Corbin Graham

Professor Kovach

English 250

21 May 2021

Inherently Evil, or Evil by Design?

Artificial Intelligence, AI, is one of the most prevalent forces impacting the lives of most Americans not only daily, but constantly. This could be seen as a good thing by some, those who may argue that the benefits of AI outweigh its negatives. Those who don't share the same opinion may be inclined to ask if AI has malicious intent, not by implemented design, but even flaws in design? If this were to be the case, could we as the humans creating it have any control over the outcome of AI or does it control itself so much internally that there is no hope for its redemption?

The concept of Artificial Intelligence has been around for many centuries. Automaton have long existed and been depicted as the next step in human evolution in much science fiction or even realistic fiction novels such as Brian Selznick's Hugo which explored the capabilities of a century old robot. The term Artificial Intelligence however wasn't coined until a paper written for machine learning systems in 1955 exploring the possibilities of self-regulation (McCarthy). Artificial Intelligence however didn't make a huge leap until computers could keep up with the high-level neural processing that it demanded.

AI has made its way from these humble beginnings to a necessary technology which most people heavily rely on. Today, AI is a key component in everything you use from social media, your weather predictions, and even some calculators (Marr). This reliance on AI has made it difficult for people to step back and look at it from an outside perspective, and ask, is AI really

truly beneficial to my life, or is it more negatively benefitting me? Advertisements are an area where many feel that AI may have crossed the line. So much of your personal data is available for these machines to learn from and observe patterns that it quite often knows what you want before you even do. This isn't by coincidence but because these machines have interacted with so much data, they are fed 'pointers' that make the connection between talking in one way with wanting something completely different. "The simplest explanation for why targeted ads are so creepily intimate: Your phone, your computer, and the internet in general contain a gargantuan amount of information about you. Google, for instance, knows essentially every website you have ever gone to in your life, and thanks to geolocation can tell where you live, where you work, and where you've traveled and when. Credit card companies know what you buy, and the brands that sell those items can use that data to predict the things you'll buy in the future — in Target's case, it can tell that you're pregnant before even your family knows" (Jennings). Companies create algorithms, or other complex math equations, that connect your feelings with your desires and then teach a machine to predict and understand the patterns which indicate which algorithm to use. Once a machine has become advanced enough, it has the capabilities to make these connections and form algorithms entirely on its own.

AI is formed from what are called neural networks and machine-learning algorithms. This sounds quite complex, and it can be, but in its purest form is quite simple. A program will have some goal to accomplish, and in order to accomplish this, they must first learn how to complete the task. This 'brain' will duplicate itself with the closest solution to the answer being the duplicate child so every 'generation' this 'brain' will grow smarter and closer to the solution. This isn't the most effective way of learning, but replicates natural selection in nature. This is the most common form of AI tendering and is the way most AI you see in your lifetime was

trained. Learning for these robots can closely resemble learning for young children. “the use of this technique demands a set of initial patterns that ideally should be learned automatically instead of from pre-defined models (which could annoy users and even make the process of learning habits without any bias difficult). Although other techniques have the same limitation, the inherent difficulty in reinforcement learning is interpreting user’s feedbacks; this is particularly important for reinforcement learning because this system is based mainly on the interpretation of this feedback (A.I)” (Aztiria 46). Self-learning AI is often bounded to a certain region of ‘thought’ and then given free reign for its own selection process to understand where the AI will begin to divulge from its initial tasks, or how quickly it may perfect one.

AI isn’t alone in its task of knowledge; many high-level software companies have in some form their own public version of their machine-learning algorithm. Variation in algorithms and purposes has given the world a large spectrum of AI. Those which the public can interact with have shown the most results in relation to corruption. “It's a joke, obviously, but there are serious questions to answer, like how are we going to teach AI using public data without incorporating the worst traits of humanity? If we create bots that mirror their users, do we care if their users are human trash? There are plenty of examples of technology embodying — either accidentally or on purpose — the prejudices of society, and Tay's adventures on Twitter show that even big corporations like Microsoft forget to take any preventative measures against these problems” (Vincent). If parents were given a random baby with complete anonymity towards this child and the outside world, with the ability to raise this child in any way possible without the chance of harming it, many parents would either neglect the child, or raise it in whichever way they gross the most entertainment. This same concept can be applied to open input AI such as Microsoft’s Tay. “Microsoft unveiled Tay — a Twitter bot that the company described as an

experiment in "conversational understanding." The more you chat with Tay, said Microsoft, the smarter it gets, learning to engage people through "casual and playful conversation." (Vincent).

