Statement of Purpose

Xingjian Bai *Ph.D. Applicant*

My research interests lie at the intersection of algorithms, mathematics, and deep learning – a domain where elegant theories meet practical models. I have explored these intersections through several research projects. These include adapting classic algorithms to leverage learned predictions [1], and enhancing machine learning models through rigorous theoretical analysis, spanning the fields of adversarial robustness [2], generative modeling [3], and reinforcement learning [4]. Integrating the strength of algorithms and deep learning, my primary research objective is to develop intelligent systems with learning and logical abilities, able to perceive, think, and explore the world beyond human-level capabilities.

Algorithms were my idea of fun in high school. Participating in the Chinese Olympiad of Informatics, advancing to the USACO camp, and winning 1st Place in the Canadian Computing Olympiad National Team selection, I enjoyed learning algorithms with friends worldwide. Later, I delved deeper into algorithms with theoretical foundations, such as the Fast Fourier Transform, the Miller-Rabin test, and Euler's iterative method. These algorithms revealed a complex world behind codes, of analytical and algebraic structures I could not yet understand. I majored in mathematics and computer science at Oxford to further explore the theoretical foundations.

Studying more about machine learning (ML), I was captivated by its strong performance across various real-world applications, despite the absence of worse-case guarantees. Its performance in tasks like the game of Go and protein structure prediction well surpasses that of classic algorithms. This led me to a realization similar to what Prof. Sanjeev Arora described in his blog, Off the Convex Path: Beyond the idealized world of problems addressed by classic algorithms, there lies a broader set of challenges – erroneous, stochastic, and worst-case intractable by nature – yet they are prevalent in real-world scenarios. What, then, is the suitable paradigm to tackle these problems? Is it possible to develop solutions that are both practically effective and theoretically sound? Driven by these questions, I entered the world of deep learning from an algorithmic perspective, with the goal of integrating theoretical rigor into practical ML models.

Learning-Augmented Sorting One of my projects came about using inaccurate advice to speed up sorting, perhaps the most fundamental algorithmic task. This project followed the line of work on learning-augmented algorithms, initiated by [5], which seeks to craft algorithms with theoretical guarantees that can leverage predictions from ML models. While previous studies have introduced approximated sorting algorithms utilizing noisy signals, no one has invented provably correct sorting algorithms to cope with erroneous advice. Working with Prof. Christian Coester at Oxford, I developed a framework where a quick-and-dirty comparison function is available, besides the original exact (but slow) comparator. I drew insights from self-balanced data structures and randomized algorithms, and designed a novel algorithm that can precisely sort with a small number of clean comparisons, if given dirty comparisons are mostly correct. Subsequently, I established the optimality of my algorithm and validated its efficacy through experiments.

This work [2] will appear at **NeurIPS 2023**. The proof of my algorithm is elegant enough that I was told by a professor at a workshop that he was considering teaching it at his university. I have also discussed with researchers in applied drug discovery, who suggested that my findings could be applied to reduce the number of comparative experiments necessary for selecting molecular structures among thousands of candidates.

Wasserstein Distributional Adversarial Robustness of Neural Networks Unlike algorithms with provable guarantees, neural networks are vulnerable to worst-case attacks, making them unsafe for real-world applications. Despite the extensive literature on practical attacks and defenses [6,7], finding the right theoretical tools for analyzing their worst-case behavior is still an open challenge. Working with Prof. Jan Obloj at Oxford, we used Optimal Transport to lift the robustness analysis into infinite-dimensional space, then applied Distributional Robust Optimization (DRO) sensitivity analysis to evaluate robustness. As the only team member with an ML background, I worked on translating mathematical concepts into functional algorithms. I proposed a new loss function, Rectified DLR, along with an attack algorithm with certified bounds. Further, I integrated off-policy techniques from RL to mitigate the computational cost of minimax optimization and proposed a robust training algorithm. This work [3] will appear at **NeurIPS 2023**.

Neuro-symbolic control on diffusion models Diffusion models have all the strengths of "System 1 thinking"[1] for image generation tasks: they excel at rendering emotions, styles, and scene diversity. However, they fall short in compositionality, consistency, and generalizability, the hallmarks of a logical system. To address these limitations, I worked with Prof. Jiajun Wu at Stanford to improve the relational compositionality of diffusion models. Diverging from traditional end-to-end diffusion models, I developed a two-stage pipeline using bounding boxes as a bridge and decomposed image generation into the generation of individual objects and relations. This approach achieved state-of-the-art performance in generating multiple objects and multiple relations on synthetic datasets.

