Research Statement of Thanh Nguyen-Tang

Machine learning has achieved remarkable breakthroughs across various application domains of Artificial Intelligence (AI), including games, protein folding, natural language processing, drug synthesis, recommender systems, self-driving cars, and materials discovery. Despite the remarkable empirical success, we still lack a solid understanding of the capabilities and limitations of such AI systems. Critical foundational gaps, if left unaddressed, will ultimately impede progress in AI and undermine its potential to meet the future needs of our society.

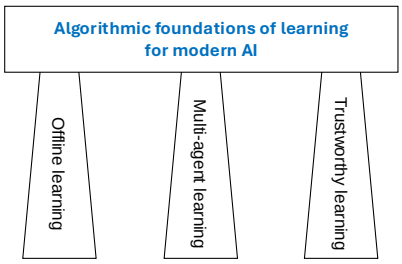

Figure 1: My research is on algorithmic foundations of learning for modern AI systems.

An overarching goal of my research is to establish **algorithmic foundations of learning for modern AI systems**, with the vision of enabling next-generation AI with better *scalability*, *explainability*, and *transferability*. For this vision and goal, I have tackled among the most challenging AI systems: data-driven decision-making, in which learners are tasked with learning an optimal decision-making model using data from interactions with an unknown environment. My work has so far focused on three key challenges of learning, emerged from practical data-driven decision-making, (1) **offline (reinforcement) learning**: learning from pre-collected offline data to mitigate expensive online interaction, (2) **multi-agent learning**: learning in the presence of multiple strategic agents, and (3) **trustworthy learning**: learning safe and robust models for adversarial environments. My approach emphasizes understanding learning through the lens of critical resources (e.g., data and computation) and designing optimal algorithms that use these resources efficiently.

1. Offline (reinforcement) learning

The growth in scale of pre-collected data suggests that data-driven decision-making should take advantage of such offline data for learning. In many cases, such offline learning is not even a choice but mandatory, due to the prohibitive costs, safety concerns, and ethical issues of *online* data collection. However, a key challenge is that the offline data distribution differs from the distribution that is induced by the target policy to be learned. Additionally, offline data often lives in high dimensions, leading to an exponentially large state space. I have established fundamental limits and capabilities of offline learning for data-driven decision-making in large state spaces, using bounded-complexity function approximation (e.g., neural networks) and novel algorithmic design.

Neural networks for offline learning of large-scale problems Neural networks are often used to approximate state value functions and generalize across large state spaces of large-scale problems. However, it remains elusive what natural problems benefit from offline data using neural networks and to what extent. In many large-scale decision-making problems, transition dynamics exhibit similarity between states, allowing for a smoothness assumption to relate these dynamics. In [1], I model this similarity using Besov smoothness – a general smoothness condition that generalizes both Lipschitz and Sobolev smoothness, and show that deep neural networks can exploit this property to learn provably near-optimal policies from uniformly covered offline data, with accuracy *independent* of the number of states and actions. My research got the attention from other machine learning experts who built on my work and studied other dynamic structures to improve sample complexity of offline decision-making using neural networks [2].

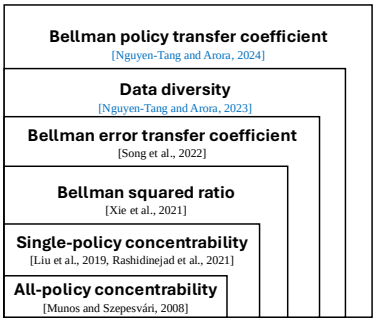

Figure 2: Learnable problem classes characterized by different notions of data coverage.

Driven by real-world needs and advances in deep learning theory, I have developed provably optimal and efficient learning algorithms for offline decision-making using neural networks and gradient-based optimization. These algorithms handle offline data with partial coverage and apply to both contextual bandits [3] and Markov decision processes (MDPs) [4], achieving competitive performance on a large-scale benchmark.

A general theory and algorithmic framework for large-scale offline decision-making Despite its significance, understanding offline decision-making in large state spaces with general function approximation remains limited. In [5, 6], I show that offline decision-making is possible for a wide and novel range of distribution shift regimes under function approximation with bounded ℓ_1-covering numbers, the most general complexity

condition thus far in the offline decision-making literature. The new regimes are captured by new notions of data coverage, including data diversity [5] and Bellman policy transfer coefficient [6], which strictly subsume previous notions (Figure 2). I also show a generic algorithmic framework that offers state-of-the-art error bounds for general bounded-complexity function approximation and, in particular, nearly minimax-optimal bounds for finite pseudo-dimension function classes. This framework unifies existing algorithms and facilitates the design of novel algorithms that use posterior sampling, providing practitioners with better insights, stronger guarantees, and more practical algorithms.

2. Multi-agent learning

Most challenging AI problems can be systematically framed as multi-agent learning, wherein multiple agents learn to act in a shared environment. Learning to make decisions in a static yet unknown environment is already hard, and it is harder still when multiple agents influence each other's learning outcomes and actions. A key challenge for multi-agent learning is non-stationarity, which is faced from the learning agent's perspective when the other agents react strategically. I have established fundamental limits and algorithmic principles of multi-agent learning in both cooperative and competitive settings, using novel analysis and algorithms built upon advances in reinforcement learning and game theory.

