UNIVERSITY OF CHICAGO

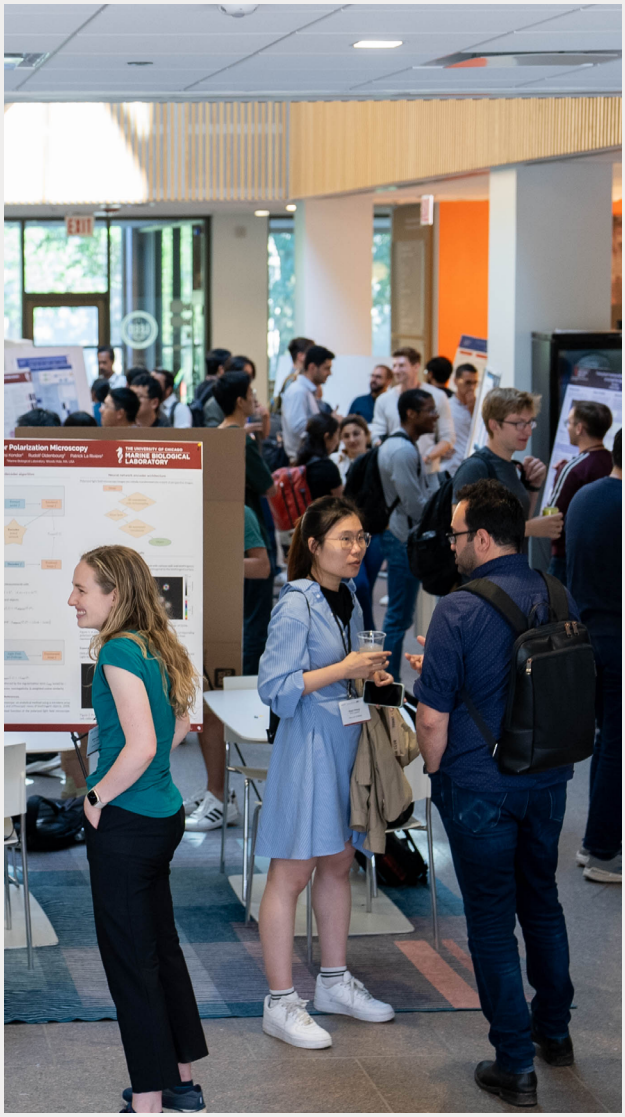
JUNE 23 - 24, 2025

MIDWEST MACHINE LEARNING SYMPOSIUM 2025

UNIVERSITY OF CHICAGO,
LOGAN CENTER FOR THE ARTS
915 E. 60TH ST.
CHICAGO, IL

MIDWEST-ML.ORG/2025

SUMMARY



WELCOME TO THE 2025 MIDWEST MACHINE LEARNING SYMPOSIUM!

By joining us for MMLS25, you are part of the growing community of machine learning researchers who are changing the field and at the forefront of cutting-edge technologies and research. This year's symposium brings together machine learning researchers from across the Midwest to engage in vibrant discussions, forge new collaborations, and showcase innovative work at every career level. The event truly reflects the collective leadership and regional spirit of the Midwest ML community.

Over the next two days, you'll experience a thoughtfully curated program of expert speakers, panel discussions, poster sessions and more spanning trustworthy ML, deep learning theory, generative AI, and AI applications in science & robotics. **We invite all attendees to engage with each other, our notable speakers, generous sponsors, and industry partners throughout the event.**

Whether you're presenting new findings, considering cross-institutional collaborations, or simply exploring fresh ideas, we hope this symposium energizes you and strengthens our Midwest ML network.

Welcome—you're part of something special!

PLATINUM SPONSORS

Schmidt Sciences

THE UNIVERSITY OF CHICAGO
DEPARTMENT OF COMPUTER SCIENCE

THE UNIVERSITY OF CHICAGO
DATA SCIENCE INSTITUTE

GOLD SPONSORS

TOYOTA TECHNOLOGICAL INSTITUTE AT CHICAGO

THE UNIVERSITY OF CHICAGO
DEPARTMENT OF STATISTICS

BRONZE SPONSORS

CHICAGO BOOTH
THE UNIVERSITY OF CHICAGO BOOTH SCHOOL OF BUSINESS

Center for Applied Artificial Intelligence

Data Science Institute
UNIVERSITY OF WISCONSIN–MADISON

Invenergy

millennium

Promega

MEAL/BREAK SPONSORS

CHICAGO BOOTH
THE UNIVERSITY OF CHICAGO BOOTH SCHOOL OF BUSINESS

Center for Applied Artificial Intelligence

MICHIGAN STATE UNIVERSITY

DAY 1 AGENDA

8 A.M
Registration & Breakfast

9 A.M
Opening Remarks

Chenhao Tan, MMLS25 Co-Chair, Associate Professor, Computer Scienc and Data Science, UChicago

