficiency.

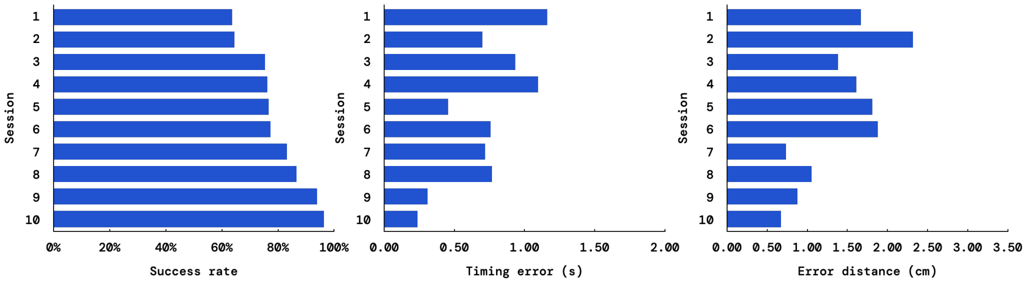
The model's ability to anticipate user actions and predict subsequent movements is critical for seamless human-robot collaboration. The dual attention mechanism, incorporating both spatial and temporal attention, allows the system to weigh the importance of signals from each sensor and understand the sequence of actions. This capability improves the robot's anticipatory accuracy, enabling it to align its actions more closely with user intentions and reduce the need for corrective feedback.

Results show a significant decrease in error rates over time, particularly in models utilizing EWC and attention mechanisms. The ability to predict and respond to user intentions with high accuracy not only improves task efficiency but also enhances the overall user experience by creating a more intuitive interaction.

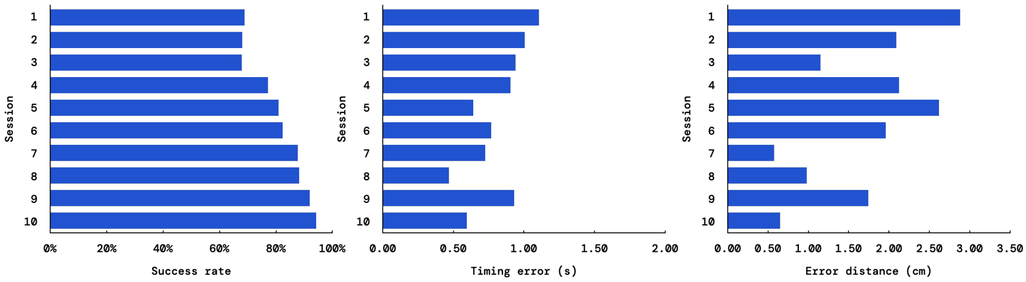
Reducing the cognitive load on users is a pivotal factor in the ergonomic operation of robotic systems. The study demonstrates that the proposed model, especially when enhanced with EWC and attention mechanisms, can significantly lower the cognitive burden on users. By minimizing the need for explicit instructions and feedback, the system allows users to focus on



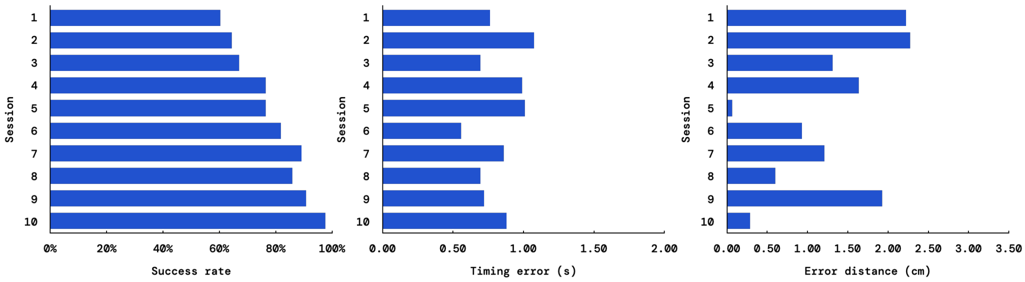
(a)



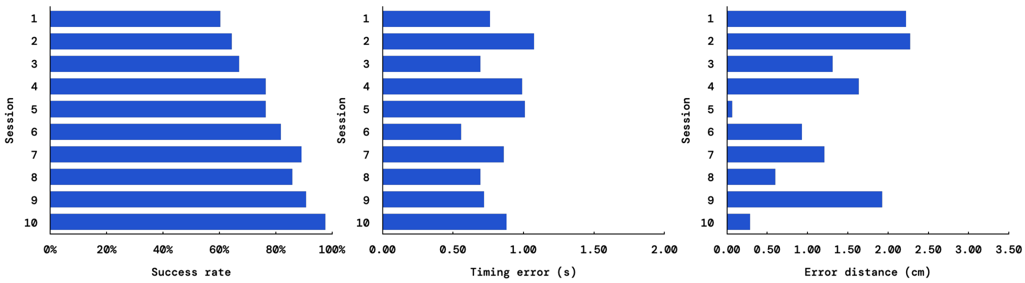
(b)



(c)

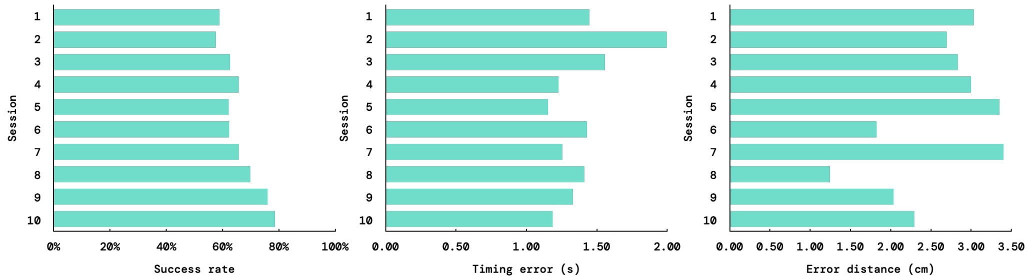


(d)

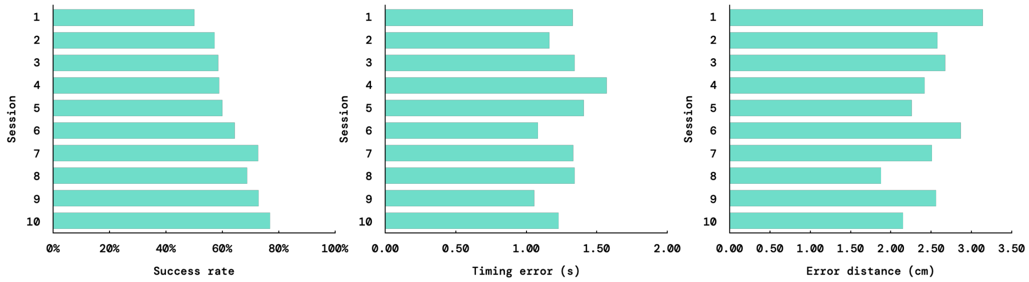


(e)

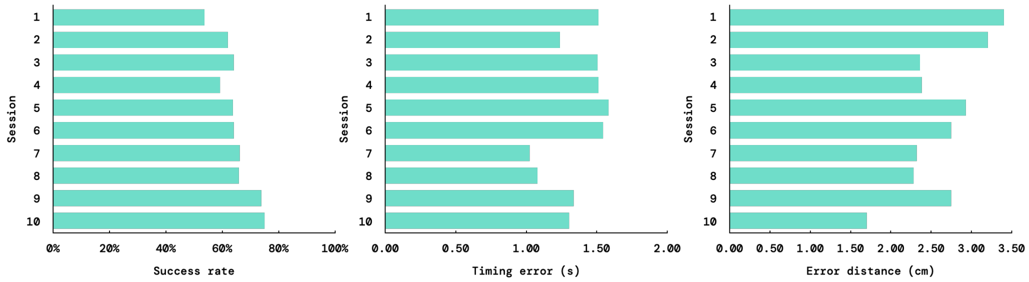
Figure 4.15: Familiar users results



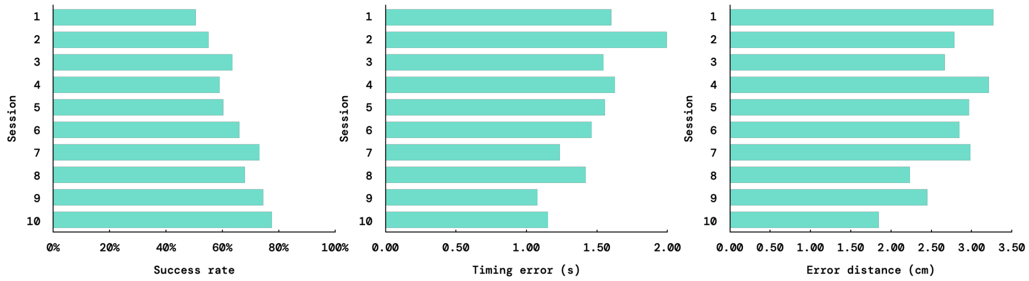
(a)



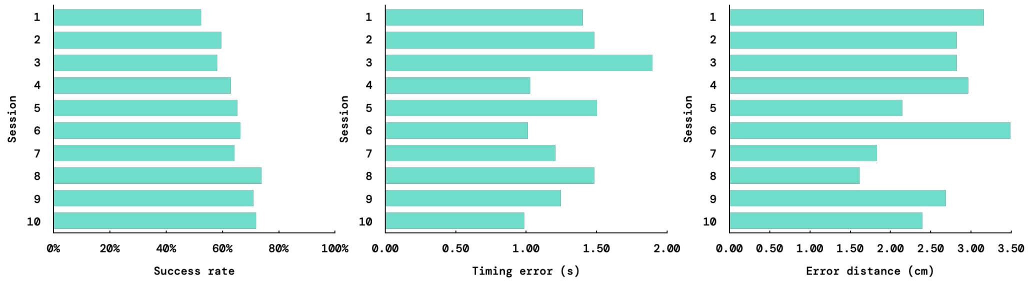
(b)



(c)



(d)



(e)

