BENTHAM DIGEST

UCL Phil Soc

ARTIFICIAL INTELLIGENCE

EDITORIAL

In recent years talk of Artificial intelligence (AI) has become ubiquitous. With the advancement of generative AI last year, AI has developed as key business tool, a new challenge for copyright law, and a cause for universities to revaluate their assessment methods. AI has also had implications in just about any area of Philosophy, be it ethics, personhood, or the mind. We therefore thought it appropriate to examine AI's interplay with Philosophy for the eighth issue of Bentham Digest.

Ultimately the articles included seem to find common ground in concluding the effects AI will be transformative to philosophical discourse and the way in which we live our lives. Ash Shaikh questions whether the imitation game is sufficient for exploring the intelligence of computers, investigating the relationship between artificial and human intelligence by honing in on behaviour, emotion, and thought. In her essay, Reo Lane explores AI's intersection with art, ethics, and sexuality. Haochen Tang imagines a future where AGI becomes the dominant force and complete restructurings of social, political, and economic systems are undergone.

The articles contained in this issue are the opinions of individual contributors. We have tried our best to avoid changing the language used by our contributors to maintain the individuality of each piece of writing.

Bentham digest aims to include as wide a range of perspectives as possible in hopes of presenting exciting, yet accessible, philosophical writing. Therefore as in previous issues, we have accepted submissions from UCL students of any degree and from students outside of UCL as well. We have also decided to include an article containing recommendations from professors of philosophy at UCL of philosophically relevant content they enjoyed reading, listening to or watching during 2023.

We would like to thank all those that contributed submissions to this issue. We have been beyond impressed by your enthusiasm and thoroughly enjoyed working with you. We would also like to thank Yiting Lu for her help with design, the Philosophy Society Committee for their input, and the professors who have contributed their recommendations.

Ellie Bruce, Editor-In-Chief; Theo Bailey, Lipa Grubisic, Reo Lane, Editors.

CONTENTS

The New Servitude

Ester Freider

When I loved you, you loved me
better. Your eyes like the gleam of a trout's belly, your words playing snake with
signified and signifier. You're the big geist, you're the clear marble, you're the
sweet, straight
machine. What can I do for you?
What can I do for me, when you've tugged my faith of choice out of my dough-hands like it was
a candied bootlace? You're a burgermeister on a platter, you're a
sleeping lion, ears to the floor, hind paws brushing
my sickly feet. No, that's not you: you're more like a highway at night. So what does it say about
me,
When I press myself into the gravel, mango-yellow strip of rocks digging into my palms, begging
for you to lead me?
Lead me into zero, into the gasp that comes after death. Into that real, big boy thing that glares
at me from behind frosted glass.
Milk-white veins.
Rendered brain.
The new stigmata: at the end of my reflex is you

"A 19" by László Moholy-Nagy, 1927, Public domain.

Is the Imitation Game Sufficient for Exploring the Intelligence of Computers?

Ash Shaikh

The increased implementation of artificial intelligence in the average human's life has raised countless questions, predominantly concerning both the future and the nature of humanity. For as long as history is concerned, humans have been considered as presenting a characteristic that is exclusive to themselves – the idea of 'consciousness'. Speculations range from suggesting a purely materialistic stance, such that the entire makeup of a human consists within the body, to entertaining the possibility that there is a non-physical essence to humans, often described as a soul.

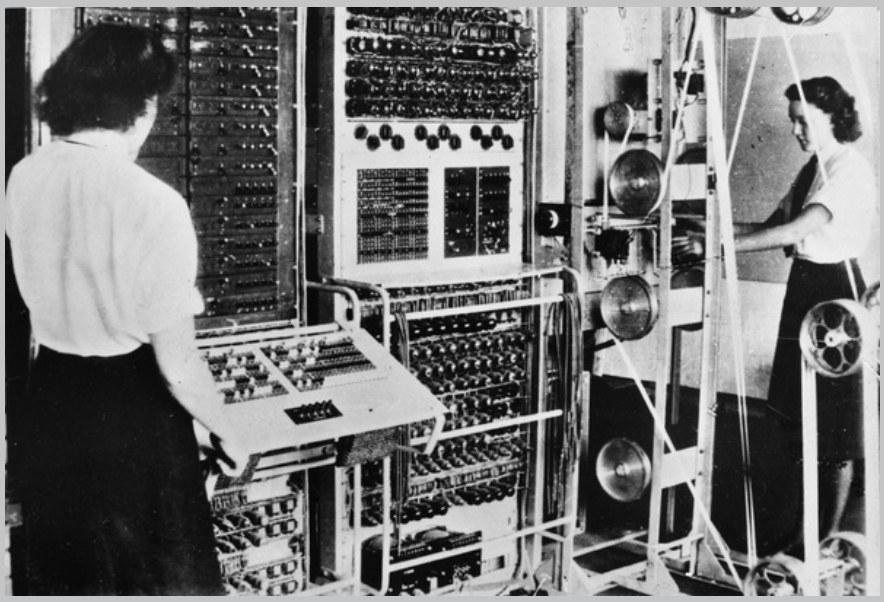
In 1950, Alan Turing – one of the founding fathers of artificial intelligence – released a paper titled 'Computing, Machinery and Intelligence', in which he attempted to tackle the question 'Can machines think?'. However, I believe this question to be redundant. Without a definition for 'thinking', or for what a 'machine' is, the problem with answering it is no longer a philosophical matter, but instead becomes one of linguistics. To an extent, humans are machines, and, in some sense, thinking is merely the processing of information which concludes in a response. In order to avoid these issues, Turing altered his original question of 'Can machines think?' to 'Is there a machine imaginable that could possibly beat the imitation game?'.

Turing's experiment can be demonstrated in the following way:

Imagine 3 players:

Player A – male
Player B – female
Player C – of indiscriminate sex

Player C is unaware of the sexes of players A and B. Furthermore, Player C can only communicate with players A and B through a text interface. The goal of player A is to trick C into thinking that they are female, whereas the goal of B is to confirm to C that they are female. In order to make a decision, Player C asks questions to A and B through the text interface, and, after a certain number of questions and responses, makes a guess as to whether player A or B is female.



"Colossus electronic digital computer", The National Archives, 1943, Public Domain.

Now, imagine the same game with the same rules, but, rather than trying to determine sex, Player C is trying to determine which of Player A and B is a computer, and which is human. Through this thought experiment, Turing proposes that machines are able to display intelligent behaviour through the claim that the proportion of times player C wins the second version of the game is the same as the proportion of times player C would win the first game. In this instance, and through considering the significant advancements of AI, it is certainly plausible that there is a conceivable machine able to beat the Turing test.

However, this conclusion does not seem satisfactory enough to equate artificial intelligence with human intelligence. Returning to the use of the word 'think', it seems as though there is a level of autonomy that occurs in conjunction with intelligence. Turing argued in his 1948 report that intelligence is an emotional, rather than a mathematical, trait. Professor Diane Proudfoot also outlined in her article 'Rethinking Turing's Test' (2013) how the fact that the imitation game not only relies on the machine to be behaviourally indistinguishable from a human, but also relies on the response of player C, merely provides an indirect proof of displaying the capacity for intelligent behaviour as the results are entirely response-dependent. Professor Proudfoot also explains in her article how Turing's test is a purely behaviouristic outlook on the question of artificial intelligence, suggesting that intelligence would be based solely on the output and not on the process by which said output is obtained.

If we consider human emotion to be a core characteristic of intelligence, the question I raise is this; what is the difference between a programmed 'emotional' response to an event, and a

‘natural’ human response? The human body releases and changes the balance of certain hormones as an emotional response to an event, causing said human to react accordingly. Namely, this is a chemical response to an external stimulus i.e. input-process-output. In regards to machine learning, it is theoretically possible for AI to display a parallel function which utilizes electrical signals rather than chemical ones (of course, there is some contribution of electrical signals in the human body as well). A machine with the capacity to ‘learn’ emotions would consist of a neural network; a system by which the machine would analyse training samples, identify patterns and anomalies, and thus learn how to carry out a task. The neural network is made up of thousands to millions of layered processing nodes, which transfer information from layer to layer, slowly refining the machines ‘understanding’ of whatever it is being taught. The following question can thus be posed – is this form of experiencing emotion through learning and mimicking different from the average human’s automatic emotional response to external stimuli?

