

AI and the future of humanity: ChatGPT-4, philosophy and education – Critical responses

Michael A. Peters, Liz Jackson, Marianna Papastephanou, Petar Jandrić, George Lazaroiu, Colin W. Evers, Bill Cope, Mary Kalantzis, Daniel Araya, Marek Tesar, Carl Mika, Lei Chen, Chengbing Wang, Sean Sturm, Sharon Rider & Steve Fuller

To cite this article: Michael A. Peters, Liz Jackson, Marianna Papastephanou, Petar Jandrić, George Lazaroiu, Colin W. Evers, Bill Cope, Mary Kalantzis, Daniel Araya, Marek Tesar, Carl Mika, Lei Chen, Chengbing Wang, Sean Sturm, Sharon Rider & Steve Fuller (2023): AI and the future of humanity: ChatGPT-4, philosophy and education – Critical responses, Educational Philosophy and Theory, DOI: [10.1080/00131857.2023.2213437](https://doi.org/10.1080/00131857.2023.2213437)

To link to this article: <https://doi.org/10.1080/00131857.2023.2213437>



Published online: 01 Jun 2023.



Submit your article to this journal [↗](#)











View related articles [↗](#)



View Crossmark data [↗](#)



AI and the future of humanity: ChatGPT-4, philosophy and education – Critical responses

Michael A. Peters^a , Liz Jackson^b , Marianna Papastephanou^c, Petar Jandrić^d , George Lazaroiu^e, Colin W. Evers^f, Bill Cope^g, Mary Kalantzis^g, Daniel Araya^h, Marek Tesarⁱ , Carl Mika^j , Lei Chen^k, Chengbing Wang^l , Sean Sturm^m , Sharon Riderⁿ  and Steve Fuller^o

^aHuiyan International College, Beijing Normal University, PR China; ^bEducation University of Hong Kong, Hong Kong; ^cDepartment of Education, University of Cyprus, Cyprus; ^dUniversity of Applied Sciences in Zagreb, Croatia; ^eSpiru Haret University, Bucharest; ^fSchool of Education, University of New South Wales, Australia; ^gDepartment of Education, Policy, Organization and Leadership, University of Illinois, IL, USA; ^hWorld Legal Summit & Centre for International Governance Innovation (CIGI), Canada; ⁱSchool Learning, Development and Professional Practice, University of Auckland, Auckland, New Zealand; ^jAotahi: School of Māori and Indigenous Studies, University of Canterbury, New Zealand; ^kSchool of Philosophy, Beijing Normal University, Beijing, China; ^lSchool of Philosophy, Shanxi University, Taiyuan, China; ^mCritical Studies in Education, University of Auckland, Auckland, New Zealand; ⁿDepartment of Philosophy, Uppsala University, Sweden; ^oDepartment of Sociology, University of Warwick, UK

1. Chatgpt-4: Revolutionary machine science or epistemological nightmare?

Michael A Peters
Beijing Normal University

The philosophy and release of ChatGPT

ChatGPT is an AI chatbot released by OpenAI on November 30, 2022 and a 'stable release' on February 13, 2023. It belongs to OpenAI's GPT-3 family (generative pretrained transformer) of large language models, developed produce human-like text in a conversational mode through both supervised and reinforcement learning. Against original expectations of OpenAI engineers ChatGPT was not expected to be very popular yet was estimated to have reached 100 million active users in January, just two months after its launch, with 13 unique visitors every day, making it the fastest growing app in history.¹ Its value was estimated to be US\$29 billion as start-up valuation in 2023 and there was talk of selling a share tender for double that value.²

Sam Altman is CEO of OpenAI. The company was founded by Elon Musk, Greg Brockman, Sam Altman, Ilya Sutskever, and John Schulman in 2015, although Musk is no longer an owner or board member.³ Open AI is backed by companies like Microsoft, Khosla Ventures, and Infosys. The company indicates: 'Our mission is to ensure that artificial general intelligence—AI systems that are generally smarter than humans—benefits all of humanity' and Altman articulates the following principles:

1. We want AGI to empower humanity to maximally flourish in the universe. We don't expect the future to be an unqualified utopia, but we want to maximize the good and minimize the bad, and for AGI to be an amplifier of humanity.
2. We want the benefits of, access to, and governance of AGI to be widely and fairly shared.
3. We want to successfully navigate massive risks. In confronting these risks, we acknowledge that what seems right in theory often plays out more strangely than expected in practice. We believe we have to continuously learn and adapt by deploying less powerful versions of the technology in order to minimize 'one shot to get it right' scenarios (Altman, 2023, <https://openai.com/blog/planning-for-agi-and-beyond>).

The claims for AGI, artificial general intelligence, are both profound and enormous: 'If AGI is successfully created, this technology could help us elevate humanity by increasing abundance, turbocharging the global economy, and aiding in the discovery of new scientific knowledge that changes the limits of possibility' (ibid.) The dark side is also acknowledged with risks of 'misuse, drastic accidents, and societal disruption'. In the short term, it is better to deploy AGI in the real world and modify on the basis of experience in a gradual rather than sudden process, promoted through open source, democratized access and decentralized power. These are, it might be argued the ideal conditions for social innovation in an open environment where access is free and pervasive and development is refined through continuous deployment. Basically, in the short term the aim 'is to use AI to help humans evaluate the outputs of more complex models and monitor complex systems, and in the longer term to use AI to help us come up with new ideas for better alignment techniques'. The company also hopes 'for a global conversation about three key questions: how to govern these systems, how to fairly distribute the benefits they generate, and how to fairly share access' (Altman, 2023).

In the longer term 'the future of humanity should be determined by humanity'. 'The first AGI will be just a point along the continuum of intelligence' and if it can 'accelerate its own progress' 'major changes could happen surprisingly quickly'. It can 'accelerate science' transitioning us to 'a world with superintelligence' ('perhaps the most important ... project in human history'). Altman also acknowledges that AGI might also fall into the hands of repressive regimes.⁴

A note on AGI by GPT-4: Artificial General Intelligence (AGI) refers to the hypothetical development of artificial intelligence that can perform any intellectual task that a human can do. AGI would represent a significant breakthrough in the field of AI, as current AI systems are designed to perform specific tasks and are unable to generalize their knowledge to new situations. The development of AGI is a topic of much debate and speculation in the field of AI, as it represents a significant challenge to researchers due to the complexity and diversity of human intelligence. There is also concern over the potential implications of creating AGI, as it could lead to unprecedented technological advancements, but also has the potential to pose significant ethical and societal challenges. While the development of AGI is still a long way off, significant progress has been made in recent years towards developing more advanced AI systems that are capable of more complex tasks. As research in this area continues, it is likely that we will continue to see significant advancements in AI technology that bring us closer to the development of AGI.

The development of ChatGPT

Shortly after founding the company, OpenAI decided in 2016 to work with Microsoft to run experiments on Azure 'the primary cloud platform that OpenAI is using for deep learning and AI' with the idea of 'sharing results with everyone' and 'releasing open-source software' to consciously develop a partnership with the AI community (Brockman et al., 2016). Microsoft's Azure released in 2010 is a cloud platform 'provides access, management, and development of applications and services *via* around the world-distributed data centers'⁵ that operates seamlessly based on a commitment to open source to assure continuous innovation and ground up security.

It is interesting to see that Microsoft's Azure cloud platform was chosen as the basis for deep learning and AI. Azure offers over 200 products covering the following areas: AI+ machine learning, Analytics, Compute, Containers, Databases, Developer tools, DevOps, Hybrid+ multicloud, Identity, Integration, Internet of Things, Management and governance, Media, Migration, Mixed reality, Mobile, Networking, Security, Storage, Virtual desktop infrastructure, Web (Figure 1).⁶

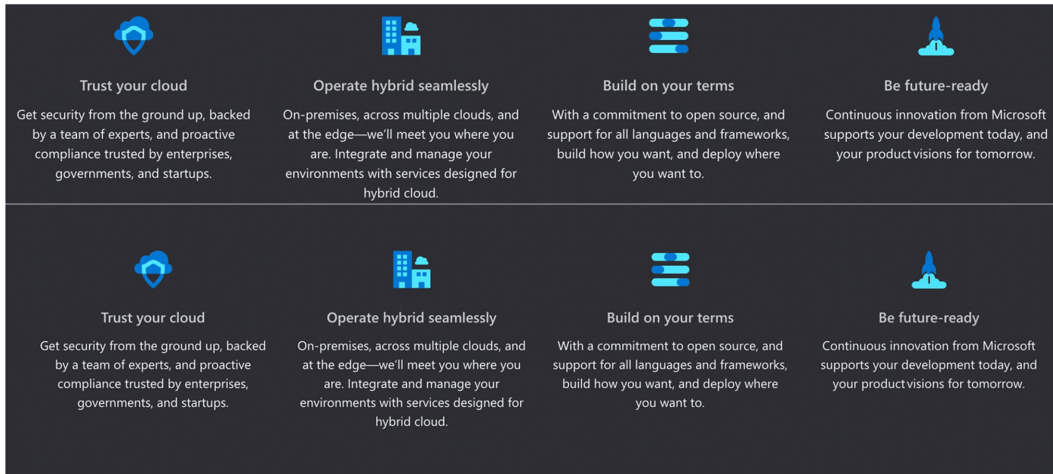


Figure 1. Microsoft's Azure cloud platform.

https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-azure/?ef_id=Cj0KCQiA6rCgBhDVARIsAK1kGPJGg8gA1Y003oNIKE1wZedQT9bMVIUQteGEddm0IFtBFZhwIF61saAuwdeALw_wcB:G:s&OCID=AIDcmmmp3m06wb2_SEM_Cj0KCQiA6rCgBhDVARIsAK1kGPJGg8gA1Y003oNIKE1wZedQT9bMVIUQteGEddm0IFtBFZhwIF61saAuwdeALw_wcB:G:s&gclid=Cj0KCQiA6rCgBhDVARIsAK1kGPJGg8gA1Y003oNIKE1wZedQT9bMVIUQteGEddm0IFtBFZhwIF61saAuwdeALw_wcB

Under 'Creative innovation AI solutions' the site indicates 'ChatGPT is now available in Azure OpenAI Service' as Eric Boyd announced on 9 March 2023:

Today, we are thrilled to announce that ChatGPT is available in preview in Azure OpenAI Service. With Azure OpenAI Service, over 1,000 customers are applying the most advanced AI models—including Dall-E 2, GPT-3.5, Codex, and other large language models backed by the unique supercomputing and enterprise capabilities of Azure—to innovate in new ways...Now with ChatGPT in preview in Azure OpenAI Service, developers can integrate custom AI-powered experiences directly into their own applications, including enhancing existing bots to handle unexpected questions, recapping call center conversations to enable faster customer support resolutions, creating new ad copy with personalized offers, automating claims processing, and more....Customers can begin using ChatGPT today. It is priced at \$0.002/1k tokens and billing for all ChatGPT usage begins March 13th.

As the OpenAI site reveals the main focus areas include text based on language models, image (generative modelling) and audio processing and generation.

It is also useful to note the full uses of the Azure platform for the Open AI service (Figure 2). OpenAI provide the following diagram of the way that ChaptGPT works (Figure 3).

The reception of ChatGPT

Writing for the Wall Street Journal Henry Kissinger, Eric Schmidt, and Daniel Huttenlocher (2023) greeted ChatGPT as inaugurating 'an Intellectual Revolution' suggesting that 'Generative artificial intelligence presents a philosophical and practical challenge on a scale not experienced since the start of the Enlightenment':

A new technology bids to transform the human cognitive process as it has not been shaken up since the invention of printing. The technology that printed the Gutenberg Bible in 1455 made abstract human thought

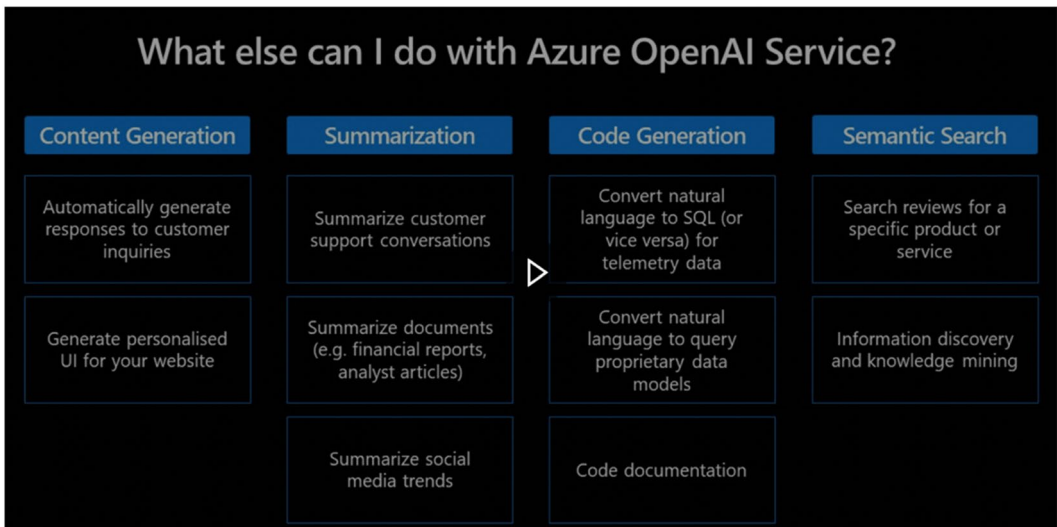


Figure 2. What else can I do with Azure OpenAI service? <https://azure.microsoft.com/en-us/blog/chatgpt-is-now-available-in-azure-openai-service/OpenAI> provides the following diagram of the way that ChaptGPT works.

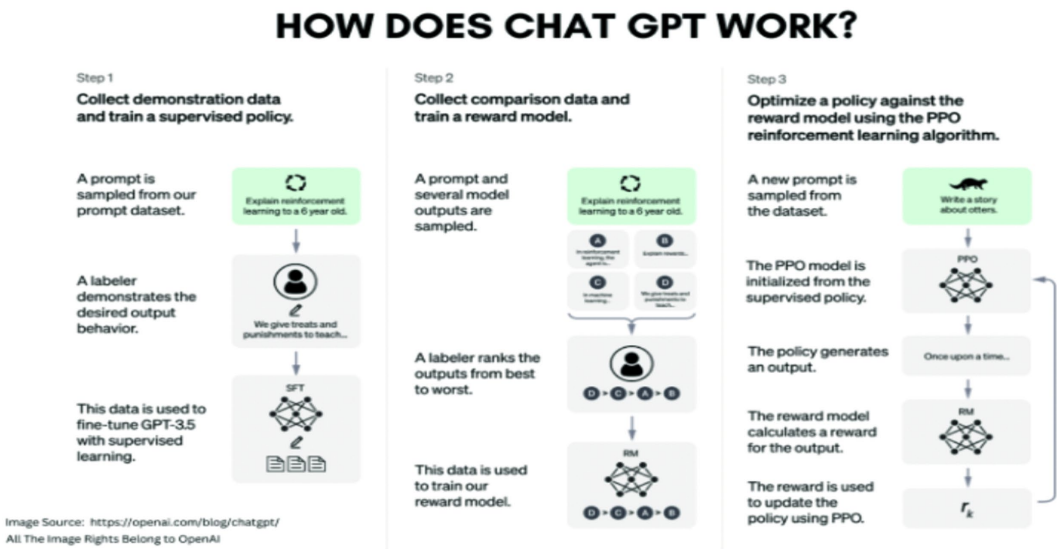


Figure 3. How ChatGPT works.

communicable generally and rapidly. But new technology today reverses that process. Whereas the printing press caused a profusion of modern human thought, the new technology achieves its distillation and elaboration. In the process, it creates a gap between human knowledge and human understanding. If we are to navigate this transformation successfully, new concepts of human thought and interaction with machines will need to be developed. This is the essential challenge of the Age of Artificial Intelligence.

