

Question1:

a)From Boddington, first read section 10.2.7 and 10.2.8. Answer exercise 5 from this section

The bio-digital technology implemented during the Covid-19 pandemic to supervise maintaining the physical distance at the workplace causes several ethical concerns from the employees and organizational leaders' standpoint.

Staff Perspective:

- **Privacy Concerns:** Workers might feel that they are being stalked as an AI follows their actions and location to other people.
- **Dignity and Autonomy:** Sustaining such type of monitoring may demotivate the workers as it gives them a feeling that they are being highly supervised.
- **Trust Issues:** Work technology could also have a really negative impact if abused in that it could break the bond of trust that exists between staff and their managers.
- **Unintended Consequences:** Such inaccuracy can result in false readings and added pressure to employees who are considered a Houston company's asset.

Management Perspective:

- **Health and Safety:** AI assists in maintaining social distancing and protects employees from the virus.
- **Efficiency:** It enhances the rate of monitoring; however, this is a consideration of morale among the workers.
- **Management Responsibilities:** For this reason, management needs to be open, respect the privacy of their subordinates, and avoid using the technology for monitoring performance.
- **Dual-Use Risk:** This meant that such an application could be used for other functions apart from the health monitoring uses that were intended, which are ethical issues.

Broader Ethical Concerns:

- **Consequentialism:** Incentive-centered, namely in whether benefits (health and safety) outweigh harms (privacy breaches).
- **Deontological Ethics:** Calls for compliance with workers' rights to privacy and freedom to work.
- **Virtue Ethics:** Determines whether AI is employed to enhance trust, the level of transparency as well as the level of respect given to the employees.

Good and Bad Points:

- Good: Contributes to safety, eradicates spread of diseases, and enhances productivity.
- Bad: Creates violation of people's privacy, the degradation of trust in such systems and potential misuse of the system for supervision purposes.

Thus, they acknowledged the potential of bringing down the rate of workplace dangers due to AI but it was also categorical that it needed to be exerted prudently in the matter of privacy, autonomy and trust. The primary assessments one should designate with respect to brands and their marketing strategy are clarity and ethical decision making.

b) From Boddington, first read section 11.4. Answer exercise 9 from this section.

The thought experiment of uploading your mind to a computer raises several important ethical considerations about the nature of consciousness, identity, and moral rights.

Moral Respect and Rights Retained:

If your mind were successfully uploaded to a computer, the moral respect you would retain might depend on several factors:

1.Sentience and Consciousness: If, due to the uploading the mind possesses sentience (the capacity to feel pain and pleasure, to empathize), it will remain deserving of moral respect. Moral consideration of animals is well emphasized with the involvement of sentience in decision making and moral status of sentient beings.

2.Intelligence and Communication: If the uploaded mind can still exercise intelligent thinking, address language and still be socially responsive, in some way it can be considered to possess some level of Intellectual liberty which might still merit moral dignity. It may be expected that if and only if it is capable of reasoning, has capabilities to form relationships, and hence qualify to be accorded dignity like every human being.

3.Personhood and Continuity: If your mind continues to be yours in the way it is now with your memories, and personality and consciousness – then you might

retain the moral rights and respect that gives you personhood, such as the rights to liberty and personal security, or the rights to privacy and non-interference.

What Might Change:

Several aspects would likely change due to disembodiment:

1. Embodiment and Moral Status: Specifically, reading builds towards the idea that embodiment is at work where there is moral status. Due to that, you may fail to have the feeling of the world and all other sentient beings by lacking the physical body that enables moral duties to be performed in the right way. This could skew your rights with reference to physical injury, as the assailant would no longer be a threat to you in the same way.

2. Physical Vulnerabilities: In that situation, being an embodied personality, you may not think much of your physical condition as a human being. This may no longer be regarded as ethical to physically cause you pain or suffering, though digitally annihilating you or shutting down your existence (e.g. erasing your consciousness) could be.

3. Dependence on the Machine: If you are a result of a brain-computer interface then your life might be at the mercy of the stability and management of that computer system. The moral respect you are entitled to may depend on the rights and personhood possessed by the system governing the machine, and by the acknowledgment of your dis-em-bodied condition by those who operate the system.

