Research Statement

Qi Zhang

Cooperative artificial intelligence (AI) equips a group of sequential decision makers, or autonomous agents, with the capability of planning and learning to maximize their joint utility. It is the core component in any AI system that involves multiple autonomous agents collectively accomplishing a shared goal. While being a promising paradigm, current solutions to cooperative AI, instantiated as cooperative multi-agent planning and learning methods, are still far more computationally expensive and data-hungry than what practice can often afford. My long-term research goal is to achieve cooperative AI that is sample- and computation-efficient, real-world applicable, and able to interact with humans.

Since my doctoral study, I have been making breakthroughs toward this goal. My past research has been following two lines of work: §1.1) multi-agent coordination via structured and learnable inter-agent communication, in the format of a) multi-agent commitments through which agents make credible and prolonged promises about their future behavior and b) language-like symbolic communication that facilitates human-AI interaction; and §1.2) identification and exploitation of various forms of multi-agent symmetries that are prevalent in cooperative AI tasks, including multi-agent homogeneity, Euclidean symmetry, and potential game structures. Given that the theory and methods developed from my research are largely independent of any particular domain of application, they have been successfully applied to several applications closely connected to the real world, including transportation management, simulated robotics, NLP for clinical decision support, etc. My doctoral thesis has been acknowledged by David J. Kuck CSE Dissertation Prize presented by the University of Michigan. My current research is supported by three ongoing and one completed NSF awards for which I am the lead/sole PI, including an NSF CAREER award.

In the next 3-5 years, I will continue making progress towards my long-term research goal via the support of my current grants (§2.1), initializing a new project on leveraging foundation models for cooperative AI (§2.2), as well as an effort to expand my applied and interdisciplinary research (§2.3). Below, I describe in detail my completed and ongoing research, followed by my future research plan.

1 Completed and Ongoing Research

1.1 Structured and Learnable Inter-Agent Communication

Multi-agent commitments [IIS-2154904]. The notion of multi-agent commitment refers to an agent making credible and prolonged promises about the consequences of its future actions to achieve coordination. For example, for multiple cars navigating through an intersection, turn signals can be viewed as credible commitments that enable the cars to coordinate their moves in concert by carefully choosing what turn signals to and not to activate. My prize-winning PhD thesis [ZDS+16, ZSD17, ZLSD19, ZDS20a, ZDS20b, ZDS21] contributed formalisms and algorithms for agents to coordinate with a general-purpose multi-agent commitment parameterization. I have continued this line of work under my ongoing NSF project (IIS-2154904, $332K, July 2022 – June 2026) that aims to scale the multi-agent commitment framework to various dimensions of cooperation complexity, including large numbers of agents, long decision horizons, high-dimensional perception and action, and high degrees of partial observability. The project has yielded successful outcomes. In my ICML 2023 paper [CZ23], we leverage Bayes networks (BNs) to introduce action dependencies among cooperative agents, which is an instantiation of multi-agent commitments that seamlessly interpolate the two extremes that have been dominating the field: fully correlated action selection that attains optimality more easily but cannot scale to large numbers of agents, and fully uncorrelated action selection that is computationally friendly yet might sacrifice solution quality. Theoretically, we justify why action dependencies are beneficial by deriving the multi-agent policy gradient formula under BN action dependencies and proving its global asymptotic convergence to Nash equilibria. We further establish its finite-time convergence guarantee in a paper for oral presentation at NeurIPS 2024 BDU Workshop [CZK+24]. In my IROS 2023 workshop paper [MZ23], we have successfully applied the framework to tactical decision-making in autonomous driving scenarios, and my future research will expand this work.

