Name: Avery Vallee-Vezina

Professor: Lucas Jurkovic

Course: Values in the Workplace (345-CXA-HR)

Due Date: April 26th

A programmer’s dog is called AI

Should AI be allowed in the workplace? In this paper I will argue for AI since it is arguably the next standard toolset modifying the way people handle tasks. When we consider how people have been using AI, it makes sense to adopt the standard. It can also perform some tasks in a fraction of time that a human counterpart would. But its also important to know that training an AI from scratch is in some way, comparable to training a dog. Ai, like a dog, in its young days requires a lot of training, to make sure that it serves its purpose and can do the role it's supposed to do. A dog can be given bad training, or a dog can learn unintentional lessons from its teacher, and either of these can make the dog not able to perform its role, and in the exact same manner AI can run into those issues. To be able to guard against those, the solution is the same as in the person training the dog, making certain that they're trained and educated in the topic.

As we’re developing AI to automate sensitive tasks about people’s information, we need to consider potential issues. Nick Bostrom and Eliezer Yudkowsky wrote in their paper “*The Ethics of Artificial Intelligence*” about an algorithm that would automatically evaluate a mortgage application. The AI would reject people who lived in poor areas which introduced a racial bias. Discovering the cause of these issues is often impossible to trace in similar AI algorithm so consequently “It will become increasingly important to develop AI algorithms that are not just powerful and scalable, but also *transparent to inspection*-” (*source 1*). This being mentioned in 2011 has become more important in the last decade, with the introduction of publicly accessible Large Language Models (LLM) like ChatGPT. It led to amazing discoveries on the capabilities of the technology but also shed light to other issues. It was discovered that you could trick the LLM to divulge dangerous information, like for example a recipe to make chlorine gas. Designing the algorithm to be transparent allows us to make dangerous information less accessible.

Being conscientious of these issues means that the people developing these LLM’s can follow guidelines to minimize risks. The BigScience RAIL License stipulates that unforeseeable new uses can arise. Currently the concerns are about cyber terroristic threats amongst other risks. BigScience initiative proposes a collaborative act for research to help mitigate some of these issues in a transparent way due to “legitimate concerns about the impact of this technology on society and our planet are being raised, including the need for AI fairness, transparency, explainability and robustness, as well as addressing issues related to privacy, accountability, addiction, manipulation, and misuse” (source 2). All these concerns justifiably scare many people in the world, but its important not to forget the purpose we have for AI. It is a tool to enhance human intelligence and that is a great thing' (*source 3*), and as of writing this paper, the current Google analytics show a consistent reference towards AI and its hypothetical ability to revolutionize major industries such as healthcare, financial services, or any new environments (*source 4*). However, the future is quite unpredictable so only time will tell.

In this paper so far, I’ve presented several ethical issues revolving with AI. The reality of the situation is, we need to be aware about these issues to be able to shape AI’s future properly. If we are successful in introducing a sort of ethics system, we will be able to fully benefit from its potential. For example, thanks to AI today, we replaced some of the old standards for animal testing during drug trials to using a small lab grown biochip instead (*source 5)*. It is said to mimic conditions in the human body to allow safer and more accurate testing, which is a step in the right direction. This is considering the alternative is less effective and has harmed animals in the past. It’s safe to argue that AI powered technologies like this one has a place in our future, but how about AI in a more common workplace? When introducing it as a tool and give staff a basic training in AI, we can expect improved task efficiency and overall increased productivity. Imagine the following scenario, an employee takes 20 minutes to make a report for a user. Instead of wasting 20 minutes making each report, the company introduced an AI to automate that task. Consequently, the algorithm will complete all the reports in a handful of minutes whilst the employee is busy working on more important tasks. With how applicable this scenario is, its safe to assume AI will present itself in all sorts of useful purposes.

When approaching the future of AI, it is beneficial to take a situationist perception. They believe that human behaviour is strongly influenced by external situational factors. For instance, people are likelier to view someone in a lab coat as a credible source because they associate lab coats to doctors and scientists which manifests itself as an authority bias. But how does this tie down to AI? Just like the analogy in the first paragraph of this paper, training AI is like training a puppy. When training either, you want to reduce distractions (or bad teachings) to ensure they can perform their assigned tasks. In AI that might present itself when the public tries to use the LLM and prompt bad entries or just by not verifying the training data before using it. This can influence the LLM and introduces issues like bias.

We have a right to be afraid of some future advancements. This fear though, does not give us the right to stop trying. Consider the Moon landing in 1969; the potential risks were said to be endless, but we still pulled through! Just because an idea is terrifying, doesn’t mean we shouldn’t do it, rather we just need to break it down and approach it knowledgably. At the end of the day, the worry of AI taking over is just an idea, not a prophecy. We already have measures in place to make it more reliable, robust, powerful, and perhaps most importantly we’re starting to implement transparency in these AI models. This is not a call to stop it’s use, but rather to highlight the basic ethical consideration of using it in any workplace. Only after understanding the considerations, can we make a beneficial future with the help of Artificial Intelligence.

**Sources**

* Source 1: The Ethics of Artificial Intelligence (2011)   
  By: Nick Bostrom and Eliezer Yudkowsky  
  Link: <https://nickbostrom.com/ethics/artificial-intelligence.pdf>
* Source 2: BigScience  
  Collaborative open science initiative  
  Link: <https://bigscience.huggingface.co/blog/the-bigscience-rail-license>
* Source 3: IBM  
  Link: <https://www.ibm.com/downloads/cas/035WYAWQ>
* Source 4: NESS
* Link: <https://www.ness.com/the-future-of-ai-how-adaptive-ai-will-transform-the-industry>
* Source 5: Biochip  
  <https://www.biodot.com/news-resources/biochip-technology-a-safe-and-economical-alternative-to-animal-testing>