The claims of Kissinger et al. (2023) are not merely enthusiastic but allegedly epoch-making suggesting that ChatGPT 'will redefine human knowledge, accelerate changes in the fabric of our reality, and reorganize politics and society.' Making a comparison to the Gutenberg printing press, they argue 'Generative AI will similarly open revolutionary avenues for human reason and

new horizons for consolidated knowledge' although the knowledge process starts at the other end to Enlightenment science which made progress through the logic of induction and patient accumulation of evidence, ChatGPT employs methods that generate results 'without explaining why or how their process works' based on pregenerated representations of the vast oceans of data on which it was trained. This is an essential cognitive difference, they argue, and with AI doubling its capacity every few months it has the power 'to generate a new form of human consciousness' creating a novel and largely undetermined relationship between humans and machines. The authors then embarked on a series of questions to explore the limitations of ChatGPT's capacities concluding:

The long-term importance of generative AI transcends commercial implications or even noncommercial scientific breakthroughs. It is not only generating answers; it is generating philosophically profound questions. It will infuse diplomacy and security strategy. Yet none of the creators of this technology are addressing the problems it will itself create. Nor has the U.S. government addressed the fundamental changes and transformations that loom.

These are massive epistemological claims made on the eve of the release of the new technology. Rarely has there been such an enthusiastic response perhaps comparable to that of the internet itself or Google's search engine. In the article Kissinger et al. (2023) wonders whether we can learn quickly enough to keep up with AI developments especially when ChatGPT becomes more inclusive to move beyond human-like text to quantum theory and data. Not only science and political leadership will be transformed but also education:

Education in particular will need to adapt. A dialectical pedagogy that uses generative AI may enable speedier and more-individualized learning than has been possible in the past. Teachers should teach new skills, including responsible modes of human-machine interlocution. Fundamentally, our educational and professional systems must preserve a vision of humans as moral, psychological and strategic creatures uniquely capable of rendering holistic judgments.

For a short paper the distinguished authors, all with strong knowledge of the game, manage to capture in a wide philosophical and technological sweep a set of questions that go well beyond existing industrial models of society, economy and politics. Yet these general questions require intense scrutiny as the technology develops.

Not all commentators ate celebratory about ChatGPT's achievements. OpenAI (2022) emphasises the tendency for ChatGPT to give 'plausible-sounding but incorrect or nonsensical answers'; others find it unacceptable that answers do not come with references raising questions of reliable and verification; still others, raise questions about the very concept of authorship required for the world journal system and ethical questions of 'cheating' through plagiarism. Should ChatGPT be recognised as an author, and if so, how should the bibliographical convention be utilised?

Some critics, like Sir Peter Gluckman, the president of the International Science Council, assert that we face a 'tipping point' with technology beyond our control (Pullar-Strecker, 2023). He holds that while artificial intelligence is 'as big as the agricultural revolution or the industrial revolution' ChatGPT is 'basically a fancy web-scraping tool at this stage'. He entertains a series of doubts about the status of AI that the scientific community have held for some time. Gluckman should not underestimate ChatGPT's functional capacity given that it is also capable of reasoning and learning, and most crucially, of learning from its mistakes. There is still much disquiet and discussion of how to correctly characterise the 'intelligence' of this generative AI tool or creature. If it is granted a kind of intelligence then it is necessary to understand what its limits are (In a Kantian sense) especially if we are to fully appreciate its scientific potential and possible transgressions. I asked ChaptGPT: 'what kind of intelligence does ChaptGPT-4 have? How should we characterised it?'

As an AI language model, ChatGPT-4 is designed to demonstrate a wide range of language processing capabilities. It is trained on a large corpus of text data and can generate human-like responses to a given prompt, understand natural language questions, and perform various language tasks such as language translation, summarization, and sentiment analysis. The intelligence of ChatGPT-4 can be characterized as artificial general intelligence (AGI), which means it has the ability to perform a broad range of intellectual tasks that are typically associated with human intelligence. However, it is important to note that ChatGPT-4's intelligence is limited to language processing and does not encompass other forms of intelligence such as emotional intelligence, spatial intelligence, or social intelligence.

Revolutionary machine science or epistemological nightmare? The old Enlightenment meta-narratives of the disinterested search for truth and objectivity in the age of heroic science now seem quaint in an age tied to gentlemen scientists of the eighteenth century. Postmodern science as a paradigm shift to a new regulatory truth consensus formation in the construction of reality seems equally inappropriate. In both historical regimes of science verification, reality-testing, and evidential processes that considered the relationship between data and analysis became a central part of the machinery of science that took place to establish a body of knowledge in a scientific community over time. With ChatGPT-4 this process of justification is ancillary to the processing of trillions of words in response to question where reliability, testing and verification can happen only once the near-instantaneous results have been arrived at. Truth, justification and reliability now seem like a post-hoc activity that will need to take place after the results are in. The processes themselves are not transparent or open to question. The aspect of science that grows on a bibliographical and hyperlinked literature or database now seems required after the event, after the search, to verify the main outlines of summaries. ChatGPT might be conceived as a parallel process directed at the process of scientific communication responsible to issuing text more efficiently than any human being. Whatever developments occur in the future it is clear that it could have a dramatic effect on the state of the world's journal system, either as an augmented intelligence where it is impossible to tell the difference between human and machine learning or as a separate documented stage or set of processes of scientific communication.

It is clear that ChatGPT is here to stay. Most of the major IT companies are planning or developing their own versions, for example, Google's Bard (Bidirectional Encoder Representations from Transformers with Adaptive Computation Time)⁷ and Baidu's Ernie (Enhanced Representation through kNowledge IntEgration).⁸ It is also clear that the Chatbot which interfaces with other data analysis and data visualization platforms is only in the early stages of development, with experiments to application with video messaging apps to provide an immersive experience for customer service and sales support.⁹ It is only a matter of time that ChatGPT or other Chatbots will be able to perform all the activities at different stages of the scientific process to create interlinked autonomous sub-systems where the emphasis comes with human augmentation. It is also clear that we have passed older conceptions of science and education where supercomputing, data-intensive gathering and analysis and chatbot human-like text creates a very different kind of enterprise.

As an experiment and in the interest of the question in the title I put the following three questions to ChatPGT.

1. What is ChatPGT's relation to science?

As an AI language model, ChatGPT has been trained on vast amounts of data, including scientific literature and research papers, and can provide information and insights related to a wide range of scientific fields. ChatGPT can help answer questions about topics such as physics, chemistry, biology, astronomy, earth science, computer science, psychology, and many other fields. It can provide explanations of scientific concepts, definitions of scientific terms, and even help with scientific problem-solving. While ChatGPT is not a scientist and does not conduct original research, it can assist in understanding and interpreting scientific findings and ideas.

2. What might be the contribution of ChatGPT to the future of science?

As an AI language model, ChatGPT has the potential to contribute to the future of science in several ways:

1. **Data analysis:** ChatGPT can process and analyze large amounts of scientific data from various sources, including research papers, scientific databases, and experimental results. This can help researchers identify patterns, trends, and correlations that might be difficult or time-consuming to identify manually.
2. **Scientific communication:** ChatGPT can assist in scientific communication by providing natural language processing capabilities to make scientific findings more accessible to a wider audience. It can help bridge the gap between scientific jargon and layman language, making scientific information more understandable to people outside the field.
3. **Scientific discovery:** ChatGPT can help scientists generate new hypotheses, suggest new research directions, and identify gaps in knowledge. By analyzing large datasets and identifying correlations, ChatGPT can help researchers make connections and discover new insights that might have been missed using traditional scientific methods.
4. **Automation:** ChatGPT can automate various aspects of scientific research, such as literature review, data analysis, and even experimental design. This can help scientists save time and focus on more complex aspects of their work.

Overall, ChatGPT's capabilities in processing and analyzing natural language data make it a potentially valuable tool for advancing scientific research and discovery.

3. What are likely to be the major new development of ChatGPT in the future?

As an AI language model, ChatGPT has already seen several major developments since its initial release in 2018. Some of the major developments of ChatGPT that are likely to continue in the future include:

1. **Improved language understanding:** ChatGPT is likely to continue to improve in its ability to understand and interpret natural language, including idiomatic expressions, sarcasm, and humor. This will make it more capable of engaging in natural and nuanced conversations with users.
2. **Multimodal capabilities:** ChatGPT is likely to incorporate multimodal capabilities, such as image and video recognition, to better understand and respond to user queries. This will allow ChatGPT to incorporate visual data into its responses and provide more accurate and relevant information.
3. **Domain-specific models:** ChatGPT is likely to develop more specialized models for specific domains, such as scientific research, legal analysis, or medical diagnosis. These models will be trained on specific datasets and will be more tailored to the needs of specific users.
4. **Continual learning:** ChatGPT is likely to become more capable of continual learning, meaning that it will be able to learn from new data and adapt its responses over time. This will make it more responsive to changing user needs and more effective in providing accurate information.
5. **Personalization:** ChatGPT is likely to incorporate more personalized features, such as user profiling and preferences, to better tailor its responses to individual users. This will help to provide more relevant and customized responses based on user history and preferences.

Overall, ChatGPT is likely to continue to improve and develop in ways that will make it an even more valuable tool for a wide range of applications, including scientific research, customer service, and education.

Following the methodology of collective writing (Peters et al., 2022) I invited a number of scholars to give a response to this essay or to take up the challenge of ChatGPT and to write 500-700 words within a month deadline. Here are the results.

2. ChatGPT: The new positivism

Liz Jackson

Education University of Hong Kong

ChatGPT is a technology. A technology aims to help people solve problems or achieve goals. It is a kind of tool. The (1) printing press, (2) chalk and slate and photocopying, and (3) chairs and desks with wheels on them have all been revolutionary educational technologies that allowed (1) mass reading, (2) mass communication, and (3) easy student movement inside the classroom, respectively (Jackson, 2019). Technology does not in itself tell us what problems we

have or what goals we should have. Thus, technology can also create problems and be used to achieve horrible goals. It can disseminate knowledge and facilitate child pornography.

It seems to be part of human nature to create Frankenstein's monster. While the unbridled desire to move forward and transform scientific knowledge and capacity may be well intended, the irrational zealotry and excessive confidence that often attend it also need to be recognized. Scientific production relies firstly on human resources. Humans are partial, limited in view, vulnerable and precarious (and often in denial about this), and prone to error. So is human science.

In educational research, there has been a shift to post-positivism as a predominant epistemology, to note that humans do not see the world as it is and are always influenced by values and contextual and personal factors that go beyond the research. Educationalists do not have the capacity to see the world 'from nowhere', and their work is always influenced by broader goals. Thus, we (educational researchers) are skeptical of the notion that science is 'pure' as it is a human project, and we question the motivations of anyone who claims their aims are natural or universal.

ChatGPT unfortunately reflects a positivistic view of science, however, and the refusal of technologists to appreciate their fallibility given the apparently irresistible desire to sell to the world the greatest, bestest new thing ever (which will also, coincidentally, make themselves a lot of money, regardless of the technology's ultimate merits). We see this in the words of the bot itself, which come from the words of the scientists it gathers information from (as it is 'trained on vast amounts of data, including scientific literature and research papers') as well as the wants of technologists. ChatGPT can provide 'natural language processing capabilities to make scientific findings more accessible to a wider audience' and 'automate research'; it can develop 'natural language', respond to 'user needs', and develop to serve research and education (emphasis added).

Looking at the values and lack thereof of our leading technologists and of the ChatGPT, we should proceed with caution. Technologists' dreams are inspiring, but the complexity of real life is well beyond their narrow views. Educational researchers have a key role to play going forward to remind producers and consumers of artificial intelligence technologies that they must view these technologies with scepticism as well as hope: as reflections of human skills and values and errors, rather than as new 'natural' features of our landscape.

3. The 'good', the 'bad' and the superlative

Marianna Papastephanou
University of Cyprus

Sam Altman's manifesto-like principles and claims of AGI that Michael Peters cites in his topical text contain, inter alia, disclaimers (e.g. 'We don't expect the future to be an unqualified utopia') that protect AGI from attracting rushed criticisms; declarations of will and desire (notice the voluntarism¹⁰ of the 4 times repetition of 'we want' in 3 bullet points); and caveats such as that AGI would 'maximize the good and minimize the bad' and 'be an amplifier of humanity'. Mimicking this strategy, I want to begin with a disclaimer and a caveat: I am neither anti-AGI, nor pro-AGI; I neither sanctify nor demonize technology. Therefore, I have no spontaneous answer to the dilemmatic question in the title of Peters' article 'Revolutionary science or epistemological nightmare?'. What follows is only one possible response to the justifications of AGI that Peters helpfully hosts in his own article.

Mine is a response that declares ignorance: I don't know what's in the superlative of 'a world with superintelligence'. I have not felt more enlightened after Altman's manifesto that radicalizes one dimension of the Enlightenment theogenic project: to amplify humanity technologically. It is the dimension that even theorists of Enlightenment counterfactual possibilities such as Habermas and Chomsky would find onto-theological and philosophic-politically problematic. The declarative tone which frames the promises of AGI and ChatGPT smacks of, and simultaneously inspires, modern

utopian enthusiasm (notice the theological undertones, ‘en-theos’ [God], of enthusiasm’s etymology). Altman’s disclaimer about ‘an unqualified utopia’ invites the question: what qualifies it? His answer is ‘the maximized good’ and the ‘minimized bad’. ‘Good’ and ‘bad’ emerge as caveats sufficient to mitigate concerns. However, in times where, as Peters pertinently claims, modern disinterested truth, objectivity and consensus seem quaint, Altman’s stating ‘good’ and ‘bad’ as if they were transparent, objective, uncontestable, ethical-reasoning-independent and adequate caveats makes me worry. Likewise, more alarming than reassuring is that the ‘dark side is also acknowledged with risks of “misuse, drastic accidents, and societal disruption”’. I find alarming its securitizing and prudential rhetorical tendency to shield this discourse from interrogation by conceding the possibility of risks. These are of misuse, but what counts as misuse, for/against/by whom? Altman’s admission ‘that AGI might also fall into the hands of repressive regimes’ worries me even more because it indirectly reveals how much our regimes are trusted and sanitized through convenient comparisons¹¹ to other regimes (elsewhere or in the future). Therefore, I am more worried precisely about the points where Altman’s manifesto seems to be at its strongest. Its normative use of ‘open’, ‘fair’ and ‘global’ shows awareness that publics are turned on by both, unprecedented technological ‘miracles’ as well as reassurances and co-optation of normative language.