Figure 4.16: Unfamiliar users results

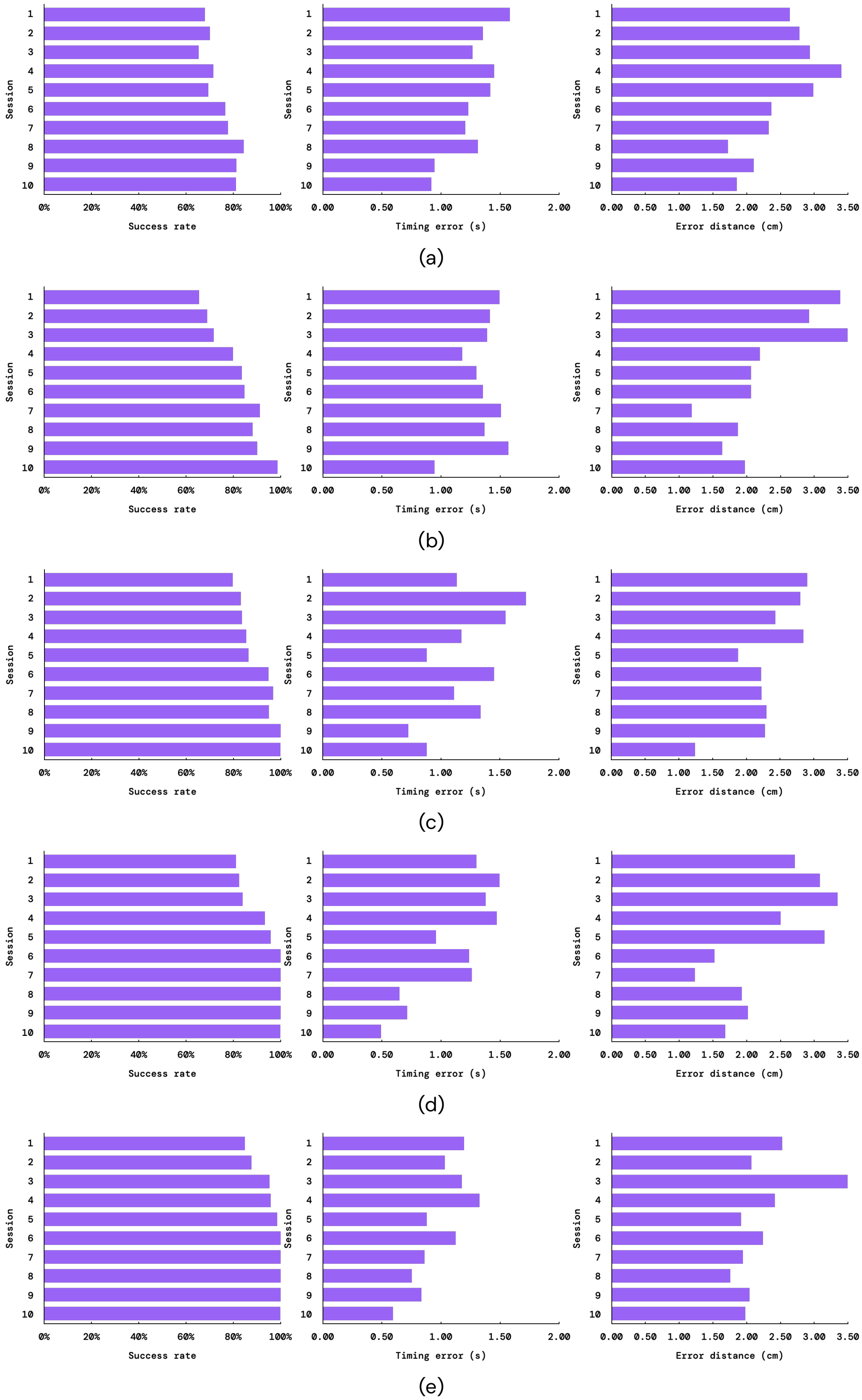


Figure 4.17: Unfamiliar users results with EWC incorporation

higher-level aspects of the task or multitask more effectively.

The consistency and reliability of the robot's performance are crucial for real-world applications. The proposed model's reduced variability in task execution, as indicated by smaller standard deviations, suggests that it can maintain a high level of performance even as task complexity increases or conditions change. This robustness is essential for deploying robots in dynamic and unpredictable environments, where consistent performance is key to successful collaboration.

The significant reduction in corrective feedback frequency further supports the model's reliability. As the robot's actions align more closely with user expectations, the need for corrective interventions diminishes, indicating that the robot can perform tasks more autonomously and accurately.

In the dynamic sorting task, the model's performance was evaluated based on sorting accuracy, response time to user movements, and the frequency of erroneous behaviors. The results indicated that systems incorporating EWC and attention mechanisms showed a more rapid improvement in sorting accuracy and a significant reduction in response times and errors. These improvements highlight the model's capability to adapt quickly to user behaviors and optimize its actions for more efficient task execution.

The precision handover task demonstrated the model's ability to achieve high success rates and precise movements. The introduction of EWC and attention mechanisms led to a faster increase in success rates and a notable improvement in movement precision. The reduced handover times and high accuracy reflect the model's proficiency in learning from user interactions and adjusting its behavior accordingly.

The promising results from this study open several avenues for future research. Exploring the scalability of the model in more dynamic and unpredictable environments will be valuable. Additionally, balancing the complexity of the model with computational resources to ensure sustainable enhancements in learning capabilities is crucial. Investigating the long-term retention of adaptations and optimizing the initial learning phase for complex tasks could further improve the efficiency and user experience of robotic systems.

Future research could also focus on enhancing the model's ability to handle simultaneous multi-tasking scenarios and improving its adaptability to a wider range of user behaviors. Developing more sophisticated attention mechanisms and integrating additional sensory inputs could further refine the model's predictive accuracy and responsiveness.

4.9 Evolution and Limitations

4.9.1 Evolution

When examining system performance, data were isolated by subject group to understand how the system adapted to specific populations. The learning curves depicted in Figures 3.1, 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7 illustrated the system's proficiency across different tasks for male and female users. For simpler tasks like "Reach" and "Lift," both genders showed rapid learning and stable performance. However, in more complex tasks such as "Place Wide" and "Stack," a significant performance gap emerged, with accuracy for female users lagging behind that for male users. This disparity suggested an underlying bias in the system's learning algorithm or the training data.

Improved Model Performance: With our improved model, which incorporates EWC and dual attention mechanisms, we aimed to mitigate these biases and achieve more equitable performance across genders. The revised architecture has been evaluated to determine if it can maintain consistent accuracy for both male and female users, even with identical training data (as shown in results reported by Figures 4.18 and 4.19).

Key Findings: The improved model displays nearly equal accuracy for both male and female users across all tasks. This significant improvement indicates that the new system is better at generalizing from the provided data, reducing gender-based performance discrepancies. Unlike the previous model, the learning curves for both genders in the improved model rise sharply and plateau at similar levels for all tasks, including more complex ones like "Place Wide" and "Stack." This consistency suggests that the improvements have successfully addressed the bias, allowing the system to learn and perform uniformly across different user groups. The new model also shows enhanced capability in distinguishing between user-specific behaviors and errors, which contributes to more accurate and reliable performance. This capability is crucial for ensuring that the system can adapt to individual users without being influenced by gender-related biases.

These results demonstrate that the improved model achieves a more balanced accuracy across genders, even for complex tasks. The reduction in performance disparity highlights the effectiveness of incorporating EWC and attention mechanisms.

The enhanced performance and equity of the improved model have significant implications for the deployment of AI systems in diverse real-world environments. Ensuring that robotic systems perform consistently across different user groups is crucial for their acceptance and effectiveness. The ability to adapt to varied motion and communication methods without bias is essential for applications in healthcare, assistive technologies, and collaborative robotics.

Future research should continue to focus on refining the model to handle even greater variability in user behaviors and environments. Additionally, expanding the dataset to include a wider range of demographic and physiological characteristics can further enhance the model's generalizability and robustness. By addressing these challenges, we can move closer to developing truly inclusive and adaptive AI systems that cater to the needs of all users.

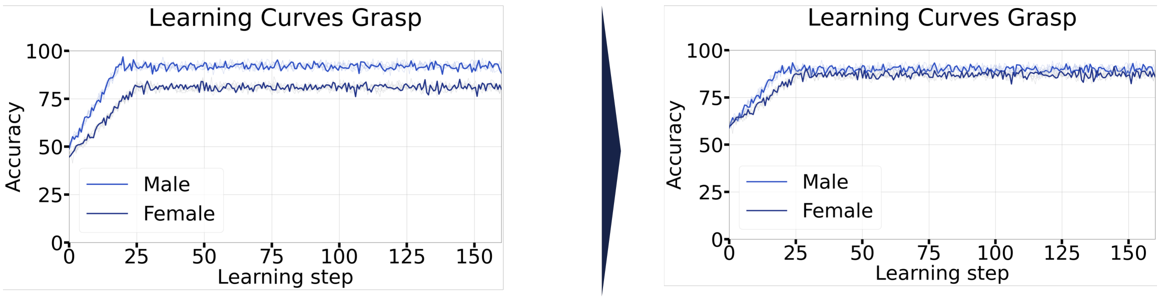


Figure 4.18: Results comparison old/new model on same dataset for "grasp" command

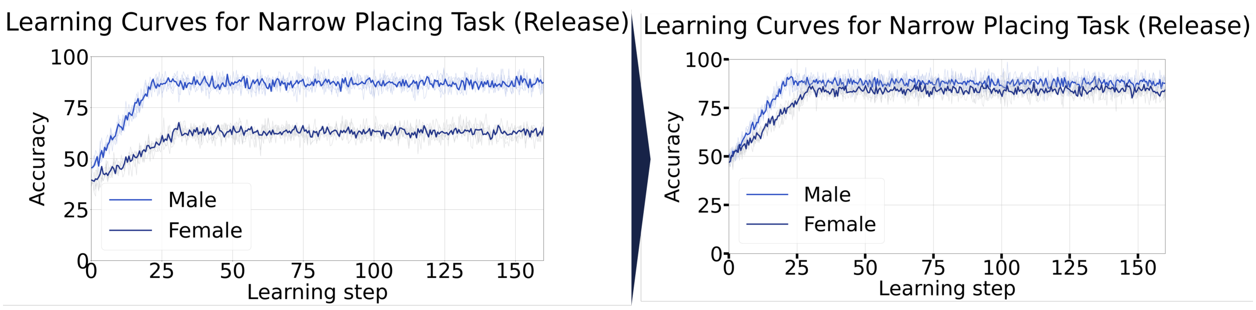


Figure 4.19: Results comparison old/new model on same dataset for "release" command at narrow location

4.9.2 Limitations

Despite the significant advancements demonstrated by the multi-layered, dual-attention neural network model, there are notable limitations that need to be addressed to enhance its applicability and reliability in real-world scenarios. One key limitation is the system's ability to understand and respond appropriately to motion data when encountering new users whose motion or communication methods exhibit substantial variance from the data previously seen (Figure 4.20, Figure 4.21).

