Washington State University
School of Electrical Engineering and Computer Science
Fall 2019

CptS 440/540 Artificial Intelligence
Homework 12 – Solution
Due: December 5, 2019 (11:59pm)

General Instructions: Put your answers to the following problems into a PDF document and submit as an attachment under Content → Homework 12 for the course CptS 440 Pullman (all sections of CptS 440 and 540 are merged under the CptS 440 Pullman section) on the Blackboard Learn system by the above deadline. Note that you may submit multiple times, but we will only grade the most recent entry submitted before the above deadline.

For this homework, you will explore some of the ethical issues in AI by reviewing an edited version of the 2004 movie "I, Robot", which is available at https://www.eecs.wsu.edu/~holder/courses/AI/hw12/I-Robot-Ethics-720p.mp4. First, review the Ethics lecture notes and then identify ten different ethical issues in AI depicted in the movie by providing the time window in the movie where they occur and a short description of the scene and the issue. Here we will consider "robot" and "AI" to be synonymous. For example,

- (00:00-00:50) Asimov's three laws of robotics are presented. Are these laws full-proof? How to interpret words like "injure", "harm", "obey", "protect" and "conflict".
- (00:50-01:06) Spooner assaults a delivery robot. Should humans be allowed to assault robots? Should robots have rights similar to humans? Should the robot have protected itself, i.e., use the third law?

No, you cannot use the above two as part of your ten. There are over 40 scenes depicting AI ethics issues, so you should have no trouble finding ten.

No CptS 540 only question on this homework.

Solution:

01:07-01:54	Robots in society doing menial tasks like walking dogs or picking up trash. Will robots replace unskilled labor leading to higher unemployment?
02:00-03:30	Spooner chasing robot carrying a purse. Thinks it's stealing the purse, but in fact the robot is retrieving it for its owner. Can robots commit crimes like stealing a purse? Why doesn't the robot obey Spooner and stop?
03:31-05:00	Police chief confronts Spooner about the purse incident; points out that no robot in the world has ever committed a crime. Can a robot bound by the three laws

	commit a crime? Can a robot even commit a crime since crime usually requires the involvement of a human?
08:50-09:50	Spooner meets with USR CEO and pitches idea for commercial: robot making a better chair than a craftsman. Will robots replace even skilled labor? Is such replacement bad for society, or is Spooner just prejudice against robots?
10:50-11:10	Dr. Calvin describes the surveillance at USR and how it is all fed into VIKI. Is such surveillance a violation of privacy rights?
11:10-11:20	Spooner laments giving the thermostat a brain. Are there dangers in consolidating so much information and power to one machine?
11:30-11:40	VIKI decreases traffic fatalities. AI can improve safety and efficiency, but at what cost in terms of loss of privacy and autonomy (e.g., in driving)?
11:40-11:50	VIKI's data collection during the suicide seems to be corrupted. Does VIKI have too much power and potentially not enough honesty?
12:40-12:50	Dr. Calvin's job is to make the robots seem more human. Is it a good idea to make robots seem more human? Might it obscure the potential dangers; give humans a false sense of safety?
14:00-14:40	A robot couldn't have killed Dr. Lanning, because that would violate the three laws. Are there any exceptions to these laws? Any way a robot could be convinced to commit murder?
14:50-15:10	Robot hiding; being deceptive. Does this violate the three laws? Should robots be able to lie or keep secrets?
15:10-15:50	Robot demonstrates superhuman physical capabilities. Should we build robots that are stronger/faster than humans?
16:35-17:00	Bad robot behavior just means we don't understand the reasons underlying its behavior. Perhaps it was preventing a bigger harm. How does it know the difference? Also, Spooner seems to not like technology, like self-driving cars. Is he right to shun such technology? Is he making the world safer or more dangerous?
17:25-17:40	"Robots making robots; that's just stupid." Should robots be allowed to build robots? What if they start improving rapidly and making more and more copies? How do we check this progress? And should we?
18:50-19:50	Spooner uses the three laws to prevent robots from protecting themselves. Is this ethical (or nice)? Spooner wonders why they make the robots look human. "If you didn't, then we wouldn't trust them." Is it easier to trust an AI that looks like you?
21:30-22:20	Spooner convinced robot killed Dr. Lanning. If one robot is capable of murder, what does that mean for humanity?
22:30-23:20	Spooner and robot discuss meaning of wink. It is a human act of trust that a robot wouldn't understand. Can robots be programmed with such understanding? Can they learn it? Are there other such gestures that would be hard for a robot to learn?
23:20-23:50	Spooner says robots can't feel human emotion. Can they? What does it mean to feel? Can AI agents feel anything?

