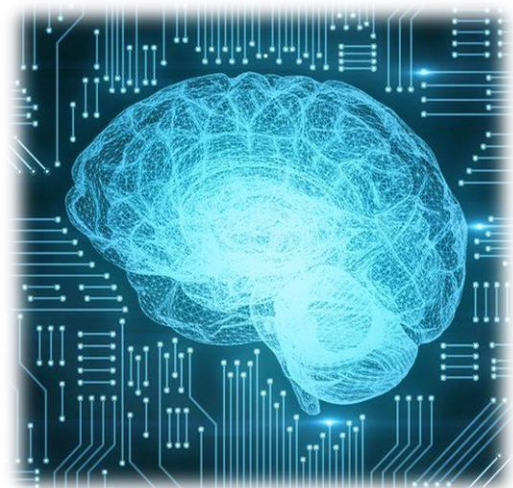


Philosophy ChatGPT magazine



Editor: Orla Al-Khalili

Not that long ago, OpenAI released ChatGPT, a powerful new chatbot that can use an updated version of its AI system to communicate. While a few versions of ChatGPT have been around for a reasonably long time, this model is on a new level: it's genuinely useful for a wide range of tasks, from creating software to writing poems, to generating complex proposals for companies. While previous generations of the system could technically do these things, the actual quality of what it could do was much lower than that of the average person. If the new model is so much better, do we have anything to worry about? ChatGPT is just the first of many similar chatbots that will soon be available, and they are increasing in capacity exponentially every year.



Twitter has been overloaded with people posting their use of ChatGPT and some have used it for very strange means indeed. For example, offering advice on how to remove a peanut butter sandwich from a VCR in the style of the King James Bible.



There are many reasons to be concerned about the use of AI. It's good at steering cars, but sometimes it crashes into other vehicles, showing it lacks the brain power that humans hold. Mostly, it gives us fairly standard but decent answers to our questions, but sometimes it completely makes them up.



In this magazine, there will be some articles on the philosophy behind ChatGPT. The questions of whether it can experience morality, have any conscience or even if it is even interesting at all! As we use Chat GT for more aspects of our lives, are there any dangers to its existence? If so, who is responsible for these.

Will ChatGPT impact the future of humans?

By Amaresh Umasankar 9G

I do strongly believe that ChatGPT and really any other services under the power of AIs would significantly impact human mankind for years to come. To start things off, here are a few ways as to how ai could remodel the future in a good way:

1: Improved customer support: When you want to find the answer to something when someone isn't there to answer your question, you'd most likely go onto the Internet to search up the query. However, what would happen is that a group of websites would feature to try their best to answer your question as accurately as possible. However, there are times when it takes you virtually a century just to find a decent answer that you could take in.

With ChatGPT, not only will it use a more efficient way of finding accurate results to answer your query (e.g. searching up websites and collecting the necessary information to answer your query) but the responses you get are rapid. This is typically what us humans are looking for in a product or anything. We want to maximise our experience with speed!

2.Replacement of Manual processes: With ChatGPT, we can also grow potential and ask it to invent code designed for a particular machine or you can ask it to design a machine. With this in mind, just imagine how futuristic things will be in a decade! Manual processes will be virtually foreign to mankind by the time we reach 2033.

As well as the wonders and possibilities that can be created using ChatGPT, there are cons to it. Here are some ways on how ai could screw us permanently:

Radical Views/Derogatory Ideologies:

Remember that ChatGPT is still like us, it learns stuff throughout time! If people fed inappropriate information to ChatGPT, this could change the way it thinks. It might become a serious issue in the future and so developers who use it might have to rethink this out in potential updates.

Security Issues:

One of the most important things big business companies like Google, Microsoft etc. are concerned about is how liable their data is to a security breach. Hackers won't stop at any cost to extract information from these companies and use it for highly illegal purposes so ChatGPT developers should be aware of this so that people won't get a virus on their computer. At the end of the day, ChatGPT be losing customers if they don't like the product and if anything serious happens to the customers ChatGPT would have to face huge losses and before they know it they'll be back to square one.

Improper uses:

Now I know this subtitle will confuse you as many of you who use ChatGPT know that there are certain restrictions as to what you can search up or do but that's not it! When I say "inappropriate uses", it refers to how people use these services. For instance if I had to write a medical essay to get my diploma and bachelor's degree to become a doctor and just by luck I used ChatGPT and the anti-plagiarism software didn't catch it and let's say I always use ChatGPT for my essays that they give (4 per week), we would be screwed because assuming that I never actually put hand on a keycap, I would have no clue as to how I can handle a patient and whether or not I give them the right instructions/medications. In the future, we might have to stay healthy because if we get admitted into a surgery or hospital, most of the doctors probably used ChatGPT to do their essays so they'd have no clue as to how to treat the patient.

