SHAWN SPROUSE

INTRODUCING

RUOXI JIA

- › Postdoctoral fellow, University of California, Berkley, 2018-2020
- › Ph.D., electrical and electronics engineering, University of California, Berkley, 2018
- › B.S., microelectronics, Peking University, 2013

Enabling Responsible Data Science

Machine learning and AI make life easier, but do they make life safer? ECE assistant professor Ruoxi Jia doesn't think so, but she's working to change that.

When she was a Ph.D. student, Jia analyzed sensor data from buildings to more efficiently control features like lighting and air conditioning. But this line of research made Jia feel uneasy.

"I realized that requiring sensing and data analytics in smart infrastructures could represent a significant privacy threat," explained Jia. "If you monitor the occupancy of the building, it's easy to infer more personal information, like an occupant's habits, interests, and relationships."

Her discomfort led to interest, which set Jia on a new path—enabling responsible data science— which is her research focus as a new ECE assistant professor.

PRESERVING PRIVACY

Jia's research on privacy includes understanding what privacy means in different social and application contexts, using rigorous mathematical tools to characterize privacy, and developing techniques for managing the trade-off between privacy and data.

In her recent work, she has been exploring ways to defend machine learning systems from model inversion attacks, where an attacker aims at reconstructing training data from machine learning model parameters. Mounting such attacks on face recognition models, for instance, can expose private face images used for training the models.

While the vulnerabilities to model inversion attacks are well-understood for simple models, Jia's group is the first to demonstrate that deep neural networks are also susceptible to such attacks. To make machine learning systems more robust to inversion attacks, her group is using information theory to assess the amount of private information memorized by models, and designing new algorithms to limit the unwanted memorization.

MACHINE LEARNING SYSTEMS AT RISK

Privacy isn't the only threat associated with machine learning in the real world. In traditional computer systems, you can build an explicit boundary between the system and the outside world, explained Jia, but that's not possible in a machine learning system whose most crucial ingredient—data—directly comes from the outside world. "This means there are many potential attack surfaces, and it can be hard to deal with the security problem," said Jia.

For instance, a malicious attacker can inject bad data into the training dataset to manipulate the behavior of the machine learning model. Imagine the machine learning model is used in an autonomous car to recognize objects on the road. Through such attacks, the attacker can mislead the model to recognize a stop sign as a speed limit sign, which leads to disastrous consequences.

Jia and her team are developing defenses that can help mitigate attacks. They are building new machine learning algorithms that are robust to bad changes in the training data.

Jia's interest in securing deep neural networks extends to issues of intellectual property. Her team is investigating methods to insert watermarks into deep neural networks to secure the model and prevent plagiarism.

THE VALUE OF DATA

Data is the foundation of machine learning. Jia is also interested in improving the quality and robustness of machine learning models by designing incentives for people to contribute good data, and for companies to collaborate and share good data. But this presents a fundamental quandary—how should we value data?

"Data is very different from any other commodity," explained Jia. "Unlike a physical object, the same piece of data can be copied, shared, and utilized by different entities at the same time. The value of data really depends on how and when it is used."

Jia and her team are investigating principled methods to value data, and using the data value scores to inform data markets, improve data quality, and enable strategic economic data collection.

In Jia's ideal future, we're recompensed for our data, which is secured from attack and used to constantly improve the quality of our lives.

> "Unlike a physical object, the same piece of data can be copied, shared, and utilized by different entities at the same time. The value of data really depends on how and when it is used."
>
> –Ruoxi Jia

SHAWN SPROUSE