**1: Prep**

How much of the talks did you watch before the class?

* 90-100%

**2: Discussion**

How much did you engage in discussion?

* I listened actively, invited others to participate, referred to specific sections of the source when appropriate, and added to the conversation by building on what others said

**3: Reflection**

To outline “What happens when our computers get smarter than we are?”, Nick Bostrom accentuates that super intelligent machines (SIM) are nearing sooner than we anticipate because “artificial intelligence is getting smarter by leaps and bounds.” Bostrom describes SIM to have algorithms that are able to learn faster than humans, which will be the last form of AI that needs to be made. While these thinking machines are bound to make extraordinary breakthroughs, Bostrom highlights that the primary concern is the extent to which these machines will preserve humanity and our values; this was my group’s main discussion on this talk. One of the fundamental solutions that Bostrom accentuated is the importance of avoiding anthropomorphizing. There are many dangers that we discussed will arise from anthropomorphizing SIM including making people believe that it not only senses and shows emotion, but actually feels it, which may lead to its befriending. Convincingly, this blinds people from SIM’s legitimate possibilities and limitations. Furthermore, anthropomorphizing may lead to SIM controlling people when we are supposed to be controlling it, which poses the question: if SIM attains a point where humans are unable to control it, how do we ensure that these machines are still safe by preserving humanity and being on our side? While there is no way of answering this question because SIM is yet to be fully established, a valuable consideration when building these machines is to ensure that we align SIM’s goals with ours such to ensure that they don’t harm humans, regardless of whether they take control or not.

In “This app knows how you feel – from the look on your face,” Rana el Kaliouby presents a new AI application that opposes Bostrom’s assertion to avoid anthropomorphizing. Kaliouby introduces and demos her new application that is able to “read facial expressions and matches them to corresponding emotions,” and did this by training and testing a deep learning algorithm on more than 12 billion emotion data points. She recognizes that this application doesn’t have emotional intelligence…yet, but her goal is to enable that because she believes emotion-empowering technology will allow people to reclaim their emotions through technology, since technology is what is currently separating us from it. While she outlines many applicable use cases for this technology, there are many challenges and potential of misuse that we discussed. For example, it can be misused to gauge human emotions towards advertisements that are presented to users on the internet in order to understand their preferences, which introduces new privacy issues for regular users. Furthermore, such technology may make users worse off because of inaccurate decision making that stem from factors beyond facial expressions, such as human judgment, which can’t be achieved through artificial intelligence. Thus, since Bostrom implores people to avoid giving SIM human characteristics such as emotional intelligence, while Kaliouby proposes the contrary, the unsettled question from our discussion was: whose proposition’s benefits outweigh the other in this regard? How can we create emotionally intelligent machines that we can control, so that we ensure SIM does not control us?