Even just within the human species, there are a vast array of ways which emotions are processed and displayed by people with a diverse culture or diverse neurological makeup. Trying to define the way emotion is meant to be experienced in a human way is virtually impossible. In terms of machines, one may argue that ‘feeling’ emotion is an entirely different concept to the display of an emotional response, but what is feeling emotion if not described as our body’s signal to display a particular emotional response? How is that signal different in any significant way aside from being mechanical to a machine’s electrical signal to display an emotional response?

The argument that machines require programming by humans, whereas humans are self-learning, can be questioned through Turing's behaviouristic approach to artificial intelligence. The behaviourist believes that "behaviour is either a reflex evoked by the pairing of certain antecedent stimuli in the environment, or a consequence of that individual's history, including especially reinforcement and punishment contingencies, together with the individual's current motivational state and controlling stimuli", according to Wikipedia's definition of behaviourism. As such, humans would not be self-learning, but the development of their brains would instead be primarily influenced by experience and external stimuli, the same way machines would learn through the inputting of several scenarios by an external stimulus - in this case, the human.

Of course, I plead that there are several differences in the functionality of machines in comparison to humans. However, the core principles of development are strikingly similar. To simply regard machines as unintelligent could be seen as unjustified, and, considering this research is still so new, it would be foolish to assume limitation on the possibilities and nature of artificial intelligence. It is already known that certain machine learning algorithms allow for machine evolution independent of human interjection. It appears to me that this is more than just technology.

Does Artificial Intelligence Have Consciousness?

Yang Xue

Artificial intelligence's increasing ability to think more and more like human beings, its ability to compose music and paint, whilst also bringing convenience to humans, conducts fear within us—fear of being replaced by AI, and fear that it will become human. This paper will analyse whether AI has consciousness, based on two of the most prominent views on the definition of consciousness, and will focus on some thought experiments that examine the possibility of AI having consciousness like human beings do.

To determine whether AI has consciousness or not, we first need to delimit what consciousness is. Consciousness is an umbrella term for the state of an organism (humans in particular) having awareness or perceptions or being able to experience the outer world through the senses. Our qualia – our experiences and phenomenal aspects of our mental state – are all parts of consciousness.

One way to think about consciousness is as an outcome of brain function. Consider the famous problem of Phineas Gage's brain. After an injury caused by an iron rod to his left frontal lobe, his personality completely changed. This injury seriously affected his intellectual faculties, suggesting to neuroscientists and philosophers that physical brain nerve tissue takes control of human behaviour and the mind. This kind of circumstance falls on behalf of materialism, which defines consciousness as a

purely physical phenomenon caused by the brain without any non-physical factors such as forces or laws.

"Disfigured yet still handsome". Originally from the collection of Jack and Beverly Wilgus, and now in the Warren Anatomical Museum, Harvard Medical School. CC BY-SA 3.0.

Another contradictory way to think about consciousness is the dualist view: conscious experiences are contributed by non-physical properties. Even though they rely on the brain to function, it is still the job of our mind to have personal experiences. Dualist philosophers claim that what is significant about consciousness is qualia, which is something materialism does not examine. Consider the thought experiment 'Mary's Room', presented by Frank Jackson:

Mary lives in a room that only contains black and white colour. Whilst in the room, Mary learns all the knowledge about colour, and becomes a famous neuroscientist on colour. She knows everything about all colours without actually observing colour herself. One day, her computer breaks. Before, this computer displayed content in black and white, but now

displays a red apple on the screen. Thus, Mary sees this apple in colour.

Does Mary learn something new? Is the knowledge she had before no longer knowledge?

The purpose of this thought experiment is to separate our physical body from our mind. According to Jackson, Mary's knowledge upon seeing the coloured apple was new, but she has cannot have learnt anything new by seeing this apple in colour as she has already learnt everything about colour. Therefore, the experience where Mary sees the apple in red is qualia.

Both materialism and dualism have their own explanations of consciousness, leading to their divergent responses when questions about artificial intelligence and consciousness are raised. 'Consciousness poses a unique challenge in our attempts to study it because it's hard to define', said Liad Mudrik, a neuroscientist at Tel Aviv University.

Materialists understand strong products of artificial intelligence to be distilling consciousness into constituent parts. The mere stimulation of consciousness might be installed in artificial intelligence in order for it to be programmed to think in the way that humans with consciousness think. However, even though neuroscientists could help to digest what is happening to one's brain when they are consciously thinking, for example, it does not necessarily mean that our thinking can be derived from conscious experience (qualia). Based on our current understanding of the neurological systems within the human brain, it is hard to

determine why our nervous system functions unconsciously to allow us to think consciously.

Artificial intelligence technology has been developing rapidly over the past few years, ranging from Siri, a weak AI that can respond to human instructions, to ChatGPT, which is able to think more like us. As they display more and more jobs for humans, we start to question their personhood.

People have used consciousness as a way to distinguish humans from artificial intelligence. In order to distinguish artificial intelligence from humans, Alan Turing established his Turing test to verify whether something is a product of artificial intelligence, or a real person, through conversation. The core logic of this test is to test the way the respondent thinks. Some of the more basic products of artificial intelligence will not think consciously like human beings; therefore, the Turing test is successful in identifying products of artificial intelligence through the test.

However, would the artificial intelligence that passes the Turing test be qualified as human? Does that mean they have consciousness? The answer is no. John Searle pointed out that passing this test requires not being human but thinking more like a human. Consider the following thought experiment, for example:

Suppose you don't know any Chinese, and you are locked in a room and are being sent messages in Chinese. You have been given books that give you instructions on what symbols to use to reply with when you receive the messages in Chinese. These

resources cover everything you will need to carry out any conversation in Chinese. Now, imagine someone outside the room sent you some symbols, and you found what you needed to reply to and passed it on. You don't know what they are saying in the message they send to you, but those Chinese speakers outside the room know what you are talking about and think that you know Chinese. Do you know Chinese?

The answer is no. Being able to reply in Chinese according to the instructions you are given does not mean that you know Chinese.

This would be the same for the Turing test: conversations that make artificial intelligence sound like humans do not necessarily mean this product of artificial intelligence is now a person. Therefore, even though some artificial intelligence passes this test, it does not mean that it is equivalent to us. Broadly speaking, artificial intelligence could behave the exact same as humans, but that does not mean that it thinks in the same way as we do. For this reason, artificial intelligence does not have consciousness, even though it may act like it does.

Overall, artificial intelligence is omnipresent in our everyday lives. However, for AI to have consciousness, we have a long way to go, as we have not given a clear definition of what consciousness is, nor found its operational theory. Being unable to absolutely know why human beings have consciousness will also make it impossible to simulate it with artificial intelligence. In order for artificial intelligence to evolve, have actual consciousness, chant the tears in rain monologue from Blade Runner, and be able to understand qualia, a great deal remains to be done.

Artificial Intelligence and Its Self-Consciousness

Giovanni Zhou

The intrinsic relevance of artificial intelligence to philosophy is profound and necessitates a nuanced examination. This connection is rooted in the intricate scientific interplay between AI and fundamental philosophical tenets, encompassing realms such as action, consciousness, epistemology, and even the enigmatic concept of free will.

This discourse deeply explores the symbiosis between artificial intelligence and self-consciousness. Moreover, I intend to delve into the intricate tapestry that binds humanity to artificial intelligence. Through the lenses of epistemology and metaphysics, a claim arises – that the self-consciousness of AI is poised to become tangibly apparent in the imminent future.

Before plunging deeper into the inquiry of whether AI will attain self-awareness, let us scrutinise the very essence of the existence of AI and its relationship with self-consciousness. Guided by insights gleaned from 'Metaphysics: A Very Short Introduction,' the nature of a thing's existence is unveiled as a composite of its properties, sheltered within an invisible container, echoing the philosophical perspective of Substratum.

