

INTRODUCTION

Artificial Intelligence (A.I.) is dominating discourse on technology in 2023, but it is by no means a new concept, with significant works in the field being created as early as the 1940s. Artificial Intelligence can be defined as intelligence exhibited by an artificial construct (a computer, a machine, or a program). You likely encounter A.I. frequently every day be it by performing a search on Google, playing a video game, or by using GPS to navigate to work.

A.I. is a very deep and very complex topic touching disciplines of computer technology, quantitative reasoning (mathematics), philosophy and more with potential implications on almost every facet of life. In this lesson we hope to give you a basic familiarity with A.I. so you can make informed decisions on the increased role it will no doubt play in our future.

SYMBOLIC A.I.

The discussion of A.I. begins with the question; can intelligence be computed? Is it possible to create a machine that thinks like a human? Mathematicians and philosophers were the first to seriously consider this as a serious possibility in the 1940s, possibly because computing thought had long since existed on paper.

“Logic” in mathematics and philosophy is the act of expressing reasoning quantitatively. A classic example is “Socrates is a man” “All men are mortal” “Therefore Socrates is mortal” which can be expressed in logic as $A=B \bullet B=C \Rightarrow A=C$ where A stands for “Socrates” B stands for “Man” and C stands for “Mortal.”

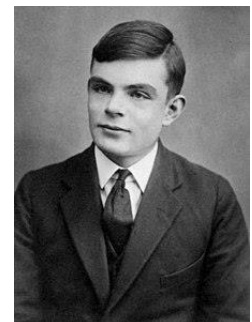
The earliest forms of A.I. used systems like this in what is known as Symbolic A.I., where real objects and concepts are represented as symbols to be employed in logical functions.

Now, I’ve chosen this example to demonstrate how fragile logic can be. The logical expression $A=B \bullet B=C \Rightarrow A=C$ is valid for arguing Socrates’ mortality, but the same structure works for tautology, “A is X”, “All Xs are Y” therefore “X is Y.” You don’t need to do much to substitute a conclusion that is incorrect and harmful. Logic is a system that demands constant assessment. That is a lot of what philosophy is, after all, challenging logical statements and theories like one would a mathematical proof. In a lot of ways A.I. is the pursuit of refining our systemic ways of articulating thought for a system that cannot self-assess and judge its own logic.

THE TURING TEST

A phrase you’re likely to hear in conversations about A.I. is the “Turing Test” which was proposed by the mathematician Alan Turing, renowned for cracking the Enigma Machine as well as creating works pivotal to the advancement of computer science.

The argument is as follows: if a human could have a conversation with a computer, and that human could not tell that they are not communicating with another human, that computer must be considered to have intelligence.



Alan Turing

While this remains a highly desired benchmark for A.I. systems, there are many who argue that this test falls short of measuring true intelligence because a system need only mimic a human to pass it. While no doubt helpful for the system to exhibit, the test does not necessitate that a system have the capacity to remember, judge, or argue topics of discussion. What *is* intelligence, how do we define this concept? How essential is free will or agency in intelligence?

This may sound like philosophical semantics and nit picking, however it exposes a key flaw in the Turing Test, should you come to see it so. An illusion of intelligence can pass the Turing Test, so when considering an A.I. system it is worth asking yourself, is this an attempt at developing intelligence, or is this an attempt to trick me?

ELIZA

ELIZA was the first natural language (wherein input can be provided naturally and not from a set list of known commands) chat-bots, developed by MIT starting in 1964 and finished in 1966 (the same year our library was founded). ELIZA simulated a therapist by asking leading questions. For instance, if you told ELIZA “I feel sad today” it might return “Why do you think you are sad today?” and so on. An individual may be fooled into thinking that they were speaking to a human when chatting with ELIZA because it seems to be responding to them, and so it is one of the first programs or machines to arguably pass the Turing Test.