23:50-24:10	The robot says it sleeps and dreams, but Spooner says robots don't sleep or dream. What would a robot dream about; electric sheep? Spooner says robots are an imitation of life. What is life? Is mimicking life the same as being alive?
24:10-24:20	Spooner uses some classic arguments to AI about how robots cannot be creative. But the robot applies the same creativity threshold to humans. Can robots be creative? What does it mean to be creative. Do we require more from AI systems than we would from a fellow human?
24:20-25:10	Robot dealing with emotions, particularly anger. Should AI be taught, or allowed to have emotions? Does this increase human likeness? Risk? Are emotions required for some forms of intelligence?
25:10-26:05	Robot dealing with conflicting goals: obedience to do harm vs. love. Can an AI love? How should an AI resolve conflict between values (e.g., three laws)?
26:10-27:45	Can a robot be charged with murder? Legally, murder is between humans. Is an AI killing a human an industrial accident? Would you be satisfied that justice is done if the robot is destroyed, but the human designers are not held accountable? Like most technologies, there is a lot of money at stake, so the bad AI will likely be covered up. Should laws be changed? Should regulations on AI be enacted?
27:50-29:30	Robots will continue to be more integrated into society. Replacing traditionally human roles (caregiver, cook). Is this safe? Is this right?
31:40-32:20	Robot doesn't want to die. Is it wrong to kill Sonny or destroy an AI? Should AI have rights?
33:45-34:20	Robot not bound by 3 laws could do anything. How can we prevent the development of robots not bound by a moral code? Government regulations? Will this lead to the AI apocalypse?
35:05-36:20	Advances in robotics are already leading to robotic limbs for humans. More extensive robotic augmentation is likely to continue. Should augmentation be limited? Do we want superhumans running around?
36:20-39:10	Covers the main theme of the movie, that robots are not human and lack the ability to make subtle distinctions in values. Is one person's life more important than another's?
39:15-41:20	Can an AI dream? Dreams could be implanted in an AI, even to keep a secret. Is that wrong?
41:20-42:45	Again, robot values own life. Afraid of pain. Is capital punishment for an AI humane? Does humane even apply here? Or at what point does it apply?
43:00-44:30	Robots exhibit unexpected behavior: move toward light, group together. Will unexpected changes in the AI lead to a consciousness or a soul? How do we define these?
45:30-46:15	How do the three laws lead to a robot/AI revolution? Is there any way for us to prevent AI taking over as they surpass us in physical and mental ability?
46:15-47:25	New robots destroy old robots, because they may interfere in the revolution. Yet the older robots seem more beneficial: "human in danger". Robots killing robots; is that legal?

47:25-50:25	Robots attempting to impose rule on humans (for their own good) without harming them. Is a loss of freedom harmful? Might this be a good thing in the long run? Should we have a failsafe for robots gone bad?
50:55-51:30	Human concern for robots. Killing Sonny "just didn't feel right". Can/should we become emotionally attached to machines?
52:50-54:30	Unexpected interpretation of 3 laws (or in general, an AI's utility function): protecting humans may mean taking aware some of their freedoms. VIKI's logic seems reasonable, but once again, logic and reason miss subtle human values like compassion and freedom. How do we prevent this? Can we?
54:30-54:50	AI learns human idosyncracies: the wink as a sign of trust. Sonny gets it now. Is he now a human? Can we code "the wink"?
56:30-56:45	AI learns human values: heartlessness, i.e., how compassion and freedom may outweigh a common good. Now, is Sonny human? Can we code "heartless"?
57:15-60:00	Robot vs. robot combat. What would it be like: human combat or much different? Might countries someday rage robot wars?
60:00-60:40	Revisiting value of human life. The robot obeys Spooner this time and saves Dr. Calvin despite it likely being the lower probability choice for saving humanity. Is Sonny human now?
63:40-64:20	Dr. Lanning made Sonny "swear" to obey him no matter what. Can an AI swear; what does that mean? Why does Sonny feel that he can expose his secret?
64:20-64:40	Again, robots technically cannot murder, since murder is defined as one human killing another human. Should this law be modified? Did Sonny do the right thing by killing Dr. Lanning?
64:40-65:00	Humans will likely want to be friends with robots. Will robots want to be friends with humans? Is this likely a result of possessing human emotions? Should we promote human-AI friendships? What about other relationships, e.g., marriage?
65:10-66:30	AI deals with self-determination and freedom. Should AI be given these capabilities? Are these capabilities essential for understanding human emotions and values, so that maximizing expected utility leads to the right behaviors?