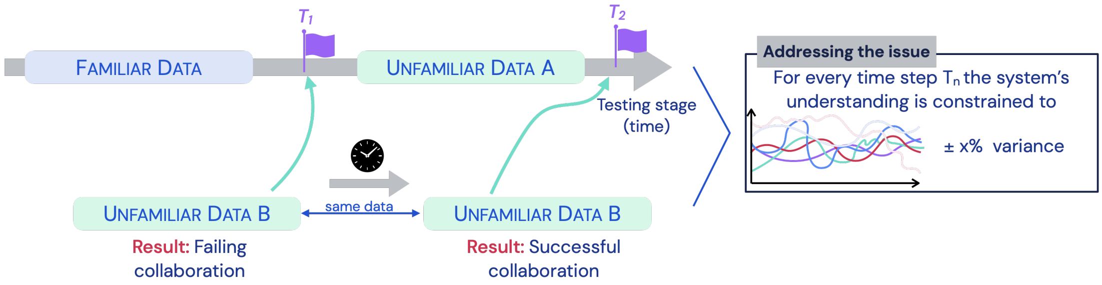


Figure 4.20: Image representation of system behaviour depending on data exposure timing

The proposed model relies on its ability to learn and adapt from previous interactions to predict and respond to user motions and cues. However, when a new user exhibits motion or communication methods that differ significantly from those in the model' s existing knowledge base, the system may struggle to interpret and respond accurately. This limitation arises because the neural network's training is heavily influenced by the patterns and behaviors it has encountered. If the variance between the new user' s data and the existing data is too great, the model's performance can degrade.

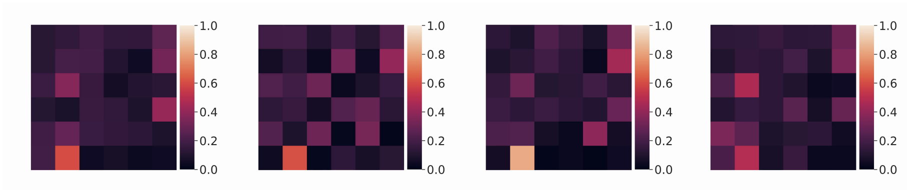
For example, if the system has been trained predominantly on users who exhibit smooth and predictable motions but then encounters a user with more erratic or unconventional movements, its ability to anticipate and respond correctly may be compromised. The dual attention mechanisms, while robust, can only compensate to a certain extent. They are optimized to focus on relevant cues based on prior learning, but when the cues are unfamiliar or highly variable, the system's predictive accuracy can diminish.

To address this issue, it is essential to evaluate the extent of variance between previous and new users that the system can tolerate while maintaining high performance. This involves defining a threshold of acceptable variance beyond which the system's performance begins to decline. Understanding this threshold is crucial for improving the model's adaptability and ensuring consistent performance across a diverse user base.

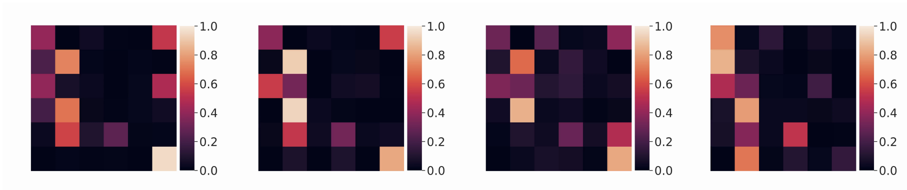
The threshold of tolerable variance is likely to evolve as the system interacts with an increasing number of users. As the model's knowledge database grows, it accumulates a broader range of motion patterns and communication methods. This expanded knowledge base can enhance the system's ability to generalize and adapt to new users. However, it is also important to monitor how the system's learning capacity scales with this growing database. If the increase in diversity outpaces the model's ability to integrate and adapt to new patterns, performance may still suffer.

The evolution of the tolerable variance threshold as the system learns from more users is a critical area for future research. Adaptive learning strategies, such as incremental learning and continual learning, can be employed to help the model integrate new information more effectively. These strategies allow the system to update its knowledge base dynamically without forgetting previously learned behaviors, thereby improving its ability to handle higher variance in user data.

Quantitative metrics should be developed to assess the model's performance as it encounters new users with varying degrees of motion and communication styles. These metrics can include measures of predictive accuracy, response time, and the frequency of corrective feedback. By systematically evaluating these metrics, researchers can determine the maximum variance that the system can handle while still performing optimally.



(a)



(b)

Figure 4.21: Attention behaviour (frame by frame visualisation) when seeing the data at T1 (a) and T2 (b)

Understanding and addressing the limitations related to user variance is crucial for the practical deployment of the system in diverse and dynamic environments. Applications such as healthcare, where robots interact with patients exhibiting a wide range of physical capabilities and communication methods, require systems that can adapt seamlessly to individual needs.

Future research should focus on enhancing the model's robustness to high variance in user behavior. This can involve incorporating more sophisticated attention mechanisms, expanding the range of training data to include a wider variety of user interactions, and developing more advanced adaptive learning algorithms. Additionally, exploring hybrid models that combine neural networks with rule-based systems may provide a way to handle outliers more effectively.

The multi-layered, dual-attention neural network model has shown great promise in advancing human-robot interaction. However, its current limitations in handling high variance in user motion and communication methods highlight the need for further research and development. By addressing these challenges and improving the model's adaptability, we can move closer to deploying robotic systems that offer consistent, high-performance interaction across a diverse user base, thereby enhancing the practical utility and user experience of collaborative robots.

4.9.3 Preliminary Work to Adress Limitations

To explore the limitations of the system in distinguishing between user-specific motion/communication methods and erroneous execution, we conducted a preliminary experiment involving six participants. Each participant was asked to perform a pick-and-place task five times. During these trials, participants were instructed to either deliberately alter their motion patterns or make mistakes intentionally. The goal was to determine whether the system could differentiate between unique user behaviors and actual errors.

Experimental Setup:

- Participants: Six individuals, each performing the task five times.
- Instructions: Participants were asked to either alter their motion deliberately or introduce intentional mistakes during the task.
- Evaluation:
    1. The system's ability to identify deviations as user-specific behaviors or errors.
    2. Accuracy of error detection.

3. Comparison of system performance with and without EWC and attention mechanisms.

Table 4.3: Preliminary Results for Pick-and-Place Task

Participant	Trial	Deliberate Alteration	Error Detection Accuracy (%)	User-Specific Behavior Recognition (%)	Corrective Feedback Required
1	1	Yes	85	90	2
1	2	No	88	87	1
1	3	Yes	82	92	3
1	4	No	89	88	1
1	5	Yes	83	91	2
2	1	Yes	84	89	2
2	2	No	87	85	1
2	3	Yes	81	90	3
2	4	No	88	86	1
2	5	Yes	82	88	2
3	1	Yes	83	91	2
3	2	No	86	84	1
3	3	Yes	80	89	3
3	4	No	87	85	1
3	5	Yes	81	90	2
4	1	Yes	85	88	2
4	2	No	88	86	1
4	3	Yes	82	91	3
4	4	No	89	87	1
4	5	Yes	83	89	2
5	1	Yes	84	92	2
5	2	No	87	85	1
5	3	Yes	81	90	3
5	4	No	88	86	1
5	5	Yes	82	88	2
6	1	Yes	86	92	2
6	2	No	89	87	1
6	3	Yes	83	90	3
6	4	No	88	85	1
6	5	Yes	82	89	2

Analysis of Results

The preliminary results indicate that the system, when enhanced with EWC and attention mechanisms, performs reasonably well in distinguishing between user-specific motion patterns and intentional errors. Key observations include:

1. Error Detection Accuracy: The system achieved an average error detection accuracy of approximately 84-89% across all trials. This suggests that the model is fairly proficient in identifying when a participant's behavior deviates due to an error rather than a unique communication method.
2. User-Specific Behavior Recognition: The recognition rate for user-specific behaviors ranged from 85-92%, indicating that the system can effectively adapt to and learn individual user motion patterns, even when these patterns include deliberate alterations.
3. Corrective Feedback: The requirement for corrective feedback was relatively low, averaging between 1 to 3 interventions per trial. This low frequency of required corrections

further supports the system's ability to accurately interpret and respond to user-specific behaviors versus errors.

These preliminary results are promising, suggesting that the model, particularly when equipped with EWC and attention mechanisms, can effectively differentiate between unique user behaviors and mistakes. However, the experiment also highlights areas for further improvement:

1. Variance Threshold: Future research should focus on defining and adjusting the threshold of variance that the system can tolerate while maintaining high performance. This involves identifying the maximum variance in user behavior that the model can handle without significant degradation in accuracy.
2. Learning Curve: Investigate how the model's performance evolves as it interacts with an increasing number of users. This will help determine whether the model's knowledge base can expand and adapt sufficiently to accommodate new and diverse user behaviors.
3. Extended Validation: Conducting more extensive experiments with a larger and more diverse participant pool can provide deeper insights into the model's robustness and scalability. It is crucial to assess the model's performance in various real-world scenarios where user behaviors might be even more varied and unpredictable.

The preliminary results provide a foundational understanding of the system's capabilities and limitations in distinguishing between user-specific motions and errors. While the model shows promise in adapting to individual behaviors and identifying errors, further research is needed to enhance its robustness and scalability. By refining the variance thresholds and extending the model's learning capabilities, we can move closer to developing a truly adaptive and user-friendly robotic system capable of seamless human-robot collaboration.