Learning against adaptive opponents While most of the literature focuses on learning equilibria, equilibria are not all we need [7]. In strongly reactive systems (e.g., stock markets) where the opponent is adaptive to the learner's past strategies, the learner needs to exploit the opponent to maximize their return. For adaptive opponents, perhaps the only performance measure that makes sense is policy regret [8], a counterfactual notion that evaluates a competing strategy on the sequence of events that would have been generated if the competing strategy were followed. In [9], I initialize the first study on learning against adaptive opponents in Markov games using policy regret. I establish statistical limits for exploiting an adaptive opponent in games by showing the necessity of imposing constraints on memory, stationarity, or regularity in the opponent's responses. I showcase an algorithmic principle to exploit the weakness of an opponent who adapts more structurally. My work will benefit various learning-based systems with hierarchy information, such as human-robot collaboration, autonomous systems, and mechanism design.

Learning to collaborate to solve similar tasks faster One of the main advantages of multi-agent learning emerges in collaborative settings, where agents can share their experiences to learn their similar tasks faster and better. Despite the ubiquity of collaborative settings, a formal understanding of how and when learners of similar tasks benefit from sharing their respective experiences is still in its infancy. In [10], I formulate this question in the context of contextual linear bandits. I design a novel, computationally efficient, and nearly minimax-optimal distributed learning algorithm based on upper confidence bounds, that adaptively coordinates a set of agents to share their respective experiences while the agents are solving their own tasks. This paper will benefit federated learning and distributed collaborative systems.

3. Trustworthy learning

A key challenge in modern AI systems is ensuring generalization across diverse environments. Existing methods often lack robustness to variations in the test environments. To address this challenge, several recent works leverage adversarial training to optimize for worst-case scenarios, which often exhibit poor average-case performance, limiting their practical utility. I have developed scalable algorithms that better balance robust and average performance for data-driven decision-making, using distributionally robust optimization and distributional reinforcement learning (RL).

Robust decision-making under uncertainty Decision-making problems often assume that the training environment matches the deployment environment, which is unrealistic in many practical settings. In [11], I address this by introducing a robust regret objective, aimed at maximizing expected return against the most adversarial distributions over environments. Given the continuous action space, I model reward similarity using Gaussian processes, solve a tractable distributionally robust optimization problem with Thompson sampling on the surrogate reward, and show that this approach converges to an ϵ-suboptimal robust policy in a finite number of steps, despite the continuous action space. This work establishes the first framework for computing distributionally robust policies under uncertainty in continuous action spaces.

Scalable methods for distributional RL Distributional RL learns the full return distribution for each policy, enabling the use of risk-sensitive measures (e.g., CVaR, CPT) for robust generalization. Existing methods represent the return distribution via order statistics but fail to account for their properties during learning. In [12], I introduce a novel method using free particles to simulate return distribution samples based on statistical hypothesis testing. My algorithm sets a new record in the Atari game benchmark, is widely adopted by RL practitioners, and is featured as an exercise in the distributional RL textbook [13].

4. Future directions

I have made significant contributions to the foundations of previous learning settings, though we have only scratched the surface of these areas, and there is much more to be done. Beyond these ongoing topics, I will describe my new research agenda, which fits well within my expertise, inspires multi-disciplinary collaborations (including fellow faculties, postdocs, graduate students, undergraduate students, and high school students), and expands over 2-year, 5-year, and 10-year research plans.

Transfer learning The techniques I have developed in offline learning apply to harness pre-collected data to improve the decision-making performance for the *same* task. While this fundamental setting is already challenging, a generalized paradigm emerges in practice and awaits further research. Moving forward, I will focus on the fundamental question of using data from a source task to improve the decision-making performance w.r.t. a target task for which (active) data collection is limited or unavailable. Answering this question will benefit application domains where it is costly to acquire new data (e.g., AI in medicine, genomics, insurance industry, smart cities), and offer opportunities to re-think many fundamental aspects of machine learning from modern challenges. Building on my previous work, I aim to tackle this research question by exploring the role of task similarity and function approximation for transfer learnability. The emphasis will be on adaptivity, i.e., how to design a learning algorithm that automatically adapts to all task similarity scenarios, instead of specifically designing for different scenarios. Algorithmic approaches to transfer learning, such as representation learning [14, 15, 16] that I have co-developed, will also be useful to study the foundations of transfer learning.

Learning meets games The world is moving toward the coexistence of multiple agents that learn from their interactions. Data input to machine learning algorithms can be generated by self-interested agents, and machine learning is employed to address complex data-driven decision-making problems in economics, such as mechanism design. The theoretical foundations of these problems lie at the intersection of learning and game theory. With my future lab, I will continue to contribute to bridging the gap between machine learning and game theory. Five research agendas aimed at addressing this gap were proposed by other experts in the field almost two decades ago [7]. Despite numerous developments ever since, the field is still in its infancy. For example, we lack a solid understanding of how to act optimally in the presence of other agents who can adapt and learn. This question underlies much of the applications in collaborative settings, where multiple agents are centralized and coordinated to achieve a team goal, or in strategic settings, where agents have their own interests and behaviors. My work [9] has provided an initial yet important step toward addressing this question by studying algorithmic performance through a counterfactual notion of regret, providing statistical limits and algorithmic design insights. A critical next step is to explore which properties of the opponent's learning algorithm (e.g., algorithmic stability) are sufficiently general yet exploitable by a learner, and how to address large-scale problems using function approximation.

