OPTML@Michigan State University
Department of Computer Science & Engineering
428 S. Shaw Lane
East Lansing, MI 48824, United States

Reference Letter

To Whom It May Concern,

I am delighted to write this recommendation letter in support of **Mr. Yihua Zhang's application on the IBM Fellowship Award**. He has been working under my supervision as a Ph.D. student since January 2022, focusing on his thesis titled "Scalable and General Optimization Strategies for Trustworthy Machine Learning: Foundations and Applications", with an anticipated graduation date of August 2026. Yihua is **the best Ph.D. student** that I have mentored/co-mentored in my career (around 50 graduate students from MIT, CMU, UofM, Northeastern, Syracuse, UMN, UT-Austin, RPI, and MSU). This includes my past mentees Dr. Kaidi Xu (IBM Research intern in 2020 and currently an Assistant Professor of CS at Drexel University since 2021), Dr. Ren Wang (RPI-IBM AI Horizon Scholar in 2019-2020 and currently an Assistant Professor of ECE at Illinois Institute of Technology since 2022), and Dr. Tianlong Chen (IBM Fellowship Award recipient in 2021 and currently an Assistant Professor of CS at The University of North Carolina at Chapel Hill since 2024).

1. Summary of Yihua's qualifications and my evaluation

To date, Yihua has published **11 first-authored papers** in top-tier ML/AI conferences, including CVPR, ICML, NeurIPS, ICLR, and ICCV, as well as in IEEE Signal Processing Magazine. Among his many accomplishments, he was recently recognized with the MLCommons Rising Star Award in 2024. He is also the lead student author of the **MSU-IBM collaborative work** titled "Distributed Adversarial Training to Robustify Deep Neural Networks at Scale", which was honored with the **Best Paper Runner-Up Award** at the 38th Conference on Uncertainty in Artificial Intelligence (UAI) in 2022. Additionally, he has been honored with the **Outstanding Reviewer Award** at CVPR 2023-2024 and NeurIPS 2022-2023, as well as the **NeurIPS Scholar Award** in 2022-2023. I offer my **strongest support** for Yihua's application, based on his exceptional academic performance, leadership in collaborative research projects, resilience in overcoming challenges, ability to transform setbacks into successes, proficiency in multitasking, maturity in oral presentations and communication, and outstanding mentoring skills. These qualities, combined with his immense potential, make him a highly promising candidate for a successful faculty in his career plan.

2. Detailed evaluation of Yihua's academic performance

I would like to highlight Yihua's outstanding academic achievements, with a particular focus on his significant contributions to advancing the algorithmic foundations of trustworthy and efficient machine learning. Following this, I will emphasize **five** of his most representative publications, in which he played a *leading role* as the primary author and engaged in *collaborative research with IBM Research*.

First, in [ICML'22] '*Revisiting and Advancing Fast Adversarial Training Through The Lens of Bi-Level Optimization*',Yihua for the first time developed a general and scalable robust training framework by leveraging bi-level optimization (BLO). The significance of his contribution can be justified below. It becomes known that today's deep learning (DL) systems are vulnerable to adversarial attacks. Although developing methods to secure DL against adversaries is now a primary research focus, it suffers from two limitations: lack of optimization generality and lack of optimization scalability. Prior to Yihua's work, nearly all the existing work adopted min-max optimization (MMO) as the algorithmic backbone of robust learning. Yet, the MMO principle requests the defender and the adversary to share the same objective function type, leading to a

poor generality of robust learning against diverse adversarial attacks. By contrast, Yihua showed that different from MMO, BLO enables us to customize its lower-level objectives to incorporate different types of adversaries in DL, and is capable of tackling both generality and efficiency challenges in robust DL. As of August 10th, this work has garnered **74 citations**.

