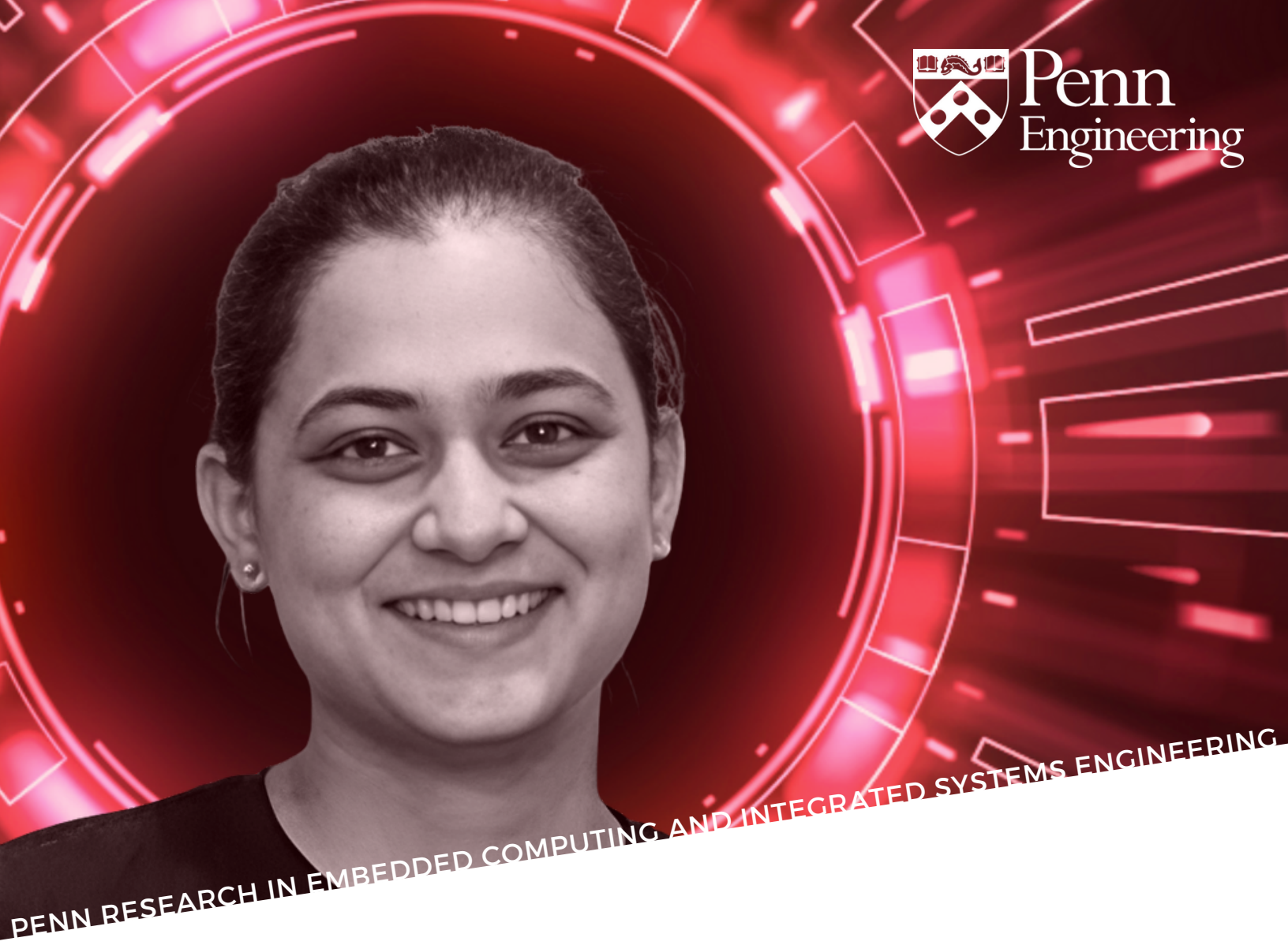


Penn Engineering

PENN RESEARCH IN EMBEDDED COMPUTING AND INTEGRATED SYSTEMS ENGINEERING

IN THE SPOTLIGHT

6 NOV 2023

RAMNEET KAUR

BUILDING MORE ACCURATE DEEP LEARNING SYSTEMS

Just like humans, artificial intelligence systems can make mistakes.

Research by PRECISE Center research scholar Ramneet Kaur aims to better detect scenarios where errors are likely, leading to more accurate performance by safety-critical cyber-physical systems.

After earning her Ph.D. from Penn Engineering in December 2023, Kaur, who is advised by PRECISE Center Director Insup Lee and Professor Oleg Sokolsky, will begin working as an Advanced Research Scientist at the Stanford Research Institute in Menlo Park, California.

“Developing machine learning models that can operate in open world, including detecting, and adapting to novelty, is a critical goal on the path to building intelligent systems that can work alongside humans to solve complex problems while being reliable enough to handle the unexpected,” Kaur said. “My work on Out-of-Distribution (OOD)/adversarial detection; predicting OOD performance of machine learning models in novel environments, and monitoring for the system-level failures in OOD scenarios is a step towards open-world learning for safe deployment of AI in the real-world.”

