Strong AI: Prospects and Control
Talk and Debate

Thursday, April 28, 2016
6pm
CIT Building, Room 368

1. Talk by David McAllester, Toyota Technological Institute at Chicago

Professor McAllester received his B.S., M.S., and Ph.D. degrees from the Massachusetts Institute of Technology in 1978, 1979, and 1987 respectively. He has served on the faculty of Cornell and MIT and has been a fellow of the Association for the Advancement of Artificial Intelligence (AAAI) since 1997. Since 2002 he has been Chief Academic Officer at the Toyota Technological Institute at Chicago (TTIC). He has received 20 year "test of time" awards from the Association of the Advancement of Artificial Intelligence (AAAI) and the International Conference of Logic Programming.

Abstract: There continues to be public concern over the near-term possibility of strong Al -- human level intelligence in machines. This talk will present the case that, while it is very difficult to make any such prediction, near-term strong Al is plausible. The talk will review recent rapid progress in computer vision and the recent success of AlphaGo over Lee Sedol. The talk will then argue that machine understanding of natural language is now more of an engineering problem than a deep scientific mystery. Getting machines to replace software engineers is harder. None the less, it seems difficult to rule out near term (ten year) automation of software engineering, especially given recent progress in the foundations of mathematics. The talk will also consider approaches to the control of superintelligent machines. The frameworks of rational choice, game theory, and reinforcement learning will be reviewed. An argument will be made for the "servant mission" where each machine is given the mission (objective) of behaving within the law to fulfill the requests of a certain human individual. This transfer's moral responsibility from the machine its master. It also creates a large society of distinct superintelligent machines hopefully avoiding the "singleton scenario".

2. Debate on Probable and Possible Risks of Artificial Intelligence

Recent developments in artificial intelligence have attracted public attention to the topic of what dangers they might give rise to in the long term. This debate will provide the Brown community an opportunity to hear from two informed but somewhat contrasting voices on this topic. They will present their views and engage in a broader discussion with the attendees. Of particular importance is the question of whether artificial intelligence poses a significant risk to the wor1d, and, if so, whether there are steps that we can take in the near term to mitigate them.

The event is co-sponsored by Brown Effective Altruists, the Brown Digital Literacy Forum {DigLit), and the Humanity Centered Robotics Initiative (HCRI).

For more information on this talk and the HCRI Speaker Series, contact hcri@brown.edu or visit hcri.brown.edu.