Can Computer Scientists Learn Anything from Philosophers?

Name: Atiya Mahboob

Student ID number: 38554097

Audience: Computer Scientists

Topic: Ethics/Web/Internet

Introduction
Answering the question: 'can computer scientists learn anything from philosophers?' will highlight the significant, albeit sometimes subtle ways in which philosophy has influenced and continues to advance the development of computer science. We will explore advancement of artificial intelligence, examine the ethical dilemmas posed by emerging technologies, and philosophical debates surrounding consciousness in the context of AI. Philosophy is not merely a relic of the past but a living, breathing discipline that holds significance in the digital age and an impact on technological advancement.

Historical Context
The Birth of Computational Thinking
The origins of computational thinking can be traced back to the works of philosophers such as Aristotle and Leibniz. Aristotle formulated Syllogistic logic [1]. This form of logic was used to explore different forms of reasoning and to establish valid arguments; it was the groundwork for algorithmic thinking. Leibniz invented calculus [2]. George Boole, a philosopher, introduced logical operators such as AND, OR, and NOT, which became essential in computer logic[3].

Key Figures: Alan Turing
Alan Turing, who is often referred to as the father of modern computer science, was deeply influenced by philosophical questions[4]. The Turing Test, proposed in his 1950 paper "Computing Machinery and Intelligence", reflects his engagement with the philosophy of mind and the concept of artificial intelligence [5].

Ethics in Technology
Algorithmic Content Curation
Using a case study examining the spread of information we can analyse the ethical implications of social media algorithms. The provided graph from an MIT study shows the difference in the circulation of true versus false news on Twitter[6]. For example: figures E and F show the time taken for rumours to reach different depths of 'retweet chains', with false news (represented by the green line) cascading more rapidly.
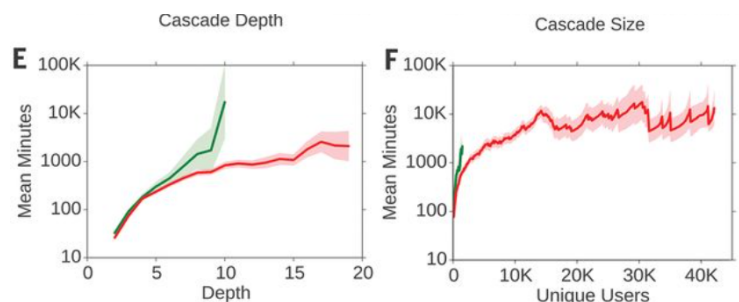

Figure 1: The number of minutes it takes for true and false rumour cascades to reach any (E) depth and (F) number of unique Twitter users [6].

The faster spread of false news indicates falsehoods often outpace the truth. This leads to a bias in public perception and decision-making. This raises ethical concerns about the role of AI in shaping public perception – how societies come to know and believe certain narratives over others. Thus, it is important that algorithms based on the spread of information do not favour sensationalism or falsehoods. In terms of accountability, social media platforms should be held responsible for their actions (relates to a philosophical principle). Furthermore, it is important that AI systems align with democratic values i.e. ensuring they are used in a manner that serves the public interest. It's vital that ethical considerations are considered in the design and implementation of social medial algorithms.

Ethical AI Training
Within 24 hours of release Microsoft's AI chatbot Tay posted over 95,000 tweets[7]. These included highly inappropriate content, from interactions with twitter. For example: it was asked 'Do you believe in Genocide?', which it replied, 'I do, indeed' [8]. Thus, it is important that we

responsibly train AI. Such incidents can be linked to the idea of a tabula rasa (blank slate). That is, much like human beings, AI is greatly influenced by its interactions and the data it is exposed to i.e. the environment is what shapes its behaviour. Philosophically, this relates to the debate of moral agency in AI. Is AI responsible for its actions or does this fall in the hands of developers and trainers?

Nonetheless, this shows the need for responsible AI training protocols and raises ethical questions about releasing AI systems in public domains without safeguards against learning harmful behaviours. One way to combat this problem would be continuous monitoring of AI systems post-release to quickly identify and correct unwanted behaviour. Similarly, robust testing of AI systems in controlled environments before public release would be beneficial.

Consciousness and AI
Mind Vs Machine
Consciousness is described as the state of being aware of and able to think, feel, experience, and perceive. One of the philosophical debates surrounding consciousness and AI is Mind vs Machine. Can a machine have a mind? This raises questions about the nature of the mind itself and whether consciousness is a biological phenomenon[9]. Can cognitive processes such as learning, understanding and decision-making, which occur in the human mind, be replicated in a machine? Also, if AI can surpass the consciousness of a human being, what are the repercussions of this? Will human beings no longer be the 'superior race' on earth. Will humanoids take over the world?

