Maneesh Wijewardhana 2023-01-29

Week 4

From this week's readings, I want to dive deeper into the idea of complacency when it comes to complex automated systems. When the author said "Put simply, once humans are accustomed to trust a system that's reliable most of the time (but not all of the time), humans themselves tend to "switch off," falling into a sort of "autopilot" mode where diffidence, complacency and overtrust set in" (Zerelli 80), it made me question why this occurs and what we as humans can do about it. Just because some automated systems are complex and seem to do what they are supposed to, it does not mean that we can handwave the fact that humans get complacent with them.

We see tons of headlines like "Google AI researchers working with Northwestern Medicine created an AI model capable of detecting lung cancer from screening tests better than human radiologists with an average eight years of experience" and "MIT CSAIL and Massachusetts General Hospital published a paper in the journal Radiology about a system capable of predicting the onset of breast cancer five years in advance using mammography exam imagery (https://venturebeat.com/ai/confidence-uncertainty-and-trust-in-ai-affect-how-humans-make-decisions/). It is interesting to point out that these achievements not only fuel the acceptance of AI systems in society but also the trust that AI operators have with them. One can argue that it's in human nature to build trust and if an AI system truly is performing well all the time, it is safe for complacency to take effect. I disagree with both of these and believe that organizations should be focusing on who we let supervise these systems and to what degree.

In my opinion, no matter how complex the system becomes, we either need intelligently trained individuals and/or teams to be present and know the limitations and potential risks associated with the system so that appropriate action can be taken when necessary. I also believe that these systems should undergo regular testing and monitoring even when it seems unnecessary since as we know from past events like Tay AI, complex systems can behave unexpectedly and at unpredictable times. What truly will allow for non-complacency is a healthy balance of trust and distrust when it comes to these systems. The question then becomes, why are we putting so many resources into building complex AI systems when at the end of the day, we would still need humans to supervise? Is it not the goal for AI to take over tasks that humans just cannot accomplish biologically? Actually, the role of a human with complex AI systems is changing, we need to be trained on how to detect malfunctions, conduct testing, and practice skepticism.