Second, in [NeurIPS'22] '*Advancing Model Pruning via Bi-level Optimization*' (collaborative work with IBM Research Staff Parikshit Ram), Yihua addressed an open question of how to close the gap between pruning accuracy and pruning efficiency in the area of deep model compression. In the literature, model pruning has seen extensive research interest, with the aim to reduce model sizes by removing (or pruning) redundant parameters. Yet, the computation cost of model pruning grows prohibitively as the targeted pruning ratio increases. To co-optimize pruning accuracy and efficiency, Yihua revisited the algorithmic foundation of model pruning through the lens of BLO, which enables us to optimize the coupling between model pruning and model retraining. He developed a new bi-level pruning (BIP) algorithm, built upon the implicit gradient-based optimization theory. Unlike conventional BLO solvers, he theoretically showed that BIP is as efficient as any first-order optimization by taking advantage of the bi-linear nature of the pruning variables. He also showed that BIP can find a model pruning scheme with higher accuracy than the state-of-the-art (SOTA) pruning baselines, and is highly efficient, yielding 2-7$\times$ speedup over the SOTA methods. As of August 10th, this work has garnered **40 citations**.

Third, in [ICLR'23] *What Is Missing in IRM Training and Evaluation? Challenges and Solutions*, a collaborative effort with IBM Research Staff Parikshit Ram and IBM Fellow Kush Varshney, Yihua tackled key challenges in invariant risk minimization (IRM). This work, which addresses the need for environment-agnostic data representations to avoid spurious correlations and improve out-of-distribution generalization, identified and resolved three major limitations in IRM. First, he highlighted the overlooked impact of batch size in IRM training, suggesting improvements through batch-aware optimization. Second, he addressed the risk of false invariance due to improper evaluation environments by introducing diversified test-time environments. Lastly, he proposed a novel variant of IRM to overcome the limitations of ensemble methods, framing IRM as consensus-constrained bi-level optimization.

Fourth, in [ICML'24] *Revisiting Zeroth-Order Optimization for Memory-Efficient LLM Fine-Tuning: A Benchmark*, a collaboration with IBM Principal Research Scientist Pin-Yu Chen, Yihua advocated for the use of BP-free, zeroth-order (ZO) optimization to reduce memory costs during large language model fine-tuning. Through a comprehensive benchmarking study across five LLM families (Roberta, OPT, LLaMA, Vicuna, Mistral) and five fine-tuning schemes, he identified overlooked ZO optimization principles, including the importance of task alignment, the role of forward gradient methods, and the balance between algorithm complexity and performance. He also introduced novel ZO enhancements, such as block-wise descent, hybrid training, and gradient sparsity, offering a promising approach for memory-efficient fine-tuning of foundation models. The ZO-LLM benchmark he established has made a significant impact in the field.

Lastly, the ICCV 2023 Oral paper *Robust Mixture-of-Expert Training for Convolutional Neural Networks* showcases Yihua's ability to transform setbacks into valuable insights and extract new knowledge from challenging situations. In this groundbreaking work, he introduced the first adversarial training framework specifically designed for mixture-of-expert (MoE) models. He demonstrated that the conventional adversarial training (AT) mechanisms, originally developed for standard CNNs, are ineffective in enhancing the robustness of MoE models. To address this, he dissected MoE robustness into two dimensions: the robustness of routers and the robustness of experts. His analysis revealed that routers and experts struggle to adapt to each other under traditional AT, leading to the development of a novel router-expert alternating adversarial training framework for MoE, termed AdvMoE. This paper exemplifies Yihua's journey from initial challenges to a successful and innovative solution, highlighting his ability to carefully analyze and leverage setbacks to create high-impact, widely recognized research.

3. Evaluation of Yihua's leadership and additional skills

Expanding beyond his academic prowess, Yihua demonstrates exceptional leadership qualities within all collaborative research endeavors he engages in. He exhibits a keen focus and establishes clear, actionable objectives to augment team efficiency and concentration. Simultaneously, his aptitude as an effective communicator fosters the development of collaborative pathways for meetings and research endeavors. Additionally,

he embodies traits of humility and patience, which empower him to acknowledge mistakes, treat collaborators with respect, and exhibit considerable patience throughout the collaborative process.

First, Yihua's skilled management has guaranteed that this expansive collaboration culminates in collective success, leaving all participants satisfied with the outcomes. Due to his leadership skills, I invited him as a student chair of the ICML'22 & 23 and NeurIPS'24 Workshop on *New Frontiers in Adversarial Machine Learning (AdvML-Frontiers)*. He did an excellent job in communicating and collaborating with other workshop organizers and participants in the website design and setup, online forum building, publicity, and publication. Likewise, during the collaboration with IBM Research on the UAI'22 publication '*Distributed Adversarial Training to Robustify Deep Neural Networks at Scale*', Yihua was the student leader and responsible for coordinating meetings/discussion with IBM on algorithm implementation, experimentation, and authors' rebuttal. His efforts have made this work the Best Paper Runner-Up Award at UAI'22, a big surprise to everyone.

