Statement of Purpose

of Long Le (https://vlongle.github.io/)

My long-term vision is to have AI agents with common sense that co-exist and assist humans in our environment. To achieve this goal, I aspire to develop socially intelligent and learning-efficient robots. During college, I have had many opportunities to engage with different research projects and do an industry internship. Two major themes of my research so far have been **machine learning** and **multi-actor systems**.

Machine Learning Research: In my first semester of college, I worked with Ph.D. student **John Lalor** on using Item Response Theory for computer vision classifiers. Our work modeled the classifiers as test-takers and classification problems as multiple-choice questions to assess the learners' intrinsic capacity. This short first exposure to research motivated me to seek more opportunities as my advisor graduated from UMass. The following semester, I worked with Professor **Yao Li** on computational neuroscience. Specifically, we were interested in the problem of recovering the parameters of a neuronal population from observed firing. We formulated this parameter estimation as a supervised learning problem, which overcame many shortcomings of previous work, such as needing additional simulation or expert knowledge. In our paper [13], I experimented with 5 ML learners and showed that they could recover the parameters well while a traditional parameter estimator failed. In Summer 2020, I interned at **Facebook** with the Ads Supply Optimization Team. Our team built models that control the number of ads shown to users. Given that a high level of ads is likely to negatively impact the user's sentiment while a low level hurts revenue, our objective was to manage this trade-off. During my internship, I identified and built new contextual ad models. In a large A/B testing experiment, my models significantly outperformed the old non-contextual methods in important metrics, including sentiment-revenue trade-off (3x improvement). After graduation, I will be joining **Google** in February as a software engineer. Through this, I hope to gain more experience in applied machine learning while collaborating with their researchers on fundamental problems before entering graduate school in September.

Multi-actor Systems Research:

Game Theory: Since Fall 2020, I have been working with Professor **Yair Zick** on cooperative game theory. In our setting, players form coalitions to complete some desired tasks without fully knowing each other's skill level. With imperfect information, they have to decide whom to cooperate with and how to divide the final profit. We are interested in the dynamics and fairness of this game, i.e., would players with higher skill get paid more in the long run? So far, I have theoretically shown that it is possible in games with perfect information to reconcile players' strategic behavior with fair division. Surprisingly, this reconciliation might not be possible in games under uncertainty, even when a consistent proxy for skill is available. I have also developed AI agents to play the coalition formation game in a repeated setting. The repeated game has many episodes, each consists of several rounds. In each round, a player is chosen randomly to propose a potential coalition and revenue division. A new coalition is formed if every prospective partner agrees to the proposal; otherwise, the proposal is rejected. The overall skill level of each formed coalition is divulged to its member at the end of each episode. My agents use reinforcement learning (RL) and bandit algorithms to select proposals, responses, and update beliefs. The agent experiments revealed that the cumulative wealth of each agent is often a useful signal for skill, allowing them to form more efficient coalitions and refine their beliefs faster under uncertainty. As the next step, I plan to conduct a human experiment on the same game. Humans have been shown to display complex behavior in previous studies in games with perfect information, so it would be interesting to see their behavior when faced with incomplete knowledge.

Human-agent Teaming: Since the Summer of 2021, I have also been working with Professor **Katia Sycara** at the Robotics Institute, Carnegie Mellon University, on DARPA's Artificial Social Intelligence. In this project, we are studying human-agent teaming in an urban search-and-rescue (USAR) scenario. A team of human players with different capabilities navigates through a collapsed building, clears rubble, and rescues victims. We also develop an AI coach who observes the team and issues interventions to improve the team's performance. For the intervention to be effective, the coach needs to be socially intelligent by accurately inferring the hidden mental states such as beliefs or preferences of the players, the ability known in the literature as Theory of Mind (ToM). In the paper that I second-authored [5], we developed a ToM model to dynamically predict the players' beliefs based on observed behavior and team communication. Our ToM model achieves a human-level prediction accuracy comparable to Mechanical Turkers. Using a ToM model, I also built an intervention model to share knowledge between teammates intelligently. For example, when a medic is planning to enter a room previously discovered as empty by a teammate, my coach predicts their intent using a ToM model and issues preventative advice. In an experiment with synthetic rescuers, I showed that this intervention model improved the team's performance compared to baselines of no knowledge sharing or limited compliance with advice.