Future Directions My research goal is to develop learning systems that combine the reliability of classic algorithms with the seemingly unreasonable power of machine learning. Such systems would be more interpretable and composable than traditional ML models while surpassing classic algorithms in learning abilities.

To start with, several questions immediately pique my interest:

- **How to seamlessly integrate deep learning models with provably correct algorithms?** I believe the key lies in the interface. Previous works in learning-augmented algorithms have modeled predictions with discrete structures for simplicity in complexity analysis. However, this approach overlooks the structure of networks and consequently limits the seamless integration of ML models with algorithms. I aim to design continuous, interactive, bi-directional interfaces tailored between specific network types, such as Transformers and Graph Neural Networks, and classic algorithms.
- **How to leverage algorithmic techniques to enhance ML models?** Classic algorithms offer a wealth of powerful tools – recursion, divide-and-conquer, dynamic programming, etc. Integrating these into learning systems can significantly enhance their logical reasoning capabilities. As a preliminary approach, I plan to develop implicit neural networks, starting from [8] and [9], to embed algorithmic principles within learning systems as specifications. The primary challenge involves finding appropriate settings and robustly training these sophisticated networks.

Stanford is an optimal place for me to pursue my research goals. I am particularly excited about the opportunity to work with **Professor Ellen Vitercik**. I am keen to follow her works on the automatic configuration, selection, and design of algorithms with the help of machine learning. Working with her, I am keen to develop algorithmic pipelines that can self-adapt their hyper-parameters and structures according to the inputs, thereby leveraging hidden structures behind the tasks. Additionally, with shared interests in neural-algorithmic reasoning, I am eager to explore and improve the algorithmic reasoning capabilities of Graph Neural Networks.

I would also like to work with **Professor Percy Liang**. By leveraging algorithmic insights, I want to contribute to his work in understanding and improving the reasoning abilities of Foundation Models and in developing learning systems with strong guarantees of fairness and robustness. Working with him, I am excited about exploring the potential of LLMs in classic algorithmic settings, thereby broadening the impact of foundation models to a broader range of scientific domains.

I am also keen on working with **Professor Greg Valiant**. His emphasis on developing algorithms and theoretical bounds of data-centric tasks aligns well with my academic pursuits. Collaborating with Professor Valiant, I aim to delve deeper into the realms of algorithmic efficiency and robust ML models, exploring how theoretical insights can inform practical, data-driven solutions.

[1] as described in the book *Thinking, Fast and Slow*

References

[1] **Xingjian Bai** and Christian Coester. Sorting with Predictions. *NeurIPS 2023.* [arXiv]

[2] **Xingjian Bai**, Guangyi He, Yifan Jiang, and Jan Obloj. Wasserstein Distributional Robustness of Neural Networks. *NeurIPS 2023.* [arXiv]

[3] **Xingjian Bai** and Luke Melas-Kyriazi. Fixed Point Diffusion Models. *Under review at CVPR 2024.*

[4] Jacek Karwowski, Oliver Hayman, **Xingjian Bai**, Klaus Kiendlhofer, Charlie Griffin, and Joar Skalse. Goodhart's Law in Reinforcement Learning. *Under review at ICLR 2024.* [arXiv]

[5] Thodiris Lykouris and Sergei Vassilvitskii. Competitive Caching with Machine Learned Advice. *ICML 2018.*

[6] Francesco Croce and Matthias Hein. Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks. *ICML 2020.*

[7] Francesco Croce, Maksym Andriushchenko, Vikash Sehwag, Edoardo Debenedetti, Nicolas Flammarion, Mung Chiang, Prateek Mittal, and Matthias Hein. RobustBench: a standardized adversarial robustness benchmark. *NeurIPS Datasets and Benchmarks 2021.*

[8] Shaojie Bai, J. Zico Kolter, Vladlen Koltun. Deep Equilibrium Models. *NeurIPS 2019.*

[9] Fangda Gu, Heng Chang, Wenwu Zhu, Somayeh Sojoudi, and Laurent El Ghaoui. Implicit Graph Neural Networks. *NeurIPS 2020.*