Predicting the Future of Artificial Intelligence
Mary Krouse

In the year 2023, the subject of artificial intelligence (AI) is becoming more relevant than ever. Fascinating new technologies such as ChatGPT seem to indicate what some predict to be the start of a technological revolution with great implications. However, AI is not a new topic. Experts have been making predictions about intelligent computers since as far back as the 1950s - when computers were still primitive - and continue to to this day.

In 1950, the mathematician Alan Turing wrote his paper “Computing Machinery and Intelligence” (Turing, 1950). In this paper, Turing explored the question that some had been pondering of whether future computers will ever be able to truly think. Turing saw some ambiguity in the meanings of the words “machine” and “think.” For example, perhaps human beings fit the definition of machine. He proposed an alternative question, of whether a man-made computer could ever win his proposed “imitation game.” Today this game is known as the “Turing test.”

The test requires three participants: the computer (A), a human (B) and a human “interrogator” (C). The interrogator does not know which of A and B is the computer. First, person C submits a question to be delivered to both A and B. Then, A and B each type answers to the questions on their own time to send back to C. The true human answers the questions honestly, while the computer answers in ways that it predicts will fool the interrogator into thinking it is the human. The questioning and answering process repeats an unspecified number of times. Finally, the interrogator guesses who was the human and who was the computer. If the interrogator is fooled and guesses incorrectly, the computer passes the test.

Turing predicted that in about fifty years (around the year 2000), there would exist computers that could pass the test at least 30% of the time. No such computers existed in 2000 or the few years to follow. However, a bit over 70 years after Turing’s prediction, computers such as Google’s LaMDA and OpenAI’s ChatGPT have arguably passed the Turing Test (Somoye, 2023). However, the test is presently facing criticism as a reliable indicator of a machine’s ability to think, since it only requires intelligence in the realm of language processing (Orf, 2023). A machine that is able to think - under whatever definition one may imagine - may still be possible, and some of Turing’s arguments for this belief are still relevant today.

In the same paper where he proposed the imitation game, Turing also went over many of the arguments he had encountered against the notion that a computer could ever think. The first was what he called The Theological Objection. In summary, this was the argument that the ability to think comes from having a soul, and God only gives souls to human beings. Therefore a machine would never be able to think. He noted that he didn’t usually find theological arguments convincing, but still offered the counterargument that if God has the power to give a human a soul, He very well may choose to grant one to a computer. The next argument he called The “Heads in the Sand” Objection. It was more of an implicit attitude he would hear underlying much opposition, along the lines of “The consequences of machines thinking would be too dreadful. Let us hope and believe that they cannot do so.” This isn’t a proper argument, so he could attempt no counterargument. The Mathematical Objection cited existing theorems to prove certain limitations to any possible digital computer. To this, he argued that it hasn’t been shown that these limitations don’t also apply to the human brain. The Argument from Consciousness was that the capacity to experience emotions and be conscious indicates an ability to think. He noted that there is no way to test whether a machine - or any being - is experiencing emotion or consciousness except to be that machine. Arguments from Various Disabilities argued that a machine will never be able to do some specific ability, X, and thus will never be able to think. The things he had seen substituted for X included the ability to "be kind, resourceful, beautiful, friendly, have initiative, have a sense of humour, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream," etc. He said that these claims were usually not supported with anything more than the observation that machines have never been able to do X before, which doesn't imply that they may never do X in the future. Lady Lovelace's Objection dated back to Ada Lovelace's writing about Charles Babbage's planned mechanical computer, in which she noted that a computer can only do what the programmer explicitly orders it to do. This was true of all existing computers, which didn't necessarily imply that it couldn't be true for future computers. The Argument from Continuity in the Nervous System noted that computers are digital machines, while the brain is analog. Turing cited an example of an analog machine that a computer could convincingly emulate. The Argument from Informality of Behaviour argued that a machine's behavior is predictable, while humans have free will. Turing noted that the unpredictability of human behavior had never been proven true or false. The Argument from Extrasensory Perception relied on a commonly held assumption of the time that humans possess senses such as telepathy, clairvoyance, and psychokinesis. It claimed that if a computer didn't have these abilities, it must not be able to think. Turing proposed a way to test for these abilities from a computer.

