Finlay McCormick

Will artificial intelligence be helpful or dangerous to humanity in the future?

Abstract

The goal of this project is to come to a conclusion about artificial intelligence's usefulness to humanity in the future, based on what AI has been used for already and how it has proven helpful, the problems AI has already begun to show and then finally the thoughts and theories held by people with lots of experience in artificial intelligence. In doing this I have explored a wide variety of uses that people have already found for AI, ranging from detecting illnesses to playing board games. I will also be covering issues such as bias, lack of general intelligence and the potential danger military AI can pose. At the end of the project my conclusion is that although AI could be misused and potentially be a threat to mankind, it is still worth creating because of the amazing improvements humanity will be able to achieve with it.

Introduction

In this EPQ I intend to cover one of the most important things which humanity will have to deal with, one which could either allow humanity to ascend to what we could only view as godhood or wipe us out in ways totally unimaginable to us. This thing is artificial intelligence and it could be, in my opinion, the greatest thing humanity will ever achieve, IF we do achieve it. At this point in time, we have no idea if the more advanced types of AI (compared to what we have today) are actually possible. For example, not even the smartest minds know for sure if we can simulate a conscience, whether this be because the biology of the human brain is too complex to understand, let alone simulate, or because our minds are simply something which cannot be recreated. For this EPQ though, we are going to assume that advanced AI is possible and that it is going to be created in the future.

Over the course of this project I am going to discuss what AI is being used for and whether it is ultimately going to be helpful or dangerous to us as a species. I will be using information from various sources as well as my own personal speculations.

An important thing to mention before I start is what machine-learning, artificial general intelligence (AGI) and artificial super intelligence (ASI) is. Machine learning is a technique used to train AIs where the AI will use data to learn and improve on its own. An example of this is deep learning, which is inspired by the nerve cells (neurons) that make up the human brain (Macuga, 2017). An artificial general intelligence is an AI that is as intelligent as a human, unlike AIs we have now which are usually only 'intelligent' in one or two specific areas. An artificial super intelligence is an AI which is far smarter than the most intelligent person can possibly be, as well as having the ability to improve both its physical hardware and its software (Tegmark, 2017).

How artificial intelligence will help us

Artificial intelligence is a relatively new development in science and technology. Though it is still at a relatively early stage in its development it has seen use in many different scenarios and ways. One of the ways it has proven itself extremely useful is in the medical field. An example of this is an artificial intelligence developed to detect brain changes which may lead to Alzheimer's in the future. This is extremely helpful because, whilst there is currently no cure for Alzheimer's, there are drugs in development that are intended to help with Alzheimer's and are likely to work better if they are given to the patient early rather than later. This algorithm was initially trained using 67 MRI scans, of which 38 were taken from people who had Alzheimer's. Later on, it was tested with a second set of scans from 148 people, with 52 being healthy, 48 having Alzheimer's and 48 having mild cognitive impairment, who within the next two to nine years developed Alzheimer's. The artificial intelligence could tell the difference between a healthy brain and a brain with Alzheimer's with 86% accuracy. More importantly though, it could tell the difference between a healthy brain and a brain with mild cognitive impairment with 84% accuracy, meaning that the algorithm was capable of spotting changes in the brain that would lead to Alzheimer's, in some of the cases, nearly a decade before the symptoms began to appear (Ananthaswamy, 2017). This source is reliable because the writer has written several other books about physics/computing and it is useful because it shows how AI can be used to help humans in ways which were not possible before.

Whilst the AI mentioned in the source is very limited in its scope (as most AIs made now are) it and other medical AIs are capable of doing things that a doctor couldn't do. I believe that, whilst doctors and nurses do an amazing job at diagnosing and curing people of ailments, and specialists have an immense knowledge of their field, they simply cannot contend with an artificial intelligence in the medical field. Whilst doctors may have to trawl through a database to find several possible things that can be wrong with a patient, an AI will potentially be able to perform the diagnosis within a hundredth of the time and with far more accuracy, if it has access to the right resources. Whilst an AI like this could be extremely useful for humanity in general, it would change the role of some doctors, nurses and other medical specialists, possibly even putting some out of work. Whilst job losses aren't a good thing, it would improve the efficiency of the medical profession and allow them to treat more patients than they do now or to retrain to different jobs. The AI will also open up more job opportunities for technicians who need to keep the AI running.

