Human-Autonomy Systems: How Robots Can Become a Part of Socially Interactive Teams

STS 4500 Prospectus

School of Engineering, Computer Science Department

Bachelor of Science in Computer Science

The University of Virginia, Charlottesville

Name: John Zoscak

Technical Advisor: -

STS Advisor: Alice Fox

Projected Graduation Date: May 2024

Submission Date: 5/9/2023

Human-Autonomy Systems: How Robots Can Become a Part of Socially Interactive Teams

Overview:

Automated technologies in computer systems are at the forefront of Computer Science. As the field continues to advance, there is a growing desire to increase the scope and breadth of its applicability with the intent of enhancing it's usefulness. In robotics, there is a need for interpretive programs for action mapping related to human agents (this is known as Human Machine Teaming, HMT). This exemplifies that the enhancement of software systems for more seamless human-computer interaction is becoming much more important. As part of this trend, I intend to do a socio-technical systems analysis of HMT, and to assess methods for integrating robots into socially interactive environments by use of Theory of Mind models.

Positionality:

My goal in researching HMT is simple in principle. I grew up in a rural home near NOVA and have gone to school there my whole life. Both of my parents worked in the technology industry for my whole life. I feel very grateful to have had their vested interest in my education and support in fields of interest. I've always been fascinated by engineering and the black box of software engineering behind it all. So, when it comes to studying in higher education, I seek out engagement in the practice. Additionally, it is satisfying to interact with the observers of the software as a continuation of that inquisition. Seeing fascination in the eyes of others is a wonderful reward.

As the fields in software engineering advance, every innovation pushes the line of computing potential and possibilities even higher. I love working toward the innovation of technology and contributing to the spectacle of human advancement. I see continued advancements in the fields of high-level engineering as key in shaping the future. One of the most exemplary examples of this includes robotics and how it is integrating into society. Understanding how robotic systems are going to interact with humans is essential for these changes to be successful. Personally, I think that robotics is one of the most engaging and visibly evolving fields in computer science. This is what first drew me to software engineering in robotics. In 2019, I joined my high school's robotics team. We were a dedicated group who contributed much of our time and energy to our competitions. To this day I seek out similar dedication, engagement, and value in my work. I have come to have robotics and other related studies become a considerable portion of my undergraduate education.

Problematization:

Many workplaces (as well as the US military) are beginning to show interest in human-machine teaming. Research groups are trying to tackle this problem of developing systems that can seamlessly and implicitly cooperate with humans based on seemingly trivial indicators humans are so adept at interpreting from their teammates. This begs the question of how robots could interact with humans at such a deep social level. Generating a profile that captures personality, gestures, and historical behavior is part of it. Using that to map observed behavior with subsequent intent and physical action is the core action of these systems. The viability and scalability of "Theory of Mind" into this topic of robotics is a form of contention that makes moving forward different than many other fields.

Guiding Question:

What social aspects constitute successful teams and how can an intermediate (such as Theory of Mind) be used to cultivate social interactions amongst human-robot systems?

Projected Outcomes:

My research will analyze the literature of the field and will draw in connections between tied subjects to challenge the quality and comprehensiveness of the technologies. Furthermore, it will emphasize social cognition and social relationships in research, connect those to technological solutions, and extend this to HMT. The goal of the research is not to critique, but to help collect cross-disciplinary fields and provide direction towards future development with STS in mind.

Technical Project Description:

In my technical work, I explore the solution of generating terrains by a vector sum of 2-D textures. By utilizing a simple terrain generation methodology, I was able to see that linearly generated progressive terrains can be processed in parallel during progressive generation phases of terrain generating software. By utilizing matrix addition, the terrains can be treated as textures with variable contribution to final terrains. I was able to demonstrate that this is a viable solution for giving users greater control over the final terrain by customizing independently configured textures.
That change in structure of terrain generation will increase the accessibility, customizability, and control over terrains generated by the users of software. Tools like this help the users of them to fully realize their ideas without the restrictions of convoluted processing that is too involved for humans. As similar dynamic relationships between humans and autonomy arise, higher levels of complexity and intuitive response are necessary. Modern robotics and HMT solutions are at the forefront of this continued progression. In this line of work, understanding the human mind at a social and individual level becomes more important. Integrating such technology into our society warrants a study of social relationships and how cyber-physical systems can come to fit in them.