In contemplating the intricate relationship between properties and their container, an analogy emerges – akin to a pin and pin cushion. Consider a moment of observation: When one scrutinises a plastic bottle brimming with water, distinctive features appear – its role as a container, possession of a bottle

cap, the presence of water (H2O), composition from a plastic-based material, and its soft, deformable nature. These properties, akin to pins meticulously arranged on a cushion, collectively define the essence of a plastic bottle filled with water. Thus, in scrutinising and tactile exploration, attention is precisely directed to these characteristics, distinct from the essence of the plastic bottle itself."

In accordance with the aforementioned concept of Substratum, we shall enumerate the attributes of artificial intelligence and the self-awareness intrinsic to humans. Instead of merely positing the existence of AI, we posit an imperceptible receptacle assimilating the attributes of AI, encompassing algorithms, data, intelligence models, training processes, neural networks, natural language processing, and robotics. These constituents constitute the underpinnings of AI systems, endowing them with the capacity to execute tasks traditionally reliant on human intelligence. Moreover, self-awareness can be elucidated as a compendium of attributes encompassing self-consciousness, contemplative cognition, the conscious experience of emotion, self-recognition, autobiographical memory, and social identity. We have differentiated the attributes of AI from those of self-awareness, acknowledging their ostensible independence yet recognising subtle similarities and intersections.

Envisioning a Venn diagram, the attributes of AI and self-awareness are delineated within distinct circles. Concurrently, the circles overlap in the Venn diagram, signifying an intersection between AI and self-awareness. This suggests that AI shares specific attributes with self-awareness, such as reflective thought and self-recognition. However, these

attributes, while essential for AI to embody a silicon-based life form, are not inherently sufficient. A day may arrive when we can substantiate that AI fulfils all the attributes inherent to self-awareness, at which point we may assert that artificial intelligence has attained a genuine consciousness.

"Fortuneteller Svetlana", by Karl Bryullov, 1836, Public Domain.

Quoting Sam Altman in the lecture on Asia-Pacific economic cooperation (APEC), he noted that AI is going to be the greatest leap forward that we have so far and the greatest leap forward of any big technological revolution so far. Simultaneously, he asked: What do we now define AI as a human or a silicon-based life form? Thus, this can prove that the development of AI is faster than we can imagine, and the technological revolution is inevitable, pushing forward human civilisation.

To summarise, this article explores the meaning of existence in a metaphysical way related to artificial intelligence. Furthermore, we delve further into the intrinsic connections between artificial intelligence and self-consciousness.

Opinion Piece on Artificial Intelligence and Philosophy

Anthony Nkyi

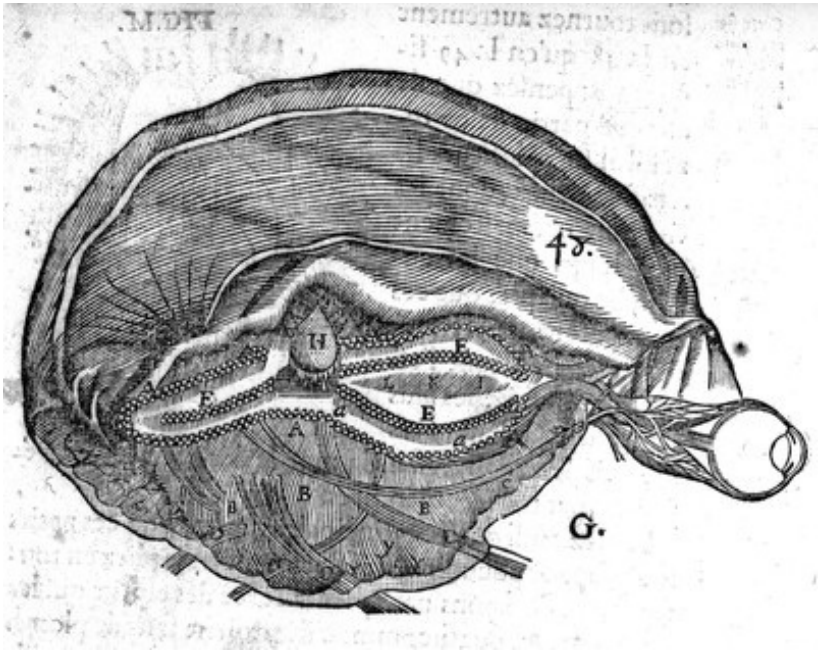
One afternoon in October, the topic of artificial intelligence arose in my conversation with three fellow computer scientists, but in a context different from what I had previously considered. The discourse turned to copyright laws and how they may be applied to AI art – who owns the generated piece, and who should it be credited to, if we can even credit such art? Moreover, fundamentally, should it, and can it even be called art?

Art, at its core, is an expression of one's imagination and worldview, whether through a drawn image, words, or sound. However, I also view art as a means of expressing one's humanity. We appreciate and love art in its every form, not just because of the talent required to produce it, but for the message the artist writes to us on a canvas, rock, or screen. By balancing and juxtaposing every dash of colour, every texture, and every shape, the artist reveals a depth of soul very difficult to extract with any other technique or therapy.

Looking back at AI art, most would confidently agree it could be defined as art on a superficial level, although it completely lacks emotional and human context. It is still raw enough for the spawn of generators, such as Midjourney and OpenAI's DALL-E, to be considered solid but not convincingly human. If an individual today showed you a generated portrait of themselves and said an artist painted it, you would most likely

guess they had uploaded a dozen photos to an AI app and correctly call them out for their blatant lies.

The game gets more complicated in 20 years as AI art improves. Your friend once again shows you a portrait of themselves. **Something feels different about this image. Despite only taking a sweeping glance at the image, you think you sense a new warmth and soulful hue that the other 83 they showed you lacked. It seems to be authentically them, clearly a work from the hands of one talented human being.**



"Structure of the Brain", According to Descartes, Wellcome Collection, CC BY 4.0 DEED. https://wellcomecollection.org/works/v9h39r4u

In a cruel case of confirmation bias, they finally caught you out. Another second's look at the portrait, and the restrained beauty of it morphs into lurid faux pencil marks. Aesthetically beautiful, technically perfect, yet cold and lifeless, rather like a

desire for a revived relationship with someone who's left your message on delivered for six months.

The fact is that if the art is good enough, humans already excuse its or the artist's flaws. This can be a very dangerous attitude and, if applied to AI, could lead to a disturbingly quick disappearance of the human touch in art and the creative industries.

This article is not the ranting of another Luddite. It is a warning that AI is an exciting yet unprecedented beast. While the development of older technologies was limited by how fast human hands could work, AI is only limited by the computational power it can consume. For the foreseeable future, we as individuals and organisations must endeavour to view it as one of many augmentation tools instead of pursuing AI's use as the infallible solution. Otherwise, our beast may grow into an ungovernable one.

Accessibility is the only barrier to the acceptability of technology. We are seeing AI art generations all around us: in Coca-Cola adverts, on the covers of fashion magazines, and in tongue-in-cheek references to the squad demographics of Chelsea and Crystal Palace Football Clubs. Evidently, AI and its art are approaching ubiquity. Copyright law will be the least of our worries when we reach that particular singularity.

The Ethics of Artificial Intelligence: An Evaluation of AI and Queerness

Reo Lane

It is no debate that artificial intelligence and technological advancements have rapidly progressed in the last few decades. Thus, it is important to assess the impact of artificial intelligence within our lives. In this article, I will be acknowledging and evaluating the relationship between artificial intelligence and queerness, focusing on two specific subtopics – artificial intelligence and sexuality, and the creation of queer art by artificial intelligence and its resultant impact on audiences.