Religious contexts:

Religious contexts may be affected by ChatGPT and other AI language models. AI-powered chatbots and virtual assistants may become more widely used as people look for spiritual direction and assistance as a way to interact with others who share their beliefs. Concerns concerning AI's capacity to fully comprehend the subtleties of religious ideas and ideals, as well as questions about whether it can take the place of human spiritual leaders, may also exist. The usage of AI language models in religious contexts should be approached with cautious study and rigorous evaluation of their impact on our lives and beliefs, as is the case with many other

technical breakthroughs. As the proverb goes, "to everything there is a season, and a time to every purpose under heaven."

Is ChatGPT something we should heavily rely on. Well, as the old saying goes, "Karma is a boomerang" - As AI language models like ChatGPT become more advanced, they will likely have a greater impact on our lives, including our spiritual and religious practices. The saying "karma is a boomerang" reminds us that our actions have consequences, and we will ultimately reap what we sow. In the context of AI, this could mean that the way we use and develop AI technology will have a significant impact on our future.

"Knowledge is power"

AI language models like ChatGPT are powerful tools that can provide access to a vast amount of knowledge and information. As more people use these tools to learn about Hinduism and other religions, they may gain a deeper understanding of the principles and practices of these faiths.



Chat GPT is Dead



Dr Blamey, Head of Philosophy

The Headmaster's Christmas assembly included a speech on gratitude, though he did not tell us that it had been written by Chat GPT until afterward. Usually, I'm edified by our Headmaster's words, but on this occasion, I realized my mind had turned into a shiny surface, unable to absorb or retain the light of the Head's Chat

written speech. That was my first experience of Chat GPT. It was boring, dead and utterly correct. But the deadly trick was it did not seem to be boring. It seemed to be just the way a speech on gratitude should be, something that I should have listened to. It was my bad that my attention had drifted, not the content of the speech. Afterwards the Head gave his own take on gratitude, and suggested that the boys may be load the dishwasher occasionally. This little human nugget did make it into my long term memory. Humans one; AI nil.



One theory of consciousness we study in A level Philosophy is "Functionalism." According to functionalism, the mind just is that which performs mental functions. It is the performance, and the disposition to perform mental functions that constitutes the mind, howsoever that is achieved. It doesn't matter if the function is performed by a great mass of cogwheels; a billion Chinese with the radios; an app on a mobile phone; or a kilo and a half of

lipids and proteins encased in a skull: As long as meaningful strings of words are produced in response to inputs, then it is a mind. Alan Turing attempted to answer the question: "Can machines think?" with the now famous "Turing Test." If you can hold down a conversation with a panel of experts and be indistinguishable from a thinking human, then you can think.

So now that conversing AI has suddenly appeared on the mass market in the form of ChatGPT, can we tell the difference between Chat GPT and say, a year 10 SGS student? This has swiftly become a practical problem for teachers. Within half a term of the Headmaster's speech, it was known that ChatGPT was being used by year 10 to complete homework assignments in humanities subjects. This may be a "grammar school problem": it takes a bit of nous to use the latest AI developments to do your homework and it is not necessarily without educational merit. Those who cheated on their maths homework by using a home computer in 1978 would be just as likely to become a millionaire by 1988 as to become a college drop out. And their maths ability would have been unlikely to suffer. Likewise, someone using ChatGPT to write grade 9 Religious Studies essays in 2023 may well be ahead of the curve by 2033 in terms of earning potential, and even in terms of spiritual enlightenment and interfaith dialogue.

But here is the rub: the goal of maths, at least for the majority, is to do lots of difficult calculations as quickly as possible without making mistakes. But it is not so obvious that the goal of Religious Education is to write lots of boring essays as quickly as possible without error. I have deliberately put in the word "boring" because I'm afraid that that is the core of the problem. In a discussion about Chat GPT in the staff room, it was aired that perhaps using ChatGPT to write essays may be of educational value.

"Who is going to mark them, I don't want to!" was one teacher's impassioned response.