Another source I found concerning health care initially surprised me. This is about AI being used for end of life planning. Due to the sensitive nature of the topic I thought an AI, especially one made now, would struggle to do this, since I imagined someone looking for help with such personal issues would prefer to deal with another person rather than a piece of software. However, this has been shown to work successfully by an AI made in Indiana, where a company called LifeFolder has launched a chatbot to help people with end of life planning. The bot, called Emily, is a free messenger bot on Facebook which helps users to make important decisions and to complete all of the legal paperwork. This paperwork

comes in the form of three documents, the Health Proxy nomination, which selects who the user wants to speak for them if they are unable to speak for themselves, the Advance Directive, which outlines what you want from your healthcare, and an organ donation statement. The artificial intelligence is designed to talk through any pros and cons and to help make the difficult conversations easier (PR Newswire, 2017). Whilst AIs like this certainly won't cause any problems for us I personally think that they are, at this time, not particularly useful for the people on the receiving end. This is because unlike the previous source where a normal doctor would not be able to do what the AI could do, there are many professionally trained people who do this work, meaning that they could lose their jobs in return for an AI which may not do the job any better. The biggest advantage this AI has is for companies producing AIs like this, such as LifeFolder, because it means they can reduce costs as they don't have to pay people to do these jobs, so they can spend their money elsewhere, such as improving the AI.

Despite the number of different companies and people working on AI, I think that one of the first places where AI will be used on a massive scale will be in warfare. Autonomous robot soldiers have existed in science fiction for a long time, but now they are close to existing in real life. Artificial intelligence being used for warfare can bring many advantages, one of the biggest being that it can replace the need for human soldiers. A war fought by mass produced combat robots is preferable to one fought by people, since the robots won't be able to feel pain, get PTSD, etc. I also think that an AI commanding lots of robots will also be able to fight a war far more effectively than humans can, due to how much faster AI can process data when compared to a human and also its direct connection to all the robots under its control will let it know what is happening on every front at all times. It will also be extremely good at tactics since an AI would be able to examine and test strategies at an extremely fast rate, either using past real-world battles or simulations.

One of the problems with military AI is once again that it will lead to fewer soldiers and higher unemployment, however, on the other hand it will create jobs for AI developers and support technicians and also create more jobs for people working in the companies which manufacture the combat robots, and for engineers, who would repair the robots when they get damaged. Another serious issue with AI and robots being used for military purposes is if the AI is used against humans. If the AI was not developed with proper safeguards covering rules of combat, avoiding civilian casualties, etc., due to the AIs lack of any code of morals, it might kill every enemy without mercy and may even target civilians. Fortunately, though, companies like DeepMind are working on ways to make AI ethical (Temperton, 2017), so hopefully this problem can be addressed. There is also the fact that, if the AI is developed well enough, it could lead to a reduction in the amount of collateral damage in war zones with fewer civilian casualties and deaths.

On a lighter note, another use which AI may have is in entertainment. So far, the biggest breakthroughs for AI in this area is in board games, specifically ones like chess and more impressively Go. Go is an ancient Chinese board game with 19 x 19 board where players place stones to fight for control of the board. Due to the size of the board there are a huge number of possibilities, far more than in chess, which is why it is so impressive that an AI

taught itself how to play and beat the world's top go player, Lee Sedol (Chen, 2016). Whilst this is an extreme example, AIs have been used to replicate players in many games, chess and go included. Being able to play with or against AI allows for people to play a game they like whenever they want since they don't need anyone else. It also makes it easier for people to practice at games, and many games with an AI opponent allow the user to select the difficulty the AI will play at, meaning that they can challenge an opponent with a consistent skill level in order to improve. However, current limitations with AI mean it cannot be used to play all games. An example is Dungeons & Dragons, a role-playing game where the scenario and even rules can change. AIs find games like these harder to play because they have difficulty finding strategies which consistently allow the AI to succeed, due to the high level of randomness (Ellis & Hendler, 2017). This problem will be overcome however when we create an AGI, because it will be able to think more like a human so will probably be able to handle games such as Dungeons & Dragons better.