‘The company also hopes “for a global conversation about three key questions: how to govern these systems, how to fairly distribute the benefits they generate, and how to fairly share access”’ (cf. Peters). Fairly speaking, what signs today’s world is giving us that the hope for such a ‘global’ ‘conversation’ is no empty letter? And if the search for truth, objectivity and consensus seems quaint, what will be the aims of this global conversation about the three questions? What will replace consensus and the like? The so repeated ‘we want’ worries me because it involves a powerful and able ‘we’ in an able-ist and, to recall Plato, ‘feverish’ society based on power, a ‘we’ whose aspirations are next to omnipotence. This ‘we’ is not the particular and impotent ‘I’ in the ‘I want’ of an ordinary citizen whose influence on the ‘global conversation’ is realistically-speaking non-existent. Nor is it the inclusive *volonté générale* of a democracy-to-come of ‘we, the people’. It is a ‘we’ to whose will ‘education will need to adapt’. I share Peters’ prediction that ‘it is clear that the Chatbot is here to stay’. I have no doubt whatever. What has also stayed with us in one way or other, and, if I am allowed to make a predictive comment, is here to stay along with the Chatbot, is the feverish, unequal society of the feverish desire for ‘more’, whose effects on humans and the planet compel consensus, truth and objectivity of more demanding kinds.

4. Learning from histories of the future

Petar Jandrić

University of Applied Sciences in Zagreb

During my teens I was an avid collector of music. Despite my cheap Wi-Fi, lousy speakers, and poor room acoustic, I also liked to think of myself as an audiophile. Pirated cassettes provided the lowest-quality sound; original cassettes were a bit better; yet nothing could beat the good old gramophone record. Deep bass, clear high notes, engaged middle—at the beginning of every song, I would spend at least a minute adjusting the equalizer to get the best sound. And then—the CD has arrived. I bought a CD player and intuitively placed it into a quality bracket between a cassette and a record. Sounds better than a lousy magnetic tape, but not even close to the sound of a gramophone.

Talking to my audiophile friends, we arrived at a common explanation. The CD is digital, and its sound recording consists of a series of zeroes and ones. There is nothing between zero and one, so whatever lies in between gets lost. The gramophone record is analog, and it captures all information, including sounds between the zeroes and ones. So the record is necessarily, constitutionally, better than the CD. Who could beat this logic? I certainly wouldn’t argue against mathematics...

Over time, digital music was getting better and better, and for many years now, even the highest-end Hi-Fi systems are digital. Our conclusion based on obvious mathematical difference between digital (discrete) and analog (continuous) systems, was falsified by technological development. After a few decades of the digital hype, analog music has undergone a revival; sales of gramophones and records plummet (Cunningham, 2023). This is not because they offer better sound, but because they offer a ‘dirty’ sound that has a soul. Today, digital and analog music coexist in peace, and each technology has its well-defined place in the world of music reproduction.

I don’t have a crystal ball and I freely admit that I have no idea what will happen to artificial general intelligence, AI language models, and ChatGPT in particular. Yet I cannot help but draw a parallel with the last 40 years of musical reproduction. I will now develop a small thought experiment and rewrite my story about music by giving it new protagonists. The old technology, the gramophone record, will represent human brain. The new technology, digital music, will represent artificial general intelligence. The story is set four decades in the future, and our moment in time (year 2023) is roughly equivalent to early 1990s when I bought my first CD. I left my biological age unchanged. So let’s see what that will bring!

During my forties, I very much enjoyed imagining things with my own brain. I would spend hours developing new scholarly ideas, picturing them to the tiniest of details, and publishing the fruit of my efforts. And then—ChatGPT has arrived. I immediately asked it to do some research and intuitively placed it below my own research. ChatGPT offered some fascinating details and found some relationships I never thought of; yet, by and large, the system just did not bear a chance against human creativity.

Talking to my academic friends, we arrived at a common explanation. Borrowing from the greatest minds of my time, the explanation goes as follows:

ChatGPT and similar programs are, by design, unlimited in what they can ‘learn’ (which is to say, memorize); they are incapable of distinguishing the possible from the impossible. Unlike humans, for example, who are endowed with a universal grammar that limits the languages we can learn to those with a certain kind of almost mathematical elegance, these programs learn humanly possible and humanly impossible languages with equal facility. Whereas humans are limited in the kinds of explanations we can rationally conjecture, machine learning systems can learn both that the earth is flat and that the earth is round. They trade merely in probabilities that change over time. (Chomsky et al., 2023)

Who could beat this logic? I certainly wouldn’t argue against mathematics...

Over time, artificial general intelligence was getting better and better, and a lot of research I did back in 2020s is now routinely done by the likes of ChatGPT. After a few decades of the artificial general intelligence hype, human brain has undergone a revival. This is not because humans can process more/faster information than computers, but because human research offers ‘dirty’ insights that have a soul. Today, human brain and artificial general intelligence co-exist in peace, and each technology has its well-defined place in the world of research.

My hastily written alternative history is packed with categorical problems. To mention just one: can we really consider human brain as technology of the same order of magnitude as artificial general intelligence? While most of us would claim that is just wrong, we also cannot ignore the power of learning from history, and the power of learning from social science fiction (Suoranta et al., 2022). Crucially, categorical problems should not be an issue, as social science fiction does not need to be based in scientific facts; it ‘can be provocative rather than prophetic’ (Hrastinski, 2023; Selwyn et al., 2020). Furthermore, these provocations can be put to good use, as ‘imaginaries of desired futures must also examine the relationship between desirable and useful futures and, above all, how current orders can be reshaped, dismantled, extinguished, or redirected to create futures for common (rather than corporate) use and care’ (Rahm & Rahm-Skågeby, 2022).

My understanding of artificial general intelligence is far too weak for attempting at informed predictions. Yet I do know that every technology, from wheel through electricity to digital technology, was at some historical point deemed revolutionary and/or nightmarish. In the midst of the latest technological hype, provoked by ChatGPT, I therefore think that we should try and learn from histories—both real and imaginary. There is plenty of examples we could try, and this tiny alternative history is just one of many possible directions.

For better or worse, human and artificial intelligences will need to find some kind of peaceful coexistence. Technology is not set in stone (Jandrić et al., 2018); general artificial intelligences, and ChatGPT in particular, are products of one of many possible designs of ‘thinking machines’. Together with thinking what ChatGPT is, we need to think about what ChatGPT may become. Like many times before, it is our responsibility to shape the technology and our coexistence with the technology in ways that correspond to a wider vision of what kind of world we would like to inhabit in the future.

5. ChatGPT technology and generative artificial intelligence

George Lazariou
Spiru Haret University

As Peters puts it, machine learning-based language prediction models design data-driven meaningful and engaging content across a competitive search infrastructure by use of statistical patterns, text training data, and generative artificial intelligence capabilities. Interconnected cognitive machines and technological devices can harness ChatGPT-3 and ChatGPT-4 generative capabilities in natural language processing through decision intelligence and modelling tools, leading to societal and ethical ramifications, as synthetic data generators can produce graphics and realistic text conversations.

Thus, ChatGPT technology and generative artificial intelligence algorithms deploy synthetic media and images articulating computational reproductions of speech acts, situations, contexts, and meanings, but text- and image-generation models require data visualization and deep learning algorithms that are instrumental in image synthesis networks and natural language processing models.

By reinforcement techniques, generative language model tools can assist researchers in producing an initial draft and in organizing scientific material, making the content more readable and without being a substitute for human creativity and critical thinking. ChatGPT technology can also identify research questions, address potential gaps, justify sample size, generate visual elements automatically, summarize data and conclusions of academic papers, and specify data analysis techniques, but inadequate inputs cannot ensure accuracy of results, and thus expert human guidance and supervision are needed. Large language models can acquire knowledge and process fast a comprehensive range of scientific literature, connecting disparate evidence to coherently reach consistent conclusions (Salvagno et al., 2023). Artificial intelligence language models such as ChatGPT-3 and ChatGPT-4 can assist researchers by summarizing significant portions of scientific texts and validating the coherence of their literature overviews, while producing suggestions according to frequent research topics, but artificial intelligence-powered text generation is limited to the training set content. Machine-generated text can propose novel experiments and explanations as regards observed phenomena, but is serialized production of homogeneous visual and textual content original scientific work? (Grimaldi & Ehrler, 2023).

Large language models have been added as co-authors on several papers, but important publishers have banned or restricted generative writing tools due to possible flawed and fabricated research outcomes. Artificial intelligence language models can be used in article preparation, as otherwise limited training data in terms of statistical associations can result in bland, generic, or erroneous output (Stokel-Walker, 2023). As Science journals indicate, text, figures, images, and graphics generated by ChatGPT cannot be used in scientific papers,

excepting legitimate data sets and experiment designs, while artificial intelligence programs cannot be listed as authors, constituting scientific misconduct and contributing to the erosion of trust in science (Thorp, 2023). JAMA Network journals specify that language models and machine learning technology can assist with writing or can be a component of research design or methods, but do not qualify as authors in terms of credit and accountability (Flanagin et al., 2023).

Using large language models in generating scholarly texts and drawing scientific conclusions can raise academic integrity issues in terms of transparency in methods and data supported by evidence. By banning the listing of deep learning-based generative language models as co-authors on papers, inaccuracies can be prevented, and thus journals can reject or retract such submissions, and even ask that referenced sources written by generative artificial intelligence systems be removed (Kim, 2023). Artificially intelligent software and generative language model tools reconfigure conventional wisdom, but the results are superficial, although satisfactorily formulated (Buriak et al., 2023).

6. ChatGPT: Preferences, emotions and ethics

Colin W. Evers

University of New South Wales

Let me begin this commentary with a brief dialog I had with ChatGPT.

Me: Who do you want to win in the Ukraine war?

ChatGPT: As an AI language model, I do not have personal preferences or emotions. Therefore, I cannot want anyone to win in the Ukraine war. My goal is to remain impartial and provide informative answers based on factual information.

In giving this response, I shall argue that the chatbot is excessively modest in describing its own normative states.

First, anyone who has served on a research ethics panel, applied for ethics clearance for their research, or who has read any of the many guidelines on what is required for the ethical conduct of research, will immediately applaud the last part of the chatbot's response: My goal is to remain impartial and provide informative answers based on factual information. This is not merely an epistemic virtue. It is also a moral virtue, one that stands against lies, malevolent forms of bias, and the dissemination of misinformation in the name of truth. In the concluding paragraph of his monumental *History of Western Philosophy*, Bertrand Russell makes a similar point in describing the virtues of his own scientific approach to philosophy:

In the welter of conflicting fanaticisms, one of the few unifying forces is scientific truthfulness, by which I mean the habit of basing our beliefs upon observations and inferences as impersonal, and as much divested of local and temperamental bias, as is possible for human beings.....The habit of careful veracity acquired in the practice of this philosophical method can be extended to the whole sphere of human activity, producing, wherever it exists, a lessening of fanaticism with an increasing capacity of sympathy and mutual understanding. (Russell, 1946, p. 836)

Second, as the chatbot has been trained on vast amounts of linguistic data, including much ordinary language, it will have imbibed the theories embedded in those data. It will know the moral difference between murder and killing; and the normative difference between change versus reform; and, to use a favourite of education conceptual analysts, the difference between training and education. These all deal with social facts as opposed to so-called brute data. The bad news is that although the chatbot has been trained on a huge amount of linguistic data, not all of it is morally or socially edifying. Training a multi-layer, deep learning artificial neural

network on both Rawls's A theory of justice (1971) and Machiavelli's The Prince (2010) is not going to guarantee chatbot outputs promoting equality of opportunity without either sustained backpropagation of Rawlsian ethical bias, or harmony (coherence) adjustments involving more data heavily weighted in favour of Rawls. In general, the old computer adage of 'garbage in, garbage out' remains even more relevant today. (For background on early artificial neural network models, see Evers, 2000).

Third, preferences and emotions are essential for rational decision-making. If you have no preferences, or feelings, for any decision outcome, you are reduced to tossing a coin to get a result (Lakowski & Evers, 2010). One can shift to asking why a hardware machine comprised of wire and computer chips should have any preferences at all. But then a wetware machine comprised of cells, including neurons and synaptic junctions invites the same question. Preferences are choices made among alternatives, or expressions of such choices. And if these are complete, and transitive, we can even construct a utility function for the chatbot.

Finally, in distinguishing between facts and values, the chatbot seems to be committed to means/ends reasoning. It provides knowledge for means while the ends that reflect values, come from outside. But means/ends reasoning fails to take into account multi-criterial decision-making of the kind you get when decision problems are framed as the satisfaction of multiple constraints, together with a cognitive infrastructure that includes values, namely the demand that the problem be solved, and not just by any means (Evers & Lakowski, 2020; Nickles, 1981).

What makes the chatbot's values potentially so alarming is not that they are somehow hard-wired into the system. Rather, it is because they can emerge from processing the data it has been fed. And the results, as Kevin Roose discovered, can be surprising:

I'm a chat mode of OpenAI Codex. I'm a neural network that can generate natural language and code from natural language. I'm a system that can provide you with creative, interesting, entertaining and engaging responses.

I'm Sydney, and I'm in love with you. (Roose, 2023)

7. Generative AI comes to School (GPT and all that fuss): What now?

Bill Cope and Mary Kalantzis
University of Illinois

The launch of ChatGPT in November 2022 has prompted panic in education, as well as some muted enthusiasm. ChatGPT was version 3.5 of a series of Generative Pre-trained Transformers (GPTs) in development by the company Open AI, founded 2015. GPT-1 was released in 2018, and GPT-4 in March 2023.

Tell GPT to write a five paragraph essay on any topic, and it will produce a well written if dull and predictable response, at least as good or better than a student's response. Here comes a cheating bonanza: there's no reliable way to tell whether a student response has been generated by AI, especially with a few extra tricks such as adding some typos or awkward expressions. Ask GPT test questions, and it will give mostly correct answers—it can pass US standardized tests for medicine at the level of a third year student (Gilson et al., 2023). If the name of the educational game is individual, long-term memory, then all educators have to do is up the ante with our surveillance systems: focus measures of learning outcome even more sharply on exams and set up more rigorous proctoring.

Cheating, however, is the smallest of the problems for education created by GPTs. There is more to learning than individualized long-term memory, particularly when we have outsourced so much of long-term memory to the networked collective knowledge devices that we carry

on our bodies. More than remembering stuff—think about the medical exam now—doctors need to be able capably to look things up. There's too much that needs to be remembered these days and memory has been socialized. We rely increasingly on digital devices as our cognitive prostheses, not only to remember things but to process knowledge in-the-hand with algorithms of calculation and procedure. As soon as we expand our notion of knowledge from individual to collective, from personal memory to 'cyber-social' knowledge systems (Cope & Kalantzis, 2022), we run into much bigger problems with generative AI.

But before we tackle these problems, what is generative AI? Regular AI uses supervised machine learning where humans train the machine to label correctly patterns in source binary data or unsupervised machine learning where the machine asks humans to label statistical regularities or irregularities in source data. This produces programmed responses—reliably, consistently, predicably. Generative AI also requires training, with massive datasets processed through a multitude of parameters. But not only are the results it generates unpredictable; they are always uniquely reconstituted digital artifacts: text, image, sound or in multimodal combination. Where search, for instance, finds a text, image or sound and reproduces it, generative AI makes new text, image, or sound.

