

Storytelling to Teach Robots Right and Wrong

Introduction

In this excerpt from Science Friday, Mark Riedl discusses how his research group uses a program called Quixote to teach robots morality and etiquette. As humans get closer and closer to developing self-aware artificial intelligence, we need to make sure that artificially-intelligent computers and robots follow rules that keep them from endangering us and that align with our social conventions. Robots need to be efficient, but not in a way that leads them to bad behaviors like cutting in line, stealing, or physically harming people. How do we make sure that efficiency does not trump morality in a machine we've programmed?

Vocabulary

ethical— acting within accepted principles of right and wrong.

etiquette— behavior that aligns with social norms (e.g. being polite).

protagonist— the main character of a story who must overcome obstacles and barriers.

artificial intelligence— software or programs capable of independent deduction, reasoning, problem solving, critical thinking, and/or creativity (e.g. playing [go](#)¹ or [telling jokes](#)²).

Listening Task

While listening to the audio excerpt "Storytelling Teaches Robots Right and Wrong," collect information that answers the following questions:

Why is there a pressing need to develop artificial intelligence that complies with societal norms?

Why did Riedl and his team choose stories as a source of cultural values?

How do machine learning systems process data?

¹ Metz, Cade. "In a Huge Breakthrough, Google's AI Beats a Top Player at the Game of Go." *Wired.com*. 27 Jan. 2016. <<http://www.wired.com/2016/01/in-a-huge-breakthrough-googles-ai-beats-a-top-player-at-the-game-of-go/>>.

² Warner, Joel, and Peter McGraw. "It's Comedian vs. Computer in a Battle for Humor Supremacy." *Wired.com*. 1 Apr. 2014. <http://www.wired.com/2014/04/underwire_0401_funnycomputer/>.

Why shouldn't robots just rely on one story to learn societal norms?

According to Riedl, why is storytelling an efficient way to teach robots?

Why does Riedl argue that stories are better than large data sets for teaching robots?

Discussion Questions

Use the boxes below to record thoughts and interesting ideas during your group discussion.

Do you think stories are a good way to teach robots to understand human behavior?

Mark Riedl argues that it is important not to “cherry-pick” the stories used to teach robots, “because by doing so, we run into the danger of unintentionally reinforcing certain behaviors.” Do you agree with Mark Riedl? Why or why not?

Science fiction writer Isaac Asimov introduced the “[Laws of Robotics](#)” to guide the programming of intelligent robots in order to keep them from turning against humans. They are:

- First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- Second Law: A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law.
- Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.
(Asimov, Isaac. I, Robot. New York: New American Library, 1956.)
- Zeroth Law: A robot may not injure humanity, or, through inaction, allow humanity to come to harm. (Asimov, Isaac. Robots and Empire. Garden City, NY: Doubleday, 1985.)

Riedl and Harrison propose that we use programs that “teach” robots to create their own set of rules by analyzing stories. Asimov proposes that we program intelligent robots with directives, or rules that they must follow. Explain which approach you think is better: Riedl’s collection of stories (teaching) or Asimov’s Three Laws of Robotics (directives).

GENERATE IDEAS: As a group, come up with a list of social norms and behaviors you think moral robots should have. Rank those social norms and behaviors in terms of importance. Choose your group’s top three social norms and behaviors and make a recommendation of stories that should be used in the Quixote program. What do you hope the Quixote program would learn from each story?

