MM AI Text Mining Assignment: Chapter 7

"On Trustworthy and Ethical AI"

1. Self-driving cars have the potential to vastly improve our lives. Automated vehicles could substantially reduce the millions of annual deaths and injuries due to auto accidents, many of them caused by intoxicated or distracted drivers. In addition, automated vehicles would allow their human passengers to be productive rather than idle during commute times. These vehicles also have the potential to be more energy efficient than cars with human drivers and will be a godsend for blind or handicapped people who can't drive. But all this will come to pass only if we humans are willing to trust these vehicles with our lives? **Do you think that you might be willing to trust your life to these vehicles? Why, or why not?**
   As these vehicles currently are, I would not trust these vehicles. Maybe in some number of years when they become more advanced I would.
2. MM enumerates a number of huge benefits that AI systems already bring to society. **Please list a few of these.**
   GPS navigation, media recommendations, email spam filters, optimizing energy use in buildings.
3. MM suggests that in the near future, AI applications will likely be widespread in health care. **Please list a few of the AI applications that she foresees.**
   Diagnosing diseases, suggesting treatments, discovering new drugs, and monitoring the health and safety of elderly in their homes.
4. What, according to Demis Hassabis, the cofounder of Google's DeepMind group, is the most important potential benefit of AI?
   Helping humanity solve problems that are too difficult for humans to currently solve.
5. In discussing the phenomenon of AI taking over jobs that humans do at this point in time, MM raises the question of whether or not this will actually benefit society. In considering the question, she lists a number of jobs that technology automated long ago, suggesting that AI may simply be extending the same are of progress: improving life for humans by increasingly automating the necessary jobs that no one wants to do. **Please list a few of the jobs that technology automated long ago.**
   Calculations, washing laundry, operating elevators, and cooling down rooms.
6. What was the AI researcher Andrew NG suggesting when he optimistically proclaimed, "AI is the new electricity."
   AI is all around us throughout everyday life to the point where we don't even notice it.
7. What major difference does MM observe between electricity and AI?
   By the time the use of electricity became widespread we mostly understood it, but that is not the case with the current type of AI which is being widely used now.

8. What is “the great AI tradeoff?”
   The ability for AI to improve our lives and the potential weaknesses that current AI has like its unpredictable errors, susceptibility to bias, vulnerability to hacking, and the lack of transparency in decision making.
9. TRUE/FALSE - Machine intelligence presents a knotty array of ethical issues, and discussions related to the ethics of AI and big data have filled several books
   True.
10. List a couple of “positives” relating to face recognition systems. List a couple of “negatives” relating to face recognition systems.
    Some of the positives are the ability to search image albums for someone, helping people with impaired vision know who’s in photos, and helping locate missing people. Some of the negatives are the lack of privacy, and the potential for an innocent person to be labeled as a criminal.
11. Present-day face-recognition systems have been shown to have a significantly higher error rate on people of color than on white people. Describe the ACLU study that strikingly underscored this point.
    The ACLU study used the 535 members of congress and compared a photo of each against a criminal database where the system incorrectly matched 28 of the 535 members with someone in the database.
12. TRUE/FALSE - Given the risk of AI technologies, many practitioners of AI are in favor of some kind of regulation. But simply leaving regulation up to AI practitioners would be as unwise as leaving it solely up to government agencies. The problems surrounding AI – trustworthiness, explainability, bias, vulnerability to attack, and morality of use – are social and political issues as much as they are technical ones. Thus, it is essential that the discussion around these issues include people with different perspectives and backgrounds.
    True.
13. TRUE/FALSE questions are often used to assess student knowledge. If a student responds with the sanctioned answer, it is assumed that they possess the sanctioned knowledge. **Please suggest an alternative use for True/False questions.**
    Perhaps adding instructions to the question to explain/give thoughts on the topic
14. In one example of the complexity of crafting regulations for AI systems, in 2018 the European Parliament enacted a regulation on AI that some have called the “right to explanation.” This regulation requires, in the case of “automated decision making,” “meaningful information about the logic involved” in any decision that affects an EU citizen. This information is required to be communicated “in a concise, transparent, intelligible and easily accessible form, using clear and plain language.” This opens the floodgates for interpretation. What counts as “meaningful information” or “the logic involved”? *Does this regulation prohibit the use of hard-to-explain deep-learning methods in making decisions that affect individuals (such as loans and face recognition)?* Such uncertainties will no doubt ensure gainful employment for policy makers and lawyers for

a long time to come. **What do you think about the highlighted question? Please say a thing or two of significance about the question.**

This regulation would affect individuals, most of the AI systems that they unknowingly use make use of deep-learning methods.

