

Personal Statement

Anonymous submission

The Complexity of Human Cognition

The moment I realized that artificial intelligence could unlock the mysteries of human perception was when I watched Geoffrey Hinton's lecture "Does the Brain do Inverse Graphics?" Hinton explained how aligning AI models with human perception not only advances technology but also deepens our understanding of ourselves. He highlighted that convolutional neural networks (CNNs) lack certain human abilities—like translational invariance and handling pose information—that allow us to recognize objects regardless of their orientation or position. This simple insight ignited my passion for AI research. Hinton's perspective revealed a profound opportunity to develop models that mirror our own perceptual behaviors to bridge the gap between human and artificial intelligence.

Central to my research is comparing human capabilities with the limitations of AI models. I see robot learning as the intersection of natural language processing, computer vision, and reinforcement learning. This convergence offers a unique opportunity to build AI systems that not only perceive but also interact with the world as humans do. By enabling robots to learn from and adapt to their environments, we can uncover deeper insights into both AI and human cognition, pushing the boundaries of what intelligent systems can achieve in the real world.

Pursuit of Aligning Model Behavior

As I advance my career in AI research, my goal is to bridge the gap between human and AI behavior by applying lessons from human cognition to robotic systems. I aim to develop deep reinforcement learning models that are safe and incorporate key human-like traits essential for continual learning, such as curiosity, adaptability, and ethical decision-making. Through this work, I aspire to create robots that perform tasks efficiently while interacting with the world in ways that are not only effective but also aligned with human values, ensuring AI systems contribute positively to society.

Teaching AI and Fostering Communities

Beyond my academic pursuits, I have a deep passion for teaching and sharing knowledge. Helping others understand complex AI concepts reinforces my own understanding and spreads enthusiasm for the subject. As a member of a

student-led AI organization, I progressed from member to Director of Curriculum. In this role, I led a team to overhaul the curriculum. We developed comprehensive slide decks, homework assignments, and learning materials to enhance the educational experience. I ensured that our curriculum was rigorous yet accessible by thoroughly preparing our instructors, leading to more effective teaching. Additionally, I dedicated significant time to creating a supportive community for students entering this field.

Pursuing a wide range of interests has broadened my understanding of AI. I actively participate in weekly lunches, seminars, and reading groups in reinforcement learning, natural language processing, and theoretical computer science. By exploring diverse perspectives within AI and related disciplines, I connect concepts across fields and drive innovative research ideas. I make time for these interdisciplinary opportunities because I want to map the breadth of AI in my mind and build bridges between disciplines.

Making Opportunities in AI Research

In pursuing AI research, I have actively sought and created opportunities where few existed. Despite often being the only undergraduate student at seminars and events, I engage enthusiastically by asking questions and contributing to discussions, driven by my passion for learning. To deepen my understanding, I enrolled in Ph.D.-level courses like Robot Learning to gain formal knowledge on advanced research topics. Despite limited undergraduate research opportunities in AI when I started, I took the initiative to demonstrate my capabilities to research labs and Ph.D. mentors. Over time, I also advocated for the inclusion of other undergraduates in research and contributed to establishing formal frameworks for their involvement. These experiences have strengthened my confidence, work ethic, and problem-solving skills—qualities essential for success in research.

Research Experience

A significant part of my journey has been my involvement in various research projects:

- Efficient Fine-Tuning in Robot Learning (Spring and Fall 2023).

During Spring and Fall 2023 at a Natural Language Processing and Robotics Lab, I investigated efficient fine-

tuning methods for robot learning tasks to enhance sample efficiency and generalization. I reviewed and optimized codebases for better performance and conducted experiments using the Meta-World benchmark to evaluate model effectiveness. Notably, I proposed integrating language as an additional information source to improve adapter fusion techniques. Our work demonstrated that training on a subset of tasks could generalize to a larger set, achieving comparable performance to task-specific fine-tuning. This experience provided me with insights into meta-reinforcement learning and adapter methods, and I developed skills in experimental design, data analysis, and critical thinking.

- Phonemic Modeling for Language Models (Fall 2023).

In Fall 2023, under the guidance of a Natural Language Processing professor, I aimed to improve language model rhyming capabilities through phonemic modeling. I assessed phonetic representations in language models by creating a task for rhyming couplet completion and developing custom evaluation metrics. We released an open-source rhyme dictionary to support future research in the field. This project enhanced my understanding of natural language processing and the significance of phonetics in language modeling, allowing me to explore the intersection of AI and creativity.

- Safe Reinforcement Learning Techniques (Since August 2024).

Since August 2024, at a Reinforcement Learning and Robotics Lab, I have been developing safe, sample-efficient deep reinforcement learning techniques for agents operating in real-world environments using human interventions. This experience has enabled me to conduct extensive literature reviews, improve my presentation skills by sharing findings at university symposiums, and gain hands-on experience with real-world robotic systems.

- Addressing Disinformation with AI (Since January 2024).

Beginning in January 2024 at a Polymathic Academy, I have studied how to address the spread of disinformation on social media platforms during the 2024 election. I developed large language model-based bots on a private Mastodon server, and by implementing Chain-of-Thought techniques, we enhanced the realism of bot interactions. Our experiments demonstrated that such bots could alter user preferences and evade expert detection, highlighting significant AI risks. This project deepened my understanding of the ethical implications associated with AI-generated content and underscored the importance of safeguards in AI deployment.

What I Expect from the Consortium

Participating in the AAAI Undergraduate Consortium would be a significant milestone in my academic journey, as I have not attended a conference of this scale. I want to make the most of mentorship from experts, refine my ideas, and enhance my research. This opportunity will allow me to collaborate internationally, meet researchers who have inspired

me, and gain exposure to current problem spaces and innovative approaches. As I prepare to graduate in Fall 2025 and apply for Ph.D. programs, this experience will help me gain confidence in my research and validate the time and effort I have invested in academia.

What I Can Contribute to the Consortium

I will contribute to the Undergraduate Consortium by sharing my vision for developing safe deep reinforcement learning models that embody human behaviors used in continual learning. My dedication to building and fostering communities will enable me to collaborate effectively with peers at the consortium. Bringing a polymathic perspective and a curious voice, I am deeply interested in the application, deployment, and real-world impact of AI technologies. I am not afraid to ask questions and seek opportunities and I am committed to enhancing the collaborative and innovative spirit of the consortium.

Conclusion

I am deeply committed to advancing the field of AI, focusing on reinforcement learning and robot learning. Motivated by the potential to uncover new insights about human cognition through AI, I aspire to pursue a Ph.D. and contribute significant research to the field. My goal is to develop AI technologies that are safe, intuitive, and beneficial to humanity. By cultivating learning communities and supporting the next generation of AI researchers, I emphasize the importance of knowledge sharing and collaborative growth. I am enthusiastic about the opportunities ahead and confident that my experiences, passion, and dedication will contribute positively to the AI community.