Second, throughout Yihua's PhD studies, his resilience and problem-solving acumen have consistently stood out. An exemplary instance of this occurred during the development of his [NeurIPS'23] work '*Selectivity Drives Productivity: Efficient Dataset Pruning for Enhanced Transfer Learning*'. This project languished for over six months with little progress until Yihua joined the team and took the leadership. With his exceptional coding skills and sharp analytical abilities, he swiftly advanced the project. Especially noteworthy is his ability to maintain composure and clarity under pressure, particularly in moments when experiments deviated from expectations. His adeptness at critical thinking during crises has proven instrumental to his overall success. Furthermore, Yihua has demonstrated an outstanding ability to adapt to new challenges, swiftly mastering novel techniques and seamlessly integrating them into his work. This is evident in his latest preprint paper on dataset and benchmark, titled *UnlearnCanvas: A Stylized Image Dataset to Benchmark Machine Unlearning for Diffusion Models*. For this project, he single-handedly spearheaded the collection of the entire dataset, crafted innovative evaluation frameworks, and devised pipelines for machine unlearning and style transfer. Additionally, he implemented over ten different methods, showcasing his exceptional adaptability and unwavering determination to surmount obstacles.

Third, one aspect that truly sets Yihua apart is his exceptional ability to juggle multiple projects simultaneously. He consistently advances multiple projects concurrently, ensuring steady progress across each one. This multitasking skill not only highlights his superior efficiency but also reflects his remarkable mental maturity, contributing to an extraordinarily productive first three years of his PhD studies. During this period, he has authored *11 first-authored papers* and contributed to over *20 papers* in total. Additionally, he has played key roles in *2 tutorials* and co-organized *3 workshops* at top-tier conferences. Furthermore, Yihua has demonstrated exemplary presentation and communication skills. He has delivered numerous oral presentations at leading AI/ML conferences and was invited to give guest talks on 'Bi-Level Optimization for Robust ML' at the INFORMS Annual Meeting 2022, as well as at CS@UCSB and ECE@UMN. At AAAI 2023, he co-presented a highly successful tutorial titled *Bi-level Optimization in Machine Learning: Foundations and Applications*, further underscoring his expertise in presentations.

Lastly, it is important to underscore that, although only in his third year, Yihua has already demonstrated considerable potential to become a successful faculty member, which aligns with his career aspirations. He has adeptly mentored 9 undergraduate and graduate students during their summer internships, guiding them to produce research papers that bolster their academic trajectories. Remarkably, each of these mentees has leveraged their research experience under Yihua's mentorship to secure PhD positions at my research lab and other globally renowned universities. A specific highlight from the summer of 2023 is one of Yihua's mentees who achieved a spotlight paper at ICLR'24 (*SalUn: Empowering Machine Unlearning via Gradient-based Weight Saliency in Both Image Classification and Generation*). This accomplishment stands as a testament to Yihua's patient, thorough, and dedicated guidance, including every aspect of the research process from conceptualization to execution. His mentorship not only sharpens his mentees' research skills but also strategically positions them for successful academic careers, underscoring his potential as a rising leader in academia.

In summary, **Mr. Yihua Zhang** is the most exceptional Ph.D. student I have mentored, undeniably **a rising star** in the field of trustworthy and efficient machine learning, as evidenced by his receipt of the

MLCommons Rising Star Award in 2024. His unique strengths make him an ideal candidate for the IBM Fellowship, particularly due to his established and ongoing collaborations with IBM. This fellowship would not only support his groundbreaking research but also alleviate financial pressures during his Ph.D. studies, especially as he supports his family, including a newborn. I strongly endorse **Mr. Yihua Zhang** for this prestigious award and am confident he will continue to make significant contributions to the field. If you require any further information, please feel free to contact me.

Sincerely,

Sijia Liu
Assistant Professor, CSE@Michigan State University
Website: `https://lsjxjtu.github.io`
Email: liusiji5@msu.edu