15. TRUE/FALSE - The infrastructure for regulating AI is just beginning to be formed. In the United States, state governments are starting to look into creating regulations, such as those for face recognition or self-driving vehicles. However, for the most part, the universities and the companies that create AI systems have been left to regulate themselves.

    True.

16. One of the stumbling blocks in regulating AI is that there is no general agreement in the field on what the priorities for developing regulation and ethics should be. At least some attention should probably be focussed on:
    - Algorithms that can explain their reasoning.
    - Data privacy.
    - The robustness of AI systems to malicious attacks.
    - Bias in AI systems.
    - The potential "existential risk" from superintelligent AI.

    MM states her own opinion that too much attention has been given to the risks of superintelligent AI and far too little to deep learning's lack of reliability and transparency and its vulnerability to attacks. But I would like for you to venture your opinion on prioritizing the consideration of issues surrounding AI. How would you prioritize the focus of attention on these five issues? **Please provide a list of all five elements, ordered from that which believe is the most pressing for consideration to that which you believe is least pressing for consideration.**

    - The robustness of AI systems to malicious attacks.
    - Algorithms that can explain their reasoning.
    - Data privacy.
    - Bias in AI systems.
    - The potential "existential risk" from superintelligent AI.

17. MM poses the question: If we are going to give decision-making autonomy to face-recognition systems, self-driving cars, elder-care robots, or even robotic solders, don't we need to give these machines the same ability to deal with ethical and moral questions that we humans have? **What do you think?**

    I think we would have to give them the same ability. The problem I have with this statement is that we still don't fully understand our ability to answer these ethical and moral questions, so how would we be able to create it in AI.

18. What are Azimov's three "fundamental Rules of Robotics"?
    A robot cannot harm a human. A robot must follow instructions given by humans unless it violates the first rule. A robot must not allow itself to be harmed unless it violates either of the first rules.
19. What was Azimov's purpose in proposing the three fundamental Rules of Robotics.
    To show that a set of moral rules in AI can not work.
20. In Arthur C. Clarke's 1968 book 2001: A Space Odyssey, the artificially intelligent computer HAL is programmed to always be truthful to humans, but at the same time to withhold the truth from human astronauts about the actual purpose of their space mission. HAL, unlike Asimov's clueless robot, suffers from the psychological pain of this cognitive dissonance: "He was ... aware of the conflict that was slowly destroying his integrity – the conflict between truth, and concealment of truth." The result is a computer "neurosis" that turns HAL into a killer. **Please suggest one significant similarity between HAL and the AI Chatbots that are now being unleashed on the world, and one significant difference between HAL and the AI Chatbots that are now being unleashed on the world.**
    Both HAL and the AI chat bots were made to interact with humans for their betterment. HAL had some form of morality, where as the AI chatbots are just running off their algorithms.
21. TRUE/FALSE - The trolley problem has become a kind of symbol for asking about how we should program self-driving cars to make moral decisions on their own.
    True.
22. TRUE/FALSE - In one survey, 76 percent of the participants answered that it would be morally preferable for a self-driving car to sacrifice one passenger rather than killing ten pedestrians. But when asked if they would buy a self-driving car programmed to sacrifice its passengers in order to save a much larger number of pedestrians, the overwhelming majority of survey takers responded that they themselves would not buy such a car. According to the authors, "We found that participants in six Amazon Mechanical Turk studies approved of utilitarian AVs (that is, autonomous vehicles that sacrifice their passengers for the greater good) and would like others to but them, but they would themselves prefer to ride in AVs that protect their passengers at all costs."
    True.
23. TRUE/FALSE - A prerequisite to trustworthy moral reasoning is general common sense, which is missing in even the best of today's AI systems.
    True.