Haifeng Xu, MMLS25 Co-Chair, Assistant Professor, Computer Science and Data Science, UChicago

9:15 A.M
Plenary:
Societal Impact and the Ivory Tower: an Adversarial ML Perspective

Ben Zhao, Neubauer Professor of Computer Science at UChicago
Time Magazine's "The 100 Most Influential People in AI" (2024)

10:15 A.M
Lightning Talks

AI Reasoning, Agents, and Model Optimization

10:45 A.M
Break

11 A.M
Trustworthy Machine Learning

Zahra Ghodsi, Assistant Professor of Electrical and Computer Engineering, Purdue:
Collaborating with Confidence: Securing Federated Learning Systems

Sijia Liu, Assistant Professor, Department of Computer Science and Engineering, MSU:
Robust Unlearning for LLMs

Han Zhao, Assistant Professor, Department of Electrical and Computer Engineering, UIUC:
Revisiting Scalarization in Multi-Task Learning

12:15 P.M
Lunch + Poster Session A

1:45 P.M
Deep Learning Theory and Optimization

Wei Hu, Assistant Professor, EECS, Computer Science and Engineering, University of Michigan:
Abrupt Learning in Transformers

Frederic Koehler, Assistant Professor, Statistics and Data Science, UChicago:
On Inductive Bias in Generative Modeling

Tianhao Wang, Research Assistant Professor, TTIC:
Structured Preconditioners in Adaptive Optimization: A Unified Analysis

3 P.M
Break

3:30 P.M
Plenary:
General Search Techniques Without Common Knowledge for Imperfect-Information Games, and Application to Superhuman Fog of War chess

Tuomas Sandholm, Angel Jordan University Professor of Computer Science at CMU
Co-director of CMU AI
and serial entrepreneur

4:30 P.M
AI2050 Special Session

Chaowei Xiao, Assistant Professor, University of Wisconsin, Madison

Huan Zhang, Assistant Professor, Department of Electrical and Computer Engineering, University of Illinois at Urbana Champaign

Haifeng Xu, Assistant Professor, Computer Science and Data Science, University of Chicago

5:30 P.M
Reception + Poster Session B

DAY 2 AGENDA

8:30 A.M
Breakfast

9 A.M
Plenary:
mCLM:
A Function-Infused and Synthesis-Friendly Modular Chemical Language Model

Heng Ji, Professor of
Computer Science at UIUC
Founding Director of Amazon-Illinois Center on AI

10 A.M
Lightning Talks

Machine Learning
Foundations
and Applications

10:30 A.M
Break

10:45 A.M
AI for Science/
Engineering/ Robotics

Mengxue Hou, Assistant Professor,
Electrical Engineering, Notre Dame:
Assured Neural-Symbolic Abstraction for Hierarchical Robotic Planning

Yiping Lu, Assistant Professor of
Industrial Engineering and
Management Sciences, Northwestern:
Two Tales, One Resolution: Physics-Informed Inference Time Scaling and Precondition

Yexiang Xue, Assistant Professor,
Computer Science, Purdue:
Embedding Automated Reasoning into Neural Generation

12 P.M
Lunch + Poster
Session C

1 P.M
Plenary:
LLM Skills and Metacognition: Scaffolding for New Forms of Learning?

Sanjeev Arora, Charles C. Fitzmorris
Professor of Computer Science
at Princeton
Director of Princeton Language and Intelligence

2 P.M
Break

2:15 P.M
Generative AI

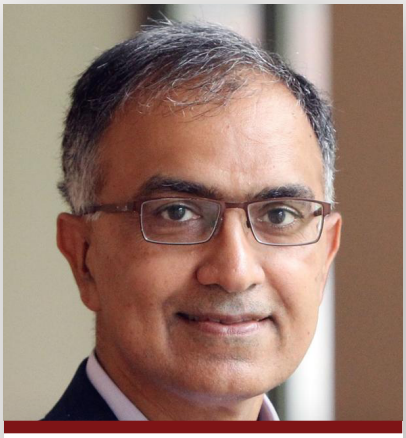
Ari Holtzman, Assistant Professor,
Computer Science and
Data Science, UChicago:
Articulating the Ineffable: What We Can't Yet (Define/Express) About (LLMs/Ourselves)