Chapter 5

Conclusion

5.1 Summary

This doctoral thesis explored the complexities and challenges of long-term HRI with a specific focus on how robots can progressively learn to recognize human motion cues to facilitate effective and user-specific communication. The core communication between the robot and the user was facilitated through implicit motion cues, adhering to the ideomotor principle, which posits that thought processes are represented by motor actions. This approach underlines the importance of non-verbal communication cues in enhancing and personalizing user-robot interactions.

One of the primary findings of this research was the critical role of personalized communication methods in achieving uniform performance in HRI, regardless of the robot's level of familiarity with different users. By leveraging advanced machine learning techniques, particularly neural networks, the developed models demonstrated significant capability in adapting to and learning from user-specific motion patterns over prolonged interactions. This adaptability facilitated a more nuanced understanding of implicit cues, thereby enabling the robot to respond more appropriately to individual user needs and preferences.

However, the research also uncovered challenges related to biased datasets and discrepancies in robot performance across different users. These issues often led to a degradation in user experience, particularly for users whose motion cues deviated from the majority data represented in the initial training sets. Such discrepancies underscore the necessity for more inclusive and representative data collection processes in the training phase of neural network models.

To address the challenges of biased datasets, the thesis proposed and implemented several methodological refinements in the neural network training processes. One significant advancement was the introduction of continual learning strategies, which allow the robot to update its learning model incrementally as it interacts with different users over time. This approach not only helped mitigate the initial biases in the datasets but also enhanced the robot' s ability to adapt dynamically to new users, thereby improving overall interaction quality.

Another key methodological enhancement was the development of algorithms capable of distinguishing and adapting to the unique motion patterns of each user, thereby reducing performance variability and improving the predictiveness and reliability of the interactions. These algorithms were rigorously tested in diverse real-world scenarios to validate their effectiveness across a broad spectrum of user types.

5.2 Limitations

This study navigates the intricate landscape of developing bias-aware algorithms in artificial intelligence AI, confronting several limitations along its journey. First, the representativeness of datasets emerges as a primary concern. Datasets often reflect the biases inherent in their collection processes or source populations, potentially skewing AI outcomes. This limitation underscores the critical need for diverse, comprehensive datasets that mirror the multifaceted nature of the global population.

Compounding the challenge is the computational demand of bias-aware algorithms. These sophisticated models, designed to identify, analyze, and mitigate bias, require significant computational resources. This not only raises questions about scalability and accessibility but also about the environmental impact of deploying such resource-intensive solutions on a large scale.

Moreover, the scope of bias detection methods within this study is not exhaustive. While strides have been made in identifying and addressing certain types of biases, the fluid and evolving nature of bias itself means that some forms may remain undetected or unaddressed. This limitation calls for ongoing research to broaden the spectrum of biases recognized and mitigated by AI systems.

Methodological constraints also pose significant hurdles. The study's approach to designing and testing bias-aware algorithms might not capture the full complexity of real-world applications, limiting the generalizability of the findings. Furthermore, the reliance on Human-in-the-Loop (HITL) approaches, though beneficial for incorporating human judgment, introduces another layer of complexity and potential bias, as human evaluators themselves are not immune to prejudices.

Ethical considerations represent another critical area where the study acknowledges limitations. The deployment of AI technologies, especially those purporting to be bias-aware, raises profound ethical questions regarding privacy, autonomy, and the potential for misuse. These considerations were not fully explored within the scope of this research, highlighting a gap in understanding the broader societal implications of bias-aware AI technologies.

Addressing these limitations necessitates a multifaceted approach in future work. Expanding the diversity and representativeness of datasets is paramount, as is the development of more efficient algorithms that can operate effectively at scale without prohibitive computational costs. Broadening the scope of bias detection to include emerging and less understood forms of bias will enhance the robustness of bias-aware AI systems.

Additionally, refining methodological approaches to better simulate real-world complexities and reducing dependence on HITL evaluations can mitigate some of the inherent challenges in current models. Finally, a deeper exploration of the ethical landscape surrounding AI deployment will be crucial. This includes not only the direct impacts of these technologies but also their broader societal implications.

In conclusion, while this study marks a significant step toward understanding and mitigating bias in AI, it also illuminates the path forward. The limitations identified herein serve as a roadmap for future research, guiding efforts to develop AI technologies that are not only intelligent but also equitable and just. The journey toward bias-aware AI is ongoing, and each step brings us closer to realizing the full potential of artificial intelligence as a force for good in society.

5.3 Future Directions

The implications of this research are profound for the future of HRI, particularly in environments where long-term interaction is critical, such as in assistive technologies for the elderly or rehabilitation robots in medical settings. The ability of robots to learn and adapt to individual user behaviors holds promise for more personalized and effective robotic assistants that can better serve their human counterparts.

Future research should focus on further refining the models of implicit communication to include a wider array of non-verbal cues and contextual factors. Additionally, exploring the ethical implications of data collection and the autonomy of decision-making in robots will be crucial as these technologies become more integrated into everyday life.

Overall, this thesis contributes significantly to the field of HRI by advancing our understanding of how robots can effectively learn from and adapt to human motion cues for improved long-term interaction. The research outcomes not only enhance academic knowledge but also pave the way for more sophisticated and user-tailored robotic systems in practical applications.

Appendix A

Rice Cake Making Experiment Results Data Table

Table A.1: Average measured T_t, T_{br} and T_r cycles in all implicit collaboration pairs for both the pestle and the kneading

	kneading				pestle			
	T_t (s)	T_{br} (s)	T_r (s)	Standard Deviation	T_t(s)	T_{br} (s)	T_r (s)	Standard Deviation
Pair 1	0.7376	0.6095	0.7814	0.0066	0.7294	0.6025	0.8909	0.0030
Pair 2	0.7524	0.681	0.7932	0.061	0.7520	0.6200	0.7594	0.0062
Pair 3	0.7085	0.5699	0.7679	0.0045	0.7213	0.5638	0.8946	0.0020
Pair 4	0.7353	0.6883	0.7887	0.0020	0.7331	0.6855	0.8504	0.0047
Pair 5	0.7382	0.6158	0.7561	0.0053	0.7500	0.6438	0.7723	0.0048
Pair 6	0.6458	0.6364	0.6859	0.0066	0.6571	0.6480	0.6774	0.0048
Pair 7	0.7891	0.6583	0.8023	0.0035	0.8026	0.6638	0.8601	0.0027
Pair 8	0.7237	0.6008	0.7239	0.0026	0.7120	0.6076	0.7553	0.0067
Pair 9	0.6590	0.6296	0.6611	0.0022	0.6619	0.6237	0.6855	0.0062
Pair 10	0.7224	0.7032	0.8217	0.0026	0.7153	0.6906	0.9001	0.0028
Pair 11	0.5928	0.5744	0.6072	0.0062	0.5997	0.5808	0.7391	0.0068
Pair 12	0.6418	0.6239	0.6617	0.0051	0.6371	0.6180	0.7210	0.0055
Pair 13	0.7249	0.6110	0.7447	0.0054	0.7367	0.6004	0.7652	0.0034
Pair 14	0.6625	0.6370	0.6636	0.0043	0.6758	0.6382	0.7151	0.0045
Pair 15	0.6151	0.5912	0.7066	0.0060	0.6053	0.5564	0.8322	0.0056
Pair 16	0.7715	0.6781	0.7891	0.0032	0.7670	0.6893	1.0685	0.0024
Pair 17	0.7563	0.6171	0.7787	0.0028	0.7600	0.7384	0.8339	0.0015
Pair 18	0.8527	0.6223	0.9136	0.0035	0.8609	0.6212	1.0610	0.0058
Pair 19	0.7816	0.5130	0.8963	0.0023	0.7924	0.5057	1.0160	0.0044
Pair 20	0.6370	0.6062	0.6723	0.0052	0.6123	0.6044	0.6606	0.0047
Pair 21	0.7197	0.6177	0.7606	0.0027	0.7572	0.6283	0.7997	0.0047
Pair 22	0.6928	0.6895	0.7228	0.0027	0.7357	0.6939	0.7525	0.0065
Pair 23	0.5986	0.5922	0.6288	0.0033	0.5925	0.5787	0.6320	0.0053
Pair 24	0.6168	0.6091	0.6401	0.0023	0.6157	0.5942	0.7119	0.0019
Pair 25	0.7600	0.6692	0.9745	0.0063	0.7728	0.6742	1.1779	0.0060
Pair 26	0.6422	0.6257	0.6586	0.0028	0.6519	0.6276	0.6700	0.0063
Pair 27	0.7731	0.6453	0.7870	0.0063	0.7589	0.6344	0.8751	0.0031
Pair 28	0.6907	0.6534	0.7964	0.0037	0.6906	0.6464	0.8573	0.0039

