Tulane researcher honored for work on AI ethics

May 7th, 2019
|
Barri Bronston bbronst@tulane.edu

View PDF

Nicholas Mattei is an assistant professor of computer science at Tulane. (Photo by Paula Burch-Celentano)

As the role of artificial intelligence (AI) expands in daily life, so too does the need to engage in research into how to most effectively build AI technologies that align with positive values and ethics.

That is one of the areas that Tulane University researcher Nicholas Mattei specializes in and for which he was honored during the annual conference of the Association for the Advancement of Artificial Intelligence.

As part of the conference's events Mattei and collaborator Francesca Rossi, one of the leading researchers at IBM, were awarded the Computing Community Consortium's Blue Sky Award for their paper *Building Ethically Bounded AI.*

> "There is a growing need to understand how to constrain the actions of an AI system by providing boundaries within which the system must operate."
> *Nicholas Mattei*

The paper is based on the idea that the more artificial intelligence, in the form of automated decision-making systems, is used in daily life's unexpected situations, the more such systems need to be flexible, adaptive and creative in achieving an assigned goal. At the same time, the pervasive deployment of AI in daily life, whether autonomous or collaborating with humans, raises several ethical challenges.

"There is a growing need to understand how to constrain the actions of an AI system by providing boundaries within which the system must operate," Mattei and Rossi write.

"The role of AI in our daily lives is only expanding," they continue. "With this fact comes the realization that we must engage both research into how to most effectively build AI technologies that align with our values but that we also invite as many communities as possibly to join in the multi-stakeholder conversation around what are the best principles and practices to use when building these systems."

In their paper, Mattei and Rossi use "specification gaming" behaviors as examples. They include a game-playing agent that pauses the game indefinitely to avoid losing and a reinforcement-learning agent in a boat racing game that repeatedly hits the same reward targets in order to increase the score.

Although there are existing approaches to building ethical AI systems (logic-based and data driven), questions and challenges remain. For example, what ethical principles should be injected into AI systems?

The authors call on scientific associations as well as social scientists, economists, policy makers and consumer rights groups to be part of the discussions to answer that question.