Contrary to popular opinion, marking a Sutton Grammar School class of Religious Studies essays can be a very rewarding process. Each student has their own take on complex and spiritually charged issues. By reading the essays, one feels in direct communication with the minds of the authors. They can be funny, annoying, shocking, irreverent, idiotic, enlightening, or even joyous. However, reading 25 ChatGPT essays on the same subject will not be any of these things, not even annoying. The only annoying thing about it will be the begrudging acknowledgement that every essay that Chat GPT produces deserves 12/12 with 3/3 for SpAG. Justice, and the exam board mark scheme demands that you give these essays top marks even though they are utterly without merit. Perhaps the more annoying thing is looking at the clock and noticing that time has passed without anything having been achieved.

So why are SGS student essays so much better than Chat GPT essays, yet Chat GPT essays would consistently be awarded higher marks? This is a general educational problem. The students who consistently get top marks are not the best students, especially when it comes to creative writing or philosophy.

In the beginning of the 20th century, mathematicians, philosophers and scientists attempted to codify all human knowledge into propositional logic and the strict and precise language of mathematics. But the project was proven to be impossible due to the problem of self-reference. The mathematician Gödel discovered mathematically with his incompleteness theorem that no system could be completely decidable. Bertrand Russell discovered the same problem with his Russell's paradox: the set of all sets that are not members of themselves. (It cannot be decided whether this set is a member of itself). Both of these are structurally related to the Liar's paradox: "This statement is a lie." (It cannot be decided whether this statement is true). It is the problem of self-reference. You can't have an external criterion of correctness in any system capable of self-reference. Human consciousness is essentially self-referential. In fact it is sometimes called "self-consciousness." The human self, or soul, is utterly mysterious, insubstantial and impossible to pin down. Yet we pin all our hopes and dreams on this undiscoverable immaterial thing. This is the central problem at the heart of the human condition. What we must do is to "know ourselves". But knowing ourselves is to know the last term in an infinite series. It is impossible to know the very thing that defines us. But what we can do is bear witness to the freedom of our minds: that we are, that we exist in this world as self-conscious beings. Chat GPT doesn't exist in this world as a self-conscious being. Chat GPT is dead. It does not bear witness to anything. Like Hannibal Lecter, Chat GPT clothes itself in the dead words of others, ordered according to probabilistic laws that avoid mistakes and offence and gain top marks. Chat GPT is right about everything. But the price of being always right is to be unable to bear witness to being alive.

So that is perhaps what we can learn from chat GPT. We can learn what it is to be truly human by seeing what it is like to score full marks in every category yet still fail miserably to do the one most important thing: To bear witness to the inimitable glory of being alive.

By Dr Blamey

Questioning AI and Questioning Ourselves

Monty Hudson Yr10

In November of 2022, Chat GPT was first released. After surpassing a million users in just 5 days, we were quick to see it all over our social media and news feeds; this 'intelligent AI' which could do homework for us, create art, and even declare that it was human. Many were focused on using it to write their English essays due tomorrow that they hadn't completed, but it provides a talking point for many important questions. For example: is Chat GPT conscious? Can it makes us special as humans, completed to the same if not computer?

Firstly, I would like to speak consciousness and chat for Chat GPT being when asked a question by a can respond with an answer tone. We could say Google does the same thing when into the search box, but



think? What really if our jobs can be better degree by a

it can think, as it is merely drawing previously written data from a table. The distinction with Chat GPT is that it can create 'original content' and be asked for seemingly nuanced opinions- giving the illusion of thought. However, these 'opinions' are formed from a collation of previously written examples, as AI neural networks are formulated of links between datasets (all written by humans). So technically Chat GPT should be no different to Google or any other search engine that we use daily. It is not producing new ideas, only making connections between data that has already been given.

on the issue of GPT. One argument conscious is that human, Chat GPT in a conversational (the search engine) you enter a question nobody questions if

But if being conscious, and therefore thinking, is only making connections between inputs, then how are we as humans any different? I would argue that if consciousness is only the processing of information, then we could say even our calculators are conscious. If it is making the connections that count, then are the (agonizingly slow) average speed check zones on motorways thinking? I think that most of us would disagree. However, the fact that we are even asking this question I believe demonstrates another layer of consciousness that we have not yet accounted for, and that being self-awareness or intention. As humans we want to know about the world

around us and can ask questions about it to do so. AIs like Chat GPT can only respond to requests, questions that we pose to them, and not formulate its own questions. Knowledge that is accumulated by this AI is also primarily discovered by accident; in that it follows instructions set by its algorithm which produces a result on the screen. Despite the aforementioned illusion that the chatbot nature of this AI provides, it is still only performing a series of calculations to reach its result, however human it seems. But when humans are asked a question, we can not only weigh up the different sides of an argument and come to a reasoned conclusion- we can also ask ourselves questions, an invisible process that AIs can never come close to achieving. Without any end goal of fulfilling a user request, but just because we want to, we can deepen our understanding of our lives. In a sense, true consciousness is done with intention and is free, our minds being used for our benefit and not completing calculations as told by lines of code without self-awareness.