One of the biggest concerns and a common issue with AI is that as it continues to develop it will start to replace more and more people in many different types of workplace. This will continue the trend of recent years where computers (office automation) and simple robotics (which are used in manufacturing) have replaced workers because the new technologies are better, cheaper, etc. for the companies. Whilst AI is undoubtedly going to cause lots of unemployment, and an estimate from the McKinsey Global Institute is that by 2030 up to 375 million people might have their jobs replaced by an AI or robot (The Economist, 2018), there are other ways in which AI is being used to help people get employed. An example of this is Seeing, a camera app made by Microsoft which narrates what it sees through the camera (Wagner, 2017). This app will make it easier for people with visual impairments, who previously would have struggled to get a job, to become employed. While this will only affect a small number of people, there are bound to be other AIs developed which will help more people to become employed and/or gain new skills, hopefully offsetting some of the unemployment caused by AI. Another advantage of AI is that the first jobs to be fully automated by AI will probably be low skill, repetitive jobs, meaning that there will be no need for humans to do these boring jobs (Ford, 2013). I think that if systems are put in place to help the people currently working jobs like these to get work after they are replaced by machines lots of people will find themselves in more enjoyable, less repetitive jobs. Also, history shows that previous waves of job losses in manufacturing, industry and office work have been offset as new job types and services have been created.

All of the things I have discussed so far have either already happened or could happen in the near future, but now I am going to talk about the far future. I think that the creation of an artificial general intelligence will be the first step to truly advanced AI. This is because of its ability to self-improve, which means that, even if the first version which we create will only be as intelligent as a human, it will be able to improve its software far faster than a human could, meaning that it could become an artificial super intelligence in a relatively short time. If our species can survive for long enough, the creation of an ASI would probably be the point in time where humanity would be launched into a world which we would today consider to be science fiction. Due to how fast ASI would be able to process and how

intelligent it would be, it could advance our technology significantly in a short amount of time. This technological advancement could lead to many great things, such as ending hunger with more efficient farming methods, helping create world peace since it would be able to understand and explain opposing arguments and come to the best compromise for everyone or maybe even allow us to develop the technology to travel between different planets, assuming it's possible.

With the technological advances an ASI could give us, it opens up the possibilities of improving human society in many unimaginable ways, leading to greater happiness and pleasure. Due to the fact that a perfect ASI would be totally fair and unbiased, it would make a great leader, which could lead to us living in a perfect utopia where a person could do whatever they want. This would be helped by the advanced technology which could allow for people to be turned into cyborgs (organic creatures with robotic implants) in order to enhance themselves, advanced virtual reality which would allow people to effectively do whatever or go wherever they want and even the possibility of being uploaded to a computer and to live forever. However a big issue with a world like this is the lack of purpose, without any jobs or research to do many people may feel their lives don't have any true meaning, although this would probably not worry everyone and may only be a concern for a small percentage of the human population (Tegmark, 2017).

How artificial intelligence will be dangerous

Whilst artificial intelligence has done lots of good for humanity already and has lots of potential for the future, there are unfortunately many dangers in creating AI. The most direct threat an AI could pose to humanity would be with its use by the military, simply because if a military AI with control over robots or even just the technology we have today were to be used against us by another country, there is not much we could do against it (without the help of our own AI). In the previous section, I mentioned that a military AI would be most dangerous if used against humans due to its lack of morals and countered that by saying that we could develop a code of AI ethics, but who's to say that all, if any, governments will bother making their military AI ethical?

Another problem is that an AI will follow its protocols no matter what. Whilst this may not sound bad, the true story of a Russian officer who, despite a newly installed missile early detection system going off, told his superiors it was a false alarm because of his gut instinct, which it was. (Editorial board, 2017). If an AI would had been in his place, it would have automatically informed the Russian command that a missile launch had been detected, which would probably have led to a nuclear strike by Russia and full scale nuclear war. Another potential scenario is where the AI, for some reason, decides to turn against humanity. An out of control AI with access to the correct tools could cause massive damage and lead to many, many deaths. This scenario is made worse when you also consider that an advanced enough AI could develop chemical or biological weapons far more dangerous than ones we have today, especially if it didn't follow any rules of war.

Whilst evil killer robots are probably what come to most people's minds when they think of the danger posed by AI, there are many other ways in which AI could cause problems for people. One of the biggest impacts would be unemployment. In the previous section, almost all of the ways that AI could be used to help people had the downside of causing unemployment. Due to how much better and cheaper an AI can perform a job than a human, many companies will want to replace their workers with machines as soon as possible. The first jobs to be automated by AI will probably be low skill jobs because of their repetitive natures. Many people think that these will be the only jobs in danger of being replaced by AI and, whilst that may be the case for the near future, as AI improves more and more it will be able to be used to automate more and more jobs, including what are currently thought of as skilled professional jobs in areas such as accountancy and legal services (The Economist, 2018).

