UC San Diego
UC San Diego News Center

October 24, 2018 | By Ioana Patringenaru

$10M Grant from NSF Establishes Center for Trustworthy Machine Learning

A team of U.S. computer scientists is receiving a $10 million grant from the National Science Foundation to make machine learning more secure.

The grant establishes the Center for Trustworthy Machine Learning at a consortium of seven universities, including the University of California San Diego. Researchers will work together toward two goals: understanding the risks inherent to machine learning; and developing the tools, metrics and methods to manage and mitigate these risks.

Kamalika Chaudhuri, a computer science professor at the Jacobs School of Engineering, will be leading the UC San Diego portion of the research.

The science and arsenal of defensive techniques emerging within the center will provide the basis for building more trustworthy and secure systems in the future, as well as fostering a long-term research community within this essential domain of technology, researchers said.

“This research is important because machine learning is becoming more pervasive in our daily lives, powering technologies we interact with, including services like e-commerce and Internet searches, as well as devices such as Internet-connected smart speakers,” said Kamalika Chaudhuri, a computer science professor at the Jacobs School of Engineering, who will be leading the UC San Diego portion of the research.

The award is part of NSF’s Secure and Trustworthy Cyberspace (SaTC) program, which includes a $78.2 million portfolio of more than 225 new projects in 32 states spanning a broad range of research and education topics, including artificial intelligence, cryptography, network security, privacy and usability. A new center-scale Frontier award headlines this portfolio by addressing grand challenges in cybersecurity with the potential for broad economic and societal impacts.

“This Frontier project will develop an understanding of vulnerabilities in today’s machine learning approaches, along with methods for mitigating against these vulnerabilities to strengthen future machine learning-based technologies and solutions,” said Jim Kurose, NSF's assistant director for Computer and Information Science and Engineering.

Chaudhuri is an expert on machine learning—especially on the foundations of trustworthy machine learning. Her research group works on problems such as learning from sensitive data while preserving privacy; learning under sampling bias; and learning in the presence of an adversary.

Researchers will pursue three different goals. They will explore methods to defend a trained model against adversarial inputs. To do this, they will emphasize developing measurements of how robust defenses are; as well as understanding limits and costs of attacks. In a recent paper, Chaudhuri and her colleagues investigated what vulnerabilities allow attacks known as adversarial examples to happen. Adversarial examples are essentially optical illusions for machine learning, which cause algorithms to deliver incorrect results—for example identifying the picture of a rifle as a turtle. Researchers showed that for certain classifiers called nearest neighbors, more training data would lead to more robust systems.

Researchers also will develop new training methods that are immune to manipulation. Finally, researchers will investigate the general security of sophisticated machine learning algorithms, including potential abuses of machine learning models, such as models that generate fake content. They will aim to develop mechanisms that prevent the theft of machine learning models.

Computer scientists taking part in the grant are already involved in a summer school program centered around trustworthy machine learning and aimed at under-represented groups. They also are holding a series of webinars on the topic for high school students.

The grant will be led by researchers at Pennsylvania State University and, in addition to UC San Diego, includes researchers at University of Virginia, Stanford University, the University of Wisconsin at Madison and UC Berkeley.

Chaudhuri is a member of the Center for Machine-Integrated Computing and Security at the Jacobs School of Engineering at UC San Diego.

MEDIA CONTACT

Ioana Patringenaru, 858-822-0899, ipatrin@ucsd.edu

UC San Diego's Studio Ten 300 offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit https://ucsdnews.ucsd.edu/media-resources/faculty-experts.