Preliminary Literature Review & Findings:

Theory of Mind (ToM) has been studied for decades now, but regardless of this, there is not much consistency in its formulation. For appropriate tests to be made, a comprehensive and consistent definition must be made before further assessments can be made. (Gurney & Pynadath, 2022) Regardless, most literature on ToM approaches the comprehension / perception of the mental states of others, including their desires, beliefs, and intentions. It is an essential social-cognitive skill in understanding others. When consistency in ToM is established, methods for assessing current technology becomes possible. As of now, there are several technological solutions that show promising growth towards quality ToM models for robots. Just a few examples include the use of Neural Networks, memory-augmentation, and constant feedback loop autonomous learning agents. (Cleland-Huang, et al., 2022; Nguyen, et al., 2023; Patacchiola & Cangelosi, 2022)

One of the biggest roadblocks to integration of autonomy is the ability of humans to trust these systems. Most literature agrees that a primary aspect of human-autonomy integration is human trust in the autonomous agents in a team. (Gebru, et al., 2022) There is a wide swath of research done in this field to help measure levels of trust (Lukyanenko, et al., 2022) and to facilitate improved trust in these systems. (Westphal, et al., 2023) Additionally, there have been studies in how to build competence affordance to autonomous systems. (Hu, et al., 2021; Westphal et al.,

2023; *Johns Hopkins APL's Cyclone Aims to Improve Collaborative Human-Machine Decision Making*, 2022) These investigations demonstrate their discovery as to how information can be presented to human-agents, and thus demonstrate the scalability of autonomous systems into team environments.

Another important development is decision support frameworks that help to incorporate autonomy in teams by organizing the agents under collective goals. (Deng, 2020) This helps remedy problems of forming cohesion. These systems also play a key part in orienting the implicit understanding between human and autonomous agents. (Aldridge & Bethel, 2023; Yang, et al., 2023) Finally, making sense of human cognition out of visual joint attention is a final part of incorporating these technologies. Current work in this doesn't have very general conclusions. (Li, et al., 2022; Perez-Osorio, et al., 2021; Chen et al., 2021)

STS Project Proposal:

In STS, we study the fields of engineering, design, and technology with relation to modern society. It is an attempt to make sense of how technology comes into being because of its conception and goals. Finally, reversing that, it then builds models to assess sociological impacts that technologies make. The goal being to make meaningful, ethical, and positive directives for people in their engineering solutions. (*Science, Technology & Society*, 2017)

In my project, I focus on precisely the sociotechnical dynamic considered in STS, but within the field of human-machine teaming. To constrain the subject, it will be considering the impact that the machines have on their teams. I will be assessing the ability of these machines to exist in that socially dependent, diverse, and trust-centric space. To do this, I will be considering the necessary related studies in ToM literature, psychological analysis of human behavior, current technical progress, and technical literature that attempts to implement theory of mind in AI models.

My approach will be modeled under value-sensitive design. Value-sensitive design (VSD) is a process for maximizing user experience by accounting for the user's values. "Along with a commitment to a deeply interaction stance, the constructs enable designers and technologists to conduct analyses of direct and indirect stakeholders; distinguish among designer values, values explicitly supported by the technology, and stakeholder values; consider individual, group, and societal levels of analysis; and carry out integrative and iterative conceptual, technical, and empirical investigations." ("About VSD," 2020) HMT and ToM may not pertain specifically to human values, but VSD remains relevant as a tool for assessing the user's mind. In the VSD process, engineers try to elicit values, represent them, assess social structure, consider design principles, and produce a value sensitive design toolkit. For the sake of my work, I intend to assess the same factors but specifically under the subject of human intent toward a shared team goal. My main reference for VSD is: ("About VSD," 2020). By using VSD, I can break down exactly what HMT systems need to do.

The research approach will ultimately produce a socio-technical systems analysis. This methodology in my research will help to draw relevant practices in psychology and sociology into the technical space. After the connection has been made, the comprehensiveness and completeness of technology in these approaches under constrained situations can be fully considered. Ultimately, conclusions will be drawn about the extent of dynamic cohesion that can be achieved between humans and autonomy. As a side effect of the subjects discussed in the

literature review, the socio-technical analysis will also consider the technical conduciveness for building trust in autonomy in HMT teams.

Barriers & Boons

Unfortunately, the fields that are discussed in this paper are subject to inconsistencies. Machine ToM lacks comprehensive and consistent verification metrics across literature. (Gurney & Pynadath, 2022) This results in the current work focusing on unclear goals and leaving behind gaps in their reviews. (Aru et al., 2023) This would be okay if the goals in each project were highly constrained, but sometimes they are not. The literature that I will be analyzing is attempting to solve similar problems but draws separate final conclusions. This makes the consolidation of current knowledge highly decentralized, leaving the review susceptible to lacking cross-disciplinary organization. Luckily, much of the literature provides the context under which their work was done, so consolidation remains possible.

The evidence-backed conclusions that current work has drawn can be highly generalized. Drawing together those conclusions while remaining faithful to their original message is difficult because I am not a specialist in these fields of work, and I also have little experience with performing STS research. For the most successful interpretation and analysis of literature, an education in psychology, Cognitive Science, Computer Science, with a background in AI and ToM would be ideal. Consequently, this review's legitimacy is contingent upon honest consolidation of current knowledge, and a highly organized cross-disciplinary review.