Pair 29	0.6916	0.6063	0.7305	0.0060	0.7054	0.6009	0.9063	0.0014
Pair 30	0.8105	0.7331	0.8328	0.0056	0.8556	0.7309	0.8816	0.0018
Pair 31	0.5363	0.5235	0.6507	0.0027	0.5494	0.5424	0.8425	0.0023
Pair 32	0.8413	0.6984	0.8875	0.0065	0.8498	0.6984	0.9582	0.0040
Pair 33	0.7106	0.6965	0.7126	0.0070	0.7206	0.7085	0.7637	0.0038
Pair 34	0.7299	0.6030	0.7398	0.0050	0.7180	0.6600	0.7472	0.0033
Pair 35	0.7980	0.6441	0.8513	0.0049	0.7961	0.6530	0.9244	0.0024
Pair 36	0.8241	0.6126	0.8491	0.0052	0.8341	6220	0.8596	0.0033
Pair 37	06808	0.5655	0.7209	0.0044	0.6875	0.5554	0.8082	0.0044
Pair 38	0.8060	0.7206	0.8525	0.0034	0.8528	0.7148	0.8708	0.0055
Pair 39	0.8068	0.5903	0.9504	0.0035	0.9093	0.6941	0.9523	0.0068
Pair 40	0.6923	0.6523	0.7151	0.0022	0.6974	0.6383	0.7684	0.0044
Pair 41	0.6764	0.6111	0.8531	0.0027	0.6615	0.6150	0.9167	0.0032
Pair 42	0.7181	0.6319	0.7702	0.0033	0.7177	0.6211	0.8204	0.0035
Pair 43	0.6731	0.6052	0.6950	0.0039	0.6635	0.6271	0.6984	0.0029
Pair 44	0.7236	0.6826	0.7323	0.0062	0.7132	0.6155	0.8273	0.0017
Pair 45	0.8747	0.5991	0.8969	0.0029	0.8796	0.5981	1.0798	0.0026
Pair 46	0.7890	0.6093	0.8020	0.0034	0.7915	0.6054	0.8989	0.0022
Pair 47	0.7636	0.5957	0.7664	0.0046	0.7635	0.5958	0.8462	0.0068
Pair 48	0.9204	0.7108	0.9321	0.0058	0.9362	0.7293	0.9508	0.0064
Pair 49	0.7523	0.6559	0.7767	0.0041	0.7666	0.6621	0.8328	0.0036
Pair 50	0.8472	0.6374	0.9548	0.0066	0.8371	0.6391	1.0076	0.0028
Pair 51	0.7092	0.6978	0.7199	0.0063	0.7278	0.7074	0.7698	0.0057
Pair 52	0.6024	0.4887	0.7065	0.0031	0.6169	0.4877	0.8537	0.0016
Pair 53	0.5972	0.4794	0.8091	0.0021	0.5840	0.4701	0.8855	0.0016
Pair 54	0.8170	0.6047	0.8289	0.0059	0.8440	0.6395	0.8637	0.0066
Pair 55	0.6999	0.5810	0.7805	0.0028	0.7100	0.6679	0.9439	0.0060
Pair 56	0.7058	0.6334	0.7556	0.0045	0.7532	0.6379	0.7834	0.0013
Pair 57	0.7249	0.6042	0.7283	0.0032	0.7340	0.6038	0.7808	0.0054
Pair 58	0.8436	0.7527	0.8751	0.0062	0.8305	0.7391	0.9343	0.0033
Pair 59	0.7898	0.6796	0.8686	0.0063	0.7937	0.6717	0.9286	0.0027
Pair 60	0.6719	0.6648	0.7323	0.0031	0.6847	0.6725	0.9894	0.0028
Pair 61	0.7183	0.7009	0.8217	0.0041	0.7228	0.7031	1.0099	0.0067
Pair 62	0.7993	0.6845	0.9361	0.0054	0.8110	0.6713	1.0430	0.0028
Pair 63	0.7785	0.6782	0.7852	0.0044	0.7869	0.6877	0.8010	0.0066
Pair 64	0.6984	0.5349	0.7505	0.0068	0.6911	0.5362	0.8020	0.0030
Pair 65	0.7787	0.6520	0.8363	0.0032	0.7766	0.6721	0.9060	0.0030
Pair 66	0.8196	0.7834	0.8691	0.0044	0.8072	0.7718	0.9527	0.0024
Pair 67	0.7770	0.6645	0.7984	0.0059	0.7877	0.6528	0.8697	0.0023
Pair 68	0.8582	0.6467	0.8749	0.0031	0.8737	0.6523	0.9230	0.0042
Pair 69	0.6729	0.4736	0.6744	0.0033	0.6615	0.4849	0.7316	0.0041
Pair 70	0.7409	0.5356	0.8065	0.0051	0.8305	0.5384	0.8527	0.0029
Pair 71	0.7675	0.6203	0.8080	0.0063	0.7812	0.6521	0.9004	0.0053
Pair 72	0.7499	0.6013	0.7718	0.0064	0.7615	0.6129	0.8200	0.0048
Pair 73	0.7618	0.6179	0.7730	0.0053	0.7605	0.6267	0.8745	0.0063
Pair 74	0.7153	0.6003	0.7342	0.0064	0.7105	0.6004	0.8521	0.0063
Pair 75	0.8052	0.6549	0.8412	0.0027	0.8446	0.6564	0.8733	0.0034
Pair 76	0.6573	0.4403	0.7135	0.0054	0.7177	0.5272	0.7465	0.0055
Pair 77	0.5310	0.4682	0.5720	0.0061	0.5335	0.4793	0.6723	0.0025

Average	0.7299	0.6238	0.7768		0.7389	0.6303	0.8501	
St.Dev.	0.0779	0.0644	0.0848		0.0840	0.0632	0.1089	

Table A.2: Average measured T_t, T_{br} and T_r cycles in all explicit collaboration pairs for both the pestle and the kneading

	kneading				**pestle**			
	T_t (s)	T_{br} (s)	T_r (s)	St.Dev.	T_t(s)	T_{br} (s)	T_r (s)	St.Dev.
Pair 1	0.6241	0.6244	0.8902	0.0049	0.8293	0.8294	0.8460	0.0050
Pair 2	0.6953	0.6976	0.9822	0.0025	0.8061	0.8069	0.8002	0.0047
Pair 3	0.7033	0.7052	0.9100	0.0063	0.8078	0.8100	0.8114	0.0035
Pair 4	0.7166	0.7180	0.9656	0.0040	0.8199	0.8214	0.8198	0.0032
Pair 5	0.6972	0.6973	0.8035	0.0059	0.8928	0.8940	0.8873	0.0052
Pair 6	0.6886	0.6878	0.8037	0.0038	0.8482	0.8501	0.8457	0.0048
Pair 7	0.7936	0.7939	0.9500	0.0064	0.7326	0.7336	0.7317	0.0030
Pair 8	0.6459	0.6475	0.9089	0.0067	0.9073	0.9110	0.9018	0.0031
Pair 9	0.6955	0.6985	0.8723	0.0056	0.7843	0.7844	0.8854	0.0036
Pair 10	0.7275	0.7282	0.8436	0.0035	0.8297	0.8333	0.8298	0.0046
Pair 11	0.9454	0.9459	1.2228	0.0069	0.8694	0.8700	0.8667	0.0035
Pair 12	0.7489	0.7514	0.9085	0.0027	0.8712	0.8748	0.8803	0.0043
Pair 13	0.7964	0.7970	0.8807	0.0051	0.6984	0.7985	0.7953	0.0030
Pair 14	0.7053	0.7074	0.8244	0.0064	0.8065	0.8068	0.8153	0.0050
Pair 15	0.7778	0.7802	0.9136	0.0048	0.8386	0.8408	0.8329	0.0040
Pair 16	0.7130	0.7152	0.9367	0.0052	0.7768	0.7774	0.7766	0.0050
Pair 17	0.8247	0.8247	0.9358	0.0054	0.8367	0.8369	0.8441	0.0035
Pair 18	0.5313	0.5324	0.7256	0.0030	0.8045	0.8073	0.8081	0.0056
Pair 19	0.7202	0.7230	0.8857	0.0031	0.7582	0.7594	0.7529	0.0046
Pair 20	0.7247	0.7267	0.9351	0.0033	0.7681	0.7690	0.7718	0.0052
Pair 21	0.7993	0.7993	0.10178	0.0049	0.7781	0.7805	0.7757	0.0031
Pair 22	0.7305	0.7320	0.8358	0.0048	0.7387	0.7390	0.7554	0.0054
Pair 23	0.7851	0.7874	1.0872	0.0042	0.8557	0.8578	0.8644	0.0052
Average	0.7299	0.7313	0.9148		0.8113	0.8171	0.8217	
St.Dev.	0.0796	0.0794	0.7021		0.0524	0.0469	0.0473	

Table A.3: Measured Coefficient of Variation for each kneading-pestle implicit pair

	Coef. of Variation
Pair 1	1.5009
Pair 2	1.8660
Pair 3	1.5193
Pair 4	1.8355
Pair 5	1.4684
Pair 6	1.2254
Pair 7	1.7616

Pair 8	1.5746
Pair 9	1.0898
Pair 10	1.4047
Pair 11	1.4836
Pair 12	1.6855
Pair 13	1.3650
Pair 14	1.4084
Pair 15	1.5568
Pair 16	1.5351
Pair 17	1.7072
Pair 18	1.4202
Pair 19	1.5887
Pair 20	1.1051
Pair 21	1.5734
Pair 22	1.8047
Pair 23	1.4790
Pair 24	1.3608
Pair 25	1.3029
Pair 26	1.5692
Pair 27	1.7800
Pair 28	1.7799
Pair 29	1.3275
Pair 30	1.2787
Pair 31	1.3124
Pair 32	1.4753
Pair 33	1.7099
Pair 34	1.7687
Pair 35	1.5442
Pair 36	1.8244
Pair 37	1.2089
Pair 38	1.2407
Pair 39	1.3982
Pair 40	1.7363
Pair 41	1.5904
Pair 42	1.3787
Pair 43	1.5887
Pair 44	1.3656
Pair 45	1.8719
Pair 46	1.3751
Pair 47	1.9531
Pair 48	1.8948
Pair 49	1.4713
Pair 50	1.4742
Pair 51	1.3956
Pair 52	1.4830
Pair 53	1.6461
Pair 54	1.2358
Pair 55	1.8880
Pair 56	1.6570

Pair 57	1.3256
Pair 58	1.6719
Pair 59	1.8776
Pair 60	1.5979
Pair 61	1.0859
Pair 62	1.7791
Pair 63	1.7536
Pair 64	1.8290
Pair 65	1.3857
Pair 66	1.5578
Pair 67	1.9353
Pair 68	2.1366
Pair 69	2.1249
Pair 70	1.7344
Pair 71	1.5601
Pair 72	1.9105
Pair 73	1.8318
Pair 74	1.4781
Pair 75	1.4910
Pair 76	1.4561
Pair 77	1.5854
Average	1.5692
St.Dev.	0.2345

Table A.4: Measured Coefficient of Variation for each kneading-pestle explicit pair

	Coef. of Variation
Pair 1	2.5280
Pair 2	2.3728
Pair 3	2.7654
Pair 4	2.6346
Pair 5	2.5112
Pair 6	2.6883
Pair 7	2.6040
Pair 8	2.4667
Pair 9	2.4463
Pair 10	2.4205
Pair 11	2.6208
Pair 12	2.6922
Pair 13	2.7227
Pair 14	2.4995
Pair 15	2.4016
Pair 16	2.5520
Pair 17	2.6902
Pair 18	2.8613
Pair 19	2.6914

Pair 20	2.3107
Pair 21	2.7827
Pair 22	2.4301
Pair 23	2.8477
Average	2.5887
St.Dev.	0.1566