Zirui Liu, Assistant Professor,
Department of Computer Science &
Engineering, University of Minnesota:
Massive Outlier Values in LLMs: Engineering and Science

Ruqi Zhang, Assistant Professor of
Computer Science, Purdue:
Toward Capable and Reliable LLMs via Probabilistic Modeling

3:30 P.M
Closing Remarks

PLENARIES



SANJEEV ARORA
Charles C. Fitzmorris Professor of Computer Science at Princeton

Director of Princeton Language and Intelligence

LLM Skills and Metacognition: Scaffolding for New Forms of Learning?

Abstract: LLMs, especially their recent "reasoning" incarnations, are capable of impressive problem solving. This talk will argue that a key role in this success is their "metacognition" capabilities ("thinking about thinking"), which we find arise spontaneously in LLMs. We'll give diverse examples of such metacognition and argue that it gives insight into how LLM training gives rise to complex capabilities, as well as how these capabilities may be enhanced in future. We will also introduce "Concept-enhanced learning", a simple setting that gives a hint about how LLM metacognition itself may emerge.

HENG JI
Professor of Computer Science at UIUC

Founding Director of Amazon-Illinois Center on AI

mCLM: A Function-Infused and Synthesis-Friendly Modular Chemical Language Model

Abstract: Everything in our wonderful world is composed of molecules. Recent advances in block-based chemistry involve the manual design of drugs and materials by decomposing molecules into building blocks—i.e., functional modules—and reassembling them into new molecules with desired functions. However, the process of discovering and manufacturing functional molecules has remained slow, expensive, and highly specialist-dependent. In this talk I will present our recent efforts at teaching computers to speak two complementary languages: one that represents molecular subgraph structures indicative of specific functions, and another that describes these functions in natural language. Unlike existing approaches that add such knowledge as a post hoc step, we have developed a tiny prototype (1B parameters) function- and synthesis-aware modular chemical language model (mCLM) which has proven to outperform ChatGPT on discovering new drugs with better functions.

PLENARIES

TUOMAS SANDHOLM
Angel Jordan University Professor of Computer Science at CMU

Co-director of CMU AI and serial entrepreneur

General Search Techniques Without Common Knowledge for Imperfect-Information Games, and Application to Superhuman Fog of War Chess

Abstract: Since the advent of AI, games have served as progress benchmarks, and most real-world settings are imperfect-information games. Meanwhile, imperfect-information variants of chess have existed for over a century, present extreme challenges, and have been the focus of significant AI research. Beyond calculation needed in egular chess, they require reasoning about information gathering, the opponent's knowledge, signaling, bluffing, etc. The most popular variant, Fog of War (FoW) chess (aka. dark chess) is a recognized challenge problem in AI after superhuman performance was reached in no-limit Texas hold'em poker. We present Obscuro, the first superhuman AI for FoW chess. It introduces advances to search in imperfect-information games, enabling strong, scalable reasoning. Most prior search techniques - such as those used to achiev superhuman play in no-limit Texas hold'em - require the construction of the "common knowledge set" as a first step, making them unusable for games with this much imperfect information. Experiments against the prior state-of-the-art AI and human players - including the world's best - show that Obscuro is significantly stronger. FoW chess is now the largest (by amount of imperfect information) turn-based game in which superhuman performance has been achieved and the largest game in which imperfect-information search has been successfully applied. This is joint work with my PhD student Brian Hu Zhang.

BEN ZHAO
Neubauer Professor of Computer Science at UChicago

Time Magazine's "The 100 Most Influential People in AI" (2024)

Societal Impact and the Ivory Tower: An Adversarial ML Perspective

Abstract: It is undeniable that computing research has the power to rapidly reshape the world we live in, and ML is literally proving this point in real time. But it is also true that we often are not aware or cognizant of the positive and negative impacts of our work. In this talk, I argue that we as researchers need to be more accountable for not just our research results, but how they may be used in downstream applications. Recognizing such impacts is arguably a very challenging task itself. Using my own experience in recent adversarial ML projects, I describe the duality of ML's impact today, both in real harms it has produced via misuse, and in protective benefits it can provide. I share some of the ethical questions we faced when considering the design and deployment of our tools Glaze and Nightshade, and our experiences through this process. Finally, I suggest some takeaways, including possible perspectives on bevaluating new research directions, as well as some concrete research questions that offer potential for positive technical and societal impact.

