The New Metaphysicians:
Philosophy, Computers, and the Pursuit of Artificial Intelligence

Justin Flaumenhaft, 2018

This summer I set out to study the philosophy of science. My initial focus was on the demarcation problem, the problem of how to distinguish science from pseudoscience. However, as my research progressed, I became increasingly interested by the philosophy of a particular area of science: the field of artificial intelligence. Computer scientists, in an effort to endow machines with ever-greater problem-solving abilities, have faced questions that have preoccupied philosophers for centuries. For example: Are thinking and understanding reducible to mechanical processes in the brain? If so, is it possible to build a machine that could replicate these processes? The answer to these questions could have profound consequences on the future of humanity. My research examined the history and philosophy of artificial intelligence in order to understand the origins of artificial intelligence and to get a glimpse of what the future might have in store.

As it turns out, philosophers were essential to the development of the first computers. Pascal invented the first mechanical calculator and Leibniz later improved on its design, by adding a multiplication and division feature. These devices were the 17th century predecessors of modern computers. A couple of centuries later, George Boole, another mathematician-philosopher, created an algebra that expanded on Aristotle's logic. In Boole's algebra "1" represents "true" and "0" represents "false." Almost a century after its advent, Boole's algebraic system would prove to be essential to the design of the first computers. Drawing on a philosophy course he took as an undergraduate, the computer pioneer Claude Shannon applied Boolean algebra to mechanical relays. This concept became the basis for digital computing.

While these pre- and post- enlightenment philosophers played an important role in the eventual creation of the computer, they also imported some of the dubious metaphysics of their time to computer science. The term "artificial intelligence" was coined in the 1950's at a time when optimism about the potential for computers was at a peak. Leading computer experts in the burgeoning field of artificial intelligence predicted that in less than a few decades computers would be able to anything that humans do. The philosopher Hubert Dreyfus of MIT was among the first to recognize that the expectations of the computer science community were unrealistically high. Drawing from contemporary philosophers like Heidegger and Wittgenstein, Dreyfus argued that the early attempts at artificial intelligence had operated under the outdated philosophical assumption that understanding consists of mechanically following rules and manipulating symbols, as computers do. The assumption that complex skills could be achieved by a system of explicit rules, Dreyfus argued, was bound to fail.

Despite being ridiculed and ostracized by his colleagues in computer science at the time, Dreyfus' skepticism appears to have been be justified. Decades later, artificial intelligence has fallen very short of the lofty expectations set for it in the fifties and sixties. The field of artificial intelligence has been hindered by long periods of little productivity known as "AI winters." In my research, I have sought answers for the limited success of artificial intelligence in the work of Heidegger, Wittgenstein, Gödel, and Hofstadter. I have examined and critiqued the arguments of public figures, including Elon Musk and Nick Bostrom, who contend that"Super"Artificial Intelligence is something we should be preparing for. And lastly, I have investigated recent developments in machine learning, which I argue are more promising than traditional AI methods (due to being less ruled-based and less rigidly programmed), but still do not bring us closer to general machine intelligence — for better or for worse.

Faculty Mentor: David Hecht

Funded by the Ellen M. and Herbert M. Patterson Research Fellowship

References:

Avery, John. *Science & Society*. World Scientific, 2017.

Dreyfus, Hubert. *What Computers Still Can't Do*. MIT Press, 2009.