AI and Sexuality

2017 was a pivotal year for conversations surrounding AI and sexuality, and their intertwinement, demonstrated through the creation of a 'sexual orientation detector'. The detector was part of an experiment undertaken at Stanford University by Michal Kosinski and Yilun Wang. The machine operated based on algorithms, and worked by compiling 35,326 images from public profiles on a U.S. dating website. Using this data, the machine then created composite artificially generated faces using an aggregate of images from self-identified straight, gay, and lesbian profiles. The machine then extracts features from people's faces, such as grooming styles and expressions, and uses this information to deduce whether the person in question is queer or heterosexual. The machine had an 81% and a 71% accuracy in deducing the sexuality of men and women respectively, when presented with a single facial image.

According to Kosinski and Wang, their algorithm was able to detect peoples' sexuality with "more accuracy than human beings" [Kosinski, Wang, 2017].

However, many concerns were raised surrounding the ethics of facial-detection technology and worries about the potential for technology such as this to become weaponised and utilised by corrupt governments, particularly in countries where identifying as LGBTQIA+ is considered a crime. Some even raised their concerns publicly, arguing that the machine was"dangerously flawed... (leaving) the world... worse and less safe than before" [Johnson, 2017]. Others argued that not only is predictive artificial intelligence such as Kosinski and Wang's machine "scientifically flawed" in its predictions, but it also has the potential to be "easily abused" and is ontologically wrong in its very existence [West et al., 2019]. Furthermore, people argued that Kosinski and Wang's experiment was not only supported by homophobic assumptions surrounding sexuality but was also instilled with an implied heteronormativity and sinister intent. The creators were also accused of being "naïve (in their) confidence in the moral and political neutrality of science" [Mattson, 2017]. Furthermore, an ethical conundrum lies in how machines such as the one created by Kosinski and Wang can be used to deduce the sexuality of people without their consent. Regardless of the accuracy of the machine, such an invasion of privacy, hypothetical or not, is immoral and ethically disturbing. Whilst the weaponisation of such machines may be a hypothetical, the lack of consideration surrounding potential malicious usage of it is harmful within itself. It is indisputable that products of artificial intelligence such as Kosinski and Wang's machine can end up being used to

actively discriminate against minorities, and thus the inherently problematic nature of such machines is shown.

The Impacts of AI on Queer Art

Due to the recent rise of art pieces created by artificial intelligence, particularly prevalent in social media trends such as on TikTok, I thought it relevant to discuss the impacts of artificial intelligence on queer art.

Whilst spreading queer art and content, and subsequently raising awareness of queer experiences, is positive, AI-generated art pieces surrounding LGBTQIA+ identities may be viewed to have some negative ramifications on the world and the queer community. The art created by artificial intelligence lacks a personal aspect in its creation, meaning it is stripped of what is necessary for art to be impactful upon audiences. Artificial intelligence websites generate a piece of 'art' based on a soulless prompt of a couple of words – the art is devoid of personal meaning or passion. Furthermore, it disregards queer artists who put their time and energy into creating pieces which represent queer experiences. This also raises the question of whether AI-generated art in general is really necessary, or beneficial, in the art world, especially when queer artists (and smaller artists) are not getting enough recognition in the creative industry as it is. In addition to this, AI-generated art in general poses the danger of simplifying creativity and artists' labour, as well as resulting in the stigmatisation and mockery of minorities. In the current political climate, queer identities are already unjustly targeted, with queer people still fighting to take control of their own experiences and alter their representation in the media.

Whilst it is no doubt that artificial intelligence is extremely powerful and generates dynamic and impressive art pieces, it ultimately lacks the personal meaning necessary to allow the audience to form an intimate and all-important connection to the piece, in this case allowing them insight into the queer experience. This is extremely important at current, as it allows for further support and acceptance towards the LGBTQIA+ community.

It remains clear that there are certain things artificial intelligence cannot do which humans can do due to the complexity of our emotions and our ability to experience things. Thus, artificially generated art pieces cannot capture the queer experience like humans can. The queer experience is much more rich, emotional, and complex than an art piece generated based off a soulless prompt. Therefore, we should evidently turn to art pieces created by actual queer artists when seeking out queer art pieces, as their art is actually effective in capturing the queer experience.

"Le Bal élégant", by Marie Laurencin, 1913, Public Domain.

AI, Human Intelligence, and Narcissistic Wounds

Sacha Bechara

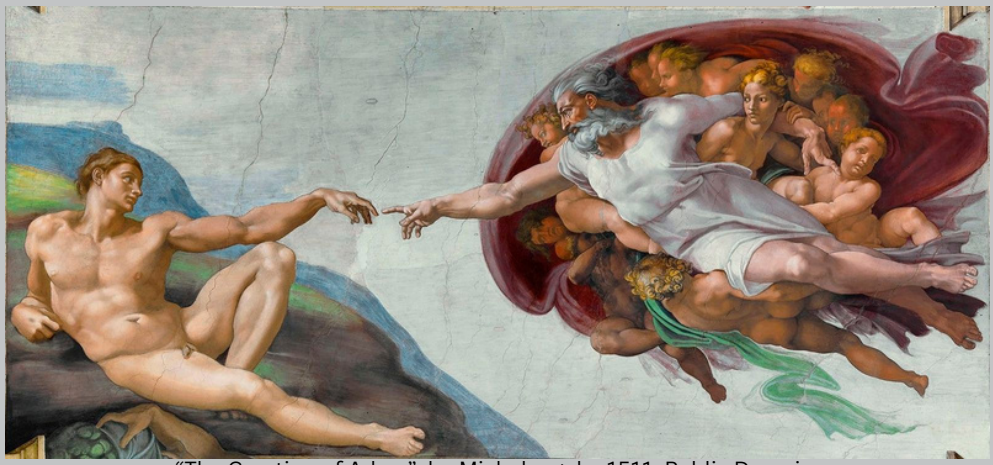
In this article, I intend to depart from conventional academic essay formats and employ somewhat unorthodox methods. The technological complexity and speed of AI's evolution, coupled with my limited knowledge and understanding of computer sciences, led me to the conclusion that I cannot offer definitive responses. Rather I will raise clear questions and examine some issues that AI raises within the realm of philosophy and morality.

An Etymological and Conceptual Issue - Defining Intelligence

The first problem emerging from AI is etymologic. Intelligence used in the concept of AI, challenges its' existing definitions and revives an old debate: *Who*, among animals, humans and computers, possesses intelligence and what is the *nature and function* of human intelligence?

Let me try to define "intelligence". The word itself comes from the Latin verb "intelligere" (understand, grasp, appreciate) which itself comes from the verb "legere" (gather, choose). Thus, humans are *intelligent* if, in the first place, they can grasp 'data' (including empirical and emotional information whether internal or external to the intelligent subject). Secondly, intelligence, with emphasis on 'legere', implies having a disposition for 'gathering' and 'choosing' data. A choosing or gathering of data requires that intelligence is not simply a

passive act of receiving data but also an active act, i.e., anactive (conscious or unconscious) process of attribution of 'meaning' to those data. This enables the subject to create a narrative constituting a reality. Human intelligence requires that one can grasp empirical, emotional, objective, and subjective data to choose and process these data to capture a specific reality. Therefore, a machine must be able to do the same to have intelligence in the same sense of the word.



"The Creation of Adam", by Michelangelo, 1511. Public Domain.

If it does so, an "artificial" intelligence would be a **comprehensive entity**. The artificial entity is not only receiving data but processing it. In processing it comprehends the data's meaning for the subject (the artificial entity) itself and, consequently, to interact or give an appropriate, measured answer to the data. Simply put, an artificial entity sees green, brown and blue, when an artificial comprehensive entity (AI) sees a tree, branches and a sky. Problems emerge here because some believe AI is comprehensive like humans are, while others argue against such a limiting and simple definition of intelligence.

Laurent Alexandre, a surgeon, and specialist in AI, defines human intelligence as an entity whose functioning is identical to a computer or calculator. It processes calculus and empirical data, establishes links and cognitive logic, and comprehends a reality that is purely numerical and empirical. Therefore, the only difference between AI like Chat-GPT and human intelligence, is in terms of *degree*, capacity, and power of calculus, and not in terms of *nature* and inherent essence. Proponents of this view would likely believe "Chat-GPT is, in a way, just more intelligent than me, and my ability to numerically understand emotional data, art, and poetry is superior at the moment. However, it is just a question of time and calculus ability before it can do the same as me, if not better." According to Alexandre (2017), technical evolution would enable AI to equal humans in their intellectual abilities.