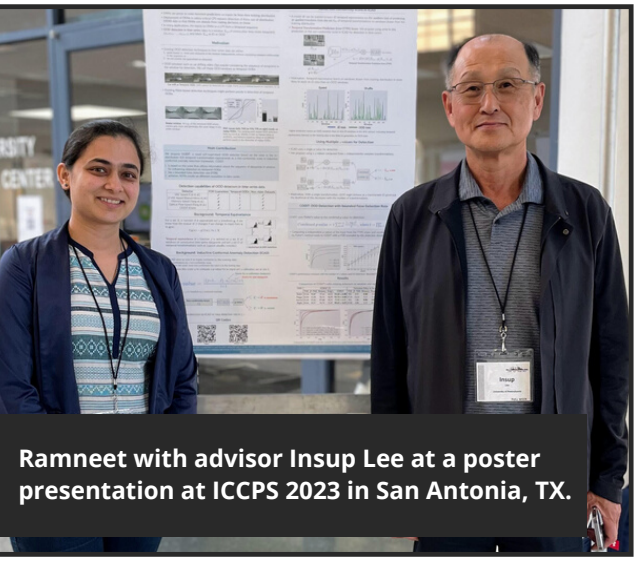
Ramneet with her husband at Wawa's 4th of July concert in 2021 (Fairmout Park, Philadelphia PA)

Kaur earned her bachelor’s degree at the College of Engineering Roorkee and her master’s degree at the Indian Institute of Technology, India. Her research interests include artificial intelligence, trustworthy machine learning, interpretability and fairness, AI alignment, resilient deep learning, open world learning, reinforcement learning and formal methods in the areas of cyber-physical systems and the Internet of Medical Things. Kaur's research has been published in top-tier cyber-physical systems and machine-learning conferences and journals, including HSCC, ICCPS, ICAA, TCPS, and AAAI. She won the best paper award in ICCPS, 2022, and best paper candidate award in ICCPS, 2023. Her ICCPS papers have been also rewarded with the Reusable and Reproducible Badge. She has served as the reviewer for multiple conferences and journals, including TCPS, ICCPS, Safecomp, RV, AAAI, and, Neurips. She is also the program committee chair of DESTION, IEEE Workshop on Design Automation for CPS and IoT. Recently, Kaur was selected as one of the CPS Rising Stars, 2023 by the Link Lab at the University of Virginia.

One of her ongoing research projects is focused on predicting safety under data distribution shifts, i.e. when machines receive data that differs from the information used to train the model.

Sometimes, machine learning components can continue to work properly despite such shifts; for example, a system designed to classify traffic signs in clear weather may be able to continue to do

so in mild rain. Kaur's research aims to monitor and raise an alarm in Out-of-Distribution (OOD) scenarios that could lead to system failures.



Ramneet with advisor Insup Lee at a poster presentation at ICCPS 2023 in San Antonia, TX.

Kaur is also working on research on predicting Out of Distribution performance. "With the increasingly high interest in using Deep Neural Networks (DNNs) in safety-critical cyber-physical systems, such as autonomous vehicles, providing assurance about the safe deployment of these models becomes ever more important," Kaur said. "The safe deployment of deep learning models in the real world where the inputs can vary from the training environment of the models requires characterizing the performance and the uncertainty in the prediction of these models, particularly on novel and Out-of-Distribution inputs."

For instance, when a machine learning model is trained to classify images of traffic signs in sunny weather and deployed in an autonomous car, it is critical to quantify how it would perform in varying weather conditions. "This shift in data distribution for these machine learning models (DNNs) is inevitable in the real world," Kaur said. "Therefore, we require a certain level of assurance for the safe deployment of these systems to avoid catastrophic failures in novel scenarios, such as the self-driving Uber crashing into the cyclist walking her bike at night."

Other research by Kaur focuses on adversarial detection, or identifying techniques that hackers use to attack machine learning systems. While advances in computing and an increase in the available data have raised the demand for deep neural network models, which mimic the processing of the human brain, these models can be easily fooled into changing their predictions due to slight changes in the inputs. "For example, it has been shown that placing a sticker on a "stop" sign can change a machine learning model's prediction to a "speed limit" sign," Kaur said. "We developed a tool for detection of adversarial attacks on DNNs developed for real-time systems, such as autonomous cars. We show the tool is effective at detecting both digital and physical attacks."

Kaur's research also includes formal verification for autonomous vehicles. She notes that as cyber-physical systems grow more complex, it becomes increasingly difficult to develop methods for ensuring their correctness. Kaur and her team developed a framework for formal verification of vehicle-to-vehicle and vehicle-to-infrastructure technologies (V2X).

"The wireless information exchange via V2X technology allow vehicles to share up-to-date information about road conditions ahead, which goes a long way in circumventing accidents arising from dangerous situations such as poor visibility, inattention of human drivers or pedestrians, or the limited mobility of differentially-abled pedestrians (e.g., on wheelchairs), etc.," Kaur says. "This is expected to help save lives, prevent injuries, enable better-informed routing choices and thereby alleviate congestion, and bolster the development of autonomous vehicles."