Multi-agent Control: In another area of research, I'm also building fully autonomous teams of rescuers. My paper [12] shows some preliminary results on a minimal environment using hierarchical RL. USAR is a difficult domain for RL due to sparse rewards, long horizons, and large state spaces. To tackle these challenges, I employed a coach that operated on a high level and gave commands to each agent, which operated on a low level. The coach also saw a high-level view of the environment abstracted as a graph while agents saw low-level objects such as walls or doors. The hierarchical abstraction in both action and state representation enabled specialization. The coach planned over a much smaller state space and learned to associate rewards with a shorter sequence of macro-actions. Meanwhile, each agent learned the local policy to execute commands. In our experiment, state and action abstractions helped our agents surpass standard multi-agent RL algorithms, including COMA and distributed option-critic. In addition to RL, I am now also looking to tackle the control problem via an operations research lens. From this perspective, agents would explore the environment and discover tasks. Then, we can use an algorithm to schedule the tasks and coordinate the rescuers optimally. This approach also has many layers of complexities that I need to handle since the tasks can have different and often ambiguous priorities, different possible player assignments, and inter-dependencies.

STEM Outreach: In college, I co-founded a club for undergraduates interested in math. Our goal is to create a welcoming space for students to meet and support each other. We have held workshops to share tips and experience applying to industry and research internships, picking classes, and study-together nights. I also have had the opportunity to TA for six classes. As a TA, I strive to foster an inclusive and open environment where my students can feel comfortable asking questions and seeking help. I further believe that engaging instruction, especially at lower-level classes, is essential in reaching out to those students who otherwise might not have considered majoring in math or CS. My favorite aspect of being a TA is thus the one-on-one interaction I have with my students during office or tutoring hours, which allows me to understand each student's unique background and customize my teaching approach to help them. I wish to contribute to MIT EECS' commitment to diversity by being a graduate TA and mentor in outreach programs such as GAAP.

Future Work: Looking to the future, MIT is a great place to work on problems that excite me. In building RL agents at Carnegie Mellon University, I have realized the key challenge of task specification, e.g., how to specify rewards or other objectives that lead to reasonable behavior. On the one hand, sparse rewards are hardly conducive to fast policy learning. On the other hand, it is difficult to properly define dense rewards to ensure desirable behavior (e.g., to avoid reward hacking). At MIT CSAIL, there is several faculty members' research on this problem from different angles, including Professor **Pulkit Agrawal**'s work on self-supervised reinforcement and imitation learning [7, 8], Professor **Julie Shah**'s work on task completion model using temporal logics [10], and Professor **Dylan Hadfield-Menell** work's on inverse reward design [4] and value alignment more generally [3]. I'd like to combine these approaches in a broad scenario that features human-robot collaboration in a dynamic environment, which would necessitate the application of those techniques simultaneously. I'm also fascinated by Professor Agrawal's work towards a general-purpose and versatile robot butler [1] [6, 11] and Professor Shah's towards intelligent machine teammates [9, 14, 15]. I believe working with both of them will allow me to blend my interests and background in robotics and artificial social intelligence. To contribute to their research, I will draw from my experience in hierarchical reinforcement learning and building AI decision support at CMU. I'm also curious about how socially intelligent robots can leverage their ToM models about humans to better understand the humans' feedback or demonstration. As a motivation, consider watching a Youtube tutorial on a complex task (e.g., cooking dishes, learning how to drive) with and without voice narration. I'd argue that having access to voice narration (i.e., one's ToM) is crucial to learning, allowing a viewer to make sense of the demonstrator's actions in terms of their mental states, such as desires, intents, and beliefs. Therefore, a question of interest is whether we can incorporate ToM into our robots to enable faster and more robust imitation learning. Professor **Joshua Tenenbaum** has done substantial work in theory of mind [2] and computational models of human learning in general, so I'd love to collaborate with him in this direction. That being said, I'm always open-minded to other research areas and directions in AI as well. After graduation with a Ph.D., I plan to continue research in academia as a faculty member.