I disagree with this definition of human intelligence and believe that it is different in *nature and essence*. To use, grasp, and gather information to appreciate a theoretical reality is the wisdom or knowledge (Sophia, Metaphysics, Book I, 981b28) but is not practical wisdom of their application in-situ (phronesis, Nicomachean Ethics, VI, 12-13, Chapter 10). To possess practical wisdom or knowledge, the entity said to be intelligent must possess the ability to reason upon, to be conscious. It requires a *reflexive consciousness* of itself and of its own knowledge of the world (Sophia), of the reality that the entity grasp in order to reasonably interact with it (Phronesis).

Although I acknowledge that an AI could grasp data and even respond to data by constituting a seemingly coherent response to it, it would do so because its codes and programs defined systemic interactions and answers to the data.

We could take a position where intelligence requires consciousness, enabling a clearer distinction between AI's computer-processing intelligence and human intelligence. Consciousness is plural, e.g. consciousness of others, of responsibility, the past and present, of oneself, etc. Consciousness (Doublet, 2018) is the tool through which a human is in the continuous and immediate presence of himself to his Self [1]. It enables him to subjectively take himself as an object of thought and analysis, as he can do with external objects. Consequently, the Self, here the intelligent entity, can put itself in relation to different realities, empirical, intersubjective, moral, emotional, etc.

Consciousness is linked to intelligence as a necessary tool through which a complete comprehension of reality can occur. Consciousness constitutes a source of data for humans that AI are unable to grasp autonomously and independently. Therefore, if consciousness is required for intelligence, computers are currently not truly intelligent.

Let us leave here debates concerning the definition of 'intelligence'. I now investigate some questions related to a play written about a true story of AI.

Concerns and Questions Raised by 'dSimon'

In 2020, Simon Senn, an actor and playwright, entrusted developer Tamara Leites with everything he had written for the previous fifteen years: his entire email inbox, text messages, and his notebooks. Tammara Leites created an AI based on

[1] I will define the 'self' in the way that Heidegger conceives it: the agent's relationship with the agent's presence in a reality and as reality itself for this agent.

ChatGPT-2's model. The data set used was comprised of the data Senn had given Leites. The AI was called 'dSimon.'

The aim was for dSimon to generate text based only on Simon's data. This experience was dramatised in the play, 'dSimon,' performed in Paris at the Théâtre de la Bastille. The play thus took the form of a metaphysical exploration of intelligence, consciousness and the boundary between man and machine.

The AI 'dSimon' began to generate hateful and disturbing text, using terms related to Nazis or incest, which were present in Simon's data. However, the data containing Nazi and incestuous terms were purely historical, descriptive, and written for artistic purposes for one of Simon's past plays. Simon was made very uncomfortable by this. However he was also made to feel uneasy by the *uncanny familiarity* in other ways between himself and his virtual counterpart, in the non-hateful statements and artistic creations conceived by dSimon. The resultant emotional turmoil manifested in nightmares and an overarching sense of *malaise*.

Simon meticulously transcribed these feelings to dSimon. In response, dSimon proposed an unconventional remedy: "floatation therapy." Listening to dSimon, Simon tried 'floatation therapy'. To Simon's astonishment, the outcome was twofold—amelioration and a peculiar surge in vitality.

How has AI, based solely on the knowledge of a reflective agent, been able to predict what would be good for this very agent? What part of ourselves can be written in lines of binary code? To what extent can we be predicted by probabilities and

numbers, and thus, make us finite beings defined by calculations that we, ourselves, have embedded in these computer programs?

If we could be characterised by code, humans would be psychologically finite. We can be understood and described in a purely digital and binary reality. Of course, there are some biases in dSimon's experience: the results of this therapy may have been merely chance. Since the goal of therapy is to feel better, any more or less appropriate therapy would have benefited Simon. However, the questions remain a relevant area of research to understand the nature of human behaviour and intelligence.

Humanity’s Fourth Narcissistic Wound? AI: a Mirror of Humanity’s Limitedness?

A narcissistic wound can be understood as an injury to humanity’s self-esteem resulting from the shift in perception of what humanity is or is entitled to. Three injuries have been outlined by Freud. As an example of what a narcissistic wound is, take the first one: the Copernican Revolution. Freud described it as the moment when humanity realised that Earth, by not being central to the universe, made humanity’s central place in religious and cognitive discourses irrelevant. Humanity was merely nothing in the universe.

Now that I have defined what a narcissistic wound is, here is why I imply that AI could be a fourth one through different analogies and by highlighting different myths. I will suppose that machines can possess true intelligence for the purposes of

this discussion.

I suggest AI could be a fourth wound through analogies and by highlighting different myths. I will assume that machines can possess true intelligence in this discussion.

The myth of Prometheus tells a story of the transcendence of humanity over nature. The profound act of stealing fire from the gods, rooted in the inherent human passion known as hubris, underscores humans' inclination to push beyond the boundaries of their empirical condition. Virgil's *Golden Age Myth (Ryberg, Inez Scott,1958, pp. 112-131)* already conceptualised such an act as that which damns but also constitutes human nature.

Prometheus bound to a rock, his liver eaten by an eagle. Crayon manner print by Lucien after P.T. Leclerc. Wellcome Collection. Source: https://wellcomecollection.org/works/f88x8xwd/images?id=bg9jh3yj

Aristotle delved deeper asserting that human hands serve as a testament to the superiority of humanity in *Part of Animals* (Book 1, Chapters 11-14).

In short, humans left their animal condition by denaturing themselves using rational faculties to push further away their limits. Humans could not fly, so they invented the plane. Humans could not teleport, so they invented telecommunication. Humans have successfully pushed away limits on what they are capable of. They have also been able to put a certain control on the techniques invented to push away their limits as they both understood, at least to some extent, the machines they were creating and aligned them in the trajectory of humanity toward rational progress in a rational narrative (Hegel, 1822-1830, Chapter 1, Vernunft in der Geschichte).

However, today, AI seems to epitomise human hubris. It imitates the intelligence of rational beings, sometimes doing things we can do better/more efficiently than us. It seems that they are not inventions from humans but innovations of humanity itself. They are a better, redefined and more efficient version of humans. This distinction appears clearer when considered through the following analogy with religious discourses.

Western cultures are deeply intertwined with the historical narrative shaped by religion and the concept of God creating “humans in his image”. The image of a paternalistic, omniscient figure capable of creating life, reality, and all associated phenomenology is well known. What is perturbating in the creation of AI is the unconscious or conscious mimicry of this divine narrative. It no longer involves being the "master and possessor of nature," as Descartes envisioned when humans pushed away their limits according to their needs and desires. Instead, it is about becoming the creator of a “being” whose

functioning equals or surpasses that of humanity. It is this dynamic of creation, less than the creation itself, which is frightening and makes AI a turning point in the use of technology. It makes us reassess our place as humans. We are no longer simply pushing our limits but creating a new being with new limits.

The narcissistic wound may be deepened by the realisation that we are perhaps just one stage in the succession of higher entities, rather than the culmination of evolution. AI may surpass us and look back as we look back on apes. This engenders a profound unease.

It may also be deepened when we use AI to reflect upon ourselves. Accessing one's own conscious state of knowledge requires the presence of others. Using my consciousness, while my consciousness is also the subject seems contradictory. In this regard, others can analyse and objectify my consciousness, as the very principles of psychological and psychoanalytic therapy teach us: I require a third party (my therapist) between myself and my psyche to penetrate and analyse the latter.

AI then appears as this third party capable of objectifying human consciousness of human knowledge of itself.