Capabilities and limitations of foundation models My previous research has taught me the crucial role of function approximation (e.g., deep networks and complexity-bounded function classes) in dealing with generalization for large-scale learning problems. Recently, a class of special neural networks known as foundation models, such as transformers, have shown remarkable performance in large-scale domains of language, image, and video. It is therefore crucial and timely to unveil the mechanisms by which foundation models facilitate learning for modern AI systems. I will focus on understanding the inductive bias induced by the special connectivity in foundation model (e.g., self-attention layers in transformers). I will investigate this problem through the lens of functional analysis and communication complexity theory to understand which functions that transformers, are (in)capable of computing, as well as from an optimization perspective to examine how training methods contribute to models that generalize well.

References

[1] **Thanh Nguyen-Tang**, Sunil Gupta, Hung Tran-The, and Svetha Venkatesh. On sample complexity of offline reinforcement learning with deep ReLU networks in Besov spaces. *Transactions on Machine Learning Research* (**TMLR**), 2022.

[2] Xiang Ji, Minshuo Chen, Mengdi Wang, and Tuo Zhao. Sample complexity of nonparametric off-policy evaluation on low-dimensional manifolds using deep networks. In *The Eleventh International Conference on Learning Representations*, 2022.

[3] **Thanh Nguyen-Tang**, Sunil Gupta, A. Tuan Nguyen, and Svetha Venkatesh. Offline neural contextual bandits: Pessimism, optimization and generalization. In *International Conference on Learning Representations* (**ICLR**), 2022.

[4] **Thanh Nguyen-Tang** and Raman Arora. VIPeR: Provably efficient algorithm for offline RL with neural function approximation. In *The Eleventh International Conference on Learning Representations* (**ICLR, top-15%-noble**), 2023.

[5] **Thanh Nguyen-Tang** and Raman Arora. On sample-efficient offline reinforcement learning: Data diversity, posterior sampling and beyond. *Advances in Neural Information Processing Systems* (**NeurIPS**), 36, 2023.

[6] **Thanh Nguyen-Tang** and Raman Arora. On the statistical complexity of offline decision-making. In *Forty-first International Conference on Machine Learning* (**ICML**), 2024.

[7] Yoav Shoham, Rob Powers, and Trond Grenager. If multi-agent learning is the answer, what is the question? *Artificial intelligence*, 171(7):365–377, 2007.

[8] Raman Arora, Ofer Dekel, and Ambuj Tewari. Online bandit learning against an adaptive adversary: from regret to policy regret. *arXiv preprint arXiv:1206.6400*, 2012.

[9] **Thanh Nguyen-Tang** and Raman Arora. Learning in Markov games with adaptive adversaries: Policy regret, fundamental barriers, and efficient algorithms. In *Advances in Neural Information Processing Systems* (**NeurIPS**), 2024.

[10] Anh Do, **Thanh Nguyen-Tang**, and Raman Arora. Multi-agent learning with heterogeneous linear contextual bandits. *Advances in Neural Information Processing Systems* (**NeurIPS**), 36, 2023.

[11] **Thanh Nguyen**, Sunil Gupta, Huong Ha, Santu Rana, and Svetha Venkatesh. Distributionally robust Bayesian quadrature optimization. In *International Conference on Artificial Intelligence and Statistics* (**AISTATS**), pages 1921–1931. PMLR, 2020.

[12] **Thanh Nguyen-Tang**, Sunil Gupta, and Svetha Venkatesh. Distributional reinforcement learning via moment matching. In *Proceedings of the AAAI Conference on Artificial Intelligence* (**AAAI**), volume 35, pages 9144–9152, 2021.

[13] Marc G. Bellemare, Will Dabney, and Mark Rowland. *Distributional Reinforcement Learning.* MIT Press, 2023.

[14] Austin Watkins, Enayat Ullah, **Thanh Nguyen-Tang**, and Raman Arora. Optimistic rates for multi-task representation learning. *Advances in Neural Information Processing Systems* (**NeurIPS**), 36:2207–2251, 2023.

[15] Haque Ishfaq, **Thanh Nguyen-Tang**, Songtao Feng, Raman Arora, Mengdi Wang, Ming Yin, and Doina Precup. Offline multitask representation learning for reinforcement learning. In *Advances in Neural Information Processing Systems* (**NeurIPS**), 2024.

[16] Austin Watkins, **Thanh Nguyen-Tang**, Enayat Ullah, and Raman Arora. Adversarially robust multi-task representation learning. In *Advances in Neural Information Processing Systems* (**NeurIPS**), 2024.