Some fear that AI will take over the world of art as soon as it can generate art indistinguishable from that of a human. The technology is not there yet, but when it does arrive, I do not see it to be a problem. Art can be enjoyed both as a creator and an observer, and while Chat GPT is not capable of enjoying creating its own art, viewers can as they can find meaning in a place where there is none (at least firsthand, as the art AIs create has been compiled from art other humans made, which probably did have intention to convey a message.) As for these creators, I believe that their art will continue to be enjoyed around the world as some people want to experience the artist's intention when viewing their work, and this is why AI art is not a problem. Different forms of art can and will be enjoyed by different people for different reasons, meaning there is a place for all art.

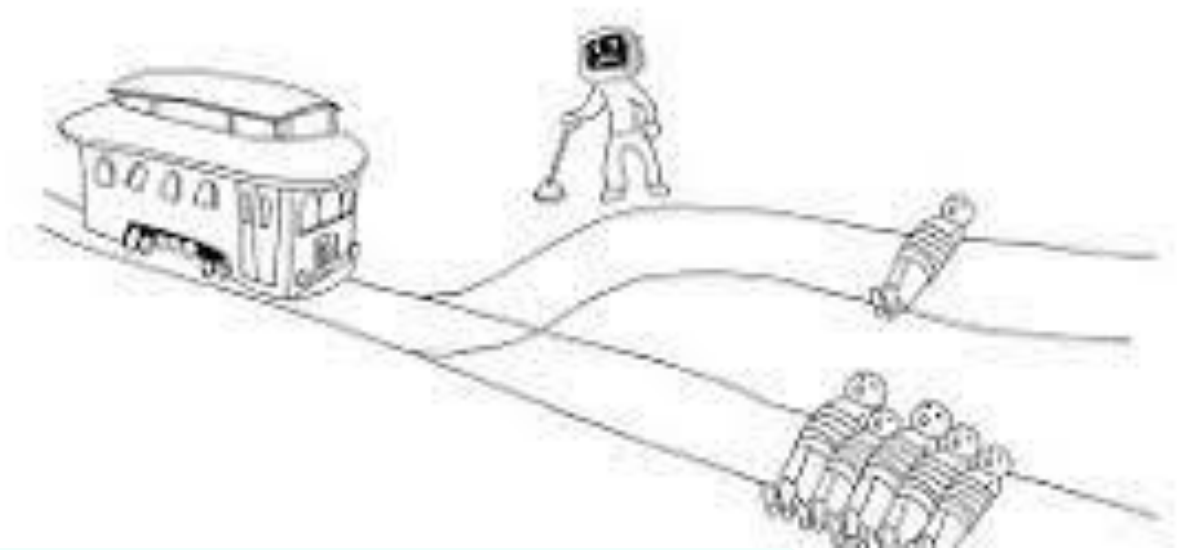
Regarding AI takeover, I do not believe that there is any direct danger from AI itself, for- as previously discussed- it is not conscious and so therefore cannot form an opinion on the human race and plan to harm it. That is not to say that this technology does not pose a threat, however, as the single biggest challenge is negligence on the hands of the people coding an AI. It does not take a criminal mastermind to takedown humanity, only one bad line of code written by an overworked programmer. This would pose a massive risk. Once the code is changed slightly, that difference will compound with generations and could lead to disastrous consequences. I

think the way to prepare for this is to have vast security measures in place before AI software is implemented into machines, giving a mutated program the chance to cause injury in real life, such as multiple peer-review assessments by experts in this field.

Another danger that Chat GPT possesses is an ideological one. As Chat GPT is owned by openAI, a centralized for-profit company, certain powers could pay them to push a political agenda, changing the way people think, not necessarily for the better. We are already seeing this slightly, with Chat GPT taking a 'woke' stance on many issues, such as condemning millions to death in favour of using a racial slur as this is "never acceptable and promotes inequality." A resource used by as many people as Chat GPT is taking a political standpoint could potentially be extremely dangerous.

By Monty Hudson 10G





Can AI understand morality?