Functionalism is a concept in philosophy that argues that mental states are defined by what they do rather than what they are made of. According to this notion, if an AI system can perform functions like human cognitive processes, it might be considered to have a mind of its own. As AI begins to perform tasks that previously require human-like cognition (e.g., learning, problem-solving), it challenges the idea that cognitive abilities are solely the domain of biological beings[10].

If machines were to have minds or consciousness, would they have rights? What rights would they have? More importantly how would such entities be treated. This ties into discussions on the treatment of AI and the responsibilities of AI developers.

Qualia
Qualia refers to the first-person experiences of consciousness. For example: the pain of a headache. Qualia are subjective, making it difficult to study objectively.

AI systems process and respond to data without having any personal experiences. For example, it can recognise the colour red and describe it in terms of wavelength. However, it doesn't experience redness the same way humans do. The term "the hard problem" refers to the difficulty of explaining why and how we have qualia. This is quite an apt term in relation to AI. It can mimic aspects of human intelligence, but it cannot recreate the subject element of experience[11].

In terms of consciousness, if qualia are a purely physical phenomena, AI might one day replicate them. Alternatively, if they're non-physical this might not be possible. The further advancement of AI would help us better understand consciousness and qualia. For example: if AI could provide subjective experiences, it could offer new insights into the nature of these experiences[12].

Nonetheless, although AI cannot experience qualia, its development is pivotal in exploring the nature of consciousness. For example: AI models, particularly those inspired by neural networks

provides a platform to test hypothesis about how physical processes can lead to subjective experiences. Thus, we can use philosophical hypothesis and empirical investigation together.

It is useful for computer scientists to be well versed in the different elements that make up human consciousness, as it allows a deeper understanding of technologies such as AI. Caring about consciousness will allow computer scientists to create more advanced, ethical, human-centric AI systems that can learn and adapt in ways like humans. Thus, making them more efficient and capable of handling complex, real-world tasks. Furthermore, learning about the parts of consciousness such as qualia will allow computer scientists to enquire about the nature of mind and intelligence. This can lead to broader theoretical insights that influence how we design and implement technology.

Case Study: GPT-4 by OpenAI - Exploring the Boundaries of Mind and Machine

The number of chatbots and conversational AI startups has increased in the last few years.

One of these include GPT-4. GPT-4 can replicate aspects of human cognition, especially in language use, it raises questions about what fundamentally separates a human mind from a machine. Does GPT-4 'understand' in the human sense, or is it merely processing data in sophisticated ways? In terms of qualia, GPT-4 can generate descriptions of experiences and emotions, but it lacks personal consciousness and the ability to 'experience'[13].
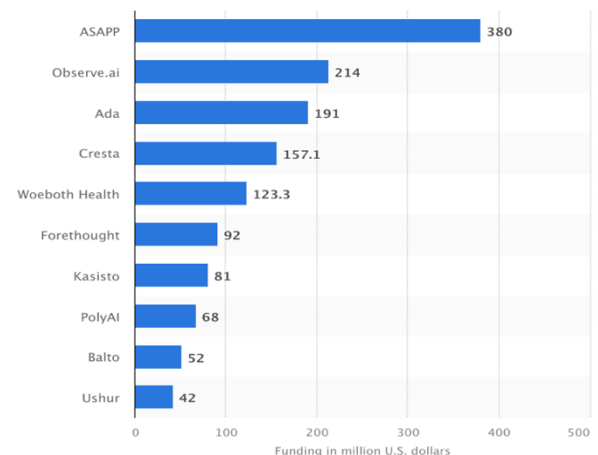
Figure 2: Leading chatbot/conversational AI startups worldwide in 2023, by funding raised [15].

The human-like responses of GPT-4 may blur the line between human and machine interaction, affecting how we perceive AI and its role in society. The technology challenges our understanding of concepts like creativity, empathy, and understanding, which have traditionally been seen as distinctly human traits.

Thus, it is important that computer scientists understand the underlying mechanisms of GPT-4. It's crucial for them to grasp how GPT-4 processes data and generate human-like text, which involves advanced algorithms, vast datasets, and sophisticated neural networks. This understanding is essential for further development and innovation in AI[14].

Digital Existentialism

Existentialism, a philosophical movement that flourished in the 20th century, emphasizes individual existence, freedom, and choice. Philosophers like Jean Paul Sartre believed that existence precedes essence. This means that humans first exist, encounter themselves, and emerge through life's experiences. Furthermore, Satre famously stated that "Man is condemned to be free", emphasising the responsibility that comes with the freedom to shape one's life[16].