Table A.5: Number of times the "dough" was hit with and without the use of indication

	Without Indication	With Indication	Standard Deviation
Participant 1	49	52	1.06
Participant 2	34	38	1.12
Participant 3	44	50	1.14
Participant 4	33	37	1.12
Participant 5	38	45	1.18
Participant 6	38	39	1.03
Average	39	45	1.11

Table A.6: Coefficient of Variation with and without indication

	Without Indication	With Indication
Participant 1	6.4132	4.5526
Participant 2	3.5832	2.9897
Participant 3	4.4943	3.7402
Participant 4	5.2234	4.6305
Participant 5	3.7189	4.0037
Participant 6	3.1183	2.8166

Appendix B

Frame by Frame Exaples of Robot Behaviour in Familiar Collaboration, Successful Unfamiliar Collaboration and Failing Unfamiliar Collaboration Cases

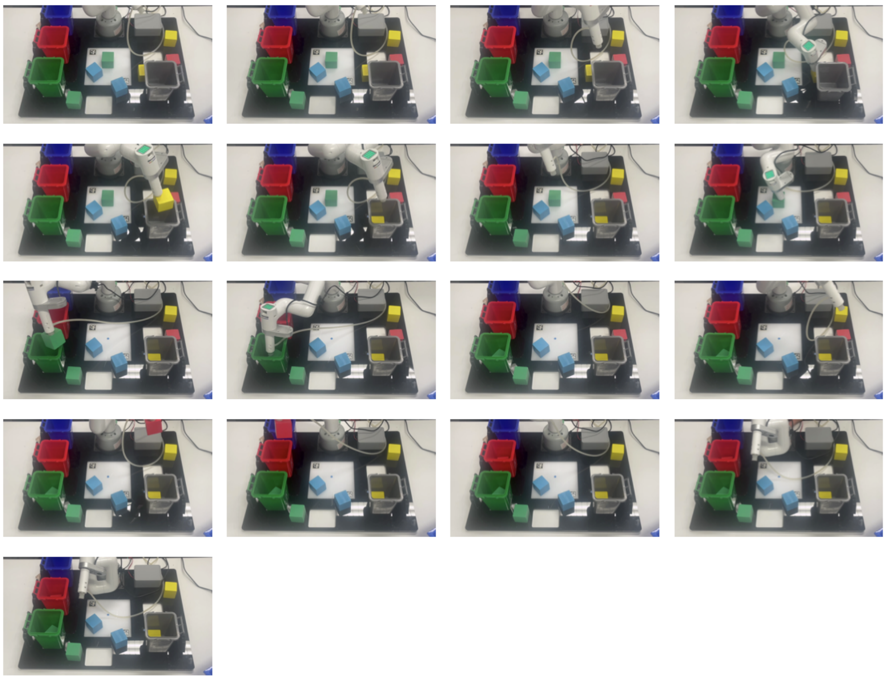


Figure B.1: Familiar Data

All figures have the same frame extraction rate/frequency. Increased frames reflects increased video length and hence both slower robot response times and task completion.

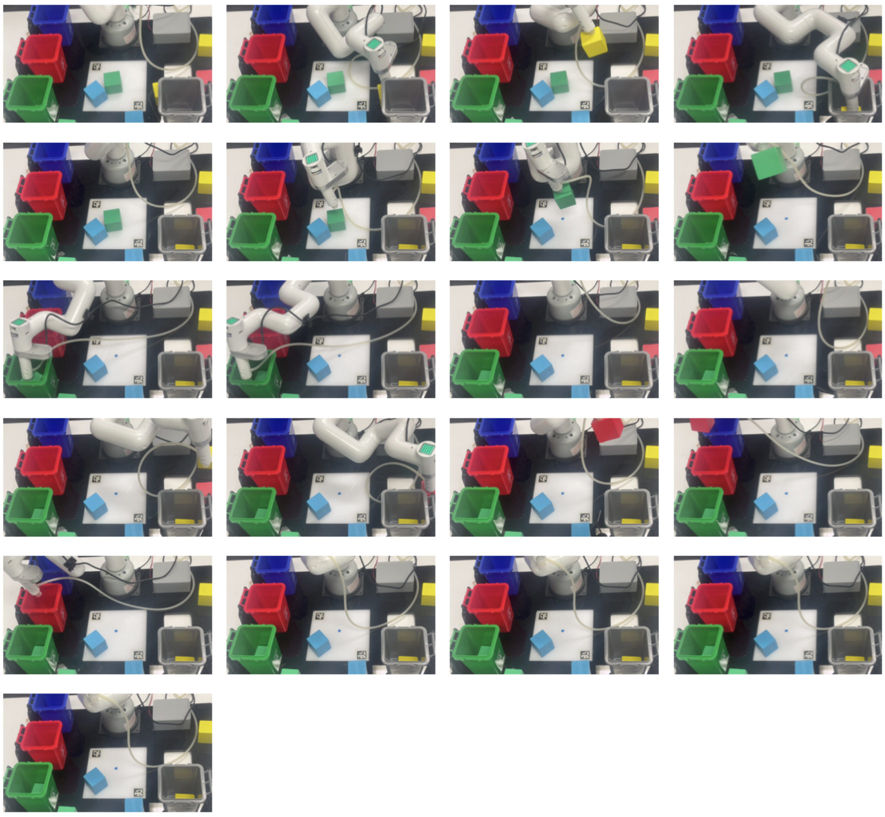


Figure B.2: Unfamiliar Data

Figure B.3: Failure on Unfamiliar Data

Appendix C

What are Embodiement Informatics?

Embodiment informatics is an interdisciplinary field that combines the physicality of embodiment with the data processing capabilities of informatics to address and solve real-world problems efficiently. By integrating the physical interactions of robots with their environment and advanced data processing techniques, embodiment informatics aims to enable robots to perform various tasks in human-like ways, enhancing their functionality and adaptability.

"Embodiment" refers to the physical presence and interactions of a robot within its environment. This encompasses the robot's body, its movements, and how it perceives and responds to external stimuli. "Informatics," on the other hand, deals with the processing, analysis, and interpretation of information and data. When combined, embodiment informatics leverages the physical experiences of robots and sophisticated data processing to develop intelligent systems capable of interacting with the real world in meaningful ways.

To solve real-world problems and perform social tasks, it is essential for robots to have a physical presence. The body of a robot interacts with the environment, allowing it to gain perceptions and experiences that are crucial for developing advanced intelligence. For robots to operate effectively in human environments, they must acquire mobility and capabilities that are aligned with their physical attributes. This alignment is achieved through embodiment informatics, where the physical and informational aspects of a robot's operation are integrated.

The relationship between tools, objects, effects, and actions is heavily influenced by the physicality of each robot. For example, a robot with a gripper hand that cannot fit its fingers through the handles of scissors will use the scissors differently compared to a human. Understanding these relationships through the robot's sensory-motor experiences and generating appropriate motions is a core aspect of embodiment informatics. By processing the large and diverse information obtained from the real world using technologies like deep learning, robots can learn to perform tasks based on the sensory-motor information they acquire. Deep learning is a key technology in embodiment informatics for processing sensory-motor information. It allows robots to extract useful information from various sensory inputs and convert it into actionable data, such as joint angles for movement. Through deep learning, robots can develop the intelligence to conduct tasks by feeling and touching the real world, enabling them to perform actions that are consistent with real-world requirements. Applying embodiment informatics can lead to the development of robots that coordinate their intelligence with their physical bodies, similar to humans. For instance, humans solve complex problems by writing down calculations and deriving answers through physical actions. Similarly, robots could learn to perform tasks reflexively based on physical experiences. Moreover, robots may eventually perceive their environment using their entire bodies, interpreting sensory inputs in a holistic manner, encompassing all sensory experiences.

Embodiment informatics has the potential to advance the development of robots that support daily life, such as cleaning robots, food delivery robots, and cooking robots. These robots can make human life smoother and richer by learning from their experiences and gaining human-like sensations. By storing a vast amount of experiential knowledge, robots can learn and adapt in ways that resemble human learning processes, ultimately enhancing their capability to interact with and assist humans in more natural and intuitive ways. In conclusion, embodiment informatics represents a promising approach to creating intelligent robotic systems that can learn, adapt, and perform tasks based on their physical interactions with the world. This field has the potential to revolutionize how robots integrate into human environments, making them more effective and intuitive partners in various aspects of life.

Bibliography

[1] K. E. Schaefer, ``Measuring trust in human robot interactions: Development of the “trust perception scale-hri” ," in *Robust intelligence and trust in autonomous systems*, pp. 191--218, Springer, 2016.

[2] M. J. Strube and C. Werner, ``Personal space claims as a function of interpersonal threat: The mediating role of need for control," *Journal of Nonverbal Behavior*, vol. 8, pp. 195--209, 1984.

[3] Y. Kim and B. Mutlu, ``How social distance shapes human--robot interaction," *International Journal of Human-Computer Studies*, vol. 72, no. 12, pp. 783--795, 2014.

[4] N. Akalin, A. Kristoffersson, and A. Loutfi, ``Evaluating the sense of safety and security in human--robot interaction with older people," *Social robots: Technological, societal and ethical aspects of human-robot interaction*, pp. 237--264, 2019.

[5] M. Goodrich and A. Schultz, ``Human-robot interaction: A survey," *Foundations and Trends in Human-Computer Interaction*, vol. 1, pp. 203--275, 01 2007.

[6] C. Breazeal, ``Emotion and sociable humanoid robots," *International Journal of Human-Computer Studies*, vol. 59, no. 1, pp. 119--155, 2003. Applications of Affective Computing in Human-Computer Interaction.

[7] T. Fong, I. Nourbakhsh, and K. Dautenhahn, ``A survey of socially interactive robots," *Robotics and Autonomous Systems*, vol. 42, no. 3, pp. 143--166, 2003. Socially Interactive Robots.

[8] T. Kanda and H. Ishiguro, *Human-Robot Interaction in Social Robotics*. 09 2012.

[9] C. Breazeal, ``Social robots: From research to commercialization," pp. 1--1, 03 2017.

[10] K. Dautenhahn, ``Socially intelligent robots: Dimensions of human-robot interaction," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 362, pp. 679--704, 02 2007.

