Sadiq Ali - Discuss the societal risks and rewards associated with generative AI (such as ChatGPT)

In the intricate dance between human ingenuity and the digital avant-garde, the emergence of generative AI becomes more than just a technological leap; it becomes a symphony of possibilities awaiting orchestration. As Toby Ord somberly notes, "For the first time in humanity's extensive history, we now have the capacity to destroy ourselves entirely, severing our future and everything we could become." [1] This observation serves as our guide through the maze of progress we see presently. Amidst the shifting landscape of data and algorithms, this essay ventures into the nuanced terrain of generative AI, navigating the subtleties between accelerated scientific knowledge and improving social sagacity. Beyond conventional narrative, I will explore the mysteries of the ORCH OR theory, consider the philosophical implications of the Chinese Room Paradox, and scrutinise the delicate interplay between pedagogy, creativity, and AI.

In grappling with the societal implications of AI, the ORCH OR theory stands out as a captivating guidepost. Presented by physicist Sir Roger Penrose and anesthesiologist Dr Stuart Hameroff, this theory explores the roots of consciousness. It puts forward that quantum computations within microtubules in brain cells contribute to our subjective awareness. [2] While its application to artificial intelligence remains speculative, it unveils a philosophical discussion regarding consciousness and whether machines can in fact, genuinely possess a form of self-awareness. Furthermore, as generative AI enters domains like education, journalism, and content creation, the ORCH OR theory urges us to scrutinise its potential impact on human cognition and creativity. It challenges us to reflect on whether the pursuit of machine sentience might inadvertently reshape our societal fabric, influencing our perception and interaction with information, art, and one another.

The ‘Chinese Room Paradox’, introduced by philosopher John Searle, adds another layer of complexity to our exploration. This paradox challenges the idea that a computer, merely following instructions, can truly understand or possess consciousness. Searle's argument prompts us to question the nature of intelligence and whether generative AI, like ChatGPT, can genuinely grasp the meaning of the information it processes. In the Chinese Room scenario, a person in a sealed room follows instructions to manipulate Chinese characters without understanding the language itself. [3] Searle uses this to argue that, similarly, a computer may process information without comprehending its meaning. This philosophical inquiry intertwines with the societal risks and rewards associated with generative AI, adding depth to the ongoing discourse.

It seamlessly intertwines with the realm of pedagogy, a domain where the transformative influence of AI becomes visible. Inviting a nuanced reflection on the potential implications of integrating generative AI into higher education, especially in University education.

‘By 2023, the global AI in education market is expected to reach approximately $3.68 billion’ [4]

As universities grapple with the integration of new and improving technologies into education, a dilemma emerges, should they embrace or ban generative AI tools? ChatGPT itself is a conversational AI that uses Natural Language Processing (NLP), communicating with the user, and even ‘answers follow-up questions, admits its mistakes, challenges incorrect premises, and rejects inappropriate requests’ [5]. This predicament mirrors historical debates surrounding technological innovations, as faced during the industrial revolutions throughout the 18th and 19th centuries. In much the same way that the calculator

transformed mathematical computations, generative AI tools have found their way into higher education, offering both promises and pitfalls.

Some argue that embracing ChatGPT as an essay-writing tool parallels the integration of calculators into mathematics education. These tools, when used responsibly, can enhance creativity and push the boundaries of imagination, especially in design education where novel tools historically sparked innovation. However, this is not straightforward for educators. Concerns about academic integrity, plagiarism, and the authenticity of students' work have prompted some institutions to ban generative AI tools. Schools however are not being provided sufficient guidance in the form of rules and advice regarding the uses of generative AI applications in education. ‘Some 40% of the educational institutions that reported having guidance, said the guidance was not written and had only been communicated orally,’ [6] further illustrating the ad-hoc nature of policy responses in education.

New AI plagiarism detection systems, such as anti-ChatGPT and GPTZero, signal an attempt to address these concerns. However, the fine line between using generative AI as a support tool and relying solely on it for academic tasks raises questions about the essence of learning and creativity in education - how do we strike a balance? Moreover, as generative AI improves, the ability to detect their use becomes a more difficult feat. Compared to traditional chatbots, ChatGPT is based on GPT-3, which is the third of the GPT series that is more advanced in terms of scale (175 billion parameters, compared to 1.5 billion in GPT-2) [7]. This fails to consider GPT-4, the next iteration, with a larger dataset, ~2 trillion parameters, enhanced capabilities, and more human-like text generations. In fact, compared to GPT-3, GPT-4 has scored 40% higher on its internal factual performance benchmark [8]. It is clear that generative AI are improving rapidly, however, as aforementioned, the comparison with

calculators, which did not render the teaching of mathematics redundant, highlights the need for a nuanced approach.

In the educational landscape, the Chinese Room Paradox prompts reflection on the essence of learning and creativity. Can AI, even if capable of sophisticated responses, replace the intrinsic human qualities of critical thinking, curiosity, and empathy? The paradox cautions against over-reliance on AI tools, encouraging this necessary balanced approach to their integration into higher education. While generative AI may provide valuable support, the Chinese Room Paradox invites educators and institutions to consider the irreplaceable role of human insight, intuition, and genuine understanding in the learning process.