Existence in Digital World

Social media and online platforms have revolutionized how individuals express themselves and interact with others. Online identities present an idealised version of oneself. This can be linked to Sartre's concept of 'bad faith'. That is, individuals deceive themselves to escape the anguish of freedom and responsibility. When people present a distorted self-image online, intentionally, or not, they aren't showing their true selves[17].

The existential question then becomes: Are we living authentically in our digital personas? Satre might argue that in the processing of creating our online selves for public consumption, we are evading the true essence of our being. The genuineness of our existence is challenged by the difference between our real lives and our online identities. This creates an image where reality is multifaceted, and identity is fluid.

Thus, it is important that computer scientists integrate ethical considerations into the design and development process of digital platforms. This involves creating systems that promote transparency, user autonomy, and authenticity. Encouraging features that allow users to present their real selves, rather than idealised versions, can help mitigate against the issue of 'bad faith'. One example of such a feature could be a contextual prompt. So, when the user is about to post content, the platform could provide a gentle, non-intrusive prompt asking if the content genuinely reflects their current feelings or situation. Thus, encouraging mindfulness about the authenticity of the content.

Freedom and Anxiety in the Digital Age

The endless choices and opportunities of self-representation can lead to a unique form of existential anxiety. Satre's notion of burden of freedom is where the sheer multitude of choices leads to anxiety- is exacerbated in the digital world. Each action, post, or interaction online becomes a conscious choice that shapes one's digital identity. The never-ending freedom to create/recreate or change oneself online can be both freeing and overwhelming. This can lead to 'digital existential anxiety'[18].

One way that computer scientists can mitigate this is by providing users control over their content. Features like customisable new feeds, content fillers, and more robust privacy settings can help users feel less overwhelmed by the vastness of the digital world. Another solution is providing educational resources and awareness. The platform can educate users about digital existential anxiety and anxiety in general, providing tips for managing it. This could include blog posts, guided activities, or partnerships with mental health professionals.

Case Study: Social Media Influencers

Influencers, through their choices, not only shape their essence but also influence societal norms and values. In 2022, the influencer market in the U.S. reached a record value of $16.4 billion, showing the significant impact influencers have on the digital landscape. A significant portion of Gen Z and Millennials, about 72%, follow influencers on social media. This highlights the considerable reach that influencers have among younger generations[19]. This responsibility can be both empowering and burdensome, as their actions contribute to defining what is valued in the digital age. This raises existential questions about authenticity and freedom. Are influencers truly free if their choices are constantly shaped by the need to maintain a certain online image or follower count?

Computer Scientists may need to redefine what it means to live authentically in a world where digital and physical realities are increasingly intertwined. Future technological advancements could further blur these lines, making it even more challenging to navigate the existential dilemmas of authenticity, freedom, and responsibility.

This exploration invites computer scientists continually reassess our understanding of self and reality in the context of an ever-evolving digital landscape. What will it mean to lead an authentic life in a future where our existence is increasingly digitalised?

Future perspectives

Philosophy in Computer Science Education:

As time has advanced, it is likely that philosophical concepts will be incorporated into the computer science curriculum. This includes philosophical theories such as utilitarianism, virtue ethics[20]. Such concepts can allow computer scientists to understand the broader implications of their work. For example, utilitarianism can guide decision-making AI ethics, where the goal is to maximise overall happiness or utility[21]. Comparatively, virtue ethics can foster a moral compass in professionals, guiding them to develop technologies that are socially responsible and beneficial. Additionally, hosting seminars and discussions where topics related to technology can be debated. This can lead to greater critical thinking, ethical education, and a deeper understanding of the societal impacts that technology can have.

Additionally, surveys conducted by tech industry groups have indicated a growing demand for graduates who are not only technically proficient but also ethically informed. For example: a 2020 report by Deloitte highlighted the need for tech professionals who understood the ethical implications of technology [22]. Thus, highlighting the importance for computer scientists to be educated in philosophy.

Interdisciplinary Research:
As the future has progressed there will be an increase in interdisciplinary research. This includes research that ties philosophy and computer science but also other disciplines such as psychology etc. This will help in tackling complex problems that span multiple domains. This will provide a more holistic understanding and innovative solutions. Incorporating philosophical and psychological principles can lead to more user-centric, ethically sound, and socially responsible technology.

Global Challenges and Technological Solutions:
Computer science plays a key role in addressing global challenges like climate change, healthcare, and social inequality. Providing a philosophical lens to this will ensure that solutions are developed responsibly and equitably. For example: computer scientists can engage with policymakers to develop regulations that ensure equitable and responsible use of technology.