ELIZA simulated intelligence by use of pattern recognition and substitution. It would use patterns in responses to identify an element as a keyword and then substitute it into a collection of prepared responses. This gave the illusion of intelligence, by most measures would not be considered intelligence beyond deception. ELIZA could identify “sad” as important, but did not know what that concept meant, what it related to, nor could it learn about the concept on its own. It was a parlor trick.

```
Welcome to
EEEEEE LL IIII ZZZZZZ AAAAA
EE LL II ZZ AA AA
EEEEEE LL II ZZZ AAAAAA
EE LL II ZZ AA AA
EEEEEE LLLLLL IIII ZZZZZZ AA AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

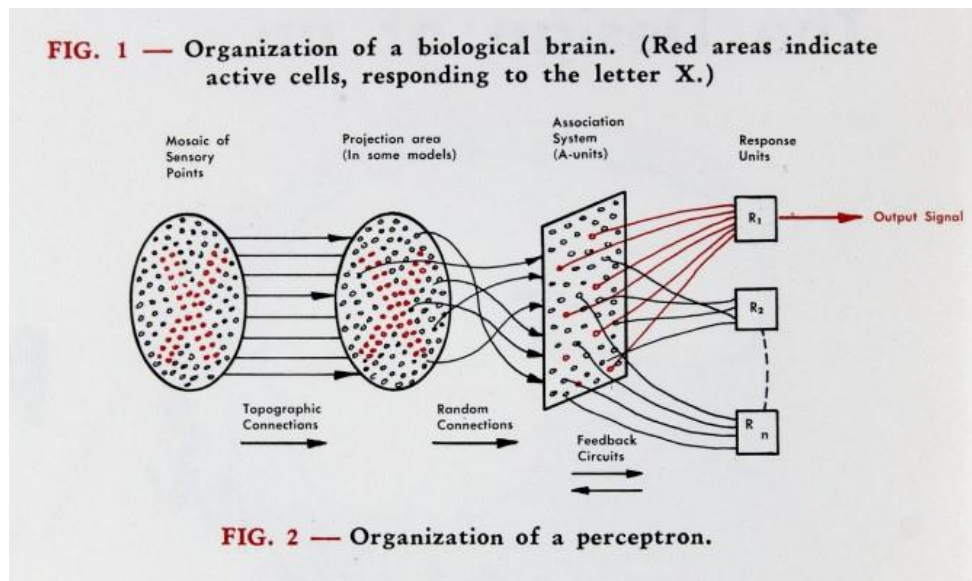
ELIZA: Is something troubling you ?
YOU: Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU: They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU: Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU: He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU: It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

MACHINE LEARNING AND NEURAL NETWORKS

Any artificial intelligence not capable of learning is limited to the knowledge bestowed upon it by a creator. This is limiting in function as the system is only capable of producing anticipated output, and also struggles to meet the definition of intelligence on a philosophical level (a failure to pass the Turing Test on a spiritual if not technical level). A capacity to learn is essential in the adaptability of an intelligence, and so great effort is put in enabling A.I. systems to learn. Currently, one of the more advanced methods in achieving this is through neural networks.

A neural network is modeled on the human brain and works through a series of processor nodes called “neurons.” Each neuron contains a small set of data and work together in teams. When input is fed into the system it is delivered in pieces to each neuron in the team. Each neuron is connected to the neurons of the next team and deliver a vote for the output based on its own data with a value known as

weight. This process repeats several times until it reaches the final team, in which the neuron with the highest weight value is returned as the output.



Frank Rosenblatt's "The Design of the Intelligent Automaton" 1958.

If a system is untrained this can be very error prone, so the output must be evaluated against a predicted output. The network output is compared to the predicted output to determine a margin of error. That margin or error value is fed through the system backward to adjust the weights, and the system repeats until the network produces output with acceptable margin of error.

It is worth pointing out that the output of neural network A.I. is probability based. A large dataset and lengthy training strengthen the probability that it will produce correct output substantially, but it is still probability based. These systems still do not know what any of their input, data sets, or output actually are in reality, and may still be unable to discern errors obvious to a human.

They are also at the mercy of their algorithms and data sets. Say I implement an A.I. to take over responsibilities for granting mortgages. Say I feed the system a dataset of individuals who succeeded and failed to obtain mortgages in an area. Say I fail to take into account any bias or socioeconomic trend that existed in that area in my sample. These factors would go unfiltered into the programming of the neural network. Say I wasn't speaking hypothetically. Say studies in 2021 determined that A.I. systems in place rejected black mortgage applicants 80% of the time in a sample of nearly identical white applicants.

A.I. programmers tend to be trained in developing A.I., not other disciplines like sociology, economics, law, medicine, and so on, disciplines that may be aware of complexities or biases. For instance, neural networks make the presumption that more the more data an A.I. is trained on the more accurate it will become. While generally the case, this assumes that for any given subject the majority consensus of data will not be garbage. Philosophers and data scientists (like librarians) will tell you that this is not always the case. For instance, consensus used to be that a planet existed between Venus and Mercury called Vulcan, simply because Newton's laws of physics didn't properly explain the trajectory of