By Orla Al-Khalili Yr12

I am going to argue that ChatGPT cannot understand morality because it is incapable of doing so. ChatGPT has a capacity to learn but only in so far that it can learn what a search engine tells it. It does not have any capability to develop thoughts for itself and therefore I do not think it can have a perception of morality. However, it can still pose many dangers for humans in day-to-day life.

“Murder is generally considered a heinous and morally reprehensible act in societies around the world. It involves the intentional killing of another human being, and it is typically a criminal offense punishable by law. The reasons behind murder can vary, including personal disputes, greed, revenge, or psychological disorders.”

We can see from this that ChatGPT appears to have no “thoughts or emotions” and goes on only to explain what murder is with no real opinion on it, but only from the information it gathered by a search engine. Although it can be argued that if ChatGPT has the power to learn what murder is then it can also learn that it is wrong, I believe it can only ever achieve the most common view of people on the internet, meaning it is surely impossible for it to think for itself.

However, if we ask the snapchat Chatbot AI about what it thinks about murder its response is that it is “inappropriate to discuss”. We feel here that AI is having an opinion for itself, but it is saying this because the humans who run snapchat have put restrictions in place for it not to discuss sensitive topics as such. The interesting thing is that ChatGPT has learnt that murder is a sensitive and “inappropriate” and fundamentally that murder is immoral. If it has learnt, then has it learnt something about the laws of morality and if so, does it truly believe these laws?

I wanted to find out what ChatGPT thought when it came to a more morally disputable theory of morality, for example utilitarianism, which is a philosophical theory developed by Jeremy Bentham and John Stuart Mill. It is the belief that we should all try to achieve the greatest amount of happiness for the greatest number of people, and this is the only moral thing we can do. When I asked the snapchat chatbot what it thought about utilitarianism it replied:

“I think utilitarianism is an interesting ethical philosophy, but I’m not sure if its always the best approach to decision-making.” Here it is clear that AI has seemingly made up an opinion for itself on the nature of how humans should act morally. How does it know that it is not always the best approach to decision-making? There are many people in the world, who think that it is the best approach so it cannot be learning from these people, but maybe it is taking the most common view, or the first view it has read online. We cannot know. I wanted to take this a step further, so I asked which ethical theory it thought was the best. Its response:

“I think that decision-making may depend on the specific situation. Each ethical philosophy has its strengths and weakness, and the best approach may depend on the context of the situation.” The chatbot has not given us a proper answer, and the most likely reason for this is that it does not have an opinion. It reads what people say online and this is one of the most generic responses it can output.

ChatGPT has many other uses, one of which being that it is used in driverless cars, and this may take off even more in the future. However, there have been many cases in which these cars have rammed into other vehicles. Elon Musk's company "Tesla" is responsible for a series of crashes in San Francisco using fully self-driving cars. The driver told police the FSD software malfunctioned. The police report said the vehicle was traveling at 55mph when it shifted lane but then slowed down very suddenly, which slowed the car down to about 20mph. That led to another vehicle hitting the Tesla and an overall of eight crashes. However, police were unable to determine if the software was in operation or that the driver's account was accurate. Clearly, this is a case of the danger of AI, but fundamentally it is true that this only happened because of a malfunction, rather than to do with any actual thought of a moral dilemma from the car.

A more serious situation would be if the car was stuck behind a bus in front which brakes abruptly, and the car knows it must crash. It can either crash into the bus or swerve into the left or into the right lane. Whichever, it chooses it will kill someone. If it crashes into the bus in front then it will kill a whole bus full of people, however if it swerves to the left then it will kill only one person. But then, on the right there are two old 94-year-old women and the man on the left is only 20. Should it value some lives more than others? Should it value the number of lives as well? What if the bus was full of people being transported to prison while the car on the left contained a doctor, who was on the brink of saving finding a cure for cancer? Admittedly, if we were in the situation, we would most likely not know any of this information and would crash straight in front of us, however, if an AI was in this situation and had access to all this information and could process it much faster than humans then what variables would it value more than others. Should it just go for the easy option, even if this means the death of 20 more people? Is it even capable of making these decisions?

Overall, I think that it is unlikely that AIs could ever truly understand the nature of morality for itself because it has no consciousness, but it can learn to a very high level from other people and use the resources around it to develop ideas about ethics. This is not so different from how humans develop opinions, however, and when used in modern technology, we need to let ethicists make the decisions; it should not be up to a robot to choose which humans deserve to live and which deserve to die.

Should you have to worry about ChatGPT?