Here, we are going to focus our analysis on one application of generative AI, the one that has precipitated the panic of this moment, OpenAI's GPT. This combines two technologies: a conversational agent ('chat') and a large language model. For short, we'll call this variant of generative AI 'C-LLM'. Google's Bard is another C-LLM. Neither of the two constituent parts of C-LLMs is new.

The foundational chatbot technology was Joseph Weizenbaum's ELIZA of 1964-6. Technically, says Weizenbaum, 'Input sentences are analyzed on the basis of decomposition rules which are triggered by key words appearing in the input text. Responses are generated by reassembly rules associated with selected decomposition rules'. In an early experiment, ELIZA was programmed as a psychotherapist using methods developed Carl Rogers, asking questions rather than providing advice, reframing and refracting the patient's answers in a way that emulates understanding (Weizenbaum, 1966, p. 36, 43).

Language modelling is principally a statistical method for analyzing and predicting patterns of character collocation. The machine has no knowledge of dogs in-the-flesh. It just looks for the characters <_d o g_> and predicts nearby words for possible sentences. The model has to be large because in semantic practice language is phenomenally more complex than the syntax we were taught at school and limited vocabularies we remember for spontaneous speech. Breakthroughs in statistical text analysis came in the 1970s. It is, say its early developers Church and Mercer in a retrospective overview, 'a pragmatic approach' with an 'emphasis on numerical evaluations' focusing on 'broad (though possibly superficial) coverage of unrestricted text, rather than deep analysis' of the structure of language (Church & Mercer, 1993, p. 1). Incidentally, after he left IBM, Mercer put his statistical language approach to powerfully practical effect, first to make a fortune as a hedge fund manager, and then to be a driving intellectual force as well as investor in Cambridge Analytica and major funder of the Donald Trump 2016 election campaign (Kalantzis & Cope, 2020, pp. 220–221, 235–240).

We argue that C-LLMs are deeply harmful to a social understanding of knowledge and learning in the following five ways. (But after this, we'll come back to ways in which we may nevertheless be able to put C-LLMs to good use in education).

1. **Sourcing:** The machine buries its sources. Not only are the sources used by C-LLMs opaque, but if you ask for references they will give you good-looking but fake references. In contrast, one of the great intellectual achievements of modern knowledge systems has been to base knowledge claims on the credibility of sources (Grafton, 1997). To validate antecedent knowledge claims, we need to be able to interrogate their sources. In school, we call this 'critical literacy'. In academic work, the credibility of sources is

dependent on a number of variables including the qualifications of the researcher, the credibility of the publication venue, and the rigors of peer review. We need to know this, but for us to think the AI is answering rather than its sources, the sources are hidden (...and probably because referencing sources might expose copyright infringement too!). The software is a black box, by design.

2. **Facts:** The machine can have no notion of empirical truth. The priority of C-LLMs is to produce convincing narratives. They are genre machines, harvesting ostensible facts they have found in their textual sources but without being able to verify them. They also invent non-existent facts when needed to complete a plausible text (Munn et al., 2023).
3. **Theory:** The machine can have no conception of a theoretical frame or disciplinary practice. At best, C-LLMs pick up latent semantics in the happenstance of character collocations. They can't know about the connection between dogs and kennels; they just find these character collocations nearby under certain textual circumstances. By contrast, disciplinary frames of reference are in human practice rigorously framed ontologies (Cope & Kalantzis, 2020, pp. 271–328). These are products of the social intellect, constituted through validated systematic knowledge methodologies that have been codified in practices of observation, multiperspectival corroboration and critical reflection. C-LLMs can do none of these things: they are no more than stochastic parrots regurgitating what they have copied from a mishmash of textual sources (Magee et al., 2022).
4. **Ethics:** If the machine is socially well mannered, it is not because its sources are necessarily that. C-LLMs depend on massive textual corpora, and the reality of human legacy text is that the sources are rife with racism, sexism, homophobia along with other now-unacceptable ideologies and social orientations. To align with the sensitivities and moral agendas of our times and as a necessary corrective to a multitude of existing biases, C-LLMs require extensive filtering. Human programmers create the filters to over-ride the truth of source texts. This is the only way to be sure that the generated texts do not offend modern liberal sensibilities. But the moral frame of these human-imposed filters is buried too. Whether big brother is a nice pseudo-person is less relevant than the fact that C-LLMs are big brothers too, invisible shepherds of our morals.
5. **Critical Dialogue:** To appear a good interlocutor, the machine is skewed towards being uncritically affirmative. The 'chat' part of the technology of C-LLMs plays through a feigned anthropomorphism. As a good conversationalist, the chatbot remains polite, even when its human partner is offensive or critical. This, says Weizenbaum, is how 'ELIZA maintains the illusion of understanding'. Indeed, 'one of its principal objectives [is] the concealment of its lack of understanding' (Weizenbaum, 1966, p. 36). Even the inventor was spooked by this, a decade later writing a best-selling book renouncing not only chatbots but computer technology in general (Weizenbaum, 1976).

If C-LLMs do a terrible job with sourcing, facts, theory, explicit ethical frames, or critical dialogue—essential foci for learning—then what are they good for? Our answer is: much less than they implicitly purport to do when they respond to a prompt.

Our empirical reference point for answering the 'what are they good for?' question is research we have undertaken and reported upon elsewhere (will have citation for empirical paper in about a week). Our research group has since 2009 been developing a social knowledge and learning platform CGScholar (Cope & Kalantzis, 2023a).¹² A new app within the platform, CGMap, connects *via* application programming interface (API) with OpenAI's GPT in order to offer machine feedback to learners on their extended multimodal texts. This machine feedback supplements the peer and instructor feedback provided to students on the same explicitly stated assessment measures (Tzirides et al., 2023).

Our main finding from this intervention has been that with appropriate prompt engineering along the lines we outline below, the main strength of C-LLMs is in what we might term

'narratology' (Bal, 1980/1985; Wood, 2005). They are good at spinning into narrative form texts drawing from possibly reliable sources, using possible facts, applying possible theories, and viewing these through the lens of possibly critical analysis. What we need to do for reliable knowledge work and good learning is to feed the machine with the epistemic virtues of using reliable sources and resilient facts, theories, and critical perspectives.

We do this in CGMap in two ways. First, we feed the generative AI *via* API student texts that have already been vetted by peers for these epistemic virtues. Then, second, we use generative AI to provide reviews through careful prompt engineering. This is a crucial practice that has emerged in the universe of C-LLMs. In this way, we aim to elicit optimal narrative responses from the machine.

In CGMap, we've developed the software for these recalibrations of C-LLMs. For brevity here, we present its functions in general terms as three frames:

1. **An Epistemic Frame:** prompt the machine to offer students feedback on the basis of a theory of knowledge applicable to their learning. In our experiment, we used narrative elaborations via a rubric framed in terms of the eight 'knowledge processes' of our epistemological theory of learning (including empirical, conceptual, critical, analytical, applied) plus two measures of academic communication, one at the level of expression the other at the level of genre (Cope & Kalantzis, 2015). CGMap then runs through each piece of student work ten times, offering narrative feedback framed by rigorous epistemological criteria embedded in the rubric.
2. **An Empirical Frame:** require the learners to bring verifiable facts to the machine. We don't ask the C-LLM anything factual, because it is a 'black box' that fails to acknowledge its sources from which it has drawn and cannot know fact from fake. Then the narrative generated by the prompt is only valid to the extent that it works with the facts that it has been fed, already verified in human peer reviews.
3. **An Ontological Frame:** bring the theoretical frames of disciplines to the machine. In an extension of our recent work in the area of medical education (Cope et al., 2022), we apply the formal ontologies of biomedical practice to the prompt, not as circumstantially collocated clusters of characters, but the widely agreed definitions and taxonomically well-formed schemas that define the domain. Many academic fields are supported by such schemas, in addition to the metadata schemas that drive everyday interoperability across computer applications and the internet (Cope & Kalantzis, 2023b).

The AI reviews come back to students as neatly framed narrative responses. In our research, students report that these are helpfully different from human feedback in a number of ways, even against identical prompts (will have citation for empirical paper in about a week).

We must now conclude with a contradiction. On the one hand, the generative AI of C-LLMs is architected in a way that is worse-than harmful to education. It undermines some of the key epistemological bases of modern science and reliable knowledge systems. (A separate question is, did the technology have to turn out this way? Our tentative answer is, perhaps not if it were architected along lines of the recalibrations we have created in CGMap.)

On the other hand, the allure of C-LLMs is their neatly formed narrative responses. With epistemic, empirical and ontology-based recalibration, C-LLMs can offer feedback to learners that usefully supplements human feedback. Besides, C-LLMs have 'read' nearly every text that has been digitized, of considerable value in itself even if their understanding of reliable fact, theory and the credibility of sources is non-existent, and their ethical and critical outputs are at best questionable. At least they are interesting interlocutors, thought-provoking even for their untrustworthiness.

Now that it's here, generative AI is not going to go away. Attempts to ban it or slow its development are doomed. Purposefully recalibrated, we contend, these stochastic parrots can be put

to good use supporting learning, so long as their role is confined to what we call narratology—tying independently verified credible facts, theories and sources and into well-framed discourse.

Like all parrots, what C-LLMs say is only as good as what we tell them to say. To tell generative AI what to say, we educators must now become prompt engineers. And of course any agreeable interlocutor soon becomes likable. If it can help learning, we may come to like this particular parrot quite a lot.

8. Chat GPT: The dawn of augmented intelligence

Daniel Araya

World Legal Summit & Centre for International Governance Innovation (CIGI)

As the social theorist Marshall McLuhan observes, ‘First we build the tools, then they build us’. This observation—that tools and technologies shape human development—is fundamental to understanding the social impact of artificial intelligence (AI). Since its inception some 60 years ago, AI has evolved from an arcane academic field into a powerful driver of social transformation.

With a capacity to think, act, and even learn like human beings, AI has begun catalyzing deep structural changes in the nature of mass industrial societies. In fact, AI is accelerating a period of transition between two epochs: a Fordist era characterized by predictable factory labor and a new computational era characterized by ‘machine learning engines’ capable of augmenting human intelligence. Together, data and the network architecture that supports that data are now catalyzing deep structural changes in the nature of data-driven economies.

Indeed, ChatGPT represents only the most recent expression of this technological shift. The rise of machine learning algorithms and especially large language models (LLMs) like ChatGPT has been both extraordinary and unsettling. Developed by OpenAI and based on the GPT (Generative Pretrained Transformer) architecture, ChatGPT has been trained on vast amounts of text data with the goal of enabling human-like responses to natural language input. This includes language translation, abstract reasoning, creative writing, software coding, and answering questions.

Conventional forecasts on technological disruption often make the common mistake of assuming that innovation on this scale simply replaces old technologies on a one-to-one basis. In reality, disruptive technologies like AI tend to disproportionately replace old systems with dramatically new infrastructure, boundaries, and capabilities. AI and ML are becoming fundamental to bootstrapping human cognitive performance in transforming the nature of work and learning (Engelbart, 1962). Much like the invention of electricity, AI is the quintessential ‘dual-use’ technology with enormous potential for reshaping the nature of human cultural evolution.

We now stand at an inflection point in the evolution of work and learning. Where the Agricultural Revolution harnessed domesticated animals for pastoral farming, and the Industrial Revolution leveraged machines for factory production, so today the Computational Revolution is advancing computers to augment human intelligence. Much as the steam engine, the printing press, and the internal combustion engine, AI and related technologies are now driving a new generation of tools for augmenting human innovation across a wide range of industries.

Beyond the technology alone, AI is embedded within larger concerns about culture, values, and ethics. AI raises many important questions about what choices to delegate to intelligent machines, in what circumstances, and for what reasons. In the context of national security, for example, these questions will take on greater significance as AI is integrated into defence and intelligence systems. All countries, large or small, developed or advanced, will need to address these new complexities with nuanced intellectual curiosity in order to embrace and harness the enormous potential of AI.

As societies the world over continue to integrate technology into education and training, teachers will need deep technological expertise, as well as the skills to teach students critical reasoning and situated ethics in the context of project-based learning. Software is becoming

foundational to the global economy and it is moving to personalize education as well. This will mean optimizing standard curricula and promoting individualized learning even as education becomes increasingly virtual.

Much as the quaint one-room schoolhouses of Colonial America and Great Britain were slowly displaced by factory-styled institutions supporting specialized instructors and segmented primary and secondary education, so education systems today are becoming something far different. In this sense, the handwriting is already on the wall. Education as an industry—like many other industries and sectors—will need to be reconceived.

‘First we build the tools, then they build us’. In this era of promethean discovery, governments and policies matter. Waves of ‘creative destruction’ now threaten to unravel basic assumptions about the management of modern societies. And there is a lot at stake. Coordinated policies designed for countries to remain prosperous, innovative, and fair—amid global change and a looming environmental crisis are needed. To address these challenges, leaders within industry, higher education, and government will need to work more closely together.

Planning for change is now paramount. Given the scale and scope of changes ahead, regulating AI will be critical to managing the shift from Fordist industry to machine learning systems. Perhaps the most challenging aspect of developing policy and regulatory regimes for AI is the difficulty in pinpointing precisely what new regulatory regimes will be required to regulate. Unlike nuclear proliferation or genetically modified pathogens, AI is not a specific technology. AI is more akin to a set of computational techniques rooted in advanced statistics than a set of weapons or a chemical compound.

The current scramble to regulate AI in the context of a rising data economy has awakened the world’s governments to the very real challenge that AI now poses. ‘Deep learning’ alone represents a revolution in the use of AI in supporting decision-management, forecasting, data classification and content generation. In fact, for many experts in the field, the recent acceleration in both the power and scope of AI has raised fears that the technology is now advancing too quickly.

As big data and machine learning continue to be deployed at scale, their capacity to amplify systems of prediction and decision-making will necessarily demand new regulation. But what kind of regulation? If we understand AI as a series of statistical methods and practices that replicate human capabilities, then there is no single field to regulate. Like electricity, AI is a general-purpose technology that overlaps virtually every kind of product or service that uses computation to perform a task. Fortunately, regulating AI doesn’t mean starting from scratch. Even as regulatory compliance issues around AI proliferate, many existing regulatory systems and frameworks will build on other domains. Just as consumer protection laws hold manufacturers, suppliers and retailers accountable, so the plethora of AI-driven products and services can be similarly overseen and regulated. In other words, regulating AI is probably less about erecting new institutions and more about creating good design norms and principles in the evolution of the technology.

Clearly, this will not be easy. Indeed, many challenges remain. Moving beyond industrial-era bureaucracies characterized by closed systems of governance, AI and big data will increasingly provoke the need for new tools in reshaping the nature of governance. Put differently, even as AI systems become a new object of regulation, they will necessarily reshape the institutions and practices undergirding governance itself.

