**PHIL\*3370 Ethics of Artificial Intelligence**

**Winter 2023 Midterm exam**

**1.**  The continuity thesis is a philosophical concept that states that there is a fundamental constancy between how human beings have thought and acted in the past and how they think and act today. In other words, despite the technological advancements and social changes that have occurred until today, underlying patterns of human behaviour remain constant and can be stated that human beings today are no different that their ancestors. A clear example of the continuity thesis can be seen when we look at the world’s major religions such as Christianity, Islam, Hinduism, and more. These religions have been practiced by millions of people throughout history and today is no different. While the specifics of religious beliefs and practices have evolved to adapt to our changing environment, the basis of them like the connection with a higher power, desire for a community, and others have remained constant.

In general, I believe that the continuity thesis is not plausible. While it is true that human beings have not changed in the fundamental sense and many thoughts/actions are rooted in deep-seated patterns, it is also true that the conditions in which we live as well as certain social norms can shape our behaviour over time. A good example of this is seen in gender equality over time. One article I read about gender roles reads “This demonstrates how gender roles were created based on the needs of a society. However today, the majority of the jobs that are viewed as important, such as being a lawyer, doctor, politician, business executive, etc., are attainable by both sexes. This means that past gender roles should not apply anymore, because both sexes are now equally capable of contributing to society” (<https://www.oneworldeducation.org/our-students-writing/gender-roles-in-modern-society/>). The author explains how just because traditional gender roles have been practiced for so long, they can still be changed since men and women can do many of the same necessary tasks. If the continuity thesis was deemed true, would men and women around the world advocate for equal gender rights? Would the definition of gender have changed into a thought of behaviours and identifications along a spectrum instead of something black and white? Social expectations of each gender changed over time and new movements are in place to promote equality which is where the idea of continuity thesis becomes discontinuous.

**2.** The usage of generative AI such as ChatGPT concerning writing essays comes with it a myriad of positives and negatives and I believe we must evaluate them properly to take advantage of these tools given to us. We have seen in class how false information can be easily portrayed as fact to the user specified in the cases of logical puzzles or even mathematical concepts. These generative AI models will output non-important details that may or may not relate to the question at hand and come to a false conclusion which the user has a high chance of perceiving as fact if not careful. In the case of writing essays, this is detrimental because as a user, you may want insight into a topic in which the AI has a high chance of contradicting itself or flat-out not making logical sense. Another weakness that comes to mind after discussion in class is the quality of work produced by generative AIs. Given how in-depth certain topics especially in philosophy go into, generative AI's are just incapable of providing rich discussions about said topics. At most, the user will receive surface-level points that have no detail and would be unacceptable if used within an essay. Despite this, there are some strengths of generative AI that if utilized correctly, can benefit the user.

Based on my own experience, giving generative AI topics in which I’m not fluent allows me to kickstart the writing process for a given idea. For example, if I do not understand the concept of cognitive bias, I can simply ask a tool like ChatGPT to explain this to me and give me definitions, examples, and other small details. What this allows me to do is be able to brainstorm ideas to write about and think for myself, about how I would go into detail with each. It simplifies the need for a search engine and to filter through results myself but instead gives me a central location to get ideas from. This requires good judgement skills and research to make sure the generative AI is giving accurate definitions but nonetheless, makes it easier to learn about new concepts and gain good talking points for writing essays. It should always be left as an exercise for the user to expand upon the ideas that generative AI’s provide us to write meaningful, content-rich essays.

**3.** The “Double Standard” argument about algorithm transparency refers to the idea that there are discrepancies in the level of critical observation and accountability that is applied to different algorithmic systems and technologies. The computer algorithms that determine a person's credit score or even the type of disease they may have from a set of symptoms have a significant impact on that person's life so it would make sense to know exactly how the algorithm came to its conclusions. This argument leans heavily towards transparency because the inner workings of an algorithm may be kept secret due to trade secrets or intellectual property factors whereas traditional systems are easily accessible and subject to the public. What creeps up due to the lack of transparency in this context is the lack of accountability. Who are we to blame when an algorithm falsely states someone's likelihood of re-offending if we do not know how the algorithm came to that conclusion in the first place? To ensure fair and unbiased results, the “Double Standard” argument says that these systems must be made more transparent so that the decision-making process by the algorithm can be audited, evaluated, and improved.