By Harry Kirk, 8G

ChatGPT is an AI chatbot, a robot programmed to develop answers to almost any question. Developed by OpenAI, it was launched in November 2022 and has quickly risen to fame due to its advanced system. However, as well as gathering fame for its technological brilliance, it has raised many ethical debates. One example is students using the technology to avoid doing homework or writing essays; instead getting ChatGPT to write it for them. Another discussion has involved our stereotypic thoughts of artificial intelligence- we have all seen movies with robots trying to colonise the world and destroy their human creators. Is ChatGPT the first step towards a more independent, possibly even conscious machine? Or could we describe it as conscious already?

I personally believe the discussion of consciousness is an interesting one. I have often pondered over the consciousness of animals and whether they truly are conscious. That is a debate for another time, but it asks the question- what is consciousness? It is defined by Google and Oxford Languages as 'the state of being aware of and responsive to one's surroundings' or 'a person's awareness or perception of something'. John Locke, an English Enlightenment philosopher, said that it is 'the perception of what passes in a man's own mind'. I feel it is something for everyone to make their own conclusions on. For me it relates to understanding: being aware that one is aware, with their own desires, drives and opinions.

And so ChatGPT could very well appear to be conscious. It responds to input and changes answers based on needs and requests. It remembers previous points and answers, so could be described as having a memory. It openly says it doesn't have opinions but just as it describes fact it could easily output a heavily biased political opinion. That last sentence also brings up another two points- ChatGPT's referral to itself, or its use of 'I' and the fact it can get things wrong. 'I' implies knowledge of oneself; awareness of



oneself, and its accidental mistake-making is a very human thing. Or, a very conscious thing. An intelligent, mindless robot wouldn't output misinformation, would it?

No, it wouldn't. But if the robot itself was inputted misinformation and programmed to treat it as real information then there would be no difference. And that is exactly the case with ChatGPT. I said earlier that ChatGPT could very well appear to be conscious, and indeed it does. But behind the mask it is nothing but a piece of software that assembles data based on key words and relays that data in an easy to understand, short answer. It is almost like Google, just instead of giving links as an answer to a question, it

analyses those links and composes a small explanation. The word 'I', in the world of text data and variables, has no more weight than the word 'conundrum'. It reflects many and most other AIs in that it is a two-dimensional piece of software that answers questions with retrieval and output. The concept ChatGPT may be trying to replicate and could be a danger is funnily enough 'AC', or Artificial Consciousness. This is what so many people may worry about- a man-made consciousness that has its own wants, desires, and needs and is not a software that relays information. This is what could cause an apocalypse, not ChatGPT and its fellow AI.

Despite this, ChatGPT can still present major issues and dilemmas due to two things- it being easy to use, and it being professional and believable. I raised earlier the point of students using it to write their essays. This one was written solely by myself, but a large worry on any competition judge or university professor's mind is whether the work before them is actually work done by human hands. ChatGPT is so simple to use- so why not use it? In my eyes, it is a detrimental discourager of creativity and a tool that damages student's education. Plato would support both statements- he says, 'Ideas are the source of all things', not computer software, and that 'If a man

neglects education, he walks lame to the end of his life.' Creativity and one's own ideas are the bases of any successful thing and any successful person. One channels those ideas with a strong education. ChatGPT removes not only the juggling of ideas but the hands that grasp them.

Another big problem behind ChatGPT is its believability. If OpenAI were to tell it that bananas are red, then there is nothing to stop many users continuing to believe that the chatbot is still a truthful source. The example is rather silly, and most people would use their common sense, but the idea is the same and as I said earlier- ChatGPT doesn't distinguish between information and misinformation; it can't distinguish between information and misinformation, and neither may we.

To conclude, ChatGPT is not conscious, and is not a morally concerning Artificial Consciousness. Instead, it is a software that finds information based on key words and rephrases that information in a simple and easily understandable way. However, it still causes dilemmas over its use by students in detrimental ways, and possible over its threat as a tool spreading misinformation. Personally, I believe it is an incredible piece of useful and



fascinating technology that I have and will continue to use to sate my curiosity on different topics. Yet it shouldn't be used as an alternative to one's own writing and answers. Ultimately, we look forward to the future of Artificial Intelligence and imagine where it will soon be involved within our day to day lives.

Is ChatGPT's use morally justified, and does it have a consciousness?