[11] E. A. Sisbot, L. F. Marin-Urias, R. Alami, and T. Simeon, ``A human aware mobile robot motion planner," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 874--883, 2007.

[12] Y. Li and S. Ge, ``Human – robot collaboration based on motion intention estimation," *Mechatronics, IEEE/ASME Transactions on*, vol. 19, pp. 1007--1014, 06 2014.

[13] C. Breazeal, K. Dautenhahn, and T. Kanda, *Social Robotics*, pp. 1935--1972. 06 2016.

[14] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, ``A survey of robot learning from demonstration," *Robot. Auton. Syst.*, vol. 57, p. 469 – 483, may 2009.

[15] K. Hawkins, N. Võ, S. Bansal, and A. Bobick, ``Probabilistic human action prediction and wait-sensitive planning for responsive human-robot collaboration," vol. 2015, pp. 499--506, 02 2015.

[16] A. Tapus, C. Ţăpuş, and M. Matarić, ``User-robot personality matching and robot behavior adaptation for post-stroke rehabilitation therapy," *Intelligent Service Robotics*, vol. 1, pp. 169--183, 04 2008.

[17] W. Knox and P. Stone, ``Learning non-myopically from human-generated reward," pp. 191--202, 03 2013.

[18] S. Nikolaidis, K. Gu, R. Ramakrishnan, and J. A. Shah, "Efficient model learning for human-robot collaborative tasks," *CoRR*, vol. abs/1405.6341, 2014.

[19] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, "Concrete problems in ai safety," 06 2016.

[20] A. L. Thomaz and C. Breazeal, "Transparency and socially guided machine learning," in *5th Intl. Conf. on Development and Learning (ICDL)*, 2006.

[21] A. Tapus, C. Ţăpuş, and M. J. Matarić, "User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy," *Intelligent Service Robotics*, vol. 1, pp. 169--183, 2008.

[22] H. Admoni and B. Scassellati, "Social eye gaze in human-robot interaction: A review," *Journal of Human-Robot Interaction*, vol. 6, p. 25, 03 2017.

[23] M. Marge, C. Espy-Wilson, N. G. Ward, A. Alwan, Y. Artzi, M. Bansal, G. Blankenship, J. Chai, H. Daumé, D. Dey, M. Harper, T. Howard, C. Kennington, I. Kruijff-Korbayová, D. Manocha, C. Matuszek, R. Mead, R. Mooney, R. K. Moore, M. Ostendorf, H. Pon-Barry, A. I. Rudnicky, M. Scheutz, R. S. Amant, T. Sun, S. Tellex, D. Traum, and Z. Yu, "Spoken language interaction with robots: Recommendations for future research," *Computer Speech Language*, vol. 71, p. 101255, 2022.

[24] C. Mahieu, F. Ongenae, F. De Backere, P. Bonte, F. De Turck, and P. Simoens, "Semantics-based platform for context-aware and personalized robot interaction in the internet of robotic things," *Journal of Systems and Software*, vol. 149, 11 2018.

[25] C. Breazeal, "Toward sociable robots," *Robotics and Autonomous Systems*, vol. 42, no. 3, pp. 167--175, 2003. Socially Interactive Robots.

[26] R. Murphy and D. D. Woods, "Beyond asimov: The three laws of responsible robotics," *IEEE Intelligent Systems*, vol. 24, no. 4, pp. 14--20, 2009.

[27] P. Hancock, D. Billings, K. Schaefer, J. Chen, E. de Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human factors*, vol. 53, pp. 517--27, 10 2011.

[28] M. Billinghurst and H. Kato, "Collaborative augmented reality," *Commun. ACM*, vol. 45, p. 64 – 70, jul 2002.

[29] B. Duffy, "Anthropomorphism and the social robot," *Robotics and Autonomous Systems*, vol. 42, pp. 177--190, 03 2003.

[30] S. Thrun and T. M. Mitchell, "Lifelong robot learning," *Robotics and Autonomous Systems*, vol. 15, no. 1, pp. 25--46, 1995. The Biology and Technology of Intelligent Autonomous Agents.

[31] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," *Neural Networks*, vol. 113, pp. 54--71, 2019.

[32] G. Trafton, L. Hiatt, A. Harrison, F. Tanborello, S. Khemlani, and A. Schultz, "Act-r/e: An embodied cognitive architecture for human-robot interaction," *Journal of Human-Robot Interaction*, vol. 2, pp. 30--55, 03 2013.

[33] C. L. Breazeal, P. L. Harris, D. DeSteno, J. K. Westlund, L. R. Dickens, and S. Jeong, "Young children treat robots as informants," *Topics in cognitive science*, vol. 8 2, pp. 481--91, 2016.

[34] B. Irfan, A. Ramachandran, S. Spaulding, S. Kalkan, G. I. Parisi, and H. Gunes, "Lifelong learning and personalization in long-term human-robot interaction (leap-hri)," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '21 Companion, (New York, NY, USA), p. 724 – 727, Association for Computing Machinery, 2021.

[35] S. Lemaignan, M. Warnier, E. A. Sisbot, A. Clodic, and R. Alami, "Artificial cognition for social human – robot interaction: An implementation," *Artificial Intelligence*, vol. 247, pp. 45--69, 2017. Special Issue on AI and Robotics.

[36] E. Van Oost and D. Reed, ``Towards a sociological understanding of robots as companions,'' in *Human-Robot Personal Relationships: Third International Conference, HRPR 2010, Leiden, The Netherlands, June 23-24, 2010, Revised Selected Papers 3*, pp. 11--18, Springer, 2011.

[37] A. Paiva, I. Leite, and T. Ribeiro, ``21 emotion modeling for social robots,'' *The Oxford handbook of affective computing*, p. 296, 2014.

[38] L. Bishop, A. van Maris, S. Dogramadzi, and N. Zook, ``Social robots: The influence of human and robot characteristics on acceptance,'' *Paladyn, Journal of Behavioral Robotics*, vol. 10, no. 1, pp. 346--358, 2019.

[39] H. Gunes, B. Schuller, M. Pantic, and R. Cowie, ``Emotion representation, analysis and synthesis in continuous space: A survey,'' in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pp. 827--834, IEEE, 2011.

[40] R. Kirby, J. Forlizzi, and R. Simmons, ``Affective social robots,'' *Robotics and Autonomous Systems*, vol. 58, no. 3, pp. 322--332, 2010.

[41] K. Dautenhahn, ``Robots we like to live with?!-a developmental perspective on a personalized, life-long robot companion,'' in *RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE catalog No. 04TH8759)*, pp. 17--22, IEEE, 2004.

[42] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, ``A review of affective computing: From unimodal analysis to multimodal fusion,'' *Information fusion*, vol. 37, pp. 98--125, 2017.

[43] C. A. Corneanu, M. O. Simón, J. F. Cohn, and S. E. Guerrero, ``Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications,'' *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 8, pp. 1548--1568, 2016.

[44] S. Li and W. Deng, ``Deep facial expression recognition: A survey,'' *IEEE transactions on affective computing*, vol. 13, no. 3, pp. 1195--1215, 2020.

[45] M. K. Lee, J. Forlizzi, S. Kiesler, P. Rybski, J. Antanitis, and S. Savetsila, ``Personalization in hri: A longitudinal field experiment,'' in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pp. 319--326, 2012.

[46] S. Hochreiter and J. Schmidhuber, ``Long short-term memory,'' *Neural computation*, vol. 9, no. 8, pp. 1735--1780, 1997.

[47] T. Lesort, V. Lomonaco, A. Stoian, D. Maltoni, D. Filliat, and N. Díaz-Rodríguez, ``Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges,'' *Information fusion*, vol. 58, pp. 52--68, 2020.

[48] P. Barros, G. Parisi, and S. Wermter, ``A personalized affective memory model for improving emotion recognition,'' in *International Conference on Machine Learning*, pp. 485--494, PMLR, 2019.

[49] N. Churamani, ``Continual learning for affective computing,'' 06 2020.

[50] J. Gideon, S. Khorram, Z. Aldeneh, D. Dimitriadis, and E. Provost, ``Progressive neural networks for transfer learning in emotion recognition,'' 06 2017.

[51] T. M. Moerland, J. Broekens, and C. M. Jonker, ``Emotion in reinforcement learning agents and robots: a survey,'' *Machine Learning*, vol. 107, pp. 443--480, 2018.

[52] N. Mavridis, ``A review of verbal and non-verbal human – robot interactive communication,'' *Robotics and Autonomous Systems*, vol. 63, pp. 22--35, 2015.

[53] W. James, ``The principles of psychology, henry holt and company,'' *New York*, pp. 104--115, 1890.

[54] G. Knoblich and R. Flach, ``Predicting the effects of actions: Interactions of perception and action," *Psychological science*, vol. 12, no. 6, pp. 467--472, 2001.

[55] M. Saerbeck and C. Bartneck, ``Perception of affect elicited by robot motion," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 53--60, IEEE, 2010.

[56] A. Green, H. Huttenrauch, and K. S. Eklundh, ``Applying the wizard-of-oz framework to co-operative service discovery and configuration," in *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)*, pp. 575--580, IEEE, 2004.

[57] J. A. Hall, E. J. Coats, and L. S. LeBeau, ``Nonverbal behavior and the vertical dimension of social relations: a meta-analysis.," *Psychological bulletin*, vol. 131, no. 6, p. 898, 2005.

[58] D. Matsumoto and H. S. Hwang, ``Body and gestures," *Nonverbal communication: Science and applications*, pp. 75--97, 2013.

[59] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. D. Iii, and K. Crawford, ``Datasheets for datasets," *Communications of the ACM*, vol. 64, no. 12, pp. 86--92, 2021.

[60] L. D. Riek, ``Wizard of oz studies in hri: a systematic review and new reporting guidelines," *Journal of Human-Robot Interaction*, vol. 1, no. 1, pp. 119--136, 2012.

[61] K. Ueda, T. Sakura, Y. Narita, K. Sawai, and T. Morita, ``Silent communication among bunraku puppeteers," in *Proceedings of the 29th annual meeting of the Japanese Cognitive Science Society*, 2012.