INVITED SPEAKERS

Trustworthy Machine Learning

Zahra Ghodsi, Assistant Professor of Electrical and Computer Engineering, Purdue
Talk Title: Collaborating with Confidence: Securing Federated Learning Systems
Abstract: Artificial Intelligence (AI) is increasingly implemented in distributed settings thanks to its ability to process large amounts of data and its power to enable a wide range of applications. Networks of intelligent devices can therefore work collaboratively to facilitate new directions in several domains such as distributed healthcare and transportation. Deploying AI successfully in the distributed or federated setting requires collaboration of a large number of devices which belong to different parties. This collaboration, however, raises security concerns relating to privacy of assets and robustness in the presence of accidental or intentional errors. In this talk, I outline the challenges in developing secure and privacy-preserving federated learning frameworks where the data or even the identity of participants can be sensitive. I highlight the need for designing new holistic solutions where requirements such as privacy and robustness must be simultaneously guaranteed. I conclude by briefly discussing the lessons learned and future research directions.

Sijia Liu, Assistant Professor, Department of Computer Science and Engineering, MSU
Talk Title: Robust Unlearning for LLMs
Abstract: As generative AI systems continue to evolve, the ability to selectively remove information from trained models, known as machine unlearning, has become increasingly essential for ensuring regulatory compliance, enforcing ethical constraints, and mitigating the retention of harmful or sensitive content. This talk focuses on a pressing challenge in this space: the robustness of unlearning in large language models (LLMs). We examine how current unlearning methods remain vulnerable to relearning attacks and post-unlearning fine-tuning, where previously removed knowledge can be partially recovered from a small subset of forgotten or auxiliary data. From an optimization perspective, we introduce a novel connection between robust unlearning and sharpness-aware minimization (SAM), showing that promoting flatter loss landscapes through smoothness-based optimization enhances a model's resistance to relearning. This draws a natural parallel to principles from adversarial robustness. The talk concludes with a discussion of open challenges and future directions for embedding unlearning into the AI lifecycle, ensuring long-term safety, compliance, and trustworthiness across the data, model, and optimization stack.

Han Zhao, Assistant Professor, Department of Electrical and Computer Engineering, UIUC
Talk Title: Revisiting Scalarization in Multi-Task Learning
Abstract: Linear scalarization, i.e., combining all loss functions by a weighted sum, has been the default choice in the literature of multi-task learning (MTL) since its inception. In recent years, there has been a surge of interest in developing Specialized Multi-Task Optimizers (SMTOs) that treat MTL as a multi-objective optimization problem. However, it remains open whether there is a fundamental advantage of SMTOs over scalarization. In this talk, I will revisit scalarization from a theoretical perspective. I will be focusing on linear MTL models and studying whether scalarization is capable of fully exploring the Pareto front. Our findings reveal that, in contrast to recent works that claimed empirical advantages of scalarization, when the model is under-parametrized, scalarization is inherently incapable of full exploration, especially for those Pareto optimal solutions that strike the balanced trade-offs between multiple tasks. I will conclude the talk by briefly discussing the extension of our results to general nonlinear neural networks and our recent work on using online Chebyshev scalarization to controllably steer the search of Pareto optimal solutions.

INVITED SPEAKERS

Deep Learning Theory and Optimization

Wei Hu, Assistant Professor, EECS, Computer Science and Engineering, University of Michigan
Talk Title: Abrupt Learning in Transformers
Abstract: Training Transformers on algorithmic tasks frequently exhibits an intriguing "abrupt learning" phenomenon in their training dynamics: an extended performance plateau followed by a sudden, sharp improvement. In this talk, I will present several empirical observations aiming to uncover universal characteristics and underlying mechanisms behind such dynamics.