There is one case where I believe that a double standard could be justified and it has to do with companies competitive advantage. The fact of the matter is that a company that has invested significant time and money to develop a proprietary algorithm will keep it as intellectual property to not harm its competitive advantage in the industry. A clear example of this is generative AI models that are being developed currently. “Like [ChatGPT](https://the-decoder.com/chatgpt-is-a-gpt-3-chatbot-from-openai-that-you-can-test-now/), Bard is an AI optimized for conversation. It is based on [Google’s LaMDA language model family](https://the-decoder.com/googles-ai-model-lamda-could-become-assistant-2-0/), which was first introduced in May 2021. However, unlike ChatGPT, Bard can access the Internet and reference current information and use it for replies” (<https://the-decoder.com/google-launches-chatgpt-competitor-bard/>). Due to the massive success of ChatGPT, companies such as Google are well on their way to developing their version to gain more users. Had the creators of ChatGPT open-sourced their code and data, the rate of innovation would not be as impressive and we as users, would not have had many options. The tradeoff here is that by limiting algorithmic transparency, companies have more incentive to innovate and provide value to their customers while on the other hand, users do not have a clear-cut explanation as to how these algorithms come to their conclusions. Competition is what drives innovation and I believe as a user, I should have multiple options to choose from and be able to evaluate the pros and cons of each to come to my own decision.

**4.** It seems as though the elevator case we discussed in class vindicates these conditions of “knowledge” and “control” for assigning moral responsibility. Knowledge in this context describes how the company that designed the elevator algorithm must know the potential consequences of its actions to be held accountable. This company was aware that its algorithm was being used to determine the location of the elevator which implies that it knew about the edge cases of its malfunctioning. When we look at the control aspect of this situation, we say that the company must have control over the outcome of its actions to take responsibility. In the case of the elevator, the sole company that designed and created the algorithm must have control over its functioning as well as control over its malfunction. The company had both knowledge of the consequences of malfunctioning as well as the control over its design to mitigate it too. Whether the developers of this algorithm did not think of the edge cases of a person needing medical care and therefore did not have the knowledge, or did think of the edge cases but did not implement mitigations, they are still morally responsible for the outcome.

Still, with modern deep learning methods comes an increase in difficulty in assigning moral responsibility when it comes to knowledge and control. One factor is how complex and non-linear these systems are which makes it difficult to understand exactly how they arrived at a particular decision. To go back to the elevator case discussion, one can argue that the lack of transparency therefore knowledge of the algorithm at hand, makes it difficult to assign responsibility for the decision that it took that led to the death of another individual. Another difficulty that arises is how these models are trained and with what types of data. As we know, humans are deeply biased and make mistakes all the time, these same mistakes and biases can reflect in the models that we create since they are trained on these types of data. This leads to discriminatory or unethical outcomes that can be hard to determine who is exactly morally responsible. Are we to blame the developers for not thinking of millions of edge cases that could go wrong? Or what about the entirety of humanity for being biased and unethical at times which bled into the models we are trying to train for our benefit? With increasingly complex systems and the presence of biased data leads to poor knowledge and control within said systems and overall makes it difficult to assign moral responsibility when an unfavourable outcome happens.

**5.** When we talk about the term “social selves” it refers to ourselves perceived by others. This means they are mainly constructed through social interactions and through sharing information about one’s self. Privacy in this case can be looked at as controlling the information that we share about ourselves with others which ultimately determines how we are perceived as well. An example of this can be seen with people and their anonymous social media profiles. They may have a different persona online than in person and they decide to separate the two personalities to protect their privacy and portray themselves differently. However, in today's digital era with massive tech companies at the forefront of collecting data, people demand the need for privacy as it impacts the formation and expression of one’s social self.

Contemporary data collection practices seem to be threatening this relationship between social selves and privacy due to a few reasons. With being online having a big impact on one’s ability to function in society, it is the biggest target for larger companies to threaten this needed privacy whether that be directly or indirectly. A personal example of this that comes to mind has to do with online shopping and targeted advertisements. For most shopping purposes, I will use an online platform such as Amazon or AliExpress, both of which collect my information on a variety of factors. Things like my browsing history and purchases get collected and are used to then deliver targeted advertisements or even share them with other organizations. The most difficult part of this is that all of these platforms require you to sign up with an account in which case make you agree to their terms. Of course, the logical answer to this is to disagree with their terms but at that point, I would not have access to the platform and this is where it seems like bad faith in the company. Why should I have to agree with a big corporation to track my every move on the application just for me to use it? This also ties directly with the real-world discussions I will have with others about a product as it is very common to see similar product advertisements online after the fact which violates my privacy even more. Personally, the knowledge that my shopping history and even discussions in the real world are being collected and potentially shared leads me to self-censor myself and reduce the information that I’m willing to share as a whole and this ultimately limits my ability to make informed decisions about my purchases.