

DISTINGUISHED LECTURE SERIES

Thursday, October 30
Cleary Alumni & Friends Center



SCHEDULE OF EVENTS

10:30 a.m. **Registration** | Cleary Alumni & Friends Center | UWL Campus

11 a.m. **SYMPOSIUM**

Building Embodied Autonomous Agents

In this talk I will show how we can design modular agents for visual navigation that can perform tasks specified by natural language instructions, perform efficient exploration and long-term planning, build and utilize 3D semantic maps, while generalizing across domains and tasks. I will first introduce a novel framework, Self-supervised Embodied Active Learning (SEAL), which learns both action and perception in a fully self-supervised manner by building and leveraging 3D semantic maps. I will show how SEAL can be used to close the action-perception loop: By actively moving in training environments, agents improve the object detection and instance segmentation performance of a pretrained perception model; in turn, the improved perception enhances performance on object-goal navigation tasks. Next, I will present a new approach to Embodied Instruction Following, which uses structured representations to construct a semantic map of the scene and perform exploration guided by a semantic search policy. I will demonstrate that incorporating explicit spatial memory and semantic search provides stronger and more general representations for state tracking and goal-directed behavior, even without expert trajectories or low-level instructions.

4:30 p.m. **Registration** | Cleary Alumni & Friends Center | UWL Campus

5 p.m. **KEYNOTE**

Towards Internet-Scale Training for Multimodal AI Agents

The rapid progress of Large Language Models (LLMs) has opened the door to building language-guided agents that can carry out complex, multi-step tasks on behalf of users, much like human assistants. Developing agents that can perceive, plan, and act autonomously has long been a central goal of artificial intelligence research. In this talk, I will introduce multimodal AI agents capable of planning, reasoning, and executing actions on the web. These agents not only comprehend text but can also navigate and interact effectively in visual environments. I will present VisualWebArena, a novel framework for evaluating multimodal autonomous language agents, and describe an inference-time search algorithm that enables agents to explicitly explore and perform multi-step planning in interactive web settings. Finally, I will demonstrate how we can build an automated data pipeline for Internet-scale training of such agents. This pipeline generates web navigation tasks across 150,000 live websites, executes LLM agents on them, and automatically evaluates their success.

Russ Salakhutdinov

Earned his PhD in computer science from the University of Toronto, where he was advised by the Nobel Laureate Geoffrey Hinton. After spending two post-doctoral years at MIT, he joined the University of Toronto and later moved to CMU. He also served as a director of AI research at Apple and is currently serving as a VP of Research at Meta. Russ's primary interests lie in deep learning, machine learning, and generative AI. He is an action editor of the Journal of Machine Learning Research, served on the senior program committee of several top-tier machine learning conferences including NeurIPS, ICLR, and ICML, was a program co-chair for ICML 2019 and general chair for ICML 2024. He has authored over 250 research papers and his work has received over 200,000 citations according to Google Scholar. He is an Alfred P. Sloan Research Fellow, Microsoft Research Faculty Fellow, a recipient of the Early Researcher Award, Google Faculty Award, and Nvidia's Pioneers of AI award.

For further information about the lecture contact:

Kenny Hunt, Ph.D.

Computer Science Department
UW-La Crosse
1725 State St.
La Crosse, WI 54601
608.785.6822
khunt@uwlax.edu