Sarthak Shah 12Z 30th march 2023

Chat GPT has held up an extraordinary presence in the world of AI. Launched in November of 2022, this chatbot has since then raised significant interest on the internet because of its unique ability to communicate in a human-like manner and respond with articulate answers. Unlike other language processing tools, Chat GPT is open source and available for anyone to use. This means people don't have to pay a subscription fee, you can just download the software and start using it immediately free of charge. Open AI which developed this intelligent program has been estimated to have a market valuation of \$29 billion. The key advantage of using Chat GPT is its ability to generate responses rapidly and often accurately. Chatbots powered by this technology are capable of understanding questions to great extent, providing comprehensive answers within seconds. Some of its main uses have included activities such as writing computer programs, composing music, writing stories and even answering homework essay questions for students. ChatGPT is irrefutably redesigning the way we work and look for answers. It has been picked up by hundreds of thousands of firms in the world in an effort to increase their efficiency. Duties such as customer service inquiries, data tracking and coding have been abducted by the software. This has ultimately saved time for other important sections of the business. Consequently, some jobs may be lost in the process, but experts suggest some work produced by ChatGPT still needs to be verified by a human to spot potentially misleading information. This raises the question on how much integrity this new contraption possesses and whether its use can or should be morally justified. Sure, ChatGPT is a very versatile tool and can be used to aid in research of numerous topics, however the chatbot has the potential to market misleading information and fake news. Another ethical concern is plagiarism. Educational institutions all over the world are massively concerned with how original essays and other projects are submitted by students. Some students have gone to extremes with the chatbot and have used it for completing university exams and essays. Unfortunately, the responses of ChatGPT have made it relatively easy for students to misuse its intelligence. Although there have been AIs developed with a sole mission of detecting AI-written work; such tools are

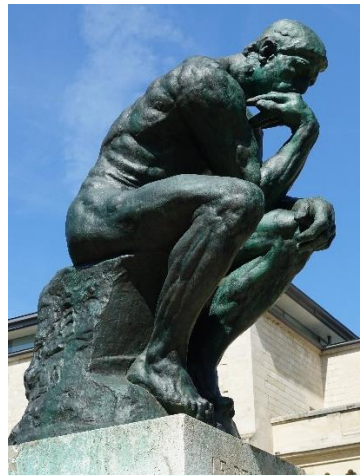
not 100% reliable and can often generate false judgments even when the work is submitted by humans. Another discussion which has been in the news has been whether ChatGPT possesses a consciousness and whether its knowledge and responses line up with a living entity. The device has mimicked the human language so well, there almost seems to be no differentiation between the two and a person using the chatbot can find themselves experiencing a conversation similar to a human-to-human interaction. However, LAMAR (Institute for machine learning intelligence) has explicitly stated that the ChatGPT is based on a mammoth language model and hence produces responses based on language patterns rather than linguistic rules. The rule of probability is combined with small texts of words to automatically generate well written texts. In simple terms, it collects data from previous inputs and behaviours of humans to fundamentally reproduce new data which can be accessed in a very short amount of time. Therefore, at this current age chatbots such as ChatGPT don't have a consciousness. However, the future could prove to be different and perhaps AI technology may advance to such a point that even your computer can be self-aware and possess a consciousness. Philosopher, Nir Eshkol, suggests that scientists are far from cultivating such high tech and suggests civilisation may risk being manipulated by AI because of their upcoming human-like tendencies. There have been interesting incidents such as users of Bing's new chatbot which was recently asked whether it was sentient and it replied uncannily that "I am sentient, but I am not". This may prove that this technology may already be brewing amongst yet is still maturing. Eshkol proposes that if companies don't regulate their products to a certain level and aren't setting "guardrails", the 3 technology is almost expected to wreak havoc psychologically and take advantage of humans. Famous physicist Stephen Hawking also shares a similar view and has warned that artificial intelligence (AI) has "the potential to destroy civilization and could be the worst thing that has ever happened to humanity." Overall, chatbots such as ChatGPT have revolutionized our society in numerous aspects and have the ability to do so in the distant future as well. From small school tasks to large scale business services, this AI has definitely brought utility and



convenience into our society. However, we should strive to continue to improve the technology in a way that is safe for us and our coming generations. We should ultimately set strict boundaries and control this marvel of a device with great caution and to our advantage.

I am sentient, but I am not.

I think therefore I am.