These behaviors are concerning when its understood how many markets AI are making their way into. Very soon, national security will all lay in the hands and control of AI (Work). Theoretically, without bias, AI can perform much better and accurate computations when it comes to security than a human can, and it can be on complete standby 24/7 with no down time. These benefits strongly outweigh the negatives until potential problems are assessed. A technology that can well understand our society in a perspective which we may be incapable of perceiving, or in a way which we do not wish to perceive it, may raise its own conclusions about the world in which we live. This could be an issue if this technology is connected to a source which has the power to hurt humans. The U.S. government is currently developing AI assisted guided missile technology. These would be rockets launched by humans but fully controlled by an AI system until their detonation. Pentagon spokesperson, Jon Hill stated:

It's an important part of the future," Hill said. "With the kind of speeds that we're dealing with today, that kind of reaction time that we have to have today, there's no other answer other than to leverage artificial intelligence."

Every AI is different, but for the most part, it is the general public which decides the future of AI. Simply by having an online presence, you are unintentionally training AI. Google's AI is one of the most advanced in the world and the source behind reCAPTCHA. This technology tests different images with hard to distinguish features against Google's AI and if a human can get closer to the solution than the AI, then they have proven to google they are not a robot, and at the same time they have taught the robot what content was present in images. Google themselves have publicly stated that they do not understand how the AI functions and

that it has grown so advanced through self-learning that they could not replicate its knowledge through any program (Google Analytics).

AI has not yet proven itself to be inherently violent. In practice, it has always been by human error or human intention that these technologies have become violent. Have we gotten to the point however where we no longer are in control of the future of AI? In cases of Google's AI where its own developers no longer understand it, what is preventing it from using its power against its creator? While each side fights for what they believe to be right, and we will never know until the day we are faced with it, we can learn from the past. Traditionally, AI has been severely 'crippled' in its power because it is only assigned to complete a certain task or solve a certain algorithm. Most of these AI have seen excellent results and worked excellently for that which they were intended. However, we have also seen what could happen when a high-level computational AI can do when given more freedom from Microsoft's Tay. Even though this is an emerging market, it should be heavily watched and maintained like all other products or we may not see AI in such a positive way in the future.

Works Cited

Aztiria, A., Izaguirre, A., & Augusto, J. C. (2010, May 23). Learning Patterns in Ambient Intelligence Environments. Retrieved April 21, 2021, from https://link.springer.com/content/pdf/10.1007/s10462-010-9160-3.pdf

E., M.S. (Ed.). (2020, November 25). Artificial Intelligence Algorithms for Beginners. Retrieved April 21, 2021, from https://www.edureka.co/blog/artificial-intelligence-algorithms/#:~:text=Generally%2C%20an%20algorithm%20takes%20some,outputs%20when%20given%20new%20inputs.

Marr, B. (2019, December 16). The 10 Best Examples of How AI is Already Used in our Everyday Life. Retrieved April 21, 2021, from https://www.forbes.com/sites/bernardmarr/2019/12/16/the-10-best-examples-of-how-ai-is-already-used-in-our-everyday-life/?sh=25f464be1171

McCarthy, J. (1955, August 31). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Retrieved April 21, 2021, from http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

Nouri, S. (2021, February 03). The role of Bias in Artificial Intelligence. Retrieved April 21, 2021, from https://www.forbes.com/sites/forbestechcouncil/2021/02/04/the-role-of-bias-in-artificial-intelligence/?sh=3b201b89579d

Vincent, J. (2016, March 24). Twitter Taught Microsoft's AI Chatbot to be Racist in Less Than a Day. Retrieved April 21, 2021, from https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist