Marcus Domingo

ENGH-302-N12

Summary/Discussion #4

October 24th, 2016

Concerns about AI

Dietterich and Horvitz (2015) outline the direction and risks of artificial intelligence (AI) in the world today. They begin to describe all the encounters humans have with AI every day such as GPS, Siri, Cortana, Google Now, and Facebook (para. 1). Dietterich and Horvitz give credit all the achievements that AI have made in its role in science and medicine. They state the future of AI is promising because of the part AI plays in human life today. That “over the longer-term, advances in machine intelligence will have deeply beneficial influences” on the world around us (Dietterich & Horvitz, 2015, para. 2). Although progress is steady, Dietterich and Horvitz point out that the thoughts of machines taking over the world is bogus. The ability for an AI to recreate itself a thousand times over then control humanity is a little absurd, but Dietterich and Horvitz do say relying on AI with dangerous resources do pose risk (para. 5). From this, Dietterich and Horvitz describe five sets of risks that pertain to AI. The first set being errors that were programmed during the design phase of the AI, thus leading to dangerous outcomes and even death (Dietterich & Horvitz, 2015, para. 7). The second set risks are the vulnerability of AI to cyberattacks (hackers and other malicious software). The third set of risks are an AI system taking instruction literally rather than how us humans would interpret. Dietterich and Horvitz give the example of a self-driving car going 125-mph to get us to a destination as fast as possible (para. 9). The fourth set of risks balance of power between AI and humans and switching over who is in control at specific points in real-time. This arises when humans don’t pay specific attention to the state in which the AI is in and when human control takes over its from poor misunderstanding that something may get messed up. The fifth set of risks are the “influences of increasingly competent automation on socioeconomics and the distribution of wealth” (Dietterich & Horvitz, 2015, para. 12). Dietterich and Horvitz state that for AI to flourish there must be attention brought to the challenges at hand as well as scholarly work for the long-term challenges. Concluding that the computer science community must put forth efforts to guarantee that AI will behave safely.

Dietterich and Horvitz (2015) heavily stress the risks of AI which supports half of their thesis, the other half being the successes and promises of AI and its future. Categorizing five different sets of risks and stating them was a strong aspect to the article. Laying out where the concerns and the risks lie ahead for the future of AI is what greatly relates to the perspectives of giving AI too much of a role in humanity. Each set of risks were distinct and didn’t relate to or depend on each other in any way which also showed great thought and expertise in the study of AI. Dietterich and Horvitz believe that AI has the potential of being safe for delicate information and human reliance once these risks have been put to a minimum.

Dietterich and Horvitz (2015) points out the risks of AI and says there are promises of safer AI solutions in the future. Schölkopf (2015) heavily talks about Q-learning being the gateway to real world problems, and this shows a potential solution to the risks that Dietterich and Horvitz state. Davis and Marcus (2015) express different approaches to AI learning commonsense and the challenges that come with them. Dietterich and Horvitz, similarly to Davis and Marcus, explained the challenges in AI but with their relationship to humanity. Thus showing that there are multiple challenges that are face throughout the field of AI that aren’t just one sided. Gil, Greaves, Hendler, and Hirsh (2014) relate to Dietterich and Horvitz in the sense of giving AI data. Gil et al. describe the AI being able to process the data towards scientific discovery, while Dietterich and Horvitz point out the risks of allowing AI to process sensitive information. Do we rely on AI too much? All the authors above believe that the future of AI is promising and that there are plenty of benefits to come.

References

Davis, E., & Marcus, G. (2015). Commonsense Reasoning and Commonsense Knowledge in Artificial Intelligence. *Communications of the ACM*, *58*(9), 92–103. https://doi.org/10.1145/2701413

Dietterich, T. G., & Horvitz, E. J. (2015). Rise of Concerns about AI: Reflections and Directions. *Communications of the ACM*, *58*(10), 38–40. https://doi.org/10.1145/2770869

Gil, Y., Greaves, M., Hendler, J., & Hirsh, H. (2014). Amplify scientific discovery with artificial intelligence. *Science*, *346*(6206), 171–172. https://doi.org/10.1126/science.1259439

Schölkopf, B. (2015). Learning to see and act. *Nature*, *518*(7540), 486–487.