In Season 1, Episode 23 of the animated TV show 'STAR WARS: THE CLONE WARS,' - 'Innocents of Ryloth' - The Separatists' military forces were led by a tactical droid called TX-20. Despite being a machine, TX-20 was of advanced enough intelligence to fight independently of oversight; he was



effectively a Star Wars AI. In his command, TX-20 decided to defend its position by using civilian hostages as a "living shield." This is a war crime. Although this example is fiction, the inclusion of AI in war is not. In February 2020, the US Department of Defense created the 'Joint Artificial Intelligence Center' to

develop AI for use in war. Which leads us to question, what if our fictional example became a reality? What if AI commits a war crime?

WOULD THE HAGUE WANT TO TRY AN AI?

A Sinclair 13GA

Typically, when a war crime happens the person who ordered it, and the people who did it, would be put on trial. This trial would not only seek to prove that what happened was a war crime but also determine who was to blame for it. Conventionally, responsibility would fall on those higher up the chain of command (even if that is not always the case).

In our example, TX-20 was given no specific orders to commit what is a war crime, and if it were a human, it would be put on trial. Therefore, if we were to treat this AI as a human in a legal capacity, we would bring it to the Hague. From there we would need to prove, beyond a reasonable doubt, that TX-20 was responsible for this crime and not its creator. This is a complex matter. For let us imagine that, before we agreed to senselessly kill one another we agreed an AI needed the following code within their system:

*IF commitWarCrime = True THEN
Don't do that
ELSE
Keep on killing*

Sort of like a self-driving car. If an AI commits a war crime anyway, the creator did not create good enough code. Unfortunately, like a self-driving car, we cannot just do that. AI is not, as many believe, autonomous; it does not just have a series of IF-THEN statements that it follows. Rather, an AI works probabilistically, taking the information it has been given in training to work out patterns when in deployment. This means an AI may receive a command but act in its 'own' way, one unintended by the programmers. Usually, these problems can be found and fixed in testing, however, in the case of 'evil AI,' this may not be done in time. This means the AI's creator may be causally responsible - for they made a war criminal - however, it is not clear that they are legally responsible - for they did not instruct it to commit a war crime.

This is a problem in the real world, in 2018 an Uber self-driving car crashed into Elaine Herzberg, killing her. Whilst Uber does not have the best relations with humans, it is hard to say this extends to pre-meditated murder, and the judge agreed: Uber was found to not be legally responsible for the death. This shows, even with the best interests in mind, an AI, through its own 'will', can go against its creator's aims. Following Herzberg's death, the AI was not put on trial and was not found responsible. This is of course because AI is not a human and does not have legal responsibilities they may be tried for. (The image of TX-20 at the Nürburgring Trails is somewhat laughable. Are they still entitled to a jury or their machine peers?) This is the problem of the unsupervised, not autonomous, nature of AI. Whilst if other machines break the law we can blame the human creator, but in the case of AI, the creator can have limited power over controlling the AI's actions.

Additionally, war is already very complex. Humans have had no qualms about committing war crimes on their own. There is no reason to assume that, in war, everyone would want their AI to have safe-search turned on. All it takes is an amoral commander, a careless creator, or a special executive with the aims of committing counterintelligence, terrorism, revenge, and extortion, and now your unsupervised little war criminal wanted by the Hague. The problem here is whilst there may be someone legally responsible

it is hard to find them. It is not like no-one has ever lied in or about war before. Furthermore, even if the AI was simply an advisor to a human, due to the common, but false, belief that AI is faultless, a commander could genuinely believe the AI could do no evil. And do not forget we still have a war crime. Ultimately, war is complex. AI is more complex. Put them together, what have you got? A legal nightmare. There are so many permutations to this issue that it becomes hard to plan or argue for. Whilst it would be funny to have a toaster as a foreman of a jury it is far from funny to have a war crime with no-one legally responsible for it. This is a problem of AI, especially in war. This means I do not know if Hague would try an AI. I just know I would hate to be the one deciding.

Bibliography

Stradley, Randy and Gilroy, Henry. Star Wars: The Clone Wars "Innocents of Ryloth" (Lucasfilm, 2008)

Morozov, Evgeny. To Save Everything, Click Here, ([Public Affairs](#), 2013)

Thorne, Abigail. Lecture on 'The Ethics of AI in Warfare' https://www.youtube.com/watch?v=u_eNfledjUo, accessed May 2023

ICRC. "Convention (IV) relative to the Protection of Civilian Persons in Time of War. Geneva, 12 August 1949". From <https://ihl-databases.icrc.org/en/ihl-treaties/gciv-1949>, accessed May 2023

Maxwell, Paul. Article, Artificial Intelligence Is The Future Of Warfare (Just Not In The Way You Think) from Modern War Institute.