Language-like inter-agent communication. Another form of inter-agent communication is language-like symbolic messages, which is certainly more viable for possible interaction with humans and, just like human language, potentially stronger at compositionality and generalization. My AAAI 2019 paper [ZLSD19] explored the possibility of training a pair of agents to learn to communicate through discrete symbols. We began by developing a new benchmark test where an agent learns to provide instructions upon visual inputs for another to assemble towers from blocks, including tens of thousands of task configurations that are systematically generated by a grammar. This is particularly useful for measuring generalization ability. We then investigated three learning paradigms, supervised learning, contextual bandit learning, and reinforcement learning (RL) from delayed rewards. We found that RL yields the highest quality communication, exhibiting the best generalization ability and giving evidence for a qualitatively greater use of frequent N-grams that indicate a greater degree of compositionality. My group has recently started developing methods to better enable inter-agent communication in language form by leveraging pretrained (vision-)language models. Our work presented at the CoRL 2023 LangRob Workshop [SZ23] has taken our first step in this direction and shown Flamingo, a vision-language model architecture, can connect both visual inputs and language instructions/descriptions to propose meaningful subgoals for a hierarchical RL agent.

1.2 Exploiting Multi-Agent Symmetries [CAREER IIS-2237963]

It is widely believed by scientists that our universe follows certain symmetry patterns and principles that lead to profound implications, such as the existence of conservation laws. When done properly, AI can and has already benefited tremendously from exploiting these symmetries, with perhaps the most well-known example of convolutional neural networks being translation invariant to the input images. Supported by my NSF CAREER award (May 2023 – April 2028), my research seeks to identify and exploit symmetry structures in cooperative AI tasks, particularly those that involve multiple decision-making agents. Symmetries in those tasks prevalently exist in the following forms, and my research has yielded fruitful results on them all.

Multi-agent homogeneity. The first form is permutation invariance which exists if agents are homogenous in terms of their effects on state dynamics and reward function. While there exists prior work that has leveraged this form of multi-agent symmetry through algorithmic techniques such as mean-filed approximation and permutation-invariant centralized critics, we are the first to formalize the theory of multi-agent homogeneity for a finite number of agents, from which we have developed a consensus-based decentralized actor-critic method where the consensus update is applied to both the actors and the critics while ensuring convergence, with my group's work published at ICLR 2022 [CLZ22].

Euclidean symmetries. The second form is Euclidean symmetries, which exist as long as the problem instance is situated in an Euclidean space (e.g., the physical world), as is the case for a variety of applications such as multi-robot systems, video games, materials design, etc. Although prevalent, multi-agent Euclidean symmetries are relatively underexplored, due to the challenges of 1) finding proper representations suitable to distribute Euclidean symmetries among multiple agents and 2) developing architectures capable of exploiting continuous Euclidean symmetries. In my ICML 2024 paper [CZ24], we have overcome these two challenges: 1) we have formulated a subclass of Markov games to rigorously describe multi-agent symmetries by mathematical group transformations performed on states and agents' observations and actions that are represented by Euclidean point clouds; and 2) we have exploited those properties by leveraging recent advances in Euclidean message passing neural networks as the architecture for multi-agent actor-critic algorithms, which are capable of preserving equivariance under all continuous Euclidean transformations. The resulting performance exceeds the prior state-of-the-art on popular multi-agent RL benchmarks with a remarkably large improvement. In a parallel work accepted to NeurIPS 2024 [LCZ24], we have leveraged data augmentation, an alternative to equivariant architectures, for RL-based robot locomotion. Our key contribution is novel limb-based transformation on the robot state representation, enabling rich rotation symmetries to be augmented to achieve superior data efficiency.

Markov potential games. The third form is the potential game structure, where any agent's unilateral behavioral change results in the same amount of change in its utility as for any other agent. Originating from normal-form games, such a structure has recently attracted AI researchers who generalize to the sequential setting as Markov potential games (MPGs). In my ICML 2022 workshop paper [CZD22], we have theoretically established the convergence properties of several classic single-agent policy gradient dynamics for MPGs, including softmax policy gradient, its regularized version, and natural policy gradient tailored for MPGs with best-response (BR). In this work, we are also the first to extend the notion of price of anarchy (PoA), originally from normal-norm games, to MPGs, which measures the quality of a product policy as the ratio between its value summed over all agents and the largest possible value achieved by any product policy, providing bounds on PoA that depend on the MPG's structural assumptions such as smoothness. While most prior works consider only "toy" problems like matrix games, we have further strengthened the theoretical results with an empirical study on high-dimensional benchmarks for deep multi-agent reinforcement learning (MARL). The results have shown the promise of incorporating BR dynamics into deep MARL: our decentralized BR variant significantly improves the performance of state-of-the-art deep MARL algorithm on some of the hardest benchmarks. This opens up a promising direction that has been overlooked by the MARL community. The work is under submission for a top conference [CZDZ24].

