

What AI Still Can't Do

The starting point of my talk, with a title paraphrasing Hubert Dreyfus' 1992 book, is a set of predictions made in 1957 by the Nobel Laureate Herbert Simon, in his keynote talk for the Operations Research Society, regarding the achievements in the area of AI within 10 years: (1) The chess world champion would be AI. (2) AI would discover and prove an important theorem. (3) AI would compose music of considerable aesthetic value. (4) Most psychology theories would be computer programmes. The first one of these happened in 1997, taking four times longer than the timeframe proposed by Simon, the rest are nowhere near. What is so weird about AI that Simon, who was one of the foremost experts in the field, got his predictions so wrong? Many other AI experts made equally mistaken predictions. I believe it is crucial to understand what is at the heart of this problem, as it may be the main reason for the high failure rates of AI projects. Although I start with something from more than half a century ago, my talk is about the future. I am trying to figure out how to think right about AI in order to make the best possible use of it. I contrast human thinking and learning with AI data processing and machine learning in order to show that the human mind and AI work on entirely different principles and modes. Based on that, it is fairly trivial that humans and AI are good at fundamentally different things; I argue that while AI is good at handling complicated issues, humans excel in dealing with complexity. Finally, I turn the initial upside down and argue that it is not that we do not understand AI, we do. What we do not understand is the human mind.