The big fear is that the people who get replaced by AI will struggle to find new jobs especially since the work which will be safest from automation by AI will be more creative and non-routine jobs, which lower skilled individuals could struggle to adapt to (Ford, 2013). This source is reliable because it has been peer-reviewed meaning that other AI professionals have read and accepted it. Another more general social problem that would arise from higher unemployment is an increase in the wealth gap, as low-skilled, poor paying jobs get taken over by AI, more people will find themselves unemployed and unable

to improve their standard of living. At the same time, there is a risk that, in the business world, companies will consolidate into a handful of monopolies and conglomerates because they have the best AI technology and take an increasing share of global profits and rewards for their staff and shareholders.

Another worry some people have for AI is about how it uses the data given to it. The concern is that AI could be given biased data which will mean that, especially with our increasing dependence on computers and use of AI, one group of people may find themselves being treated unfairly compared to another group of people because of an in-built bias in the AI algorithms. An example of this is a system called COMPAS which is used by some judges to predict how likely it is that a defendant will reoffend and from that decide if they get parole. The company that made this was called Northpointe and kept the workings of their algorithm a secret, but ProPublica, an American non-profit newsroom that does investigative journalism (ProPublica, n.d.), did an investigation and found that it may be bias against minorities (Knight, 2017). This source is reliable because not only was it written in the MIT technology review, a magazine published by the Massachusetts Institute of Technology, but it was also written by the senior editor for the magazine.

The problem of bias is compounded by the fact that most of the emerging machine learning techniques are very complex meaning that, even with access to the algorithm and the data given to the AI, it is hard to understand the AIs workings, thus making it harder to truly know if an AI is biased or not. If AIs continue to be biased, the issue with AI causing unemployment could also become worse, since a biased AI could make it harder for certain groups of people to find work in a world where it is much harder to get a job anyway, due to automation. This could be especially bad for already poor communities of minorities. This is because if an AI was biased against a particular minority, there would be increased unemployment in that community, which could lead to a concentration of crime in the area and lead to other social problems.

One of the other problems with creating AI is how it could be misused. For example, the big brother scenario where the government watches what everyone does in books like 1984 (Orwell, 1949) in order to make sure nobody is doing anything that the regime wouldn't like. With enough cameras, an AI would be able to monitor what all people are doing at all times and would also be able to stop anybody from doing anything against the government. This could greatly empower authoritarian regimes and would undoubtedly appeal to many dictatorships and governments who do not care about their citizens' privacy. The level of surveillance available to states would make it so that it is nearly impossible for someone living in that regime to do anything about it, like escape, since the government would know from the AI what was being planned.

An AI could also be used by governments to make propaganda. Whilst this wouldn't be as effective today because our AI doesn't have a good enough understanding of how humans think, in the future an AI would be able to find the most effective propaganda and mass produce it using robots. A third problem with governments using AI is that they would try to gain advantage over other countries which would most likely cause some of the other problems I have mentioned about AI. I think this is because in an effort to stay ahead, many

governments, particularly some of the authoritarian states that are not as powerful as countries like America, would do anything to get an advantage over them, which could lead to them rushing their AI development. A rushed AI would be more likely to have lots of bugs, which could just lead to it causing harm to the government or to its people, or it could potentially glitch and launch a missile. Either way, it would be dangerous.

Further into the future, when ASI has been created, AI would potentially be more dangerous than it could possibly be before. In contrast to the utopian scenario in the last section, it could just as easily present a dystopian future. Due to how intelligent an artificial super intelligence would be, it could quite easily mislead and manipulate us, allowing it to do things which we may not want it to because it could trick us into thinking that it should be allowed to do whatever it wants to do, or it could manipulate us into letting it do what it wants. On a darker note, it is possible that, in the process of achieving whatever goal we design the ASI to have, it could view humans as something which get in its way and decide to eliminate us. If this were to happen, chances are there would be very little we could do to stop it and it would probably end with our species going extinct or most of us being killed and a few of us being kept alive similarly to how animals would be kept in a zoo (Tegmark, 2017).