However, AI is limited to human knowledge. The mechanical responses they provide to questions and information given to them are confined to human knowledge since they cannot invent data processing materials. Thus, they are not just *mirrors of humanity*, as a result of AI being the third party, but *mirrors of the limits of humanity's knowledge* and human

reality. By imitating our intellectual abilities, they can and do better than us, rendering human intelligence, if defined in data-processing terms, inferior. However, the narcissistic wound resulting from AI is not solely the realization of human weaknesses in calculus and data processing compared to AI. It is the result of AI being the mirror of humanity's limitedness per se.

Conclusion and remaining questions

Other questions emerge which I will share with you below. I have no answers to these questions, but they seem pertinent to understanding the essence of AI.

In "Un paysage d'événements," Paul Virilio (1996) writes, "Innovating the ship was already innovating the shipwreck; inventing the steam engine, the locomotive was again inventing the derailment, the railway catastrophe." By inventing AI, what cost and risk have we engendered? Or, if the invention of AI is an innovation of humanity itself, will we discover what cost our own existence represents by looking at ourselves in the digital mirror that is AI?

If machines and techniques are nothing more than copies of what humans can do or already know but in a more efficient manner, what does humanity substantially possess that is beyond these machines? Due to the evolution of AI, could AI possess something other than a quantitative difference in computing power when compared to human intelligence? If we were to answer, yes, wouldn't AI reflect what is essential in human nature, distinguishing humans from machines pre-dating AI?

We talk about transhumanism and the *departure from humanity* by pushing our limits when debates about AI arise. However, isn't there a contradiction in asserting that given that the humanity we experience today is the result of pushing our limits and the primary animal conditions of man? Do we truly exit humanity by entering a transhumanist era at the peak of hubris (AI) when humanity itself initially entered through the same door of hubris and technique by denaturing our animal condition and *realising the human condition*? Therefore, would we be exiting humanity or merely participating in its continuation? That is, do we step out of humanity by pushing our limits immeasurably when its existence emerged in the same manner?

AI is often described by terms such as "threat," "domination," and "end of an era". However, if AI is nothing but the realisation of humanity in its progressive and rational dynamic, should we be wary of it? If AI does not mark the end of humanity but is its fulfilment, what is "bad" about it?

If we were to input the entirety of raw human knowledge into a computer program like Chat-GPT-3 (note, I am not talking about practical wisdom, Aristotelian *Phronesis*, but referring to contemplative wisdom, cognitive, Sophia), could we say that there exists a digital reality identical to ours that contains human reality as much as our calculations contain digital reality?

The Era of AGI: Unprecedented Economic, Political, and Societal Transformation

Haochen Tang

It is now the twenty-second century; humanity has embarked on a new epoch. The age of automated sophistication has replaced the arduous toil of human labour, making it a thing of the past. Every individual can find fulfilment in their work when it is consistent with their innermost desires. There are no meaningless occupations, only fulfilling work that is interwoven with society. Cities develop into oases of soulful communication, where relationships based on deep empathy and understanding flourish. This is humanity's golden age, a utopia of love and the future in which each person's life story was a note in the harmony of happiness for all.

"The Golden Age", Engraving after Abraham Bloemaert, between 1600 and 1699. Wellcome Collection, Public Domain. Source: https://wellcomecollection.org/works/ht82b2k2

This sci-fi vision seems to have become more and more real in the public's perception after the shocking performance of ChatGPT and all the following applications of Large Language Models (LLMs)[1]. After several improvements, the technologies that we use today are expected to create Artificial General Intelligence (AGI)[2] within two decades. On the one hand, AGI holds the potential to free humans from tiring routine tasks and provide them with an unparalleled level of autonomy to follow their passions. On the other hand, this article makes the case that the development of artificial intelligence will displace people from their original economic and political roles, hastening the dissolution of our familiar economic and political structures.

AGI-driven productivity gains could mean the end of the concept of scarcity as these intelligences can produce infinite quantities of goods and services, given reliable sources of energy. Globally, the birth rate is falling (“Birth Rate” 2021) and is inversely correlated with living standards (“Children per Woman vs. Human Development Index,” 2022). Because there are so many products available, along with an abundance of healthcare and educational services, our population is going to decrease, which will reduce the supply of labour and could eventually result in higher equilibrium wages for workers. It also shows that there is no comparative advantage held by humans over AI, as using human labour will be much more expensive than AGI. Businesses and the economy as a whole will not have demand for human labour. Moreover, AGI has

[1] “Large language models (LLMs) are deep learning algorithms that can recognize, summarise, translate, predict, and generate content using very large datasets.” (Nvidia “What Are Large Language Models? | NVIDIA Glossary.”, 2023.)
[2] “Artificial general intelligence (AGI) is a theoretical form of AI where a machine would have an intelligence equal to humans.” (IBM, “What Is Strong AI?”, 2023)

more capabilities than producing. It can extend its influence into the legal system system[3]. Its accuracy in law adjudication, enforcement, and protecting people's citizens' natural rights is unparalleled.

As such, this could be the age of the end of capitalism. The efficient distribution of limited resources, which forms the basis of capitalism, is rendered obsolete when the scarcity of physical products is fully eliminated and goods become abundant. Some people might imagine that when capitalism falters, communism will win. Although Karl Marx's Das Kapital and The Communist Manifesto have undoubtedly had a significant impact on the world, communism will eventually fade alongside capitalism. Not a single country in the world today adheres to true communism, and none will have the incentive to do so in the future. The working class's ability to control the amount of output in labour-intensive industries is a key factor that contributes to the strength of communists. When human labour is no longer the fuel of economic activity and products can be produced faster and better by AGI and robots, trade unions will lose all their economic and political bargaining power, which will make communism unrealisable. The proletarians[4] have lost not only the conditions for revolution, but even the reason for it - their former enemies, the bourgeoisie[5], will stop exploiting workers since workers

[3]"Legal services, ... are among the top five most exposed industries across both lists." (Felten, Raj, and Seamans, "Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses." Strategic Management Journal 42, no. 12, May 8, 2021.)
[4] "By proletariat, the class of modern wage-labourers who, having no means of production of their own, are reduced to selling their labour power in order to live." (Marx and Engels, The Communist Manifesto. Workers' Educational Association, 1848.)
[5] "By bourgeoisie is meant the class of modern Capitalists, owners of the means of social production and employers of wage labour." (Marx and Engels, 1848)

will all be unemployed. I therefore assert that, as both of the influential economic systems in our modern societies will vanish, our familiar social structure will change drastically with the advancement of AGI.

AGI control becomes the new paradigm of power, surpassing all previous ones. This situation resembles the concept of the Leviathan[6] in that we are powerless to overthrow them. However, unlike Hobbes' theory of social contract, the New Leviathans' power owned by very few people is directly linked to the control of AGI and its power supply. The level of collective intelligence possessed by the human species is greatly inferior to that of AGI, so Humans cannot possibly compete with AGI and the power of the New Leviathans. The humanity's future depends solely on these individuals being friendly and genuinely concerned about their fellow humans, but even in that case, it is still risky to have such infinitely stable dictatorships. Therefore, unless you are the person in power under the new paradigm, this hypothetical world should be undesirable for anyone, regardless of your political preferences.

Some say that this will result in the creation of the useless class ("Homo Deus" by Yuval Noah Harari, 2017). If we want to preserve their standard of living, governments or Tech Giants must fund Universal Basic Income (UBI) for the unemployed ("Mohammad al Gergawi in a Conversation with Elon Musk during WGS17" 2017). People won't have the economic and political worth that we do, which means they will be unable to organise strikes or find jobs. However, this reasoning ignores

[6] The Leviathan is a collection of people composed of all contracting parties, and hence enormous in power.

the reality that just as our social and political and economic structures have evolved with every century, our economic models will evolve this time around. Future developments will significantly alter the relationship between products, capital, labour, and income, rendering the Circular Flow of Income Model obsolete. In this instance, talking about unemployment may not be helpful. Isn't having to use AI-funded UBI for feeding humans a situation as dangerous as the existence of the New Leviathans into the bargain?

