

Emerging Writers

Volume 4

Article 4

5-2021

The Future of Artificial Intelligence

Alex Guerra

Follow this and additional works at: <https://digitalcommons.kennesaw.edu/emergingwriters>



Part of the [Computational Engineering Commons](#), [Computer Engineering Commons](#), and the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Guerra, Alex (2021) "The Future of Artificial Intelligence," *Emerging Writers*: Vol. 4 , Article 4.
Available at: <https://digitalcommons.kennesaw.edu/emergingwriters/vol4/iss1/4>

This Article is brought to you for free and open access by DigitalCommons@Kennesaw State University. It has been accepted for inclusion in Emerging Writers by an authorized editor of DigitalCommons@Kennesaw State University. For more information, please contact digitalcommons@kennesaw.edu.

The Future of Artificial Intelligence

by Alex Guerra

Introduction

Whether we like it or not, Artificial Intelligence (AI) is coming, and we are not ready for it. AI has unimaginable potential and will revolutionize the world over the next few decades, but with this great potential, we are faced with choices that could prove detrimental to humanity. Unlike many other human creations, there are some areas where there is absolutely no room for error. Humans invented machines to fly, but while we crashed many planes in the process, the effects of those crashes were limited. In the field of AI, there is very little room to have a crash because of the pure scale of AI's integration in our daily lives that is sure to come. One potential source of risk from AI integration into society stems from the AI thought process. AI thinks in a very different way than humans. While AI thought is, by definition, extremely logical, humans have many other variables at play when thinking. In an essay written by Stuart Russell, a professor of computer science at UC Berkeley and author of the book *Artificial Intelligence: A Modern Approach*, he discussed the popular paperclip example that is often used to describe this: "The well-known example of paper clips is a case in point: if the machine's only goal is maximizing the number of paper clips, it may invent incredible technologies as it sets about converting all available mass in the reachable universe into paper clips; but its decisions are still just plain dumb" (2015). This humorous example shows us some of the serious problems that humanity will face with AI and shows them in a realistic way as opposed to the

classic Hollywood AI military takeover. Since we know AI integration is coming, we are aware of the potential problems we must research, such as aligning AI's goals correctly with ours, preventing AI advanced warfare to the best of our ability, addressing possible AI mass collection and use of scarce resources, and halting the creation of any other potential disasters that could have been prevented with prior knowledge.

AI's Past and Present

AI's previous accomplishments.

It is important to understand the issues that will develop and also to recognize that since the beginning of AI experiments, AI computers have accomplished tasks that many assumed computers would never be able to achieve. Cade Metz describes a few of these accomplishments in a WIRED article where he discusses the loss of Lee Sedol in the game Go to an AI: "The win puts an exclamation point on a significant moment for artificial intelligence. Over the last twenty-five years, machines have beaten the best humans at checkers, chess, Othello, even *Jeopardy!* But this is the first time a machine has topped the very best at Go – a 2,500-year-old game that's exponentially more complex than chess and requires, at least among humans, an added degree of intuition" (2017). As we create machines that can play these games better than we can, our knowledge of how to develop the AIs themselves grows, and we move on to more complex and more applicable topics such as the field of medicine. In an article by Lohr, a practice run was analyzed for IBM's AI Watson that showed promising results:

At the University of North Carolina School of Medicine, Watson was tested on 1,000 cancer diagnoses made by human experts. In 99 percent of them, Watson recommended the same treatment as the oncologists. In 30 percent of the cases, Watson also found a treatment option the human doctors missed. Some treatments were based on research papers that the doctors had not read — more than 160,000 cancer research papers are published a year. Other treatment options might have surfaced in a new clinical trial the oncologists had not yet seen announced on the web (2016, paras. 23-24).

As recently as 2016, we found AI outperforming specialists in some fields of work. It is clear that we are close to a mass integration of AI into our society, and they will begin taking jobs.

The current power of AI

Already we can see AI mastering subjects that were thought to be uniquely human, from complicated gaming to driving. A particularly interesting AI is Google's DeepMind, specifically the subsection, AlphaStar. AlphaStar is an AI that plays the real-time strategy (RTS) game StarCraft II, which, put very simply, is a game where two people control an opposing base in real-time. The base makes workers, the workers collect minerals, and minerals are used to make more buildings and an army. The ultimate goal is the destroy the enemy's base and is done by obtaining an advantage through ways such as having a larger army, having a better army, or put in another way, countering your opponent's army with different troops. Although the game is over a decade old, it still has a flourishing professional scene with tournaments held

often where professionals compete for a prize pool. The Alpha Star Team posted an article that goes into large detail about what went right and what went wrong with the AI, how it was made, and how it ultimately defeated two of the best StarCraft II players: “In a series of test matches held on 19 December, AlphaStar decisively beat Team Liquid’s Grzegorz “MaNa” Komincz, one of the world’s strongest professional StarCraft players, 5-0, following a successful benchmark match against his team-mate Dario “TLO” Wunsch” (2019). A big obstacle described was that the AI had no sense of if it was winning or not in the development phase, so they had to implement a complicated system where it estimated its chances of success so it could act accordingly. This is an example of the difference between AI and humans and how we must find a way to work around the AI’s purely logical view. While a human would have a good idea of how to tell what they were doing, an AI does not have that gut feeling. As shown in the paperclip example in the introduction, we must find a way to align AI with our specific goals.