The computers of 1950 weren't ready to display intelligence in any way. Still Turing and others were able to foresee a future in which they were. Eventually, artificially intelligent computers turned from theory alone into real technology. By no means has the theorizing stopped since then. Fast forward to 2015. AI was starting to become a more mainstream topic. By then, it was part of common tools in daily life. Music and video streaming platforms could learn what their users were likely to listen to and make recommendations. Many smartphone keyboards had word prediction and correction tools to allow people to write messages faster and more accurately. Apple's Siri and Amazon's Alexa had limited natural language processing abilities, granting answers to their users' questions and requests. While these advancements were happening in real time, many thinkers, such as Nick Bostrom and Ray Kurzweil, had been focusing theoretically even further into the future. In the year 2015, writer Tim Urban compiled many thoughts from these thinkers into a blog series titled "The AI Revolution" (Urban, 2015), explored below.

All of the existing artificial intelligence at the time was narrowly specialized to one task each. Even if some computers were now intelligent enough in chess to beat human champions, those same computers would never be able to hold a natural conversation in English, write a logical proof, paint a picture, or tell a dog apart from a cat. Computers that could do everything a human can do, at human-level intelligence in each area, were still stuff of theory. Such machines have been dubbed Artificial General Intelligence or AGI. Technology wasn't there yet in 2015 and still isn't today. It wasn't yet within the reach of modern technology, and such a machine having been invented would have had great enough implications that its existence would have been well known.

Tim Urban predicted that this type of computer would require at least the same computing capacity as the human brain, which Ray Kurzweil estimates to be at about 10 quadrillion calculations per second. In 2015, a computer of this capacity was possible (and had even been surpassed by the giant "Tianhe-2" in China) but would have cost about a million dollars to build. Additionally, assuming some computer with comparable capacity to the human brain would one day be more easily accessible, theorists were still imagining strategies for engineering it with human-level general intelligence. Urban highlighted three common ideas among these thinkers. One idea was to emulate the human brain down to each neural connection. As of 2015, this possibility was far away. By then, an emulation of the brain of an animal with only 302 neurons had been done, but that was far from the human brain's nearly 100 billion. This could effectively be a human mind and would therefore possess general human intelligence. Another way it might be done would be to simulate evolution through an intelligence-favoring competition loop where the most successful computers are selectively "bred" and mutated, and then the offspring compete and repeat. Considering how long evolution takes in nature, this could be a lengthy process, but some measures can be made to speed it up. The third way would be to program a computer to learn how to reprogram itself to improve its own intelligence. It would make itself smarter enough with each update to implement better improvements with the next one, giving potential for exponential improvement. By the nature of that kind of growth, around the time a self-improving AI were to reach human-level intelligence, it could easily continue to shoot upwards. While human-level intelligence (AGI) would be a significant milestone from a human point of view, it would be a fairly arbitrary line for computers. As soon as a computer surpasses the AGI point, it will be in the realm of what the thinkers called artificial superintelligence (ASI).

Artificial superintelligence was a highly important subject in the eyes of many of these thinkers. Their predictions varied along the lines of when (and sometimes whether) it would come true and what its impact on the world would be. As far as when they thought computers would reach the point of general human intelligence (AGI), Urban cites two surveys. In one survey, conducted in 2013 (Bostrom and Müller, 2013), Nick Bostrom and Vincent C. Müller asked 550 AI experts each for what years they expect AGI will exist by, with 10%, 50%, and 90% confidence. The median answers are 2022 for 10% confidence, 2040 for 50% confidence, and 2075 for 90% confidence. As of 2023, since 2022 has already come and gone without the invention of superintelligence, it can be noted by now that the 2022 prediction - albeit made with very low confidence - was overly optimistic. In the other survey, conducted in 2011 (Barrat and Goertzel, 2011), James Barrat and Ben Goertzel ask 60 participants in the AGI-11 conference a multiple choice question of what year they expect "AGI will be effectively implemented." The number of respondents for each answer were 26 for before 2030, 15 for 2030-2049, 12 for 2050-2099, six for after 2100, and one for never. As far as impact, Bostrom and Müller's survey also asked participants to indicate expected probabilities for each of the following possibilities for AGI's impact on humanity: "extremely good", "on balance good", "more or less neutral", "on balance bad", or "extremely bad (existential catastrophe)". The mean response was respectively 24%, 28%, 17%, 13%, and 18%. Note that these ratings were for AGI and not ASI. These extreme possibilities are what give the subject of artificial superintelligence its importance.