Frederic Koehler, Assistant Professor, Statistics and Data Science, UChicago
Talk Title: On Inductive Bias in Generative Modeling
Abstract: There has been a lot of work on understanding the inductive bias of learning via gradient descent and related algorithms. For example, many fascinating phenomena have been discovered in supervised settings such as linearized neural networks, matrix factorization, logistic regression, etc. There are, relatively speaking, fewer such examples which have been worked out in the case of generative modeling and density estimation. I will discuss one such example where we were able to rigorously analyze --- for variational autoencoders --- and the role that the data distribution plays in this setting.

Tianhao Wang, Research Assistant Professor, TTIC
Talk Title: Structured Preconditioners in Adaptive Optimization: A Unified Analysis
Abstract: We present a novel unified analysis for a broad class of adaptive optimization algorithms with structured (e.g., layerwise, diagonal, and kronecker-factored) preconditioners for both online regret minimization and offline convex optimization. Our analysis not only provides matching rate to several important structured preconditioned algorithms including diagonal AdaGrad, full-matrix AdaGrad, and AdaGrad-Norm, but also gives an improved convergence rate for a one-sided variant of Shampoo over that of original Shampoo. Interestingly, more structured preconditioners (e.g., diagonal Adagrad, AdaGrad-Norm which use less space and compute) are often presented as computationally efficient approximations to full-matrix Adagrad, aiming for improved optimization performance through better approximations. Our unified analysis challenges this prevailing view and reveals, perhaps surprisingly, that more structured preconditioners, despite using less space and computation per step, can outperform their less structured counterparts. To demonstrate this, we show that one-sided Shampoo, which is relatively much cheaper than full-matrix AdaGrad could outperform it both theoretically and experimentally.

INVITED SPEAKERS

AI for Science/Engineering/ Robotics

Mengxue Hou, Assistant Professor, Electrical Engineering, Notre Dame
Talk Title: Assured Neural-Symbolic Abstraction for Hierarchical Robotic Planning
Abstract: To enable a smart and autonomous system to be cognizant, taskable, and adaptive in exploring an unknown and unstructured environment, robotic decision-making relies on learning a parameterized knowledge representation. However, one fundamental challenge in deriving the parameterized representation is the undesirable trade-off between computation efficiency and model fidelity. This talk addresses this challenge in the context of underwater vehicle navigation in unknown marine environments. To improve fidelity of the reduced-order model, we develop a learning method to generate a non-Markovian reduced-order representation of the environmental dynamics. Such abstraction guarantees to improve the modeling accuracy. Further, taking advantage of the abstracted model, we develop a Large-Language-Model-guided hierarchical planner to translate human specified missions directly to a set of executable actions with low computation cost.

Yiping Lu, Assistant Professor of Industrial Engineering and Management Sciences, Northwestern
Talk Title: Two Tales, One Resolution: Physics-Informed Inference Time Scaling and Precondition
Abstract: In this talk, I will introduce a novel framework for physics-informed debiasing of machine learning estimators, which we call Simulation-Calibrated Scientific Machine Learning (SCaSML). This approach leverages the structure of physical models to achieve three key objectives:
- Unbiased Predictions: It produces unbiased predictions even when the underlying machine learning predictor is biased.
- Overcoming Dimensionality Challenges: It mitigates the curse of dimensionality that often affects high-dimensional estimators.
- Inference Time Scaling: Improve the machine learning estimation by allocating inference time computation.
The SCaSML paradigm integrates a (potentially) biased machine learning algorithm with a de-biasing procedure that is rigorously designed using numerical analysis and stochastic simulation. We dynamically refines and debiases the SCiML predictions during inference by enforcing the physical laws. Our methodology aligns with recent advances in inference-time computation—similar to those seen in the large language model literature—demonstrating that additional computation can enhance ML estimates.
Furthermore, we establish a surprising equivalence between our framework and another research direction that utilizes approximate (linearized) solvers to precondition iterative methods. This connection not only bridges two distinct areas of study but also offers new insights and algorithms into improving estimation accuracy in physics-informed machine learning settings.