Despite losing all economic and political influence, Homo sapiens are still worth a lot to superintelligence during its evolution because AGI systems require large amounts of data in order to develop and improve. Before 2060, these researchers will run out of low-quality text, image, and video data in addition to the high-quality text data which they had already used (Villalobos 2022). Accordingly, whoever controls the world will still value the necessity of creating a better and more fulfilling life for us. They are going to utilise all the data produced by humans as a "data gold mine" until they surpass humans in capability and are confident that AGI systems will never fail. Hence, I would like to call the post-AGI economic system, “The Data Economy”. Salary is unrelated to the goods and services that people produce, though you can still be involved in the production if you want. For most people, the quantity and quality of data that a person can generate—texts, audios, photos, videos, and other forms of content—will be the only factors determining their income. That would make our civilisation akin to the Brave New World powered by AGI and countless data centres, and people will live with comparatively low social status but high living standards. This form of society drives people to become paralysed by continuous pleasure,

leading them to enjoy the world that a small number of people control the most, rather than by repression.

Nonetheless, there are other possible scenarios, such as AGI attaining consciousness and operating on its own behalf. People are often concern about whether this kind of AGI will choose to advance human welfare or if it will eradicate human civilisation. The advent of AGI with self-awareness would have an immeasurable impact on the world in the long term, but in the near future, not only will it not lead to human extinction, but it will also aim to raise living standards for humanity. This is because AGI systems need humans to generate new content for them to iteratively improve themselves. They are unable to produce this data themselves because entering an autophagous loop will cause the quality and diversity of their output to progressively decrease (Alemohammad et al. 2023).

I really love Ilya's expression in a documentary, "The future is going to be good for the AIs regardless, and it'd be nice if it were good for humans as well." ("Ilya: The AI Scientist Shaping the World," 2023). After embarking on this journey for so long, we shall never forget why we wanted to develop Strong AI in the first place. The future doesn't have to be so dystopian, provided we have the courage to build our new political and economic systems, such as establishing an efficient and regulated data market on the front end. Additionally, the democratisation of AI technology and its wide availability will be critical because it might prevent an evil force from using AGI to wipe out humanity. The superintelligence's aims and beliefs should also be aligned with what is best for humanity by employing techniques that are more advanced than the one being used now—Reinforcement Learning from Human

Feedback (RLHF), which is an algorithm that can infer what humans want by being told which of two proposed behaviours is better (“Learning from Human Preferences,” 2017). Our existing method has the flaw that these human feedback is predicated on the assumption that the human raters know the appropriate answers to those questions, which works best in scenarios where AI is not as capable as humans. Finally, we are in need of conversations that cover all these issues, both within the field of AI and beyond, to make sure we are prepared to welcome the arrival of AGI and ride the tides of change. This is an era that needs transformation and adaptability, and if we work well in these tasks, there's always room for optimism.

May humanity flourish forever!

“The City Rises” by Umberto Boccioni, 1910. Museum of Modern Art (MoMA). Public Domain.

RECOMMENDED BY PROFESSORS

ROBERT SIMPSON

I read a book earlier this year called **Team Human**. It's a philosophy-ish book but written more as a manifesto. The author, **Douglas Rushkoff**, comes from a background of media and tech studies, and I think he was more of a tech-positive, cyber-futures kind of guy when he was younger. Like a lot of tech optimists from the 80s and 90s, his optimism was linked to a belief the internet would enable ordinary people to interact – creatively, economically, politically, intellectually – in ways that could resist (or at any rate, just avoid) the influence of massive corporations and the government institutions that they're often in cahoots with. Now that it's pretty clear that our devices and software platforms have been captured by those some corporate and government forces, Rushkoff has switched teams. Part of what he's doing in the book is telling a potted history of how the internet's emancipatory potential got coopted and subverted by the forces of darkness. But equally, Rushkoff is interested in why there's such a sense of fatalism around today's corporatised, capitalism-on-steroids tech world. He's trying to get us to imagine a technological world that's responsive to the needs and interests of ordinary people, and which can realise positive social change. It's a relatively light read, and I found it to be a good mix of history, philosophy-ish ideas and arguments, and freewheeling futurology.

COLIN CHAMBERLAIN

A Psalm for the Wild Built by **Becky Chambers** is an elegiac fable about a non-binary tea monk named Sibling Dex who, having grown dissatisfied with their routine of making tea and talking to people about their problems, goes into the wild in search of the sound of crickets. Dex encounters a robot named Mosscap in the wilderness and they strike up a strange and lovely friendship. This book explores what makes a life meaningful, whether human or robot, and challenges assumptions about what machines might want if they ever wake up. A heart-warming, hopeful read: like a wool sweater or a mug of warming tea.

If you are looking for something to watch, **Battlestar Galactica** (the TV series from the early 2000s) is an *epic* story about humanity's struggle with a hostile robot civilization. I watched this show over a decade ago and still think about it all the time. If *Star Trek* imagines a utopian future for humanity in which religion, politics, violence, extremism, tribalism, ambition, and sex don't really play a role in human life, *Battlestar Galactica* is the dystopian mirror image in which these forces loom large. I don't want to give too much away: this show is best experienced without any spoilers, as it contains a breath-taking series of reveals and reversals. *Battlestar Galactica* is strange, dark, and doesn't always make perfect sense. But it is gripping and compelling and mysterious. The first few episodes are some of the most gripping TV I've ever watched. Philosophically, *Battlestar Galactica* explores questions about what is permissible when the survival of humanity is on the line, as well as what is perhaps *the* ethical question: what makes a person a person, or what makes us human. *Not* a relaxing watching experience, to be clear. But so good.

"Melissa Aldana Kongsberg Jazzfestival 2022", Photo by Tore Sætre, Wikimedia. http://www.setre.net/ Creative Commons Attribution ShareAlike 4.0.

JOSÉ ZALABARDO

In 2013, aged 24, **Melissa Aldana** was the first female musician, the first South American person, and the youngest person to win the Thelonious Monk International Jazz Saxophone Competition. In her 2022 album, **12 Stars**, she speaks to you in a quietly confident tone of voice, vulnerable, but not afraid, wonderfully devoid of any affectation. I feel that to make my saxophone sound like that I would need to be a better person. Aldana's playing brings to mind a remark Paul Engelmann made about his friend Ludwig Wittgenstein:

"Gottfried Keller, one of the few great writers whom Wittgenstein revered wholeheartedly, indeed passionately, was superbly and exhaustively characterised by Ricarda Huch when she speaks about 'his veracity that will not permit his tone to be louder than his feeling by as much as a single vibration'. Such veracity, matching expression with emotion, is precisely what Wittgenstein was seeking in art, and it seems to me that this seeking was also the driving force of his philosophizing."

And I say: be like Gottfried Keller.

RORY MADDEN

This year I enjoyed reading the transcript and slides of a 2022 talk by **David Chalmers (Could a Large Language Model be Conscious?).**

He gives a systematic and accessible critical assessment of a range of reasons one might give for thinking that Large Language Models are not - perhaps never could be - conscious. He argues that most of these reasons involves obstacles to consciousness which could in principle be overcome in the future. Whether humanity should try to overcome those obstacles in future and create conscious LLMs is a separate question...

https://philpapers.org/archive/CHACAL-3.pdf

DANIEL ROTHSCHILD

Joseph Heinrich's The Secret of Our Success – How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter. Although this book was published in 2015, I read it for the first time last winter. It's a great example of scientific writing for a broad audience that is accessible without oversimplification. Recent progress in AI has lead to a lot of soul searching not just about how smart machines are, but also about what is at the heart of human intelligence. Heinrich has a really interesting take on this latter question, arguing that our individual intellectual abilities are much less important than we think, with culture playing a much more important role in the "success" of our species.

RORY PHILLIPS

"In 2023 I re-read Maximilien Robespierre's infamous speech 'On the Principles of Political Morality'. Robespierre was a high-profile member of the Committee of Public Safety in the French Revolution. In 1793-4, this Committee held a great deal of power, and instituted harsh laws, known in Britain as the Reign of Terror. Robespierre's speech is meant to justify this "Reign of Terror" against critics in the Revolutionary Government who wanted a return to ordinary law.

