STATEMENT OF PURPOSE

Xingjian Bai   *DPhil 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 theories.

Studying more about machine learning (ML), I was captivated by its strong performance across various real-world applications, despite the absence of worst-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 exact comparisons, if given dirty comparisons are mostly correct. Subsequently, I established the optimality of my algorithm and validated its efficacy through experiments.

This work [1] 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 in his algorithmic course. I have also discussed with researchers in applied drug discovery, who suggested that my findings could be used to reduce the comparative experiments needed 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 neural networks' worst-case behavior is still an open challenge. Working with Prof. Jan Obloj at Oxford, we used Optimal Transport to lift the robustness problem into infinite-dimensional space, then applied tools in Distributional Robust Optimization (DRO) sensitivity analysis. 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 [2] will appear at **NeurIPS 2023**.

Neuro-symbolic control on diffusion models   Generative diffusion models have the strengths of "System 1 thinking"[1]: 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 approaches, I developed a two-stage pipeline using bounding boxes as a bridge and decomposed image generation into the generation of individual objects and relations. On CLEVR, my model achieved the best performance in generating images with multiple objects and relations.

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

To start with, several questions immediately pique my interest:

- **How to design classic algorithms to leverage a spectrum of "dirty" predictions?** In multiple algorithmic and practical scenarios, such as sorting, we often rely on a potentially costly evaluator (e.g., comparator, attacker, or scoring system). ML models could be trained to approximate such evaluations, albeit with errors, resulting in a range of "dirty" predictors with varying accuracy and cost. My goal is to develop algorithms that can utilize this spectrum of dirty predictions to produce provably efficient outputs.

- **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 pipelines 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.

Oxford is an optimal place for me to pursue my research ambitions. It is a place I am already familiar with, and have already had research interactions with many researchers. I am particularly excited to continue working with **Professor Christian Coester**. Building upon our prior work in sorting, I plan to explore other settings where dirty subroutines can be leveraged. I am also keen on learning about the recent development of mirror descent [10, 11] and applying it to various online settings, such as explorable heap selection [14]. Further, I want to tackle NP-hard problems under the setting of algorithms with prediction, identifying predictions that can be realistically obtained by ML models, and then designing algorithms with guarantees to leverage these predictions. I would also like to continue to work with **Professor Christian Rupprecht**, my fourth-year project supervisor. With him, I would work on building machine learning models, possibly in the field of Computer Vision, endowed with nice theoretical properties. Specifically, I aim to further investigate fixed-point implicit neural networks [8, 12, 13] and explore ways to merge algorithmic insights into learning systems.

---

[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.*

[10] Sebastien Bubeck, Michael B. Cohen, Yin Tat Lee, James R. Lee, and Aleksander Madry. K-server via multiscale entropic regularization. *STOC 2018.*

[11] Sebastien Bubeck, Michael B. Cohen, James R. Lee, and Yin Tat Lee. Metrical task systems on trees via mirror descent and unfair gluing. *SIAM J. Comput. 2021.*

[12] Shaojie Bai, Zhengyang Geng, Yash Savani, J. Zico Kolter. Deep Equilibrium Optical Flow Estimation. *CVPR 2022.*

[13] Davis Gilton, Gregory Ongie, Rebecca Willett. Deep Equilibrium Architectures for Inverse Problems in Imaging. *IEEE Trans. Comput. Imaging. 2021.*

[14] Sander Borst, Daniel Dadush, Sophie Huiberts, Danish Kashaev. A Nearly Optimal Randomized Algorithm for Explorable Heap Selection. *IPCO 2023.*