1.3 Highlighted Applied Research

Transportation [CNS-2213731]. My group has applied the framework of multi-agent commitments to tactical decision-making in simulated autonomous driving, where agents are multiple autonomous vehicles controlled by algorithms to perform tasks such as merging in the traffic on a highway and navigating through a roundabout that requires coordination among them. Our work presented at a CoRL 2023 workshop [MZ23] develops methods that prescribe an AV's future trajectory as its commitment, so that other AVs can better understand its intent and plan accordingly, which we show is critical to achieve efficient and safe highway merging. This work was in parallel with my completed NSF CCRI project (CNS-2213731, Oct 2022 – Mar 2024), a planning grant that envisioned a large-scale research infrastructure to operate in parallel with real-world transportation systems in real-time and with high fidelity, in a manner to support AI-based discovery of more effective traffic management strategies.

NLP for clinical decision support. Clinical health, with a focus on mental health, is another domain where I have applied my research outcomes. Different from most other domains, healthcare requires AI solutions to be particularly explainable and transparent, which often needs the solutions to abide by clinical rules. Thus, the main theme of my research applied here is to explicitly guide AI solutions with clinical knowledge and process. In my ECML 2021 paper [RZGS21], we took relational contextual bandits as the problem, where the contexts were represented using predicate logic clauses, and designed a UCB-like algorithm that is biased towards explicit relational knowledge. The relational nature of this work makes the method well-situated for knowledge-intensive tasks like clinical decision-making. In another line of work [RGZS22, RZG$^+$23], we layered mental health clinical process structures on the outputs of large language models when tasked to analyze text for mental health concerns, enabling clinician-friendly explanations of the language model predictions.

2 Future Research Plan

2.1 Continuation of the Ongoing Projects

For my active grants, there are numerous exciting ideas yet to be explored, which can lead to top conference papers and new funding opportunities. I below highlight two such directions for my CAREER project.

Approximate symmetries. Thus far, we have been assuming the knowledge of exact symmetries existing in a multi-agent problem of interest. In practice, exact multi-agent symmetries are often hard to find and verify. For example, if a region that a group of ground robots occupy varies in friction, even slightly, then exact rotation symmetries will not exist. Therefore, it is critical for our framework to identify and exploit approximate symmetries. As a first step for a rigorous formalism, I plan to define approximate multi-agent symmetries and prove properties therein in a way that generalizes those for exact symmetries in my completed work. The intuition is that, if the approximation errors are reasonably small, then properties such as equivariant optimal values/policies should also hold reasonably well and therefore can be exploited.

Exploiting symmetries for offline RL. My work on symmetries in RL has been primarily focused on the online setting (i.e., optimizing behavior while interacting with their environment). Our future direction is to exploit symmetries for offline RL, where agents aim to learn the best possible behavior from a fixed, existing dataset. A major challenge of offline RL is insufficient coverage of the dataset, i.e., the dataset does not contain enough samples to represent the whole state-action space well, which in turn causes difficulty in learning optimal values and policies from it. Exploiting symmetries existing in the decision-making task is a promising yet underexplored way to effectively mitigate the issue, because the symmetries reduce the effective state-action space size for optimal control or equivalently produce new, symmetric samples at no cost. My CoRL 2023 workshop paper [LZ23] shows promise, where we improve the performance of existing offline RL algorithms on classic single-agent continuous control tasks by training an abstract critic on symmetric data samples. We are now extending the current success to more complex multi-agent symmetries.