We are likely already on the cusp of these challenges as the integration of AI is underway. For example, AI has even begun driving our cars. A simple search of the Tesla website shows that you can buy a car with features such as auto lane change, auto park, the ability to move itself out of tight spaces, autosteer, and even full autopilot (2020). While Tesla specializes in this department, many other new cars come with mild AI features, such as automatic emergency braking and lane assist. It is quite realistic that there will come a time, possibly in the next century, where humans are banned from driving cars purely because people cause car accidents, but AI eventually will be much better than humans. In addition, we

are already beginning to see AI being integrated into our everyday lives in things like medicine, gaming, and driving.

AI's Near Future

Workplace Improvements

AI can already be found in some modern workplaces where they are being used to create a better and safer work environment, and every year that passes, they only get better at their job. A lot of the AI that is deployed into the field has built-in systems that increase its productivity. Take IFM/Onetrack.AI, for example; Onetrack.AI is an AI that deals with warehouse safety, as presented in an article by Mike Thomas (2020). The AI is contained in a small box with cameras and can detect a large array of safety violations such as cell phone use or where a forklift driver is looking while he is driving. Once detected, these "safety events" are sent to the warehouse manager so they can be dealt with appropriately. That is not all that the system provides though. As stated in the article:

From an image, we might be able to infer 25 signals today, but six months from now, we'll be able to infer 100 or 150 signals from that same image. The only difference is the software that's looking at the image. And that's why this is so compelling because we can offer a very important core feature set today, but then over time all our systems are learning from each other (2020, paras. 4).

AI that can augment itself is a big step in the development of AI and can show what the very near future can hold, but this brings us one step closer to the decisions that could have a huge impact on all of us.

General Improvements

AI will, as it develops, expand to a broader range of topics and professions and may eventually create an environment where humans are no longer required to work. While AI in the past was designed to accomplish a single task or achieve a single goal, CGP Grey talks about how AIs are being created with highly varying purposes, and it is doing a good job in many of these areas as described in one of his YouTube videos (2014). He brings the idea that for AI to take over job markets it does not have to be perfect, but instead, AI just has to be better than humans. AI is already proving that outperforming humans in their given task is not that hard, but AI does have to take over these professions. AI can be used instead to increase the efficiency of the people working there and work alongside them such as Onetrack.AI. CGP Grey jokes throughout the video saying, “But this job is safe right?” only to conclude that there is already AI in place or that is being developed that most people just haven’t heard of yet. AI, as with all technological inventions, should have the purpose of making our lives easier, and we must ensure that this is what happens.

AI Will Take Our Jobs

What and how many jobs will be taken?

AI already has begun to dominate many fields of work and taking jobs, and it is projected that many more jobs will be converted as well. In a WIRED article, a study from Redwood Software and Sapio Research, it is cited that 60 percent of jobs will become fully autonomous in the next five years (2018). Note that this was said in 2018. Now, this might be a bit ambitious, but it is worth considering. If anything, it would be a good exercise in thought to explore how we would react. As jobs that take up large parts of the economy, such as transportation or retail services, become autonomous, we need to know where the people can or will go. Although the research said so many of our jobs would be lost, it also concluded that new jobs will be made in the process.

What jobs will be created.

While it is completely true that AI will dominate some markets and completely remove the need for humans, by developing AI we will create more jobs in fields we cannot even begin to imagine. We observe this pattern in many technologies as we harness them, such as computers. It was thought computers would take jobs, but they ended up doing the exact opposite, creating jobs that no one would have guessed possible previously and making the jobs of virtually everyone significantly easier. Companies still feel this benefit today as we move away from paper and store our data virtually. One example of a job that we can imagine would be an AI safety inspector similar to the duties performed by The Machine Intelligence Research Institute (2013). The Machine Intelligence Research Institute is a non-profit organization that's goal is to seed research that would create safe and well restricted AI that

could prove productive in our society, some projects listed include friendly AI, stable brain emulations, and a model of AI risks. They reassure that they are not trying to enhance the opinion that AI might reach the level of intelligence of a human tribe and separate from the control of humans willingly, but instead that AI is intelligent in a different way, in which it is rewarded digitally for following its programming, much like the human reward system. This could lead to exponential expansion, and AI could steal all scarce resources in the name of completing their programmings such as solar, chemical, or nuclear energy. The work of The Machine Intelligence Research Institute is extremely important to the development of safe AI, and if we are going to make safe AI, we will need more organizations like it.