Usefulness for Considering AI Moral Respect:

This thought experiment may help to think of the moral respect that may be due to AI, particularly when an AI system appears intelligent communicative and may even be sentient. We might apply the concept of continuing rights to an individual mind if we can imagine a disembodied human mind capable of retaining certain rights, albeit through continued mental experiences, so long as its signifiers are complex, similar treatment should be given to any form of AI showing proofs of sentience or complex interactivity.

For example, if an AI was to have a type of consciousness or an emotional experience then like the uploaded mind it would be deserving of moral respect and certain rights (specifically the right to continue to exist and not be wiped off the face of the earth). The answer to whether AI should be treated with moral concern could be pegged on its ability to reciprocate or to show intent to extend its existence.

In conclusion, augmenting your brain into a computer may bring transformation in the way one perceives moral status, but if consciousness is preserved, then most of the moral

rights and concerns should endure. What is next is similar to the questions we are likely to encounter as individuals as we seek to extend rights to AI as its ability's enhancement.

2a) From the Walker Smith reading (in the Oxford Handbook), read the section starting on p. 677. Address and answer this question (found at the end): “a key ethical question is whether individual human authority—even at the potential cost of other lives—is itself a value that belongs in the technical analysis”. Give examples in your discussion

The basic ethically decisive question that in whose name has authority even if it means potential sacrifice of life as a value is substrata in technical analysis. It primarily concerns the tension between agency and technological options or otherwise including decision-making that might result in enhanced safety or operating efficiency.

Examples and Analysis:

1. **Automated Emergency Braking in Vehicles:** An example given in the reading is the Berlin truck attack of 2016. The vehicle's automatic emergency brake system was said to have saved the lives of many by bringing the truck to a halt within a short distance. However, this system works autonomously and there might be questions regarding the effectiveness of a human input such as when the driver pulled the break lever. Although the system was designed to bring the car to a halt in order to spare the remaining individuals unnecessary loss of life, it did this on its own without reference to the driver's wishes, which might create a conflict between the action of the car, even with driverless technology, and the driver. In this case the technological solution will probably serve the purpose of avoiding more deaths by preventing the truck from continuing its operation. But it raises the question: The possibility arises: should human authority intercede or overrule the system if there is the potential for technology to fail or make a “mistaken” decision? Was assigning human authority over computers potentially helpful a have contributed to the minimization of loss in some situations?
2. **Boeing 737 MAX Crashes:** The crashes of two Boeing 737 MAX planes in 2018 and 2019 are another true example of the tragic fall of the fight between human control and the use of Artificial Intelligence. In both cases, the pilots could not get an automation system to stop trying to force the front of the plane to go down after the

pilots attempted to fix it themselves. This creates the question about should human authorities have been given more credibility when the automated system was not operational. The system in question was intentionally built into the 737 MAX to detect imminent staking and automatically manipulate the angle of the airplane to prevent it. However, the failure of that system with pilots unable to override the system in a timely manner led to worst case scenario. Even these trained pilots have found themselves defenseless by a system that should help them, not overpower them and take full control. There is where the ethical question arises – should we develop such systems that would keep human authority as the ultimate decision-maker when it comes to lives? While the action of the pilots helped one escape the system and perhaps save lives, the technological error did not allow this.

3. **Truck Platooning:** Other examples discussed include an example like automated highways, where several trucks would form a platoon and move in a close formation. The system is likely to not disengage until the cars establish safe distance, which could override a driver's wish to take over in a crisis. As would be expected, the central ethical issue here is whether human control should be subservient to the system's benefit regardless of the distress of the human driver or his/her inability to act during certain incidents. Here, the safety and driving efficiency advantages of platooning—for example, less fuel used as well as more harmonious traffic flow—might be well worth the inconvenience of giving up control to the system. But this raises concerns: to what extent do we celebrate or even allow the human driver to overrule the system, for instance if he will prevent an accident? The question raised the problem of the autonomy of the technological process and the degree of human intervention that is still acceptable.

Conclusion:

This brings us back to the ethical question about the value of human authority decisions over the loss of other... In many instances, automation can be described as a lifesaver, needed for such essentials as automatic brakes or flight controls. Nevertheless, these systems can be faulty or deviant, which points to the necessity to retain human control and command. Consequently, this question boils down to how much automating a decision should be left in the hands of a programmer and how much in the hands of an algorithm. Technology generally has the potential of offering great value but it is not recommendable to rely solely on technology, especially where the issues at stake are critical such as lives of people.

2b) From the Blasimme and Veyena reading (in the Oxford Handbook), read the section starting on p. 711. First, describe an AI system that could be used in public health. Next, describe one potential benefit of such a system and describe one potential problem with such a system.

AI System in Public Health:

One practical AI that may be implemented for public health is a disease surveillance algorithm. This system processes huge quantities of health-oriented information, including stream data from electronic-health, and recording hospitalizations, to tweets. Major potential is that AI can look for the pattern of disease spreading, estimate future trends for EPIDEMICS as well as analyze tendencies of people's health-related behaviors. For instance, one way of utilizing AI is through natural language processing (NLP) where, by scanning information on social media related to food poisoning signs or food contamination, then the food chain could be alerted on possible occurrences. It used also comes in handy in determining the trends of disease in regions/ populations hence quicker identification of threats to public health.

Potential Benefit:

A potential advantage of such an AI system is enhanced epidemic prediction, and, therefore, quicker response. Algorithms that integrated with big data analysis can help AI detect new health patterns much more quickly than other surveillance systems. This means that diseases and outbreaks occur early and can be detected hence early authority responsibility, better resource utilization and intervention implementation. For instance, the AI could easily begin detecting the flu from social media posts before doctors and other health officials identify the flu, thus enabling the use of measures like vaccination or counseling people to avoid crowded places.

Potential Problem:

One of the main drawbacks of applying AI in public health was bound to achieve worse such as resultant of biased data with enhanced health disparities. AI techniques commonly employ the data of previous health, as the base, regarding which the current population is unequal by social and economic status. For example, if data discriminated against low-revenue families or some ethnicity, the AI may build prejudiced models and does not appreciate the requirements of such consumers in terms of health care services. Hence, some communities may be left out in the health interventions, or receive inferior healthcare, thus worsening the inequalities in the health sector. For instance, an AI system

could conclude that people from the lower socioeconomic status do worse after treatments because they have constrained access to medical care or unpredictable work schedules; thus, the AI-based system might advise that these patients not be given treatments at all. It could further entrench asymmetrical patterns in healthcare use and outcomes along various axes of injustice.

3a) From the Surden reading (in the Oxford Handbook), read the section starting on p. 730. Next, explain the problem for AI in the law involving “transparency and explanation”. Give examples in your discussion.

The issue of transparency and explanation in AI within the legal context, as discussed in the Surden reading, revolves around the challenges that arise when AI systems are used in legal decision-making. AI models, especially those used in criminal risk assessments or other legal adjudications, face two primary challenges related to transparency:

Transparency of AI Systems:

- Ideally, AI systems should be transparent, meaning that their decision-making process can be easily understood, audited, and reconstructed. AI models should provide clear insight into why a particular decision was made. In theory, the process of determining outcomes should be deterministic, where the same inputs and models will always result in the same output, making it possible to track and verify the steps that led to a prediction or decision.
- However, in practice, many AI systems, especially those used in legal settings, lack full transparency. For example, many criminal risk-prediction systems are developed by private companies that keep their AI models, data, and algorithms confidential. This secrecy prevents the public or legal experts from fully understanding the underlying mechanisms of AI systems. Such lack of transparency undermines accountability, making it difficult to challenge or appeal decisions made by AI systems, and undermines public trust in their use in legal decision-making.

Interpretability of AI Models:

- Another problem is the interpretability problem. There exist certain AI techniques, e.g., deep learning and neural networks, which make highly accurate decisions, yet cannot make comprehensible sense. These models encode patterns in architectures of computational mathematics which is very hard or even impossible for them to explain or interpret. Lack of documentation

leads to an inability to comprehend or justify a specific move that was made. For example, in criminal risk assessments, if deep learning models are employed to forecast reoffending risk and regardless of the input data (for example, the history, or demography of the defendant), further analysis of the data, may not help explain the AI's conclusion even if the result of a complex data analysis with hidden layers. Understand or explain why a particular decision was made.