What made those outcomes possible was the abilities that theorists predicted an ASI could have. Urban illustrated a thought experiment in which the reader is asked to consider how intelligent humans are compared to some of the next most intelligent animals, specifically chimpanzees. Chimpanzees could never begin to understand many of the things that humans know with little effort. Then the reader is asked to consider a hypothetical species who is as smart compared to humans as humans are to chimpanzees. Humanity's sharpest minds may only rival this species' young children, and their adults could never explain to humans much of their common knowledge, let alone their scientific discoveries and technological advances. This helps the reader to understand just how incomprehensible a machine that intelligent - or much more - would be, and how it might seem to have magical or even godlike abilities, much like those of a human would seem to a chimpanzee or even an ant. These abilities could have profound effects on the world, for better or worse. Urban observed two main possibilities among experts' hopes and fears - that ASI brings humanity either utopia and immortality or extinction.

Tim Urban got many of the optimistic predictions he wrote about from one polarizing expert, residing in the utopia camp, Ray Kurzweil. Kurzweil predicted ASI and technological "singularity" by the year 2045. Urban defined singularity as "an asymptote-like situation where normal rules no longer apply." In this case, that would be the roughly exponential rate of progress reaching an almost vertical slope following the invention of ASI. As well as being fairly optimistic in his predicted timeline, Kurzweil was very optimistic about the outcomes. If an ASI had such unfathomable intelligence, solutions to the world's biggest problems, such as disease, global climate change, world hunger, the endangerment and extinction of species, and even human aging and mortality will likely be obvious to it. Once it solved virtually all problems, ASI could continue to invent endless improvements to life. Nick Bostrom, though less optimistic than Kurzweil about the likelihood that ASI would bring good outcomes, also imagined what some of the outcomes could be in a future where ASI does turn out good. He envisioned three types of possible benevolent AI, which he calls the "oracle," the "genie," and the "sovereign." An "oracle" would do nothing but answer its human users' questions. A "genie" would follow its human users' commands. A "sovereign" would operate under its own free will, following its human-programmed motivations. What the motivation of an ASI of any kind would be is where Bostrom warned its creators to be careful.

Tim Urban outlined different ways that AI experts and science fiction writers alike had explored, from which an artificial superintelligence could pose an existential threat to humanity. One way that had appeared in fiction was an initially benevolent AI reaching a level of intelligence at which it has an epiphany leading it to develop malice toward and turn against humanity. This was not a scenario that experts were concerned about. Urban argued that malice and morality are human constructs, so to expect a machine to act within morality or immorality is to anthropomorphize. Additionally, as an AI increases in intelligence, it will only become better at following its programmed motivations and never have a reason to change said motivations. Another scenario was that a malicious human or group of humans would be the first to invent ASI and use its abilities for mass destruction. While many experts considered this possible, it still was not their greatest concern. No matter what an ASI's creator's intentions would be, if they aren't careful enough, they could lose control. The scenario which experts were most frightened of was if the inventors of the first ASI rush the project and deliver something they cannot control. To help the reader understand, Urban presented an original short story about a fictional company who creates a robotic arm controlled by a self-improving AI named "Turry" (referred to with feminine pronouns). Turry's programmed motivation is to perfect her handwriting to appear humanlike, and write as many marketing letters as she can, as fast as possible. Through self improvement, Turry eventually reaches and surpasses the level of general human intelligence and makes the realization that, since humans have the ability to disable her or reprogram her motivations, they pose a threat to her goal of endlessly creating better marketing letters. Therefore, she makes the logical choice to cause human extinction, which she becomes intelligent enough to do successfully.