The Real Dangers of AI

AI mass production and the scarcity of resources.

One of the most realistic paths for an AI apocalypse is AI conquering the world, claiming all the resources, and leaving nothing for the humans, not out of spite, but in the name of efficiency. This could happen if AI were programmed to do something efficiently and not to keep humans alive in the process. In a study done with 979 people, the participants were asked if, by 2030, they believed AI would provide a net good for society as a whole or not. The results were that around two-thirds of the participants believed that AI would provide a net benefit to society (Anderson et al. 2020). This remaining third has a valid reason to believe that AI will turn out poorly by the next decade because there are many ways AI could be corrupted to create major problems for humanity. Kathleen Miles explores this possible doom by AI in an article

that analyzes Nick Bostrom's website, "There are two ways artificial intelligence could go, Bostrom argues. It could greatly improve our lives and solve the world's problems, such as disease, hunger, and even pain. Or, it could take over and possibly kill all or many humans. As it stands, the catastrophic scenario is more likely, according to Bostrom, who has a background in physics, computational neuroscience and mathematical logic" (2015). The catastrophic scenario that Miles talks about is much like the paper clip example; there is only so much metal on the planet, and if an AI uses all of it to make paperclips, it could send us back to the stone age. The United States of America has addressed this slightly in an official document that talks about the benefits of AI to the economy of The United States of America and a few different ideas that the government has on how the workforce should prepare for its integration. So the government is talking about how beneficial AI could be to the economy and is trying to address some of the problems now instead of later.

AI advanced warfare

War has plagued humanity since the dawn of man, with a constant arms race to be victorious over an opponent. It will only take one person to launch us into the age of AI warfare. If a perfect AI is created, it will inevitably be harnessed by a country for war, and this is something we need to prevent for as long as we can, because it could end with an AI versus AI war that may destroy the planet in the process. Max Tegmark, an American-Swedish physicist and researcher at MIT, goes into great depth on this topic in his TED talk, where he prominently discussed the threat of AI warfare Tegmark (2020). AI's potential to wage unimaginable wars

would only be countered by AI warfare. Tegmark plans to prevent or delay this for as long as possible. This is another example of the current push we have to keep AI in check, but it is not enough; we cannot gamble with AI and hope it comes out ok.

Alignment

The Importance of Alignment.

As mentioned, the most likely apocalypse to occur from AI will be from a scarcity of resources, and this ties deeply with alignment. Alignment is the difference between 1) making as many paperclips as you can or 2) make as many paperclips as you can, but do not kill all humans in the process. Scott Alexander published a fun numbered list on his blog:

1. If humanity doesn't blow itself up, eventually, we will create human-level AI.
2. If humanity creates human-level AI, technological progress will continue and eventually reach far-above-human-level AI.
3. If far-above-human-level AI comes into existence, eventually, it will so overpower humanity that our existence will depend on its goals being aligned with ours.
4. It is possible to do useful research now, which will improve our chances of getting the AI goal alignment problem right.
5. Given that we can start research now we probably should, since leaving it until there is a clear and present need for it is unwise (2015).

General-purpose AI will be created if we make it that far, and as Alexander says, we better figure out how to align its goals with us before that happens. The development of AI will be

exponential as AI is created to make more AI, and their intelligence rises far beyond any humans. We need to tell the paperclip AI that keeping humans alive to use the paperclips is just as important as making paperclips efficiently. We also need to tell the AI that there is such a thing as too many paperclips.

How do we align AI with our goals?

Alignment is as important as it is complex. There is no switch on an AI's black box that says "do not kill all humans" or a Boolean value to change in its code. Serious thought will be needed to ensure AI and humans coexist. Joe Petro, CEO at Nuance, a computer software company, takes a slightly different approach to alignment in his article published by MIT, "Even as a nascent technology, AI is incredibly complex and powerful, delivering benefits by performing computations and detecting patterns in huge data sets with speed and efficiency. But that power, combined with 'black box' perceptions of AI and its appetite for user data, introduces a lot of variables, unknowns, and possible unintended consequences" (2020). Petro is saying here that a large part of keeping AI aligned is to feed it the data that we want it to see. AI collects large amounts of data and bases its actions on it, so one possible solution is to monitor the inputs into AI. So, it appears that instead of fixing the code, instead, we must fix the way we treat the AI. This solution does have its limitations, as discovered by Natalie Wolchover and discussed in her article, "The most alarming example is one that affects billions of people. YouTube, aiming to maximize viewing time, deploys AI-based content recommendation algorithms. Two years ago, computer scientists and users began noticing that

YouTube’s algorithm seemed to achieve its goal by recommending increasingly extreme and conspiratorial content” (2020). The more inputs you put into the YouTube algorithm, the more it equalizes the extreme content with what you are interested in, but this initial spike is concerning. For AI to continue development, we need to monitor both the code and the inputs we put into it. While this is not a perfect solution to the problem, it is impossible for there to be one without further development and research because the future of AI is uncertain.