Yexiang Xue, Assistant Professor, Computer Science, Purdue
Talk Title: Embedding Automated Reasoning into Neural Generation
Abstract: Automated reasoning and machine learning are two fundamental pillars of artificial intelligence. Many real-world applications are beyond reach when reasoning or learning are applied in isolation. Reasoning without learning leads to rigid and brittle formulations, while learning without reasoning produces suboptimal models violating critical constraints, hallucinating, and behaving unexpectedly in unseen situations. This talk introduces Spatial Reasoning Integrated Generator (SPRING) for design generation. SPRING embeds a neural and symbolic integrated spatial reasoning module inside the deep generative network. The spatial reasoning module samples the set of locations of objects to be generated from a backtrack-free distribution, guaranteed to satisfy user specifications while capturing subtle utility and aesthetics. SPRING offers interpretability, allowing users to visualize and diagnose the generation process through visualizing the predictions of neural networks. SPRING is also adept at managing novel user specifications, thanks to its proficiency in zero-shot constraint transfer. SPRING is supported by our recently defined Contextual Analog Logic with Multimodality (CALM), in which predicates have analog truth values to capture subtle human preferences. CALM is grounded in multimodal environments (texts and images) with the aid of neural networks, while classic logic requires explicit efinition of symbolic representations and their groundings, which can be ad-hoc, brittle, and unscalable.

INVITED SPEAKERS

Generative AI

Ari Holtzman, Assistant Professor, Computer Science and Data Science, UChicago
Talk Title: Articulating the Ineffable: What We Can't Yet (Define/Express) About (LLMs/Ourselves)
Abstract: One of the most frustrating parts about trying to work with deep generative models is that we are often unable to satisfactorily define what they are doing and how they do it. What do models consistently miss? What do they consistently believe? How do they store new information? In addition to current concrete studies, I will make the case that LLM systems can and should be used to future-proof humans against the influence of increasingly persuasive LLMs. By helping us articulate ideas that express our deeply held individual intuitions, machine-assisted expression can help us make humans less manipulable—and helps us know ourselves better.

Zirui Liu, Assistant Professor, Department of Computer Science & Engineering, University of Minnesota
Talk Title: Massive Outlier Values in LLMs: Engineering and Science
Abstract: Deploying LLMs for long context processing and long generation scenarios are major challenges in LLM serving. A variety of compression techniques have been proposed like quantization, token eviction, and linear-attention models. However, our understanding of how LLMs internally process information is still limited. In this talk, I will highlight one widely existing but under-discussed observation: the abnormal distribution of massive outlier values in the Key and Value token embeddings within self-attention modules. We show how these extreme values are closely tied to context processing and demonstrate ways to leverage them for more efficient computation.
On the engineering side, I'll introduce our work on 2-bit KV cache quantization, which significantly improves both memory usage and inference throughput. On the scientific side, I'll discuss our new findings on the role these extreme values play in shaping model behavior.

Ruqi Zhang, Assistant Professor of Computer Science, Purdue
Talk Title: Toward Capable and Reliable LLMs via Probabilistic Modeling
Abstract: As large language models (LLMs) are increasingly deployed in complex and high-stakes applications, advancing their capabilities and reliability is essential. In this talk, I will explore how probabilistic modeling provides principled and effective approaches for moving toward more capable and reliable LLMs, with a focus on reasoning, alignment, and safety.

First, I will explore how self-correction—viewed as modeling the probabilistic relationship between initial and revised reasoning paths—can serve as a powerful strategy for improving LLM reasoning, even with limited annotated data. Next, I will introduce a framework that casts LLM alignment as a problem of probabilistic inference, and present two discrete sampling techniques for efficient inference. Finally, I will show how variational inference can be used to automatically uncover diverse adversarial inputs, providing a comprehensive, distributional characterization of model vulnerabilities.

On behalf of the entire MMLS 2025 Organizing Committee, thank you for being part of this year's symposium. We're incredibly grateful for your presence, participation, and the energy you brought to these two days of learning, connection, and collaboration.

Your engagement—through talks, posters, questions, hallway conversations, and new partnerships—is what makes MMLS more than just a symposium. It's a community. We hope the ideas you encountered and the connections you made here continue to inspire your work long after you leave. Until next year!

— MMLS 2025 Organizing Committee

Haifeng Xu (Co-chair, UChicago) | Chenhao Tan (Co-chair, UChicago) | Ce Zhang (UChicago) | Zhiyuan Li (TTIC) | Ruqi Zhang (Purdue) | Ren Wang (IIT) | Emma Alexander (Northwestern) | Chaowei Xiao (UW Wisconsin)

CONTACT INFORMATION
EMAIL: data-science@uchicago.edu
PHONE: 773-702-7612

UNIVERSITY OF CHICAGO
5801 S. ELLIS AVE.
CHICAGO, IL
60637

MIDWEST-ML.ORG/2025