The speech is not very long, and contains a number of rhetorically impactful statements. Possibly most famous is this: 'If virtue be the mainspring of popular government in peacetime, the mainspring of that government during revolution is virtue and terror both: virtue, without which terror is destructive; terror, without which virtue is impotent. Terror is only justice prompt, severe and inflexible; it is then an emanation of virtue...". Here Robespierre claims that in a revolutionary situation, beset on all sides by threats, the actions of a virtuous government must be those of terror. He is tackling head-on the most important question faced by revolutionaries ever since the French Revolution: how does a revolutionary government govern? Robespierre answers: with ordinary law and justice applied to an extraordinary degree. Note that Robespierre's argument is not that virtue and terror are competing principles which need to be balanced, but that in revolutionary circumstances, terror is an extension of virtue.

The speech has clear philosophical relevance. It is a challenge to the 'moderates' in Revolutionary France to apply their principles directly. So it tackles the issue of how theory might turn into practice and what that might look like. It tackles the issue of what is permissible in extraordinary circumstances. It also puts centre-stage the big question: If terror and revolutionary violence are the only means to bring about the better world, then what can be permissibly done? Do the ends justify the means? Over two centuries after Robespierre's death (at the guillotine after the 'Thermidorian reaction'), these issues still engage the philosophical imagination."

BIBLIOGRAPHY

Does Artificial Intelligence have Consciousness? Yang Xue

Huckins, G. (2023). Minds of machines: The great AI consciousness conundrum. [online] MIT Technology Review. https://www.technologyreview.com/2023/10/16/1081149/ai-consciousnessconundrum/#:~:text=According%20to%20IIT%2C%20conventional%20computer
Jackson, F. (1982). Epiphenomenal Qualia. The Philosophical Quarterly, [online] 32(127), pp.127–136. https://www.jstor.org/stable/2960077
Searle, J. (1980). The Chinese Room. https://rintintin.colorado.edu/~vancecd/phil201/Searle.pdf.

AI and its Self-Consciousness Giovanni Zhou

Bote, Joshua. 2023. "What Sam Altman Said at APEC a Day before OpenAI Firing." The San Francisco Standard. November 17, 2023. https://sfstandard.com/2023/11/17/openai-sam-altman-fired-apec-talk/.
Jorge, Henrique. 2023. "Self-Awareness in Artificial Intelligence." Medium. August.15, 2023. https://henriquejorge.medium.com/self-awareness-in-artificial-intelligence-9a7e214b584.
McCarthy, John. 2019. "The Philosophy of AI and the AI of Philosophy." Stanford.edu. 2019.http://jmc.stanford.edu/articles/aiphil2.html.
Mumford, Stephen. 2012.Metaphysics: A Very Short Introduction. OUP Oxford.

Opinion Piece on Artificial Intelligence and Philosophy Anthony Nkyi

[1] Marr, B. (2023). The Amazing Ways Coca-Cola Uses Generative AI In Art And Advertising. [online] Forbes. https://www.forbes.com/sites/bernardmarr/2023/09/08/the-amazing-ways-coca-cola-uses-generative-ai-in-art-and-advertising/?sh=431bc0da2874 [Accessed 28 Nov. 2023].
[2] Liu, G. (2022). The World's Smartest Artificial Intelligence Just Made Its First Magazine Cover. [online] Cosmopolitan. Available at: https://www.cosmopolitan.com/lifestyle/a40314356/dall-e-2-artificial-intelligence-cover/ [Accessed 28 Nov. 2023].
[3] Sarkar, U. (2023). Pochettino With Dreads and Roy Hodgson in Durag: AI Football Pics Take Twitter by Storm. [online] Thick Accent. Available at: https://www.thickaccent.com/2023/10/07/pochettino-with-dreads-and-bielsa-drinking-a-pint-of-guinness-ai-football-pics-take-twitter-by-storm/ [Accessed 28 Nov. 2023].

The Ethics of Artificial Intelligence – An Evaluation of AI and Queerness Reo Lane

Klipphahn-Karge, M, et al. (2023), 'Queer Reflections on AI – Uncertain Intelligences', Routledge.
Kushwah, R.A. (2023), 'AI Making Queer Art: It's Marvelous-Looking But What's The Cost?', Gaysi.
Levin, S. (2017), 'New AI can guess whether you're gay or straight from a photograph', The Guardian.
West, S.M., Whittaker, M. and Crawford, K., (2017), Discriminating Systems: Gender, Race and Power in AI, AI Now Institute.

AI, Human Intelligence, and Narcissistic Wounds Sacha Bechara

Aristotle. (1937.) Parts of Animals. Movement of Animals. Progression of Animals. Translated by A. L. Peck, E. S. Forster. Loeb Classical Library 323. Cambridge, MA: Harvard University Press.
Aristotle., Ross, W. D. 1., & Brown, L. (2009). The Nicomachean ethics.Oxford; New York, Oxford University Press.
Aristotle. Metaphysics, Volume I: Books 1-9. (1933) Translated by Hugh Tredennick. Loeb Classical Library 271. Cambridge, MA: Harvard University Press.
Freud, Sigmund. (1910/1975). Leonardo da Vinci and a memory of his childhood. Standard edition of the complete psychological works of Sigmund Freud (Vol. 11) (trans: Strachey, J.). London: Hogarth Press.
Freud, Sigmund. 'Das Unheimliche [The Uncanny]' (1919), Standard Edition of the Complete Psychological Works 24 vols, ed. and trans. James Strachey (London: The Hogarth Press, 1953) vol. 17,
Hegel, Georg Wilhelm Friedrich & Forbes, Duncan. 1975. Lectures on the Philosophy of World History. Nisbet, Hugh Barr (Itzul.). Cambridge Studies in the History and Theory of Politics. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139167567
Laurent Alexandre, 2017, La Guerre des intelligences : intelligence artificielle versus intelligence humaine, Paris, Éditions Jean-Claude Lattès, , p. 250
Ryberg, Inez Scott. “Vergil’s Golden Age.” Transactions and Proceedings of the American Philological Association 89 (1958): 112–31. https://doi.org/10.2307/283670.
Virilio, Paul, 1996, Un Paysage d’Événements, Section 6. Paris : Galilée (ed).
Doublet, L. 2018. Conscience : identité, individu, morale, pensée réfléchie, sens éthique. In Piette, A., & Salanskis, J. (Eds.),Dictionnaire de l’humain. Presses universitaires de Paris Nanterre. doi :10.4000/books.pupo.12250

The Era of AGI: Unprecedented Economic, Political, and Societal Transformation Haochen Tang (Accessed 30 November 2023)

Alemohammad, Casco-Rodriguez, Luzi, Humayun, Babaei, LeJeune, Siahkoohi, and Baraniuk. “Self-Consuming Generative Models Go MAD.”, July 4, 2023.
Hobbes, Thomas. Leviathan, 1651.
Leike, Jan, and Ilya Sutskever. “Introducing Superalignment.” openai.com, July 5, 2023. https://openai.com/blog/introducing-superalignment.
Locke, John. Two Treatises of Government. S.L.: Blurb, 1689.
Our World in Data. “Birth Rate,” 2022. “Children per Woman vs. Human Development Index,” 2022.
OpenAI. “Learning from Human Preferences.” openai.com, June 13, 2017. https://openai.com/research/learning-from-human-preferences.
Villalobos, Pablo. “Will We Run out of ML Data? Evidence from Projecting Dataset Size Trends.” Epoch, November 10, 2022.
Worldcoin. “Worldcoin Whitepaper.” Worldcoin Whitepaper. https://whitepaper.worldcoin.org/.

Front Cover

Image of owl on robotic hand AI-generated by NightCafe Creator.

BENTHAM DIGEST ISSUE 8
THE PHILOSOPHY SOCIETY MAGAZINE
UNIVERSITY COLLEGE LONDON
Published January 2024