References

About VSD. (2020). *VSD Lab*. https://vsdesign.org/vsd/

Aru, J., Labash, A., Corcoll, O., & Vicente, R. (2023). Mind the gap: Challenges of deep learning approaches to Theory of Mind. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-023-10401-x

Audrey L. Aldridge and Cindy L. Bethel. 2023. M-OAT Shared Meta-Model Framework for Effective Collaborative Human-Autonomy Teaming. In Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23). Association for Computing Machinery, New York, NY, USA, 663–666. https://doi.org/10.1145/3568294.3580169

B. Gebru et al., "A Review on Human–Machine Trust Evaluation: Human-Centric and Machine-Centric Perspectives," in IEEE Transactions on Human-Machine Systems, vol. 52, no. 5, pp. 952-962, Oct. 2022, doi: 10.1109/THMS.2022.3144956.

Chen, B., Vondrick, C. & Lipson, H. Visual behavior modelling for robotic theory of mind. *Sci Rep* **11**, 424 (2021). https://doi.org/10.1038/s41598-020-77918-x

Gurney, N., Marsella, S., Ustun, V., Pynadath, D.V. (2022). Operationalizing Theories of Theory of Mind: A Survey. In: Gurney, N., Sukthankar, G. (eds) Computational Theory of Mind for Human-Machine Teams. AAAI-FSS 2021. Lecture Notes in Computer Science, vol 13775. Springer, Cham. https://doi.org/10.1007/978-3-031-21671-8_1

Hu, Q., Lu, Y., Pan, Z., Gong, Y., & Yang, Z. (2021). Can AI artifacts influence human cognition? The effects of Artificial autonomy in intelligent personal assistants.

International Journal of Information Management, *56*, 102250. https://doi.org/10.1016/j.ijinfomgt.2020.102250

Deng, X. (Nancy). (2020). Artificial intelligence and human-robot teaming: Challenges and design considerations. In *Knowledge management, innovation, and entrepreneurship in a changing world* (pp. 103–122). Business Science Reference/IGI Global. https://doi.org/10.4018/978-1-7998-2355-1.ch005

Jane Cleland-Huang, Ankit Agrawal, Michael Vierhauser, Michael Murphy, and Mike Prieto. 2022. Extending MAPE-K to support human-machine teaming. In Proceedings of the 17th Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS '22). Association for Computing Machinery, New York, NY, USA, 120–131. https://doi.org/10.1145/3524844.3528054

Johns Hopkins APL's Cyclone Aims to Improve Collaborative Human-Machine Decision Making. *Johns Hopkins University Applied Physics Laboratory*. (n.d.). Retrieved May 6, 2023, from https://www.jhuapl.edu/news/news-releases/220630b-apl-cyclone-ai-agent-improves-human-machine-teaming

Lukyanenko, R. (2022). Trust in Artificial Intelligence: From a Foundational Trust Framework to Emerging Research Opportunities. University of Virginia, 2022. Retrieved from https://doi.org/10.18130/bxy1-7234

M. Patacchiola and A. Cangelosi, "A Developmental Cognitive Architecture for Trust and Theory of Mind in Humanoid Robots," in *IEEE Transactions on Cybernetics*, vol. 52, no. 3, pp. 1947-1959, March 2022, doi: 10.1109/TCYB.2020.3002892.

M. Westphal, M. Vössing, G. Satzger et al., Decision control and explanations in human-AI collaboration: Improving user perceptions and compliance. Computers in Human Behavior (2023), doi: https://doi.org/10.1016/j.chb.2023.107714

N. Gurney and D. V. Pynadath, "Robots with Theory of Mind for Humans: A Survey," 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Napoli, Italy, 2022, pp. 993-1000, doi: 10.1109/RO-MAN53752.2022.9900662.

Nguyen, D., Nguyen, P., Le, H., Do, K., Venkatesh, S., & Tran, T. (2023). *Memory-Augmented Theory of Mind Network* (arXiv:2301.06926). arXiv. https://doi.org/10.48550/arXiv.2301.06926

Rabinowitz, N., Perbet, F., Song, F., Zhang, C., Eslami, S. A., & Botvinick, M. (2018). Machine theory of mind. In International Conference on Machine Learning (pp. 4218–4227). PMLR

Yang SC, Folke T, Shafto P. The Inner Loop of Collective Human-Machine Intelligence. Top Cogn Sci. 2023 Feb 20. doi: 10.1111/tops.12642. Epub ahead of print. PMID: 36807872.

Z. Li, L. Terfurth, J. P. Woller and E. Wiese, "Mind the Machines: Applying Implicit Measures of Mind Perception to Social Robotics," *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Sapporo, Japan, 2022, pp. 236-245, doi: 10.1109/HRI53351.2022.9889356.