2.2 Foundation Models for Sequential Decision Making

Motivation and research plan. Recently, there has been phenomenal success of language and vision foundation models pretrained on broad data at scale and then adapted to various downstream tasks. However, the progress has been relatively slow to extend such success to the realm of sequential decision making, due to unique challenges therein such as dealing with domain-specific data modality, dynamics, and rewards, as well as long-horizon credit assignment and lack of high-quality data. These existing works on leveraging foundation models lie at two extremes: they either do not even attempt to leverage the language abilities in pretrained large models (instead train task-specific transformer-based architectures from scratch) or are constrained to text-based decision making. I plan to develop novel formalisms and methods that will enable us to combine the best of both worlds, i.e., to adapt pretrained foundation models in a task-specific manner for their commonsense knowledge, language modeling, and reasoning abilities in a way that benefits a wide range of sequential decision making tasks. Specific research tasks will include 1) developing a language-rich history encoder from pretrained foundation models as the core component to achieve more sample-efficient online RL, 2) synthesizing programs as agent policies for non-language-rich tasks, and 3) extending the methods developed in 1) and 2) to the multi-agent setting to facilitate inter-agent communication, possibly in a natural language form that enables interactions even with humans.

Preliminary results and funding. My CoRL 2023 workshop paper [SZ23] has shown that vision-language models can be effectively repurposed to propose subgoals in long-horizon, language-rich navigation tasks. My EMNLP 2024 paper [CZZ24] has developed a method that efficiently leverages large language models (LLMs) for textual bandits. Treating LLMs as static bandit policies, our method dynamically selects either an LLM or a standard bandit learner at each decision round, leveraging LLMs' general knowledge to warm-start bandit learners while relying on the learners to achieve long-term adaptability for domain-specific optimality. These preliminary results prove the promise of this research and have helped secure an NSF award (IIS-2425005, $580K, January 2025 – December 2027), a collaborative proposal with me as lead PI.

2.3 Interdisciplinary and Applied Research

Robotics. Plenty of methods from my research, even though designed to be domain-agnostic, have achieved superior performance on continuous control tasks of simulated robotic locomotion, including those involving multiple robotic agents. I plan to transfer such success to intelligent robotic systems living in physical worlds. Such an endeavor will be in close collaboration with robotics researchers, with possibilities open to various types of robots (e.g., surface and underwater robots) and use cases (e.g., smart agriculture, and planetary exploration). I will primarily target NSF and NASA for my applied research in intelligent robots.

Human-AI teaming. Existing approaches to multi-agent teams either assume a predetermined and fixed team configuration, address teaming from the perspective of an individual agent, or rely on optimizing with simplified agent/task attributes, which fall short of accommodating the rapid development of modern agentic assets. As a future direction, my research aims to adaptively select, train, and refine a subset of an excessive amount of learning agents with heterogeneous perception and actuation capabilities. If successful, the effort will drastically help unlock the potential of leveraging increasingly complex and abundant agentic resources for open-ended multi-agent missions. This research is well-suited for funding opportunities from

DoD agencies. In the next few years, I will target the research for DoD early career programs (Young Investigator Research Programs, DARPA Young Faculty Award). This research will be also promising to promote interdisciplinary collaboration with people from human-robot interaction, human-centered AI, physical AI, etc.

Cyber-physical systems. With my current achievements and preparation in applied AI research for intelligent transportation, I am confident that I will bring about additional funding opportunities in transportation and cyber-physical systems in general. As an immediate next step of my NSF planning grant, my team is targeting the NSF CIRC program, for a 3-year Medium project that will be for up to $2M in total. Another possibility is to go for a research-intensive project, which can be funded by, for example, NSF Smart and Connected Communities, or DOT funds. Both projects involve in-depth interdisciplinary collaboration with researchers from AI, transportation systems, computer and network systems, operation research, etc.

References

[CLZ22] Dingyang Chen, Yile Li, and Qi Zhang. Communication-efficient actor-critic methods for homogeneous markov games. In *ICLR*, 2022.