- For example, in criminal risk assessments, if an AI system uses deep learning models to predict the likelihood of reoffending, even if we examine the input data (such as a defendant's history or demographics), we may not fully understand why the AI came to a particular conclusion, especially if the decision-making process involves complex hidden layers of data analysis.

Explanation of Legal Decisions:

- One constitutional principle that is common to most legal orders, especially in civil-liberty countries, is that legal decisions especially those affecting citizens' rights, must be made rationally or reasoned and communicated to the public by legal officers. There are also some limitations insofar as interpretation is concerned, and these limitations come in the following scenarios, if judges or other legal authorities makes decisions based on the recommended AI, the parties to the said decision may not be provided with a satisfactory holistic and coherent explanation of the same. Just informing the layman that there are mathematical formulas or statistics to back up an AI decision may not suffice. There is a need for law to be explained in a manner that is acceptable both, socially and legally as well as comprehensible to the society. A defendant's desire of being told a decision was based on an inclusion of an amount arrived at by an AI might not make him or her happy if there is no human decision that makes social and legal sense.
- The reading emphasizes that, while AI systems may provide detailed computational justifications for decisions, they may lack the humanistic elements of legal decision-making. This can reduce the acceptability of decisions made by Artificial Intelligence among the citizens of a given country.

Examples:

- Criminal risk assessments: An example of an AI system is COMPAS, used in the U.S to distinguish the probability of reoffending, yet the prosecution strategy of the AI is usually not made clear to the public. Such lack of

transparency makes it difficult for a defendant, his rights or future decisions to be influenced by an AI-driven risk assessment.

- AI in sentencing or parole decisions: While using the AI developed options to recommend on a parole clause, the complication of the AI models applied poses a real challenge in presenting the specific factors considered by the judge or the defendant relating to the decision made. If the model is black box, it cannot be clear whether the choice was made rightfully and fairly, i.e., using the proper attributes of the applicants.

Thus, the dilemma for AI in the law, especially, as relates to explainability and accountability, is that the opaqueness of the advanced AI models, and the interpretability of some of the applied methods present challenges in making certain that determinations made through the aid of AI in the law are comprehensible, reasonable and understandable and, where necessary, appealable that is, these aspects align with core doctrines of the law.

3b) From the Zeide reading (in the Oxford Handbook), read the section starting on p. 798. Next, explain a major policy issue for personalized learning systems in public education. Give examples in your discussion.

While reading Zeide, the major policy concern that relates to the local implementation of personalized learning systems in public schooling is the struggle between the democratic sovereignty of local governance systems on the one hand, and the obscurity of AI driven personalized learning technologies on the other hand.

Explanation: Adaptive learning systems which are largely information systems that enable learning content and learning environments to be adapted to meet the needs of learners are based on the incorporation of AI. Though this system will certainly improve education since the learner receives specific instruction about what to work on while teachers adjust lessons according to the learners' speeds, the system raises a lot of concerns on issues to do with openness and integrity.

This has resulted in one of the biggest concerns, which is the lack of explainability of how these systems make the various decisions that they make. Unlike conventional pedagogy whereby books and syllabi are analyzed and discussed in the open domain, AI solutions work in closed compartments and are opaque. Because the process of personalization is

based on algorithms, educators, parents, or policymakers cannot fully understand that. On this account, it becomes difficult to assess where and how decisions as to the learning path of a specific learner are being made, thus eroding the democratic values of the public education where parents, local communities, and educators are part of the decision processes on matters affecting the curriculum.

For example, in the case of adopting a set of textbooks for use in teaching and learning, community members have laid down procedures including formal reviews and hearings. But for personalized learning systems, there is no such participatory procedure and openness. A learning system may, for instance, suggest specific educational content knowing the student's data (or preferences, learning modality or past performance) but if the system is opaque, stakeholders will find it hard to comprehend or challenge the basis on which these decisions were made. However, if these systems make decisions such as promoting stereotypes or exclusion of the minority it becomes difficult for the public to counter such aspects since they have no mechanism of understanding the system decision making process. This policy issue involves managing on the one hand the advantages that flow from the application of AI in education; self-paced learning, personalized content, on the other hand, is on the other hand managing the democratic values of openness, participation, and responsiveness in decision regarding education.