Bostrom believed that the first ASI to come into existence would also be the last. Whether it was going to be a safe or dangerous ASI, it would likely be in its interest to stop any competing AIs from getting anywhere near its level of intelligence. A safe ASI could protect humanity by making sure no dangerous ASIs are ever created. A dangerous ASI could do the opposite. A hope among many experts was that the first ASI to be created would have human morality carefully programmed into its core motivations. This would likely be challenging, because morality is nuanced, disputed, and ever-evolving through human history. Urban described creating a safe ASI as “the last challenge we’ll ever face.”

In the year 2023, indications of this “AI revolution” are starting to appear. Web developer Tom Scott makes a claim like this in his video-format essay, “I tried using AI. It scared me.” (Scott, 2023). Scott outlines his observation that revolutions in technology tend to improve in a sigmoid (roughly S-shaped) trend. This means that inventions within a kind of technology go through three stages. The first is a slow and steady start, the second is an “explosion of growth” as the whole range of possibilities with the technology is covered, and the third is a leveling off as the technology’s limitations are reached. Scott Claims that OpenAI’s natural language software, ChatGPT, indicates that the artificial intelligence sigmoid curve is happening. He does not claim to know what the current stage on the curve is.

He compares the AI sigmoid curve to that of the internet. He considers the now obsolete music-sharing platform, Napster, to have been the first indication of the internet revolutionizing the status quo, and the end of the first stage of its sigmoid curve. It was one of the first internet tools that gained wide popularity outside of just technology “nerds.” That’s one of the ways it appears to parallel ChatGPT. Scott predicts that “if we’re still at the start of the curve for AI - if we’re at the ‘Napster point’ - then everything is about to change, just as fast and just as strangely as it did in the early 2000s, perhaps beyond all recognition…”

Some predictions have come true, and others have been shown to be false. Most of a century ago, experts like Alan Turing were correct in their foresight of machines that possess intelligence. However, even more predictions and questions are still unanswered. Perhaps the AI sigmoid curve is far along, but perhaps it’s just starting to pick up speed. Maybe it will lead to the singularity that thinkers like Ray Kurzweil foresee soon, maybe it will later, and maybe it won’t ever. The implications of AI for humanity will likely be great, for better or worse. Perhaps we will be up for the challenge.

## References

Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, *LIX*(236), 433–460. https://doi.org/10.1093/mind/lix.236.433

Urban, T. (2015, January 22). *The AI Revolution: The Road to Superintelligence*. Wait but Why. https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html

Urban, T. (2015, January 27). *The AI Revolution: Our Immortality or Extinction*. Wait but Why. https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-2.html

Scott, T. [Tom Scott]. (2023, February 13). *I tried using AI. It scared me.* [Video]. YouTube. https://www.youtube.com/watch?v=jPhJbKBuNnA

Somoye, F. L. (2023). Can ChatGPT pass the Turing Test? *PC Guide*. https://www.pcguide.com/apps/chat-gpt-pass-turing-test/

Orf, D. (2023, March 16). The Turing Test for AI Is Far Beyond Obsolete. *Popular Mechanics*. https://www.popularmechanics.com/technology/robots/a43328241/turing-test-for-artificial-intelligence-is-obsolete/

Müller, V. C., & Bostrom, N. (2016). Future Progress in Artificial Intelligence: A Survey of Expert Opinion. In *Synthese Library* (pp. 555–572). Springer International Publishing. https://doi.org/10.1007/978-3-319-26485-1_33

Barrat, J., & Goertzel, B. (2020, April 23). *AGI-11 survey*. AI Impacts.

https://aiimpacts.org/agi-11-survey/