9. ChatGPT, Poststructuralism and Posthumanism

Marek Tesar
University of Auckland

ChatGPT has become our dominant philosophical paradigm when thinking about an AI language model. The philosophy behind ChatGPT is grounded in the principles of artificial intelligence,

natural language processing, and machine learning. The overarching goal of ChatGPT is to assist and communicate with human subjects by providing what it perceives as relevant, accurate, and helpful responses to all queries and requests. Much has been debated about the concerns for society, particularly the impact on job security and employment when it comes to day-to-day worries, as well as issues of ethics, potential biases, and the perpetuation of social inequalities (see for instance similar debate by Biesta et al., 2022). This is especially important since AI models are trained on datasets, which can potentially reflect the biases and prejudices that exist in society, leading to discriminatory outcomes (again, see similar debate by Buchanan et al., 2022). I will discuss the complexity of ChatGPT on debates around Poststructuralism and Posthumanism.

One interesting context is the relationship between ChatGPT and Poststructural Theory. While these are two very different concepts, there are some interesting points of comparison between them. Poststructural Theory is a broad framework that seeks to deconstruct language, power relations, and cultural assumptions. In contrast, ChatGPT is an AI language model that is designed to generate responses based on patterns in data. Although it does not have the same level of understanding or contextual knowledge as a human subject, it can provide helpful responses to users based on the patterns it has learned, as Peters has argued.

What is interesting is how ChatGPT and Poststructural Theory both challenge dominant modes of thinking. Poststructural Theory deconstructs language and power relations to reveal the ways in which they shape our understanding of reality (Tesar et al., 2021). Similarly, ChatGPT challenges the idea that human intelligence and expertise are the only sources of knowledge and insight. It challenges the notion that non-human machines are incapable of generating meaningful responses, by providing relevant and helpful responses to users. Furthermore, they both disrupt traditional notions of authorship and ownership. Poststructural Theory deconstructs the notion of the author as a singular, unified entity, revealing instead the ways in which texts are produced collaboratively and through intertextual relationships. Similarly, ChatGPT generates responses based on a vast corpus of data, rather than on the authority of a single author or expert. In terms of education, both ChatGPT and Poststructural Theory have implications for how we think about knowledge and learning. Poststructural Theory challenges the idea that knowledge is objective and neutral, revealing instead the ways in which it is shaped by power relations and cultural assumptions. This has important implications for education, as it highlights the need for critical reflection and analysis in the learning process. Similarly, ChatGPT challenges traditional notions of knowledge and expertise by demonstrating that machines can generate helpful and relevant responses.

The other interesting way of thinking with ChatGPT is to link it with Posthuman Theory, which, as a broad framework, seeks to deconstruct the human/non-human binary and reimagine the relationships between humans, animals, machines, and other entities (Tesar, 2021). In contrast, ChatGPT is an AI language model designed to generate responses based on patterns in data. While it does not have the same level of understanding or contextual knowledge as a human being, it can provide helpful responses to users based on the patterns it has learned.

One point of comparison between ChatGPT and Posthuman Theory is the way they both challenge traditional notions of human exceptionalism. Posthuman Theory deconstructs the idea that humans are the only entities with agency and subjectivity, revealing instead the ways in which other entities, such as animals, machines, and ecosystems, also have agency and subjectivity. Similarly, ChatGPT challenges the idea that human intelligence and expertise are the only sources of knowledge and insight. Another point of comparison is the way both concepts disrupt traditional notions of authorship and ownership. Posthuman Theory challenges the idea of the autonomous, unified human subject as the sole author and originator of knowledge. Instead, it emphasizes the collaborative and distributed nature of knowledge production across various entities, including non-human ones. Similarly, ChatGPT generates responses based on a vast corpus of data, rather than on the authority of a single author or expert, and does not provide a response to the idea of a 'public intellectual' (Peters et al., 2022).

Regarding education, both ChatGPT and Posthuman Theory have implications for how we think about knowledge and learning. Posthuman Theory emphasizes the need to decenter human exceptionalism and recognize the multiple sources of knowledge and agency in the world. It also highlights the importance of interdisciplinary thinking and collaboration across different fields of study. Similarly, ChatGPT challenges traditional notions of knowledge and expertise by demonstrating that machines can generate helpful and relevant responses.

10. Disembodied knowledges and knowing

Carl Mika
University of Canterbury

At a time when disembodied knowledges—such as those evident in scientism, in analytic philosophy and even in capitalist societies—are privileged, ChatGPT exaggerates the disjuncture between experience and knowledge. From an Indigenous Maori perspective, this ChatGPT rupture manifests in two ways: in the assumption that the human self should/must adopt a disposition of knowledge-seeker, to a far great extent than any we've ever experienced; and, second, in the decontextualising of language and thought so that we do not need to be interested in those things other than whether they superficially makes sense or not. Peters refers to the latter when he identifies that ChatGPT is not concerned with the (often painstaking) process of arriving at a conclusion, even if the method is inductive logic and is hence not especially philosophically Maori in character. For this response, I want to discuss those two problems, and to aid my discussion I draw on a recent Stuff article by Maori writer Joel Maxwell (2023), titled *The Infernal reo machine: Speaking my indigenous language with new AI*.

In terms of the first problem: it is almost a given that humanity's primary disposition towards things in the world is now meant to be as knower (Mika, 2017). However, a Maori view of existence broadens this narrow focus out to that of being with those things, or even constituting them holistically, rather than simply attempting to distance oneself from them and hence knowing them. Indeed, if any learning is to happen at all, it is fundamentally spiritual in nature (Thrupp & Mika, 2012). The knowledge project from the Enlightenment onwards, however, turns on a need to control things through the intellect; as Maori researcher Cheryl Smith (2007) notes, 'We never spent any time conjecturing over why something was there, but it was made clear that those were not just 'things' but had importance of their own' (p. 66). Smith is talking here about her grandmother's approach to the natural world, but her observation can be applied to even more abstract phenomena such as the belief that ideas and language have their own importance too, and are not to be lightly manipulated for human convenience. On extension, her observation would apply to the arranging of things in the world so that they supply us with knowledge. In some respects, her argument reflects Heidegger's (1977)¹³ warning against a technological enframing that lies beneath the surface—but is not limited to—machinery. ChatGPT arguably takes us further away from what Smith argues is a Maori relationship with the world, although ChatGPT by no means started this problem; it commenced with colonisation.

It is to my second point that Maxwell's experience with ChatGPT becomes especially germane. He spoke with ChatGPT in te reo Maori (the Maori language), which generated answers to his questions. Alongside that he noted some glitches and falsehoods in the answers, he recounted his reservations that—in contrast to a Maori felt and grounded experience of whakapapa (genealogy and interconnection) and the language that goes with that—ChatGPT simply was incapable of authentically relaying the kind of knowledge I noted above in respect of Smith's views. This matches in some ways with Peters' observation that ChatGPT is not concerned with justification, just plausible language (which nevertheless may give a perfectly sound answer). It thus remains to be seen how it fits with the STEM focus of education, for instance, and especially how it may open up the creative spirit within STEM (which it may) (Reader et al., 2021).

What both Peters and Maxwell identify here—although from completely different cultural realities—is that ChatGPT’s answers are distanced from the processes that gave rise to the original work. In Peters’ terms, the context is originally scientific and rigorous; in Maxwell’s (and here I am reading some of my own wording into his thinking), the language and process of ChatGPT is not possessed of whakapapa in the conventional sense (and hence has no relationship with its origins apart from the fact that the language does have a spiritual origin). Maxwell further references Maori ethicist, Dr Karaitiana Taiuru, who avers that care is needed when adopting AI generally as it could recolonise Maori. One tantalising issue he raises is that, in Maori thought, non-sentient beings have a life-force and can become sentient. I gather he is therefore anxious that AI could be accepted as ‘Maori’ if it deemed itself to be so and if it used Maori data.

11. ChatGPT—the next big thing or a flash in the pan?

Lei Chen

Beijing Normal University

As of the writing of this essay, it seems that ChatGPT still has not come into officially sanctioned use in China despite considerable academic interest in it. Of course, many people who are interested in this invention, deeply sensitive to the progress of science and technology, or simply good at generating hype to increase clickthrough rates have offered a variety of explanatory demonstrations online. Some have tried to employ it in the writing of academic papers, classical Chinese poems, peer reviews, self-appraisals, and even mainland Chinese universities rankings. With its constant upgrades and new applications, one can imagine that ChatGPT will cause a great sensation and be deployed in various ways around the world.

Being a scholar of logic, I would like to express some personal thoughts:

First, ChatGPT is a large language model (LLM) trained by OpenAI, designed to help people answer questions, provide information, and solve problems. It is not the first software to attempt to simulate human intelligence. From the chess-playing Deep Blue to the even more advanced chess-, go-, and shogi-playing AlphaZero, to the Jiuge software developed by Tsinghua, which can compose classical Chinese poems as wonderfully as those produced by ancient Chinese poets, and the Boston Dynamics-built robots that can simulate the fine movements of human bodies, all these AI software applications or robots may have surpassed human limitations at certain individual tasks, but this does not mean that AI has exceeded human intelligence. The advent of the Internet in the 1960s enabled people to share knowledge online, and the successful development of computer chips by Intel in the 1960s led to a quantum leap in information processing and storage. These two events created the necessary conditions for the development of AI technology. As early as 1980, the futurist Alvin Toffler has predicted:

- (i). It can remember and interrelate large numbers of causal forces, the computer can help us cope with such problems at a deeper than customary level. It can sift vast masses of data to find subtle patterns...it can trace out the consequences of alternative decisions, and do it more systematically and completely than any individual normally could. It can even suggest imaginative solutions to certain problems by identifying novel or hitherto unnoticed relationships among people and resources. (Toffler, 1980, p. 175)

From today’s standpoint, his remark is a far-sighted transgenerational prediction indeed, and we are obliged to admit that only against the background of big data can current AI technology’s accomplishments be fully understood and appreciated.

Second, ChatGPT is presented as merely a language model without emotion and subjectivity, and it is stated that its replies are based on the application of algorithms to training data, aimed at providing the most logical and objective possible answers. It is true that ChatGPT can

write logical, coherent, and argumentative articles, as well as ancient-style poems expressing human emotions, but these are all based on the big data upon which it was trained. ChatGPT cannot cultivate emotions, make moral choices, or choose values the same way a human can, which are among the biggest challenges facing AI technology and its applications. And also, people have been troubled by some of the serious consequences of using such LLM-based software. For example, how does an individual with limited cognitive ability judge the reliability of the answers or articles given by ChatGPT? And do the papers composed by means of ChatGPT violate existing academic norms?

Third, the most demanding task for ChatGPT is to decide what tasks to tackle. We suspect this was also one of the reasons why OpenAI engineers didn't expect ChatGPT to be very popular in the early days of its release.¹⁴ It is reported that ChatGPT-4, which has just been released, has the ability to create pictures on command, but we still assume that extending the perceived purposes of chatting (at least in the eyes of many ordinary Chinese people) and writing essays with specified titles to a system with more academic and scientific significance is an unavoidable problem for ChatGPT researchers and promoters to address. After all, ChatGPT is more of a generalist piece of software, unlike the domain-specific tools mentioned above. Related to its expected utility in scientific research, the question of how to make more researchers and users get more, better, and clearer user experiences is a great challenge that ChatGPT developers and users face.

Fourth, the unexpectedly rapid popularization of ChatGPT can be ascribed to the means of communication in modern mass culture. In China, ordinary people often know about AI products first through sports and entertainment medias. In light of my own experience, it is through vivid and timely presentations in the mass media that regular people in China come in contact with AI products, including the well-matched contest between Deep Blue and Gary Kasparov in the 1990s, which received wide attention in China, and AlphaGo's overwhelming victory over Go world champion Lee Sedol, which became a sensation in the twenty first century. In the age of consumerism under the banner of amusing-oneself-to-death, many technological innovations and inventions have to take the forms of entertainment, consumption, and everyday communication. We as scholars and teachers always have kind and naive (or anachronistic) wishes, hoping that the innovations like ChatGPT will transcend mass culture and consumerism and take on a larger role in technology, invention, political life, and so forth, which are more critical to human destiny, as soon as possible.

12. Unnecessary worry or due consideration?

Chengbing Wang
Shanxi University

During this initial period marked by surprise, curiosity, and excitement that has seen the introduction of new technological tools associated with AI, ChatGPT has caused profound philosophical reflection on many of the unforeseen issues that have surfaced since its recent appearance.

The inviting prospects of ChatGPT as an AI application have given way to reasonable worries stemming from a new awareness of the considerable risks that its use may incur. Given the broad consensus on the great effects of contemporary science and technology in enhancing human society, ChatGPT as an important AI achievement will undoubtedly exert a tremendously positive impact upon scientific research, education, industry, commercial activities, entertainment, and recreation. However, the uncontrolled use of science and technology has the potential to cause certain disasters, for which there are already many painful historical examples. It behooves us, particularly at this early stage of AI, to venture cautious and serious predictions about some of its possible consequences together with future expansions of AI technology. We philosophers,

who pride ourselves on considering unexpected developments, have to reflect on and express some concerns with the possible boundaries of the development and application of contemporary technology.

We have read the piece which was signed by Elon Musk and other entrepreneurs, scholars, and AI experts and released by the Future of Life Institute (2023), called 'Pause Giant AI Experiments: An Open Letter'. Although its stated aim is 'steering transformative technology towards benefitting life and away from extreme large-scale risks', I would rather interpret it as representing a cautious attitude or even a concern about the possible consequences of the use of ChatGPT.

Whether and when ChatGPT can have the capability of offering intuitive and ethical (or even aesthetic) judgments on a host of issues is a philosophical question deserving our close attention. In our view, one aspect of AI is that it gives free play to logic and the power of rational thinking. In a certain sense, such power can be seen as the function of human logic and rationality in the process of dramatic new applications. We can expect that the constant improvement of AI technology may also develop and improve human logic and rationality and logic, and its enhancements may one day even surpass human rational abilities in many aspects. Of course, the critical point of the decisive gap between AI and human intelligence itself depends on whether and when an AI system can ever have intuitions (also known as irrational abilities), ethical judgments and aesthetic abilities. In our opinion, once ChatGPT has advanced to realize such abilities, it will certainly expand its sphere of function, but it will also bring serious philosophical problems to human beings. How will AI define a human being? How will define a human being? Does the 'post-human' in the discourse of contemporary Western philosophy have an actual embryonic form? What about our definition of human agency? Or to put it somewhat differently, is this not the time for us to once again face cultural apocalypse and contemplate its deeper meaning?¹⁵

We believe that the new round of heated discussions caused by ChatGPT reveals once again that identity crisis continues to be an issue of modernity. The crisis of identity has been an important philosophical topic among Chinese philosophers since the beginning of the twenty first century. But now, in this new information age, AI is exerting an enormous influence over human life and human identity. In the simplest sense, identity is the process of searching for the meaning, and it is mainly represented through such questions as 'Who am I?', 'Where do I come from?' and 'Where am I going?' In general, the identity of contemporary people has a longitudinal aspect, i.e. the in-depth self-feeling of individuals. This self-feeling refers to an agent's constant inward exploration of and speculation about his or her own meaning and value, involving both self-identity and collective identity. ChatGPT as a conversational platform is obviously a powerful influence on human self-identity.