Yet, the ethical landscape of generative AI tools introduces complexities not seen with calculators. ChatGPT, immersed in biased datasets, runs the risk of unexpectedly accentuating harmful biases and stereotypes, prompting ethical concerns regarding its application in education. Additionally, lingering uncertainties surround issues of copyright infringement and the legal validity of training such tools on data curated from the internet. ‘Information about the functionality of algorithms is often intentionally poorly accessible’ [9] and this exacerbates the legal problem surrounding the lack of algorithmic transparency.

The debate within academia intensifies as some, such as the academic Jim Clack, argue that generative AI ‘challenges the very concept of academic integrity’, while others emphasise its potential as an educational tool. UNESCO [10] suggests that universities should proactively explore how to use AI tools as part of the curriculum, integrating lessons on AI ethics and skills. The delicate balance between leveraging the benefits of generative AI and maintaining academic integrity calls for clear guidelines developed collaboratively with students

themselves. This conversation around generative AI in higher education prompts reflection on the essence of human abilities - critical thinking, curiosity, empathy and imagination, amongst others- that define the process of pedagogy. Generative AI, much like calculators in mathematics, may reshape the landscape but highlight the importance of preserving the innately human aspects of learning and creativity.

Creativity unfolds as a dynamic interplay in the realm of artistic creation, with the integration of artificial intelligence (AI) exemplified by cutting-edge tools such as DALL-E. Developed by OpenAI, DALL-E is a Generative Adversarial Network that has been trained on a diverse range of images and textual descriptions [11]. Its functioning involves two neural networks - the generator and the discriminator. The generator crafts images from textual prompts, while the discriminator evaluates these images for realism. Through iterative learning, DALL-E refines its ability to generate novel and coherent images in response to textual inputs from users. This sets the stage for a nuanced equilibrium between societal risks and rewards, as AI-driven platforms present a myriad of opportunities. On one hand, they emerge as powerful tools that complement and amplify human creativity. They serve as a collaborative partner, in the context of DALLE-2, sparking inspiration and extending the boundaries of conventional artistic expression. Nevertheless, the integration of AI into the artistic realm raises valid concerns about the potential homogenisation of creativity. The very essence of human artistry lies in its rich diversity, informed by unique perspectives, emotions, and experiences. Although AI can effectively recreate various art styles and produce visually convincing outputs, there exists a genuine risk that an overreliance on these tools may lead to a dilution of the deeply personal and subjective aspects that truly define art. Striking a delicate balance becomes paramount when navigating the evolving landscape of AI enhanced artwork, preserving the authenticity and integrity of human expression.

History suggests that humans consistently seek methods that are faster, easier, more efficient, and convenient to complete their tasks. [12] Consequently, the drive for progress continues to inspire humanity to explore novel and improved approaches to various endeavours. Upon realising that tools could alleviate numerous challenges in daily life, human inventions allowed for accomplishment of tasks with greater efficiency, speed, and intelligence. Creativity and innovation serve as a catalyst for human progress. Yet, as AI exceeds its capabilities year by year, what was previously a piece of science fiction creeps into reality, and its risks are accentuated. Generative AI has a strong capability to supplement human progress, whilst simultaneously bearing the ability to destroy and detriment it.

‘I strongly believe that given the technologies we are now developing, within a century or two at most, our species will disappear. I don’t think that in the end of the 22nd century, the Earth will still be dominated by Homo sapiens.’ [13]

References:

[1] Paraphrased excerpt from ‘The Precipice: Existential Risk and the Future of Humanity’, Toby Ord (2020)

[2] Hameroff, Stuart, 'Orch OR and the Quantum Biology of Consciousness', in Shan Gao (ed.), Consciousness and Quantum Mechanics (New York, 2022; online edn, Oxford Academic, 20 Oct. 2022) - URL: https://doi.org/10.1093/oso/9780197501665.003.0015

[3] "The Chinese Room Argument", The Stanford Encyclopedia of Philosophy (Summer 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.) -URL: https://plato.stanford.edu/archives/sum2023/entries/chinese-room/

[4] Al in Education Market by Technology (Deep Learning and ML, NLP), Application (Virtual Facilitators and Learning Environments, ITS, CDS, Fraud and Risk Management), Component (Solutions, Services), Deployment, End-User, and Region - Global Forecast to 2023 - URL: https://www.marketsandmarkets.com/Market-Reports/ai-in-education-market-200371366.html

[5] OpenAI. (2023). ChatGPT: Optimising language models for dialogue. - URL: https://openai.com/blog/chatgpt/

[6] UNESCO survey: Less than 10% of schools and universities have formal guidance on AI - URL: https://www.unesco.org/en/articles/unesco-survey-less-10-schools-and-universities-have-formal-guidance-ai

[7] A Complete Comparison of ChatGPT, GPT-3, and GPT-4: What's the Real Difference? - URL: https://simplified.com/blog/ai-writing/chatgpt-vs-gpt-3/

[8] Language Models are Few-Shot Learners - URL: https://arxiv.org/abs/2005.14165

[9] The ethics of algorithms: Mapping the debate - URL: https://journals.sagepub.com/doi/full/10.1177/2053951716679679

[10] ChatGPT and artificial intelligence in higher education: quick start guide - URL: https://unesdoc.unesco.org/ark:/48223/pf0000385146.locale=en

[11] OpenAI. (2023). DALLE-2 - URL: https://openai.com/dall-e-2

[12] The impact of artificial intelligence on human society and bioethics - URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7605294/

[13] Yuval Noah Harari - URL: https://medium.com/@podclips/8-yuval-noah-harari-quotes-that-will-make-you-fear-the-future-a40d4a0bc1f1