Conclusion

In conclusion, I think that AI will be helpful to humanity in the future but could also be quite dangerous. I think that AI will be helpful to us because it has already shown that it can be beneficial to humanity, whether it is diagnosing diseases like Alzheimer's more accurately and faster than a doctor could (Ananthaswamy, 2017) or helping someone to deal with a difficult issue like end of life planning (PR Newswire, 2017) that they may struggle to do without help. AI also has massive potential for the future, with many things being developed now which could greatly benefit many people, such as apps like Seeing which can help those who would struggle without the assistance of an AI (Wagner, 2017). I also think that, if done properly, military AIs could greatly reduce the number of human casualties suffered in war. In the long term, if we ever develop an ASI, humanity could become so much more than we are now. The fact that there is a chance we may one day live in a space faring society with cyborgs and virtual reality is amazing and only possible with AI.

However, AI won't just bring good to humanity, at best I think there will still be many bumps on the road to advanced AI and our future as a species. The huge amount of unemployment AI may cause as we use it to automate more and more jobs will certainly make many people's lives far more difficult and the danger a misused or glitchy AI can cause to people is immense. There is also the likelihood that AIs will be put into use without enough time to perfect and test it, which could lead to situations where AI is biased or simply doesn't do what it is meant to. Finally if we ever manage to create an ASI, there is a chance that it could use its intelligence to control us, or it could just try to kill us all.

Though despite the danger it poses, I still think that, overall, AI will be helpful for humanity in the future. I believe this because AI can do so much good for humanity, if we just do it right, and with all of the talented and intelligent people working on AI, I think we will be able to create AIs which we can use to advance not only our technology but our society and beliefs as well, and eventually become a better humanity.

Bibliography

Ananthaswamy, A. (2017, September 14). *AI spots Alzheimer's brain changes years before symptoms emerge*. Retrieved from New Scientist: https://www.newscientist.com/article/2147472-ai-spots-alzheimers-brain-changes-years-before-symptoms-emerge/

Chen, J. (2016). The Evolution of Computing : AlphaGo. *Computing in Science & Engineering*, 4-7.

Editorial board. (2017, September 18). *The night one Russian military officer may have saved the world*. Retrieved from Washington Post: https://www.washingtonpost.com/opinions/the-night-one-russian-military-officer-may-have-saved-the-world/2017/09/18/115edcbc-9c8f-11e7-9c8d-cf053ff30921_story.html?noredirect=on&utm_term=.80ae7c5b87e5

Ellis, S., & Hendler, J. (2017). Computers Play Chess, Computers Play Go...Humans Play Dungeons & Dragons. *Intelligent Systems*, 31-34.

Ford, M. (2013). Viewpoint: Could Artificial Intelligence Create and Unemployment Crisis. *Communications of the ACM*, 37-39.

Knight, W. (2017, October 3). *Forget Killer Robots - Bias Is the Real AI Danger*. Retrieved from MIT Technology Review: https://www.technologyreview.com/s/608986/forget-killer-robotsbias-is-the-real-ai-danger/

Macuga, T. (2017, August 23). *What is Deep Learning and how does it work?* Retrieved from Australian Centre For Robotic Vision: https://resources.rvhub.org/what-is-deep-learning-and-how-does-it-work/

Orwell, G. (1949). *1984.* London: Harvill Secker.

PR Newswire. (2017, August 10). *LifeFolder launches chatbot to help people with end of life planning in Indiana*. Retrieved from PR Newswire: https://www.prnewswire.com/news-releases/lifefolder-launches-chatbot-to-help-people-with-end-of-life-planning-in-indiana-300502442.html

ProPublica. (n.d.). *About us*. Retrieved from ProPublica: https://www.propublica.org/about/

Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence.* Penguin.

Temperton, J. (2017, October 4). *DeepMind's new AI ethics unit is the company's next big move*. Retrieved from Wired: http://www.wired.co.uk/article/deepmind-ethics-and-society-artificial-intelligence

The Economist. (2018). Special Report on AI. *The Economist*, 3-12.

Wagner, K. (2017, October 3). *Microsoft's Satya Nadella says artificial intelligence could create more jobs, not just eliminate them*. Retrieved from Recode: https://www.recode.net/2017/10/3/16412072/microsofts-satya-nadella-artificial-intelligence-jobs