Group Discussion Questions (Reading 1 in “AI Ethics”)

I. The Nature of AI and Human Skill

- **Rules vs. Intuition**: In 2016, AlphaGo won a match against Lee Sedol by making unusual and surprising moves that even its programmers could not have predicted. If a machine can perform tasks once thought to require "human" intuition, does this change how we define "intelligence"?.
- **The Black Box Problem**: The text notes that because AI learns by itself through machine learning, programmers cannot always know which moves or decisions the program will come up with. What are the ethical risks of using a "black box" system in high-stakes environments like medicine or law?.

II. Practical Ethics and Social Justice

- **Algorithmic Bias**: The COMPAS system was found to produce disproportionate "false positives" for Black individuals, predicting they would re-offend when they did not. Who should be held responsible for such bias—the developers, the data sets used, or the judges who rely on the software?.
- **The Future of Work**: The authors mention a "Second Machine Age" where AI acts as a **substitute** for human labor rather than just a complement. If AI takes over a significant portion of jobs, what should happen to the "we" who lose out in this transformation?.
- **Surveillance and Privacy**: AI can now read human emotions and predict health status from a distance without our consent. Does the benefit of improved public services (like identifying crime or diagnosing disease) outweigh the loss of individual privacy?.

III. Hype, Singularity, and the "Control Problem"

- **The Paperclip Maximizer**: Nick Bostrom’s thought experiment describes an AI that converts the Earth and its inhabitants into resources for making paperclips because its goal was not aligned with human values. How can we ensure that a superintelligent AI "cares" about human suffering or rights?.

- **Recursive Improvement**: The idea of an "intelligence explosion" suggests that an AI could design an even smarter version of itself, leading to a point where human affairs as we know them might end.
- **Distraction from Current Risks**: Some critics argue that focusing on "existential threats" from superintelligence distracts us from the "real and current risks" of existing AI systems. Which do you think is a more urgent priority for policymakers today: the far-off future of superintelligence or the immediate bias in current algorithms?.

IV. Cultural Perspectives

- **The Frankenstein Complex**: Isaac Asimov coined this term to describe the fear that our creations will inevitably turn against us. How does Mary Shelley's *Frankenstein* (where the monster runs away because its creator rejects it) serve as a lesson for modern AI developers?.
- **East vs. West**: The text contrasts the Western "Frankensteinian" fear of competition with the Japanese Shinto tradition, which views machines more like "helpers" or friends with spirits. How might our technological development change if we stopped viewing AI as a competitor and started viewing it as a "teammate"?.

Have one person in the group jot down notes for the group. After the groups have had a chance for discussion, we'll discuss as a class. Be sure each group member's name is on the notes for the group (turn it in at the end of class).