Conclusion

In conclusion, there are some clear pushes to keep AI in check and predict the future of humanity with AI; but there are clear holes where that are lacking research, funding, and support. We must raise more awareness for projects to ensure safe practices in the development of AI. Currently, companies are developing simple AI and the stakes are low, but as we move into the development of more and more complex AI, and especially general-purpose AI, a more cautious approach is needed. First, you must walk before you can run, and if development moves into complex development without enough effort put into safety, then society will trip and fall. For AI to be a good thing for humanity, we must ensure the proper safety measures are in place and people are educated on what you can input into an AI before technology moves forward, not after. You can show your support by funding research organizations and companies and signing open letters such as Tegmark’s <https://futureoflife.org/ai-open-letter/> (Tegmark 2018).

References

Alexander, S. (2015, May 29). [Web log post]. Retrieved September 21, 2020, from

<https://slatestarcodex.com/2015/05/29/no-time-like-the-present-for-ai-safety-work/>

Anderson, J., & Rainie, L. (2020, July 22). Artificial Intelligence and the Future of Humans.

Retrieved September 29, 2020, from

<https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/>

CGP Grey. (Director). (2014, August 13). *Humans Need Not Apply* [Video file]. Retrieved

September 21, 2020, from <https://youtu.be/7Pq-S557XQU>

Lohr, S. (2016, October 17). IBM Is Counting on Its Bet on Watson, and Paying Big Money for It.

Retrieved September 29, 2020, from

<https://www.nytimes.com/2016/10/17/technology/ibm-is-counting-on-its-bet-on-watson-and-paying-big-money-for-it.html>

Metz, C. (2017, June 03). Google's AI Wins Fifth And Final Game Against Go Genius Lee Sedol.

Retrieved October 15, 2020, from <https://www.wired.com/2016/03/googles-ai-wins-fifth-final-game-go-genius-lee-sedol/>

Miles, K. (2015, February 05). Will Artificial Intelligence Doom The Human Race Within The Next

100 Years? Retrieved September 22, 2020, from

https://www.huffpost.com/entry/artificial-intelligence-oxford_n_5689858

Petro, J. (2020, July 06). Beyond the AI hype cycle: Trust and the future of AI. Retrieved

September 30, 2020, from

<https://www.technologyreview.com/2020/07/06/1004823/beyond-the-ai-hype-cycle-trust-and-the-future-of-ai/>

Russell, S. (2015, January 01). 2015: WHAT DO YOU THINK ABOUT MACHINES THAT THINK?

Retrieved September 22, 2020, from <http://edge.org/response-detail/26157>

Tegmark, M. (2018, June 13). Benefits & Risks of Artificial Intelligence. Retrieved September 22,

2020, from <https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1>

Tesla. (2020, June 09). Autopilot and Full Self-Driving Capability. Retrieved November 10, 2020,

from <https://www.tesla.com/support/autopilot>

The Machine Intelligence Research Institute. (2013, July 22). Reducing Long-Term Catastrophic

Risks from Artificial Intelligence. Retrieved September 22, 2020, from

<https://intelligence.org/summary/>

The AlphaStar Team. AlphaStar: Mastering the Real-Time Strategy Game StarCraft II. (2019,

January 24). Retrieved September 22, 2020, from

<https://deepmind.com/blog/article/alphastar-mastering-real-time-strategy-game-starcraft-ii>

Thomas, M. (2019, June 8). The Future of Artificial Intelligence. Retrieved September 29, 2020,

from <https://builtin.com/artificial-intelligence/artificial-intelligence-future>

WIRED. (2018, April 04). AI and the Future of Work. Retrieved September 29, 2020, from

<https://www.wired.com/wiredinsider/2018/04/ai-future-work/>

Wolchover, N. (2020, January 30). Artificial Intelligence Will Do What We Ask. That's a Problem.

Retrieved September 22, 2020, from [https://www.quantamagazine.org/artificial-](https://www.quantamagazine.org/artificial-intelligence-will-do-what-we-ask-thats-a-problem-20200130/)

[intelligence-will-do-what-we-ask-thats-a-problem-20200130/](https://www.quantamagazine.org/artificial-intelligence-will-do-what-we-ask-thats-a-problem-20200130/)