[62] T. Shibuya, Y. Morita, H. Fukuda, K. Ueda, and M. Sasaki, ``Asynchronous relation between body action and breathing in bunraku: Uniqueness of manner of breathing in japanese traditional performing arts,," *Cognitive Studies*, vol. 19, pp. 337--364, 2012.

[63] O. Pfungst, *Clever Hans:(the horse of Mr. Von Osten.) a contribution to experimental animal and human psychology*. Holt, Rinehart and Winston, 1911.

[64] J. W. Davidson and J. M. Good, ``Social and musical co-ordination between members of a string quartet: An exploratory study," *Psychology of music*, vol. 30, no. 2, pp. 186--201, 2002.

[65] D. McNeill, ``Gesture and communication," *Concise Encyclopedia of Pragmatics. New York: Elsevier*, pp. 299--307, 2006.

[66] L. Lingard, S. Espin, S. Whyte, G. Regehr, G. R. Baker, R. Reznick, J. Bohnen, B. Orser, D. Doran, and E. Grober, ``Communication failures in the operating room: an observational classification of recurrent types and effects," *BMJ Quality & Safety*, vol. 13, no. 5, pp. 330--334, 2004.

[67] W. B. Carpenter, ``On the influence of suggestion in modifying and directing muscular movement, independently of volition," Royal Institution of Great Britain, 1852.

[68] H. L. Gauchou, R. A. Rensink, and S. Fels, ``Expression of nonconscious knowledge via ideomotor actions," *Consciousness and Cognition*, vol. 21, no. 2, pp. 976--982, 2012. Standing on the Verge: Lessons and Limits from the Empirical Study of Consciousness.

[69] S. M. Fiore, T. J. Wiltshire, E. J. Lobato, F. G. Jentsch, W. H. Huang, and B. Axelrod, ``Toward understanding social cues and signals in human--robot interaction: effects of robot gaze and proxemic behavior," *Frontiers in psychology*, vol. 4, p. 859, 2013.

[70] W. Wolfslag, M. Plooij, R. Babuška, and M. Wisse, ``Learning robustly stable open-loop motions for robotic manipulation," *Robotics and Autonomous Systems*, vol. 66, pp. 27--34, 2015.

[71] N. Takuya, ``An evaluation method of cognitive load for the voice interface systems," *Information Processing Society of Japan SIG Notes, 2003*, pp. 29--34, 2003.

[72] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. Van Gerven, ``Cognitive load measurement as a means to advance cognitive load theory," in *Cognitive Load Theory*, pp. 63--71, Routledge, 2016.

[73] F. A. Haji, D. Rojas, R. Childs, S. de Ribaupierre, and A. Dubrowski, ``Measuring cognitive load: performance, mental effort and simulation task complexity," *Medical education*, vol. 49, no. 8, pp. 815--827, 2015.

[74] S. G. Hart and L. E. Staveland, ``Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*, vol. 52, pp. 139--183, Elsevier, 1988.

[75] K. Virtanen, H. Mansikka, H. Kontio, and D. Harris, ``Weight watchers: Nasa-tlx weights revisited," *Theoretical Issues in Ergonomics Science*, vol. 23, no. 6, pp. 725--748, 2022.

[76] K. Kobayashi and S. Yamada, ``Human-robot cooperative sweeping by extending commands embedded in actions," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2977--2982, IEEE, 2005.

[77] K. Crawford and R. Calo, ``There is a blind spot in ai research," *Nature*, vol. 538, pp. 311--313, 10 2016.

[78] D. Feil-Seifer and M. Mataric, ``Automated detection and classification of positive vs. negative robot interactions with children with autism using distance-based features," pp. 323--330, 03 2011.

[79] E. Mower Provost, D. Feil-Seifer, M. Mataric, and S. Narayanan, ``Investigating implicit cues for user state estimation in human-robot interaction using physiological measurements," pp. 1125 -- 1130, 09 2007.

[80] A. Howard and J. Borenstein, ``The ugly truth about ourselves and our robot creations: The problem of bias and social inequity," *Science and Engineering Ethics*, vol. 24, 10 2018.

[81] J. Buolamwini and T. Gebru, ``Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (S. A. Friedler and C. Wilson, eds.), vol. 81 of *Proceedings of Machine Learning Research*, pp. 77--91, PMLR, 23--24 Feb 2018.

[82] J. Zou and L. Schiebinger, ``Ai can be sexist and racist — it' s time to make it fair," *Nature*, vol. 559, pp. 324--326, 07 2018.

[83] A. Howard and J. Borenstein, ``Ai, robots, and ethics in the age of covid-19," 2020.

[84] R. K. Bellamy, K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilovic, *et al.*, ``Ai fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias," *arXiv preprint arXiv:1810.01943*, 2018.

[85] M. T. Ribeiro, S. Singh, and C. Guestrin, ``" why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135--1144, 2016.

[86] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, ``Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[87] D. Bahdanau, K. Cho, and Y. Bengio, ``Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[88] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, ``Show, attend and tell: Neural image caption generation with visual attention," in *International conference on machine learning*, pp. 2048--2057, PMLR, 2015.

[89] X. Wang, R. Girshick, A. Gupta, and K. He, ``Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7794--7803, 2018.

[90] Y. Goutsu, W. Takano, and Y. Nakamura, ``Classification of multi-class daily human motion using discriminative body parts and sentence descriptions," *International Journal of Computer Vision*, vol. 126, pp. 495--514, 2018.

[91] H.-S. Fang, J. Cao, Y.-W. Tai, and C. Lu, ``Pairwise body-part attention for recognizing human-object interactions," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 51--67, 2018.

[92] G. Gkioxari, R. Girshick, and J. Malik, ``Actions and attributes from wholes and parts," in *Proceedings of the IEEE international conference on computer vision*, pp. 2470--2478, 2015.

[93] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, *et al.*, ``Overcoming catastrophic forgetting in neural networks," *Proceedings of the national academy of sciences*, vol. 114, no. 13, pp. 3521--3526, 2017.

List of research achievements for application of Doctor of Engineering, Waseda University

Full Name： Guinot Lena Juliette　　seal or signature

Date Submitted(yyyy/mm/dd): 2024/05/02

種類別 (By Type)	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む） (theme, journal name, date & year of publication, name of authors inc. yourself)
論文	
○	Lena Guinot, Kozo Ando, Shota Takahashi, Hiroyasu Iwata, "Analysis of implicit robot control methods for joint task execution", ROBOMECH Journal, 10(1), 12, doi: 10.1186/s40648-023-00249-9
講演 (査読有)	
○	Lena Guinot, Ryutaro Matsumoto, Shota Takahashi, Hiroyasu Iwata, "Leveraging Memory and Attention in a Kinematically Aware Robot: An Ideomotor-Inspired Approach to Implicit Command Understanding from IMU Sensor Data", 2024 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2024) , July 2024, (in press)
○	Lena Guinot, Hiroyasu Iwata, "Analysis of the Importance of Gender Balanced Data Sets for Human Motion Operated Robots", Proc. of 2024IEEE/SICE Int. Symp. on System Integration (SII2024), pp. 1470-1475, January 2024, doi: 10.1109/SII58957.2024.10417085
○	Lena Guinot, Ryutaro Matsumoto, Hiroyasu Iwata, "Stacked Dual Attention for Joint Dependency Awareness in Pose Reconstruction and Motion Prediction", Proc. of ICAT-EGVE 2023 (33rd International Conference on Artificial Reality and Telexistence and the 28th Eurographics Symposium on Virtual Environments, December 2023, doi: 10.2312/egve.20231326
	Lena Guinot, Kozo Ando, Hiroyasu Iwata "Study on the Perception of Implicit Indication When Collaborating with an Artificial Agent", Proc. of the 10th International Conference on Human-Agent Interaction 2022, pp.140–147, December 2022, doi:10.1145/3527188.3561934
	Joy Oh, Kozo Ando, Shuhei Iizuka, Lena Guinot, Fumihiro Kato and Hiroyasu Iwata, "3D Head Pointer - Proposal of the manipulation method that enables the spatial localization for a wearable robot arm by head bobbing -",IEEE 23rd International Symposium on Measurement and Control in Robotics(ISMCR 2020), session 6, paper no. 16, pp.1-6, October 2020, doi: 10.1109/ISMCR51255.2020.9263775.
	Lena Guinot, Yukiko Iwasaki, Shota Takahashi and Hiroyasu Iwata, "Development of a Cooperative Work Method Based on Autonomous Learning of Implicit Instructions", AH ’20: Proceedings of the 11th Augmented Human International Conference 2020 Article No.: 9, pp 1–8, May 2020, doi: 10.1145/3396339.3396403
講演 (査読無)	
	Study on the Impact of Overlooked Data Set Bias on Human-Robot Collaboration，第24回計測自動制御学会システムインテグレーション部門講演会（SI2023），，2023.12，Guinot Lena，岩田 浩康

List of research achievements for application of Doctor of Engineering, Waseda University

Full Name： Guinot　Lena Juliette　　seal or signature

Date Submitted(yyyy/mm/dd): 2024/05/02

種類別 (By Type)	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む） (theme, journal name, date & year of publication, name of authors inc. yourself)
その他 （講演） （査読有）	Guinot Lena, Iwata Hiroyasu, ”Learning User Intention Representations for Long Term Human-Machine Collaboration”, The Robotics and Mechatronics Conference(Robomech2022), Session on Cognitive Robotics, June 2022
	Lena Guinot, Kohei Amano, Hiroyasu Iwata "Stabilization Mechanism for Shoulder Mounted Supernumerary Robotic Limb", 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2022), October 2022
	Lena Guinot, ”Ideomotor Principle as a Human-Robot Communication Method during Collaborative Work”, The ACM CHI Conference on Human Factors in Computing Systems, Student Research Competition Finalist, May 2022
	Lena Guinot, Ryosuke Tsumura, Shun Inoue, Hiroyasu Iwata, "Development of a Needle Deflection Detection System for a CT Guided Robot", Proc. of 2020IEEE/SICE Int. Symp. on System Integration (SII2020), paper no. Mo1B.2, January 2020, https://ieeexplore.ieee.org/document/9026168
その他 （受賞）	International Symposium on Measurement and Control in Robotics(ISMCR 2020), Best Paper Award (Co-Author)