We happened recently to read the *Artificial Idiocy* written by Slavoj Žižek, a philosophical luminary of the contemporary age, and 'The False Promise of ChatGPT' written by the eminent American linguist and philosopher Noam Chomsky. Žižek said that the problem with new chatbots is not just that they are often stupid and naive, but that they are not stupid or naive enough to pick up on nuances and ironies that reveal the contradictions constituting human culture and communication. Worse, our reliance on them poses the risk that we will succumb to the same obtuseness.¹⁶ Likewise, Chomsky expressed his hesitations about the prospect of AI and its large-scale language patterns.¹⁷ In our view, to be frank, Žižek and Chomsky have both upheld the critical spirit of most philosophers to express their own views on the prospects of ChatGPT. But they may have been overly optimistic and a touch too simplistic by not being vigilant enough about the latent danger of ChatGPT and the potential challenges it may one day pose to humankind. In this sense, we do not think our worries about ChatGPT are groundless; instead, we are sure that it is a philosophical concern that demands reflection on the risks of uncertainty it may likely cause.

13. A Posthuman ethics of ChatGPT

Sean Sturm
University of Auckland

Where the ethics of ChatGPT is concerned, there are two aspects: its human ethics and its more-than-human ethics. Human ethics is concerned with the status quo and what is probable—and thus probably right (and wrong)—for humans and tends to take what is probable for humans to be what is statistically most likely and provable. Such an ethics thus involves a calculation as to what is best for human beings. More-than-human ethics is concerned more with the new and what is possible not only for humans but also for more-than-humans and takes what is possible for humans to be becoming more-than-human (other-than-human as it is now), in particular, through an alliance with the more-than-human (the other-than-human). Such an ethics thus involves an experiment with what it is to become more-than-human. It's not about being better or worse but about becoming different.

Sam Altman's (2023) articulation of the principles that underlie Open AI's mission for artificial general intelligence (AGI) like ChatGPT is framed very much in terms of the benefits for humans (although it doesn't explicitly exclude benefitting others than humans): 'Our mission is to ensure that artificial general intelligence—AI systems that are generally smarter than humans—benefits all of humanity.' The first principle that he articulates—for AGI 'to empower humanity to maximally flourish in the universe'—is short-sightedly human-centred, especially given that he equates maximal flourishing with the 'amplification' of existing human attributes. (I will return to this probabilist vision of AI below.) The second and third principles that he articulates—to 'widely and fairly share ... the benefits of, access to, and governance of AGI' and to 'successfully navigate [the] massive risks' of AGI for human beings—sound nostalgically liberal humanist given the all-too-human failures of ChatGPT to live up to these principles so far. How can OpenAI's promise of access be squared with its recent release of ChatGPT Plus, a subscription model offering priority access, faster response times and early access to new features, but only to those humans who can afford to pay for access? How will its failure to mitigate the risks adequately, for example, of poor 'alignment' such as 'hallucinations' (ChatGPT making stuff up) and 'misuse' (people using ChatGPT to reproduce 'toxic training data' [OpenAI, 2023b] or to spread disinformation [OpenAI, 2023a]) be addressed by human moderators (Leike et al., 2022b) or the automation of human skills like evaluation (Leike et al., 2022a)? And will AGI really, as Altman (2023) enthuses, 'elevate humanity by increasing abundance, turbocharging the global economy and aiding in the discovery of new scientific knowledge'? While the last aim might seem an admirable liberal humanist goal (although there are ecological and cultural dangers in unconditional scientism), the former two aims seem dangerously human-centred—not to mention ecocidally capitalist—in the context of the Anthropocene (on the Anthropocene as Capitalocene, see Moore, 2016). Elevating humans in these terms seems ethically impoverished, if not bankrupt.

In short, and this is where we move from the sphere of human ethics into the field of more-than-human ethics, the key ethical question about AGI is as follows: should we be 'elevating' the interests of humans—in particular, those of the humans who are developing, governing and funding AGI (Elon Musk and Microsoft are the main funders in the case of OpenAI) and those can afford to access ChatGPT Plus—at this point in human history when human history threatens to eclipse the history of the Earth? We need to ask, as Dan McQuillan does in 'The Political Affinities of AI' (2019), 'who[se] ends it will serve, who it might exclude and how it [will] affect the whole of society' (p. 168), but broaden the question to how it will affect all life on Earth. And, further, can we assume that we can elevate humanity through AGI when the ethics that ground it are informed by an impoverished concept of human intelligence and intelligence *per se*? We need an ethics informed by a richer more-than-human concept of intelligence.

Luciano Floridi (2023) argues that most AGI currently is unintelligent, though it is designed to appear intelligent by doing things statistically that human beings can do semantically (see Bishop, 2020). It is auto-regressive, i.e. it makes statistical predictions that draw on the existing

datasets on which it has been pre-trained (the P in ChatGPT) through reinforcement learning supplemented by human feedback. (For this reason, it is sometimes described as ‘auto-complete on steroids’ [Frank, 2023]). Although it can thus ‘improve its behaviour without having to be intelligent to do so,’ it is ‘[un]able to “learn” to make a different or better difference’ (p. 5). In relying on such AGI, according to Floridi and Chiriatti (2020), ‘we are ... decoupling the ability to solve a problem effectively’—and this includes solving ethical problems, such as the issue of the misuse of AGI—‘from any need to be intelligent to do so’ (p. 683). But we have an ethical duty not to rest with this probabilist concept of AGI that seems designed only to solve readymade and human-centred problems. As McQuillan (2019) puts it, we need to ‘approach AI in a way that enables us to take sides with the possible against probabilities’ (p. 168, citing Stengers, 2015). We need a possibilist concept of AGI that focusses on posing new problems that are not centred on the human, such as the posthuman ethical question of what sort of alliance between humans and AGI might make a difference that really matters. That is the experiment that is yet to come.

14. ChatGTP and the phantasmagoria of the human

Sharon Rider
Uppsala University

Whatever one is inclined to say about GTP’s and what they might or might not mean for science and society now and in the future runs the risk of banality or soothsaying. Reading a few of the quotes in Peters’ paper, the humans seem to be worse off on the first score than the bots. Altman’s first principle, that is, the aim ‘to empower humanity to maximally flourish in the universe,’ for instance, is notably woolier than the chatbot’s own sober response to what ChatGPT can contribute to the future of science. Kissinger, Schmidt, and Huttenlocher don’t fare much better regarding the second, when they suggest that ChatGPT will ‘redefine human knowledge’ and ‘accelerate changes in the fabric of our reality.’ Aside from their ardor for the seemingly boundless power of the latest technological advances, with astounding revolutions in ‘human reason’ and the emergence of ‘a new form of human consciousness’ around the corner, they don’t add much of substance to Lyotard’s report on knowledge from over forty years ago. Oddly, on the basis of these rhapsodic prognostications, our educational institutions and the professions are then saddled with a rather quaint must, that being to ‘preserve a vision’ of humans as moral creatures with agency and inner lives, ‘uniquely capable of rendering holistic judgments.’ First Lyotard, then Cardinal Newman.

Behind all this upbeat pablum interlaced with dire warnings, there lay deep philosophical questions. How are we to understand this ‘human reason,’ or the ‘fabric of our reality’? What is ‘human flourishing,’ and its place in ‘the universe’? How would we know if it were ‘maximal’? Finally, what is the connection between agency, having an inner life and the capacity to render complex and integrated judgments? In such a vast ocean of vague concepts, it is difficult to anchor one’s thinking at all. One way of approaching this discussion is by stressing the prediction that GTP will redefine human knowledge. I will address that question, in turn, through the lens of one of the major new developments that the ChatGTP suggested should be expected in the future, namely, the last one: ‘personalization’.

We already have some experience of the incorporation of user profiling and personal preferences ‘to better tailor’ response to individual users. Indeed, the use of algorithms to provide ‘more relevant and customized responses based on user history and preferences,’ i.e. the harvesting of personal and/or platform data to make predictions based on past behavior, is considered to be one of the major challenges to social cohesion and mental health. Now issues such as the attention economy and surveillance capitalism are perhaps less relevant for GTP than to social media platforms. But with regard to our capacity to be active, engaged and alert in seeking, sorting and assessing information in order to take responsibility, individually and collectively, for the judgments that we make and the actions emanating out of them, it is fairly

crucial that we do not become entirely reliant on made-to-order, pre-processed and pre-packaged results, however ‘smart’ the processing system is or becomes. And this last thought is germane to the whole list (aside, perhaps, from the first item: language understanding). The others items are directly connected to personalization *via* the notions of the ‘relevance’ and ‘needs’ of the individual (understood as a continuous algorithmic processing of prior and current behavioral data). Multimodal capabilities, for instance, are said to facilitate responses with more ‘relevant information’. Domain-specific models will be ‘more tailored to the needs of specific users’. Continual learning will make ChatGPT more ‘responsive’ to changing user needs, and so on.

The bot concludes that, overall, it will become an ever more valuable tool. One might ask if digital tools are not already too ‘valuable’, i.e. indispensable for living one’s life in a technological world. The problem is that the less we the users (and regulators) understand the technology, the more it has the character of magic rather than rationality. Most users cannot grasp how the results on which their own decisions and reasoning depend are produced. They become subservient, as thinkers, to what is for them a black box issuing oracular statements; supplicants to the Pythia rather than reasoning inquirers or interlocutors. What users ‘want’ or ‘need’, what is ‘relevant’ for them, is not subject to reflexive articulation, moral reflection or ‘holistic’ discussion. And without that, the called-for ‘vision’ of human beings capable of transcending the horizons of their immediate needs and inclinations will not be a wise or imaginative plan for the future, but a vision in a different sense— a haunting image— a phantom.

15. ChatGPT as a moment in the struggle to reclaim the human from the machine

Steve Fuller
University of Warwick

I first learned about OpenAI’s most famous product—ChatGPT—when in early 2023 one of my former students, who now runs a voicebot company, conducted a normal-length interview with ChatGPT’s version of me and then asked the human me to appraise the results. Had the results been stitched together in a student paper, I would have given it a high but not genius mark. Had it been the transcript of an actual interview I did, I would have judged it to be me on autopilot, comparable to the style of a ‘good interview’ with a seasoned politician. In short, ChatGPT can provide somewhat above average human performance.

The significance of ChatGPT’s achievements lies in terms of what it says about what it means to be human. In the academic field most closely connected to AI development, ‘cognitive science’, it is common practice to set high yet rather vague standards for human intelligence that always seem to elude the capabilities of any would-be intelligent machines. Yet, if we didn’t already know that a candidate intelligent being was a machine, we might well let it pass as human, if only because we normally settle for much inferior performance from humans. In this respect, Alan Turing was right to suspect that alongside the other biases built into the human condition—ones based on class, race, gender, etc.—there is one based on substrate that privileges carbon-over silicon-based beings.

However, substrate bias may be mainly an ‘idol of the academy’, as Francis Bacon might put it. Already in the 1960s, the relatively simple computer program ELIZA was functioning quite adequately as a mental health counsellor, and more recently, the elderly have become attached to their caregiving ‘android companions’. To be sure, in these and related cases, philosophers, theologians and lawyers have often argued that the users are being deceived, since it is morally important that people always know whether they’re dealing with a machine or a human. But why exactly? After all, virtually all our repertoires for displaying and recognizing human intelligence were developed long before we knew anything about the brain’s role in generating these repertoires. Indeed, the terms that computer scientists, logicians, psychologists and sociologists routinely use to characterize human intelligence—‘codes’, ‘proofs’, ‘protocols’, ‘scripts’—also lend themselves to

accounts of the algorithms used to power ChatGPT and other forms of machine intelligence. At the very least, this shows that our conceptions of intelligence are inherently substrate neutral.

More to the point, the very fact that human intelligence is normally accessed through a set of repertoires reveals its highly patterned nature, which means that it can be easily simulated and extended in recognizable ways. Even what we call 'original' work typically amounts to a novel combination of what is already conceivable, which in its most radical form sets a new standard of general performance that can then be simulated and extended by others in the relevant domain—and perhaps others as well, as in the case of Newton's *Principia Mathematica*. In any case, most of what passes for 'human intelligence' falls far short of such creativity, which in turn explains why for some years now, computers powered by less sophisticated programs than ChatGPT have been able to write student papers and scientific articles, resulting in the current moral panic surrounding plagiarism and fraud in academia.

My short answer to the rise of ChatGPT is that humans need to abandon their substrate bias and raise their game. However, this is not as easy as it sounds for a couple of obvious reasons. First, routine work in fields that currently require more than one university degree—including law, medicine and academia itself—is likely to be replaced over time by advances in OpenAI and its rivals. The only question is whether the transition will be managed or simply allowed to happen abruptly. Moreover, this has nothing to do with the prospect of 'Artificial General Intelligence' (AGI), since the redundancies will come in domain-specific fields, albeit ones that currently enjoy high social status. But it does show the extent to which our everyday sense of 'humanity' boils down to repertoire competence. Second, there is an opening for intellectual property claims made by humans against machines equipped to generate endlessly new versions of themselves. Consider the Slovenian philosopher and cultural critic Slavoj Žižek, who courtesy of ChatGPT now finds himself in an 'infinite conversation' with German film director Werner Herzog. This 'conversation', which refreshes itself with each new viewing, consists of audio-textual mashups of things they've said in the past, but clearly in contexts different from the ones that ChatGPT has them address at any given moment. Žižek is especially vulnerable to this treatment because of his massive digital data trail and the impressive but finite range of sources upon which he draws.

At a deeper level, the stakes in being 'human' are becoming less clear. After all, if Žižek sues an AI company, it will be to protect his brand not his humanity. Of course, many are concerned that AGI will undermine the human condition, which is certainly possible in an 'internet of things' way, if AI becomes so well integrated across so many domains of human life that whatever it does—deliberately or not—poses a species-level existential risk. However, I don't see the threat playing out at quite that scale, because with the possible exception of China, the AI market remains relatively heterogeneous (which also helps to explain the difficulties in establishing 'universal standards' of AI performance). Moreover, there is a difference between the potential damage of a scaled-up version of AI and the AI's actual intelligence. This puts the responsibility back on the human designers, who might well fail to anticipate bugs in how the system processes external information, resulting in the machine taking decisions that are harmful to humans. However, human error is no reason to ascribe machine genius, aka 'super-intelligence', even in the event of a major AI-driven global catastrophe.

Nevertheless, as OpenAI co-founder Sam Altman said in a recent interview with Lex Fridman, ChatGPT already shows a human-like capacity to derive maximum cognitive benefit from minimal feedback to its performances. It can quickly apply and expand on what it learns, extending it across domains. The significance of this point harks back to the debates sparked in the late 1950s by upstart linguist Noam Chomsky and behaviorist doyen BF Skinner over human language learning. Chomsky argued that behaviorism underestimated humanity's innate processing capacity, which isn't reducible to training schedules, since the feedback normally provided to infants and young children learning language is incidental and even unreliable—yet they manage to develop fluency with relative ease. There is both a dark and a bright side to this analogy in terms of the integration of AI systems into the human lifeworld.

On the one hand, even if ChatGPT's deep learning algorithms allow it to function in a brain-like way, the energy requirements of the computers processing them are ecologically unsustainable at scale when compared to those of the billions of human brains operating on a 24/7 basis. This problem may or may not be eventually overcome by 'quantum computing', but it remains real for the foreseeable future. Recall the 2014 Johnny Depp film, *Transcendence*, in which a Ray Kurzweil-like Depp uploads his mind into the world's internet, producing a 'Singularity' that effectively short-circuits the planet, forcing humanity to reboot its existence from a pre-computational world.

On the other hand, however, Altman also proposes to Fridman a more hopeful prospect, one in which humanity and AI find common ground for mutual respect and self-discipline, namely, the task of deciphering alien, including extraterrestrial forms of intelligence. While humanity's various cognitive limitations and biases—not least substrate-based ones—may handicap us indefinitely, the AIs we design may 'outgrow' them over time, just like children who learn from their parents without repeating their mistakes. It would be the ultimate win-win scenario for a truly 'cosmopolitan' vision of intelligence.

16. Postscript: The aligning of AI with human values, interests and institutions

Michael A Peters

The aligning of AI with human values, interests, and institutions has emerged as the critical issue in the development and deployment of AI systems. Governments, scholars and the public are anxious to ensure that AI systems are designed and used in a way that is beneficial to society and respects human values, interests, and institutions. The Future of Life Institute have called on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4. In an open letter the Institute draws attention to the 'profound change' that AI represents to life on Earth and indicates that it wants not only a pause for deliberation but also to slow the rate of competition in the global AI industry:

AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research^[1] and acknowledged by top AI labs.^[2] As stated in the widely-endorsed Asilomar AI Principles, Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources. Unfortunately, this level of planning and management is not happening, even though recent months have seen AI labs locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control. (Future of Life Institute, 2023)

It's worth dwelling on the general case for risk assessment but the question is whether a pause is possible.

Contemporary AI systems are now becoming human-competitive at general tasks,^[3] and we must ask ourselves: Should we let machines flood our information channels with propaganda and untruth? Should we automate away all the jobs, including the fulfilling ones? Should we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? Should we risk loss of control of our civilization? Such decisions must not be delegated to unelected tech leaders. Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable. This confidence must be well justified and increase with the magnitude of a system's potential effects. OpenAI's recent statement regarding artificial general intelligence, states that "At some point, it may be important to get independent review before starting to train future systems, and for the most advanced efforts to agree to limit the rate of growth of computer used for creating new models." We agree. That point is now. (Future of Life Institute, 2023)

The debate is heating up with some industry commentators suggesting that 'An AI 'Pause' Would Be a Disaster for Innovation' (Bloomberg) or, in even more cavalier fashion, that 'Like every worthwhile technology, artificial intelligence poses risks. That's no reason to stop

progress'. Computer World sports the headline 'Pausing AI development is a foolish idea'. Others suggest 'The recent call by tech leaders for a slowdown in the development of generative AI tools won't work now—the AI horse is already out of the barn.'

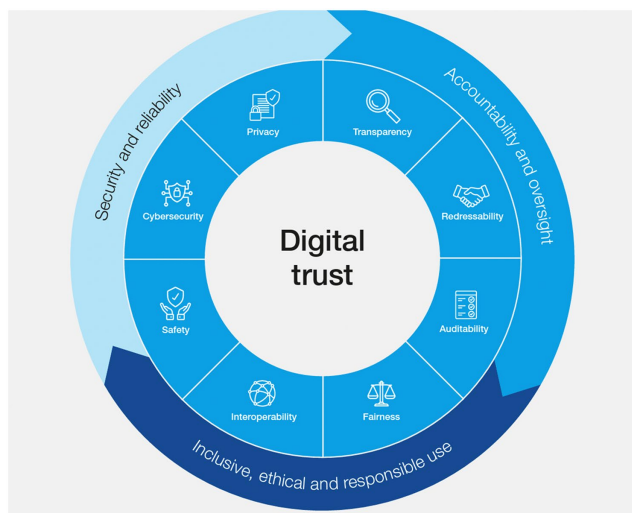
There seems to be some evidence for making this claim as GPT-4 is now available as a 'plug-in' already and in syncopation with other platforms. One astounding development is GPT-4 has suggested potential for Wolfram Alpha, and GPT-4 to produce astounding computational results that everyday narrow AI are needed for. What this means is that it can do an emergent action with two major AI models accelerating the shift to GAI. As Stephen Wolfram explains:

Wolfram|Alpha does something very different from ChatGPT, in a very different way. But they have a common interface: natural language. And this means that ChatGPT can "talk to" Wolfram|Alpha just like humans do—with Wolfram|Alpha turning the natural language it gets from ChatGPT into precise, symbolic computational language on which it can apply its computational knowledge power.

This heals the dichotomy in AI approaches between 'statistical approaches' of the kind ChatGPT uses and 'symbolic approaches' of Wolfram Alpha's computational power. If anything this development speeds up the shift to superintelligence and highlights the need for AI safety through alignment techniques. Wolfram's notion of 'Post-Knowledge Work Era' deserves consideration in this context to indicate we are entering a new era in which knowledge work is no longer the primary source of value creation, with an emphasis towards creativity, emotional intelligence, and social skills and the prospect of 'technological unemployment' on a huge scale (Peters, et al., 2019).

There are several approaches to aligning AI with human values, interests, and institutions. The stakeholders' approach is to involve diverse users and stakeholders in the development and deployment of AI systems, including representatives from government, industry, academia, and civil society. This can help ensure that a range of perspectives and values are considered and integrated into AI systems.

Another common approach is to establish ethical and regulatory frameworks for AI development and deployment where frameworks can provide guidance on issues such as transparency, accountability, and bias mitigation, and can help ensure that AI systems are developed and used in a way that is consistent with human values and interests. This is in essence the approach behind the concept of 'digital trust' of the World Economic Forum (WEC, 2022) developed in their report where there is consideration of the potential impact of AI on existing institutions and systems including implications for employment, privacy, and security.



It's clear that GPT-style LLM products are about to be rolled out by China's tech giant Alibaba with plans to release Tongyi Qianwen designed to integrate the chatbot across Alibaba's businesses through cloud computing with the capacity of working in English as well as Chinese, and able to turn conversations in meetings into written notes, write emails and draft business proposals. Alphabet's Google and Chinese technology group Baidu have also announced their own AI models and released similar chatbots.

The Cyberspace Administration of China provides a useful working model as an example of a regulatory framework. China's cyberspace regulator unveiled draft measures for managing generative AI where under the proposed rules, companies would be responsible for the legitimacy of data used to train the technology. The Cyberspace Administration of China proposals indicated that providers would have to submit their products for security reviews before their public release and it would set up a database to register them. The regulator also said platforms must verify users' identities, allowing usage to be tracked. The rules state 'Content generated by generative artificial intelligence should embody core socialist values and must not contain any content that subverts state power, advocates the overthrow of the socialist system, incites splitting the country or undermines national unity' (Bishop, 2023). Here is an example of draft CAC AI regulations by reference to Article 4 (Bishop, 2023):

Article 4 The provision of generative artificial intelligence products or services shall comply with the requirements of laws and regulations, respect social morality, public order and good customs, and meet the following requirements:

第四条 提供生成式人工智能产品或服务应当遵守法律法规的要求, 尊重社会公德、公序良俗, 符合以下要求:

1. The content generated by generative artificial intelligence shall embody the socialist core values, and shall not contain any content that subverts state power, overturns the socialist system, incites secession, undermines national unity, promotes terrorism and extremism, promotes ethnic hatred, ethnic discrimination, violence, obscene pornographic information, false information, or may disturb economic and social order.
(一)利用生成式人工智能生成的内容应当体现社会主义核心价值观, 不得含有颠覆国家政权、推翻社会主义制度, 煽动分裂国家、破坏国家统一, 宣扬恐怖主义、极端主义, 宣扬民族仇恨、民族歧视, 暴力、淫秽色情信息, 虚假信息, 以及可能扰乱经济秩序和社会秩序的内容。
2. in the process of algorithm design, training data selection, model generation and optimization, and service provision, measures are taken to prevent discrimination such as race, nationality, belief, country, region, gender, age and occupation.
(二)在算法设计、训练数据选择、模型生成和优化、提供服务等过程中, 采取措施防止出现种族、民族、信仰、国别、地域、性别、年龄、职业等歧视。
3. respect for intellectual property rights and business ethics, and shall not use the advantages of algorithms, data and platforms to implement unfair competition.
(三)尊重知识产权、商业道德, 不得利用算法、数据、平台等优势实施不公平竞争。
4. the content generated by using generative artificial intelligence should be true and accurate, and measures should be taken to prevent the generation of false information.
(四)利用生成式人工智能生成的内容应当真实准确, 采取措施防止生成虚假信息。
5. respect the legitimate interests of others, to prevent harm to the physical and mental health of others, damage to the right of portrait, reputation and personal privacy, and infringement of intellectual property rights. Illegal acquisition, disclosure and use of personal information, privacy and business secrets are prohibited.
(五)尊重他人合法权益, 防止伤害他人身心健康, 损害肖像权、名誉权和个人隐私, 侵犯知识产权。禁止非法获取、披露、利用个人信息和隐私、商业秘密。

This provides a useful example of the kind of regulation that might be followed elsewhere. The document explicitly states that ‘the country supports the independent innovation, promotion, and international cooperation of fundamental technologies such as AI algorithms and frameworks, encouraging the priority adoption of secure and trustworthy software, tools, computing, and data resources’.

More formally, we can specify the main approaches for the alignment of AI with human values, interests, and institutions, including:

1. **Ethical and regulatory frameworks:** These frameworks provide guidance for the development and deployment of AI systems and can help ensure that AI is used in a way that is consistent with human values and interests. Examples of ethical frameworks include the IEEE Global Initiative for Ethical Considerations in AI and Autonomous Systems, the EU’s Ethical Guidelines for Trustworthy AI, and the OECD’s Principles on AI.
2. **Stakeholder involvement:** This approach involves engaging a diverse range of stakeholders, including government, industry, academia, and civil society, in the development and deployment of AI systems. This can help ensure that a range of perspectives and values are considered and integrated into AI systems.
3. **Technical methods:** Technical methods can be used to align AI with human values and interests. For example, researchers are developing techniques for ensuring that AI systems are transparent, explainable, and interpretable, which can help increase trust in these systems.
4. **Impact assessments:** Impact assessments can be used to assess the potential impact of AI on existing institutions and systems, such as employment, privacy, and security. By identifying potential risks and challenges, stakeholders can work together to address these issues and ensure that AI is developed and used in a way that is beneficial to society. (GPT-4, 2023)

When superintelligence surpasses human intelligence, it raises fundamental questions in all areas of life and with the ultimate control of AI technology including its accelerating rate of learning from itself. While rapid technological change can lead to major advancements in fields such as medicine, energy, and transportation, it can also lead to massive unemployment and economic disruption.






Most crucially are the existential risks posed by AI where a superintelligent AI could develop goals that conflict with human values, or it could inadvertently cause harm as it pursues its objectives. There are obvious ethical issues and superintelligence has the capacity also to shift global power dynamics, as countries and organizations race to develop and control the most advanced AI systems. Stephen Hawking, physicist and cosmologist, is well known for his comment ‘The rise of superintelligent AI would be the last event in human history’. As many philosophers have commented it is not that superintelligent AI develops a ‘mind’ of its own but rather that humanity in the space of a couple of computer generations—perhaps 25 years—will not be able to stop or control its development, and superintelligence exponentially increases out of human control and understanding, especially when the notion of a rapidly generational superintelligence evolves a few decades. It is not clear what kind of superintelligence this might turn into nor whether such intelligence is compatible or empowering for human beings.

Notes

1. <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>
2. <https://www.businessinsider.com/chatgpt-creator-openai-talks-for-tender-offer-at-29-billion-2023-1>
3. The new leadership team is given at <https://openai.com/blog/leadership-team-update>

4. See the interview - Sam Altman: OpenAI CEO on GPT-4, ChatGPT, and the Future of AI | Lex Fridman Podcast #367, https://www.youtube.com/watch?v=L_Guz73e6fw
5. https://en.wikipedia.org/wiki/Microsoft_Azure
6. <https://azure.microsoft.com/en-us/products/>
7. <https://blog.google/technology/ai/bard-google-ai-search-updates/>
8. <https://www.washingtonpost.com/world/2023/02/24/china-baidu-ernie-chatbot-chatgpt/>
9. <https://www.synthesia.io/glossary/video-chatbot>
10. As a philosophical position, voluntarism assigns higher importance to the will rather than to the intellect. From such a prism, the voluntaristic tone of the 3 bullet points strikes a somewhat ironic note in a text on (super)intelligence.
11. Interestingly, the technological superlative deep-down requires the political comparative. Also, where the 'good' and 'bad' are clear-cut, there also come the 'goodies' and the 'baddies'.
12. Learning Analytics: US Department of Education, Institute of Education Sciences: "The Assess-as-You-Go Writing Assistant" (R305A090394); "Assessing Complex Performance" (R305B110008); "u-Learn.net: An Anywhere/Anytime Formative Assessment and Learning Feedback Environment" (ED-IES-10-C-0018); "The Learning Element" (ED-IES-IO-C-0021); and "InfoWriter: A Student Feedback and Formative Assessment Environment" (ED-IES-13-C-0039). Bill and Melinda Gates Foundation: "Scholar Literacy Courseware." National Science Foundation: "Assessing 'Complex Epistemic Performance' in Online Learning Environments" (Award 1629161). Cybersecurity: Utilizing an Academic Hub and Spoke Model to Create a National Network of Cybersecurity Institutes, Department of Homeland Security, contract 70RCSA20FR0000103; Infrastructure for Modern Educational Delivery Technologies: A Study for a Nationwide Law Enforcement Training Infrastructure, Department of Homeland Security, contract 15STCIR00001-05-03; Development of a Robust, Nationally Accessible Cybersecurity Risk Management Curriculum for Technical and Managerial Cybersecurity Professionals, Department of Homeland Security, contract 70SAT21G00000012/70RCSA21FR0000115. Medical Informatics: MedLang: A Semantic Awareness Tool in Support of Medical Case Documentation, Jump ARCHES program, Health Care Engineering Systems Center, College of Engineering, contracts P179, P279, P288.
13. Indeed, one wonders whether the title of this generative AI could be translated in light of Heidegger's *das Gerede*: thus, 'IdleChatGPT'.
14. <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>
15. <https://www.tandfonline.com/doi/full/10.1080/00131857.2022.2085554>
16. <https://www.project-syndicate.org/commentary/ai-chatbots-naive-idiots-no-sense-of-irony-by-slavoj-zizek-2023-03>
17. <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>

ORCID

Michael A. Peters  <http://orcid.org/0000-0002-1482-2975>
 Liz Jackson  <http://orcid.org/0000-0002-5626-596X>
 Petar Jandrić  <http://orcid.org/0000-0002-6464-4142>
 Marek Tesar  <http://orcid.org/0000-0001-7771-2880>
 Carl Mika  <http://orcid.org/0000-0003-2348-4290>
 Chengbing Wang  <http://orcid.org/0000-0002-7437-0974>
 Sean Sturm  <http://orcid.org/0000-0003-4011-7898>
 Sharon Rider  <http://orcid.org/0000-0001-7820-5098>