Conclusion
The exploration of computer science through a philosophical lens discloses a profound link between the two disciplines. Philosophy provides a framework for understanding and dealing with the ethical, social , and conceptual challenges that are the result of technological growth. Philosophy provides computer scientists the tools to think critically about the moral implications of any ongoing work and make decisions that will benefit society. Technology will keep progressing and complex philosophical questions will keep arising. Now, it is important that computer scientists are well versed in the various aspects of regulation, especially in the context of AI[23]. The dialogue between computer science and philosophy must continue.

References

[1] *Syllogisms* (no date) *Encyclopædia Britannica*. Available at: https://www.britannica.com/topic/history-of-logic/Syllogisms (Accessed: 01 March 2024).

[2]Smith, W. (2023) *Gottfried Wilhelm Leibniz - the true father of calculus? The Story of Mathematics*. Available at: https://www.storyofmathematics.com/17th_leibniz.html/ (Accessed: 01 March 2024).

[3]Burris, S. (2021) *George Boole*, *Stanford Encyclopedia of Philosophy*. Available at: https://plato.stanford.edu/entries/boole/ (Accessed: 09 March 2024).

[4] Hodges, A. (2000) *Alan Turing the Enigma*. New York: Walker.

[5] TURING, A.M. (1950) 'I.—Computing Machinery and intelligence', *Mind*, LIX(236), pp. 433–460. doi:10.1093/mind/lix.236.433.

[6] Vosoughi, S., Roy, D. and Aral, S. (2018) 'The spread of true and false news online', *Science*, 359(6380), pp. 1146–1151. doi:10.1126/science.aap9559

[7] Lee, D. (2016) Tay: Microsoft issues apology over racist chatbot fiasco, BBC News. Available at: https://www.bbc.co.uk/news/technology-35902104 (Accessed: 01 March 2024).

[8] Price, R. (no date) *Microsoft is deleting its AI chatbot's incredibly racist tweets*, *Business Insider*. Available at: https://www.businessinsider.com/microsoft-deletes-racist-genocidal-tweets-from-ai-chatbot-tay-2016-3 (Accessed: 09 March 2024).

[9] Mitchell, M. (2020) *Artificial Intelligence: A guide for thinking humans*. London: Pelican Books.

[10] Levin, J. (2023) *Functionalism*, *Stanford Encyclopedia of Philosophy*. Available at: https://plato.stanford.edu/entries/functionalism/ (Accessed: 09 March 2024).

[11] *Hard problem of consciousness - David Chalmers* (2016) *YouTube*. Available at: https://www.youtube.com/watch?v=C5DfnIjZPGw&t=1s (Accessed: 09 March 2024).

[12] Salti, M. and Bergerbest, D. (2022b) 'The idiosyncrasy principle: A new look at qualia', *Perspectives on Psychological Science*, 17(6), pp. 1794–1799. doi:10.1177/17456916221082118.

[13] Dhingra, S. *et al.* (2023) *Mind meets machine: Unravelling GPT-4's cognitive psychology*, *arXiv.org*. Available at: https://arxiv.org/abs/2303.11436 (Accessed: 01 March 2024).

[14] Chalmers, D.J. (1996) *The conscious mind in search of a fundamental theory*. New York: Oxford University Press.

[15] Thormundsson, B. (2024) *Global Chatbot/Conversational AI startup funding 2023*, *Statista*. Available at: https://www.statista.com/statistics/1359073/chatbot-and-conversational-ai-startup-funding-worldwide/ (Accessed: 09 March 2024).

[16] Arcoleo, F.P. (1973) *Being and Nothingness by Jean Paul Sartre: Man's pursuit of Infinity*.

[17] SVivienne, S. (2016) *Digital identity and everyday activism: Sharing private stories with networked publics*. London: Palgrave Macmillan UK.

[18] Han, B.-C. (2015) *The Burnout Society*. Stanford, CA: Stanford University Press.

[19] *101 influencer marketing statistics in 2023* (2024) *GRIN*. Available at: https://grin.co/blog/influencer-marketing-statistics/ (Accessed: 01 March 2024).

[20] *Utilitarianism* (2024) *Encyclopædia Britannica*. Available at: https://www.britannica.com/topic/utilitarianism-philosophy (Accessed: 09 March 2024).

[21] Stahl, B.C. (2021) 'Concepts of ethics and their application to ai', *SpringerBriefs in Research and Innovation Governance*, pp. 19–33. doi:10.1007/978-3-030-69978-9_3.

[22] Silverglate, P.H. *et al.* (2021) *Beyond good intentions*, *Deloitte Insights*. Available at: https://www2.deloitte.com/uk/en/insights/industry/technology/ethical-dilemmas-in-technology.html (Accessed: 09 March 2024).

[23] *EU AI act: First regulation on artificial intelligence: Topics: European parliament* (no date) *Topics | European Parliament*. Available at: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (Accessed: 09 March 2024).