[1] in his own words! [1]

References

[1] Pulkit Agrawal. The task specification problem. In *5th Annual Conference on Robot Learning, Blue Sky Submission Track*, 2021.

[2] Chris Baker and Rebecca Saxe. Bayesian theory of mind: Modeling joint belief-desire attribution. *Proceedings of the Thirty-Third Annual Conference of the Cognitive Science Society*, 01 2011.

[3] Dylan Hadfield-Menell, Anca Dragan, Pieter Abbeel, and Stuart Russell. Cooperative inverse reinforcement learning, 2016.

[4] Dylan Hadfield-Menell, Smitha Milli, Pieter Abbeel, Stuart Russell, and Anca Dragan. Inverse reward design, 2020.

[5] Huao Li, Long Le, Max Chis, Keyang Zheng, Dana Hughes, Michael Lewis, and Katia Sycara. Sequential Theory of Mind Modeling in Team Search and Rescue Tasks. *AAAI Fall Symposium on Computational Theory of Mind for Human-Machine Teams*, 2021. [link].

[6] Gabriel Margolis, Tao Chen, Kartik Paigwar, Xiang Fu, Donghyun Kim, Sangbae Kim, and Pulkit Agrawal. Learning to jump from pixels. *Conference on Robot Learning*, 2021.

[7] Ashvin Nair, Dian Chen, Pulkit Agrawal, Phillip Isola, Pieter Abbeel, Jitendra Malik, and Sergey Levine. Combining self-supervised learning and imitation for vision-based rope manipulation. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2146–2153. IEEE, 2017.

[8] Deepak Pathak, Pulkit Agrawal, Alexei A Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the 34th International Conference on Machine Learning*, pages 2778–2787, 2017.

[9] Sangwon Seo, Lauren R. Kennedy-Metz, Marco A. Zenati, Julie A. Shah, Roger D. Dias, and Vaibhav V. Unhelkar. Towards an AI coach to infer team mental model alignment in healthcare. *CoRR*, abs/2102.08507, 2021.

[10] Ankit J. Shah, Pritish Kamath, Shen Li, Patrick L. Craven, Kevin J. Landers, Kevin Oden, and Julie Shah. Supervised bayesian specification inference from demonstrations. *CoRR*, abs/2107.02912, 2021.

[11] Anthony Simeonov, Yilun Du, Beomjoon Kim, Francois R Hogan, Pulkit Agrawal, and Alberto Rodriguez. Learning to plan with pointcloud affordances for general-purpose dexterous manipulation. *Conference on Robot Learning*, 2020.

[12] Long Le, Dana Hughes, and Katia Sycara. Multi-agent Hierarchical Reinforcement Learning in Urban Search and Rescue. *Robotics Institute Summer Scholar's Journal*, 2021. [link].

[13] Long Le and Yao Li. Supervised Neuronal Parameter Estimation from Spiking Trains. *Preparing for submission to Frontiers in Computational Neuroscience*, 2021. [link].

[14] Vaibhav V. Unhelkar, Shen Li, and Julie A. Shah. Decision-making for bidirectional communication in sequential human-robot collaborative tasks. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '20, page 329–341, New York, NY, USA, 2020. Association for Computing Machinery.

[15] Vaibhav V. Unhelkar and Julie A. Shah. Learning models of sequential decision-making with partial specification of agent behavior. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):2522–2530, Jul. 2019.