[CZ23] Dingyang Chen and Qi Zhang. Context-aware bayesian network actor-critic methods for cooperative multi-agent reinforcement learning. In *ICML*, 2023.

[CZ24] Dingyang Chen and Qi Zhang. E(3)-equivariant actor-critic methods for cooperative multi-agent reinforcement learning. In *ICML*, 2024.

[CZD22] Dingyang Chen, Qi Zhang, and Thinh T. Doan. Convergence and price of anarchy guarantees of the softmax policy gradient in markov potential games. In *ICML 2022 DARL Workshop*, 2022.

[CZDZ24] Dingyang Chen, Xiaoling Zeng, Thinh T. Doan, and Qi Zhang. The effectiveness of best-response dynamics in reducing price of anarchy for markov potential games. In *Under Submission*, 2024.

[CZK+24] Dingyang Chen, Zhenyu Zhang, Xiaolong Kuang, Xinyang Shen, Ozalp Ozer, and Qi Zhang. Convergence rates of bayesian network policy gradient for cooperative multi-agent reinforcement learning. In *NeurIPS 2024 BDU Workshop*, 2024.

[CZZ24] Dingyang Chen, Qi Zhang, and Yinglun Zhu. Efficient sequential decision making with large language models. In *EMNLP*, 2024.

[LCZ24] Jinzhu Luo, Dingyang Chen, and Qi Zhang. Reinforcement learning with euclidean data augmentation for state-based continuous control. In *NeurIPS*, 2024.

[LZ23] Jinzhu Luo and Qi Zhang. Exploiting mdp symmetries for offline reinforcement learning. In *CoRL 2023 LEAP Workshop*, 2023.

[MZ23] Nishtha Mahajan and Qi Zhang. Intent-aware autonomous driving: A case study on highway merging scenarios. In *IROS 2023 MAD-Games Workshop*, 2023.

[RGZS22] Kaushik Roy, Manas Gaur, Qi Zhang, and Amit Sheth. Process knowledge-infused learning for suicidality assessment on social media. *arXiv preprint arXiv:2204.12560*, 2022.

[RZG+23] Kaushik Roy, Yuxin Zi, Manas Gaur, Jinendra Malekar, Qi Zhang, Vignesh Narayanan, and Amit Sheth. Process knowledge-infused learning for clinician-friendly explanations. *arXiv preprint arXiv:2306.09824*, 2023.

[RZGS21] Kaushik Roy, Qi Zhang, Manas Gaur, and Amit Sheth. Knowledge infused policy gradients with upper confidence bound for relational bandits. In *ECML PKDD*, 2021.

[SZ23] Jianhai Su and Qi Zhang. Subgoal proposition using a vision-language model. In *CoRL 2023 LangRob Workshop*, 2023.

[ZDS+16] Qi Zhang, Edmund Durfee, Satinder Singh, Anna Chen, and Stefan J. Witwicki. Commitment semantics for sequential decision making under reward uncertainty. In *IJCAI*, 2016.

[ZDS20a] Qi Zhang, Edmund Durfee, and Satinder Singh. Modeling probabilistic commitments for maintenance is inherently harder than for achievement. In *AAAI*, 2020.

[ZDS20b] Qi Zhang, Edmund Durfee, and Satinder Singh. Semantics and algorithms for trustworthy commitment achievement under model uncertainty. *Autonomous Agents and Multi-Agent Systems*, 2020.

[ZDS21] Qi Zhang, Edmund Durfee, and Satinder Singh. Efficient querying for cooperative probabilistic commitments. In *AAAI*, 2021.

[ZLSD19] Qi Zhang, Richard Lewis, Satinder Singh, and Edmund Durfee. Learning to communicate and solve visual blocks-world tasks. In *AAAI*, 2019.

[ZSD17] Qi Zhang, Satinder Singh, and Edmund Durfee. Minimizing maximum regret in commitment constrained sequential decision making. In *ICAPS*, 2017.