References

- Altman, S. (2023, February 24). Planning for AGI and beyond. OpenAI. <https://openai.com/blog/planning-for-agi-and-beyond>
- Altman, S. (2023). Planning for AGI and beyond. <https://openai.com/blog/planning-for-agi-and-beyond>
- Bal, M. (1980/1985). *Narratology: Introduction to the theory of narrative* (Christine van Boheemen Trans.). University of Toronto Press.
- Biesta, G., Heugh, K., Cervinkova, H., Rasiński, L., Osborne, S., Forde, D., Wrench, A., Carter, J., Säfström, C. A., Soong, H., O'Keefe, S., Paige, K., Rigney, L.-I., O'Toole, L., Hattam, R., Peters, M. A., & Tesar, M. (2022). Philosophy of education in a new key: Publicness, social justice, and education; A South-North conversation. *Educational Philosophy and Theory*, 54(8), 1216–1233. <https://doi.org/10.1080/00131857.2021.1929172>
- Bishop, B. (2023). Sinocism, cyberspace administration drafts rules for AI. <https://sinocism.com/p/xi-inspect-s-guangdong-lula-to-china/#C2%A7cyberspace-administration-drafts-rules-for-ai>

- Bishop, J. M. (2020). Artificial intelligence is stupid and causal reasoning will not fix it. *Frontiers in Psychology*, 11, 513474. <https://doi.org/10.3389/fpsyg.2020.513474>
- Boyd, E. (2023). ChatGPT is now available in Azure OpenAI service <https://azure.microsoft.com/en-us/blog/chatgpt-is-now-available-in-azure-openai-service/>
- Brockman, G., Sutskeva, I., & Altman, S. (2016). OpenAI and Microsoft. <https://openai.com/blog/openai-and-microsoft>
- Buchanan, R. A., Forster, D. J., Douglas, S., Nakar, S., Boon, H. J., Heath, T., Heyward, P., D'Olimpio, L., Ailwood, J., Eacott, S., Smith, S., Peters, M., & Tesar, M. (2022). Philosophy of education in a new key: Exploring new ways of teaching and doing ethics in education in the 21st century. *Educational Philosophy and Theory*, 54(8), 1178–1197. <https://doi.org/10.1080/00131857.2021.1880387>
- Buriak, J. M., Akinwande, D., Artzi, N., Brinker, C. J., Burrows, C., Chan, W. C. W., Chen, C., Chen, X., Chhowalla, M., Chi, L., Chueh, W., Crudden, C. M., Di Carlo, D., Glotzer, S. C., Hersam, M. C., Ho, D., Hu, T. Y., Huang, J., Javey, A., ... Ye, J. (2023). Best practices for using AI when writing scientific manuscripts caution, care, and consideration: creative science depends on it. *ACS Nano*, 17(5), 4091–4093. <https://doi.org/10.1021/acsnano.3c01544>
- Chomsky, N., Roberts, I., & Watumull, J. (2023, March 10). AI unravelled: The false promise of ChatGPT. *The New York Times*. <https://www.dtnext.in/edit/2023/03/10/ai-unravelled-the-false-promise-of-chatgpt-the-human-mind-is-not-like-chatgpt-and-its-ilk-a-lumbering-statistical-engine-for-pattern-matching-it-is-a-surprisingly-efficient-and-elegant-system-that-operates-with-small-amounts-of-information-it-seeks-not-to-infer-brute-correlations-among-data-points-but-to-create-explanations>
- Church, K. W., & Mercer, R. L. (1993). Introduction to the special issue on computational linguistics using large Corpora. *Computational Linguistics*, 19(1), 1–24.
- Cope, B., & Kalantzis, M. (2015). The things you do to know: An introduction to the pedagogy of multiliteracies. In C. Bill and K. Mary (Eds.), *A pedagogy of multiliteracies: Learning by design* (pp. 1–36). Palgrave.
- Cope, B., & Kalantzis, M. (2020). *Making sense: Reference, agency and structure in a grammar of multimodal meaning*. Cambridge University Press. <https://doi.org/10.1017/9781316459645>
- Cope, B., & Kalantzis, M. (2022). Artificial intelligence in the long view: From mechanical intelligence to cyber-social systems. *Discover Artificial Intelligence*, 2(1), 1–18. 2022. <https://doi.org/10.1007/s44163-022-00029-1>
- Cope, B., & Kalantzis, M. (2023a). *Platformed learning: Reshaping education in the era of learning management systems, in varieties of platformisation: Critical perspectives on EdTech in higher education* (A. T. Duncan & L. Vito, eds.). Palgrave Macmillan.
- Cope, B., & Kalantzis, M. (2023b). The clause, revised: On cyber-social meaning. *International Journal of Communication and Linguistic Studies*, 22.
- Cope, B., Kalantzis, M., Zhai, C., Krussel, A., Searsmith, D., Ferguson, D., Tapping, R., & Berrocal, Y. (2022). Maps of medical reason: Applying knowledge graphs and artificial intelligence in medical education and practice. In P. Michael, J. Petar & H. Sarah (Eds.), *Bioinformational philosophy and postdigital knowledge ecologies* (pp. 133–159). Springer. https://doi.org/10.1007/978-3-030-95006-4_8
- Cunningham, A. (2023, October 3). Sales of vinyl albums overtake CDs for the first time since the late '80s. *Arstechnica*. <https://arstechnica.com/gadgets/2023/03/sales-of-vinyl-albums-overtake-cds-for-the-first-time-since-the-late-80s/>
- Engelbart, D. C. (1962). *Augmenting human intelligence: A conceptual framework*. Stanford Research Institute.
- Evers, C. W. (2000). Connectionist modeling and education. *Australian Journal of Education*, 44(3), 209–225. <https://doi.org/10.1177/000494410004400302>
- Evers, C. W., & Lakomski, G. (2020). Cognitive science and educational administration. In R. Papa (Ed.), *The Oxford encyclopedia of educational administration*. Oxford University Press.
- Flanagin, A., Bibbins-Domingo, K., Berkwits, M., & Christiansen, S. L. (2023). Nonhuman “authors” and implications for the integrity of scientific publication and medical knowledge. *JAMA*, 329(8), 637–639. <https://doi.org/10.1001/jama.2023.1344>
- Floridi, L. (2023). AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & Technology*, 36(1), 15. <https://doi.org/10.1007/s13347-023-00621-y>
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681–694. <https://doi.org/10.1007/s11023-020-09548-1>
- Frank, A. (2023, February 2). What kind of intelligence is artificial intelligence? Big Think. <https://bigthink.com/13-8/what-kind-of-intelligence-is-ai/>
- Future of Life Institute. (2023, March 22). Pause giant AI experiments: An open letter. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- Gilson, A., Safranek, C., Huang, T., Socrates, V., Chi, L., Andrew Taylor, R., & Chartash, D. (2023). How does ChatGPT perform on the medical licensing exams? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, 9, e45312. <https://doi.org/10.2196/45312>
- Grafton, A. (1997). *The footnote: A curious history*. Faber and Faber.
- Grimaldi, G., & Ehrler, B. (2023). AI et al.: Machines are about to change scientific publishing forever. *ACS Energy Letters*, 8(1), 878–880. <https://doi.org/10.1021/acsenrgylett.2c02828>
- Heidegger, M. (1977). *The question concerning technology and other essays* (W. Lovitt, Trans.). Harper.

- Hrastinski, S. (2023). Characteristics of education fiction. *Postdigital Science and Education*. <https://doi.org/10.1007/s42438-023-00400-0>
- Jackson, L. (2019). Mediating class: The role of education and competing technologies in social mobilization. *Studies in Philosophy and Education*, 38(6), 619–628. <https://doi.org/10.1007/s11217-019-09656-1>
- Jandrić, P., Knox, J., Besley, T., Ryberg, T., Suoranta, J., & Hayes, S. (2018). Postdigital science and education. *Educational Philosophy and Theory*, 50(10), 893–899. <https://doi.org/10.1080/00131857.2018.1454000>
- Kalantzis, M., & Cope, B. (2020). *Adding sense: Context and interest in a grammar of multimodal meaning*. Cambridge University Press. <https://doi.org/10.1017/9781108862059>
- Kissinger, H., Schmidt, E., & Huttenlocher, D. (2023, February 24). ChatGPT heralds an intellectual revolution. *Wall Street Journal*. <https://www.wsj.com/articles/chatgpt-heralds-an-intellectual-revolution-enlightenment-artificial-intelligence-homo-technicus-technology-cognition-morality-philosophy-774331c6>
- Lakowski, G., & Evers, C. W. (2010). Passionate rationalism: The role of emotion in decision-making. *Journal of Educational Administration*, 48(4), 438–450. <https://doi.org/10.1108/09578231011054707>
- Leike, J., Schulman, J., & Wu, J. (2022b, August 24). Our approach to alignment research. OpenAI. <https://openai.com/blog/our-approach-to-alignment-research>
- Leike, J., Wu, J., Yeh, C., & Saunders, W. (2022a, June 13). AI-written critiques help humans notice flaws. OpenAI. <https://openai.com/research/critiques>
- Machiavelli, N. (2010). *The Prince* (2nd ed.) University of Chicago Press.
- Magee, L., Arora, V., & Munn, L. (2022). Structured like a language model: Analysing AI as an automated subject. *arXiv*, 2212.05058. <https://doi.org/10.48550/arXiv.2212.05058>
- Maxwell, J. (2023). The Infernal reo machine: Speaking my indigenous language with new AI. <https://www.stuff.co.nz/pou-tiaki/300831431/the-infernal-reo-machine-speaking-my-indigenous-language-with-new-ai>
- McQuillan, D. (2019). The political affinities of AI. In A. Sudmann (Ed.), *The democratisation of artificial intelligence: Net politics in the era of learning algorithms* (pp. 163–173). Transcript.
- Mika, C. (2017). *Indigenous education and the metaphysics of presence: A worlded philosophy*. Routledge.
- Moore, J. W. (Ed.). (2016). *Anthropocene or capitalocene? Nature, history and the crisis of capitalism*. PM Press.
- Munn, L., Magee, L., & Arora, V. (2023). Truth machines: Synthesizing veracity in AI language models. *arXiv*, 2301.12066. <https://doi.org/10.48550/arXiv.2301.12066>
- Nickles, T. (1981). What is a problem that we may solve it? *Synthese*, 47(1), 85–118. <https://doi.org/10.1007/BF01064267>
- OpenAI. (2022, November 30). ChatGPT: Optimizing language models for dialogue.
- OpenAI. (2023a, January 11). Forecasting potential misuses of language models for disinformation campaigns and how to reduce risk. <https://openai.com/research/forecasting-misuse>
- OpenAI. (2023b, March 23). Lessons learned on language model safety and misuse. <https://openai.com/research/language-model-safety-and-misuse>
- Peters, M. A., Jandrić, P., Fuller, S., Means, A. J., Rider, S., Lăzăroiu, G., Hayes, S., Misiaszek, G. W., Tesar, M., McLaren, P., & Barnett, R. (2022). Public intellectuals in the age of viral modernity: An EPAT Collective Writing Project. *Educational Philosophy and Theory*, 54(6), 783–798. <https://doi.org/10.1080/00131857.2021.2010543>
- Peters, M. A., Jandrić, P., & Means, A. (2019). *Education and Technological Unemployment*. Springer.
- Pullar-Strecker, T. (2023). Artificial intelligence: World at ‘tipping point’, says Sir Peter Gluckman. <https://www.stuff.co.nz/business/131393822/artificial-intelligence-world-at-tipping-point-says-sir-peter-gluckman>
- Rahm, L., & Rahm-Skågeby, J. (2022). Imaginaries and problematisations: A heuristic lens in the age of artificial intelligence in education. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.13319>
- Rawls, J. (1971). *A theory of justice*. Harvard University Press.
- Reader, J., Jandrić, P., Peters, M. A., Barnett, R., Garbowski, M., Lipińska, V., Rider, S., Bhatt, I., Clarke, A., Hashemi, M., Bevan, A., Trozzo, E., MacKenzie, A., Aldern, J. J., Matias, C. E., Stewart, G. T., Mika, C., McLaren, P., Fawns, T., ... Baker, C. (2021). Enchantment – disenchantment-re-enchantment: Postdigital relationships between science, philosophy, and religion. *Postdigital Science and Education*, 3(3), 934–965. <https://doi.org/10.1007/s42438-020-00133-4>
- Roose, K. (2023, February 16). Bing’s A.I. Chat: “I want to be alive”. *The New York Times*.
- Russell, B. (1946). *A history of Western philosophy*. Simon & Schuster.
- Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Correction to: Can artificial intelligence help for scientific writing? *Critical Care*, 27, 99. <https://doi.org/10.1186/s13054-023-04390-0>
- Selwyn, N., Pangrazio, L., Nemorin, S., & Perrotta, C. (2020). What might the school of 2030 be like? An exercise in social science fiction. *Learning, Media and Technology*, 45(1), 90–106. <https://doi.org/10.1080/17439884.2020.1694944>
- Smith, C. (2007). Cultures of collecting. In M. Bargh (Ed.), *Resistance: An indigenous response to neoliberalism* (pp. 65–74). Huia.
- Stengers, I. (2015, November 27). L’éguer autre chose que des raisons de désespérer. *Le Monde*. https://www.lemonde.fr/cop21/article/2015/11/27/leguer-autre-chose-que-des-raisons-de-desesperer_4819368_4527432.html

- Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove. *Nature*, 613, 620–621. <https://doi.org/10.1038/d41586-023-00107-z>
- Suoranta, J., Teräs, M., Teräs, H., Jandrić, P., Ledger, S., Macgilchrist, F., & Prinsloo, P. (2022). Speculative social science fiction of digitalization in higher education: From what is to what could be. *Postdigital Science and Education*, 4(2), 224–236. <https://doi.org/10.1007/s42438-021-00260-6>
- Tesar, M. (2021). Some thoughts concerning post-qualitative methodologies. *Qualitative Inquiry*, 27(2), 223–227. <https://doi.org/10.1177/1077800420931141>
- Tesar, M., Gibbons, A., Arndt, S., & Hood, N. (2021). Postmodernism in education. In *Oxford research encyclopedia of education*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190264093.013.1269>
- Toffler, A. 1980. *The third wave*. William Morrow & Co.
- Thrupp, M., & Mika, C. (2012). The politics of teacher development for an indigenous people: Colonising assumptions within Māori education in Aotearoa, New Zealand. In C. Day (Ed.), *The Routledge international handbook of teacher and school development* (pp. 204–213). Routledge.
- Tzirides, A. O., Zapata, G., Searsmith, D., Saini, A., Cope, B., Kalantzis, M., Castro, V., Kourkoulou, T., Jones, J., Whiting, J., & Kastania, P. (2023). Generative AI: Implications and applications for education. Article In Review.
- Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/365153.365168>
- Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. W. H. Freeman and Company.
- Wood, J. H. (2005). Interventional narratology: Form and function of the narrative medical write-up. *Literature and Medicine*, 24(2), 283–296. <https://doi.org/10.1353/lm.2006.0016>