Artificial Intelligence and the "Barrier of Meaning"

Neurohumanities Public Lecture

PROFESSOR MELANIE MITCHELL

Melanie is Davis Professor of Complexity at the Santa Fe Institute in New Mexico, USA. Her current research focuses on conceptual abstraction, analogy-making, and visual recognition in artificial intelligence systems. Melanie is the author or editor of six books and numerous scholarly papers in the fields of artificial intelligence, cognitive science, and complex systems. Her book Complexity: A Guided Tour (Oxford University Press) won the 2010 Phi Beta Kappa Science Book Award and was named by Amazon.com as one of the ten best science books of 2009. Her latest book is Artificial Intelligence: A Guide for Thinking Humans (Farrar, Straus, and Giroux).

Thursday May 12th 4pm - JM Synge Theatre – Arts Building Trinity College Dublin

In 1986, the mathematician and philosopher Gian-Carlo Rota wrote, "I wonder whether or when artificial intelligence will ever crash the barrier of meaning." Here, the phrase "barrier of meaning" refers to a belief about humans versus machines: humans are able to "actually understand" the situations they encounter, whereas AI systems (at least current ones) do not possess such understanding. The internal representations learned by (or programmed into) AI systems do not capture the rich "meanings" that humans bring to bear in perception, language, and reasoning. In this talk I will assess the state of the art of artificial intelligence in several domains, and describe some of their current limitations and vulnerabilities, which can be accounted for by a lack of true understanding of the domains they work in. I will explore the following questions: (1) To be reliable in human domains, what do AI systems actually need to "understand"? (2) Which domains require human-like understanding? And (3) What does such understanding entail?

SANTA FE INSTITUTE

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

wellcome

Organised by the 'Representation: Past, Present & Future' project & made possible by Wellcome Trust ISSF Neurohumanities Programme