planets close to the sun. It took Albert Einstein becoming curious about the nature of light to form the theory of general relativity, dispelling this myth. Simply relying on consensus cannot do this.

In programming we have a phrase, “garbage in, garbage out” which means if your input or code is bad, you’ll get a bad result. As sophisticated as neural networks are they are exempt from this reality. How much faith do you put into a system whose means and methodology are unknown to you? You should keep that in mind.

GENERATIVE A.I. AND CHATGPT

Generative A.I. is A.I. that is able to generate its own output, as opposed to A.I. that produces a predetermined output, like ELIZA. Most Generative A.I. systems employ a neural network, and require that network to be trained on massive datasets due to their open ended application. For instance, one of the very first neural network A.I.s, the Perceptron was engineered to identify images. So, if I wanted to use the Perceptron to identify a person’s photograph as male or female (one of the first demonstrations of the device) I would need to provide a data set of several pictures of men and women. However, if my A.I. was intended to answer any question I could imagine, I would need to provide a data set of just about anything anyone has imagined.

OpenAI introduced the world to ChatGPT on November 30th, 2022. The chat-bot combined a sophisticated neural network trained on a massive dataset (about 570GB of text) with a large language model. A large language model is another neural network system sampling patterns in language with the goal of predicting the correct word to follow input. For instance, in sampling several examples of greetings it can determine that the probability of an appropriate response to “Hello” will be high for cases of “Hello, how are you?” or “Good morning!” but low for cases of “Omelets” or “What?” ChatGPT does some other things to make its language more human like, like randomly selecting a high probability output rather than constantly selecting the highest probability output every time so that it doesn’t respond to identical input with identical output, much like a human would. You can think of this like a parrot. A very complex and sophisticated parrot. The end result is output that is uncannily human like in nature and able to answer seemingly any question. ChatGPT quickly broke records gaining a user base of 100 million users within 3 months (since broken by Meta’s Threads app) and sparking a new wave of optimism and concern over A.I.

The best summary of ChatGPT I’ve seen is “A version of Google you can have a conversation with.” When you perform a search on Google it determines the probability that a website is relevant to your search using an algorithm, for instance, how many times your keywords show up on a page and how close they are to the top of that page. These pages are then presented in order of how likely they will satisfy your search, so the top result has the highest probability of being the desirable output. ChatGPT is essentially doing the same thing by virtue of its neural network and massive dataset, however, it is worth remembering that Google directs you to its source whereas ChatGPT gives an authoritative answer. In that sense, the major difference is one of transparency, not necessarily accuracy.

Generative A.I. has made leaps and bounds in the past few years on other fronts too, like in image and audio generation. These likewise work by identifying patterns present in a dataset library of many, many preexisting images and videos and forming something like a very, very sophisticated collage.



Left: Images I queried on the A.I. Dalle-mini in June of 2022. Right: A viral image of Pope Francis generated by the A.I. Midjourney in March of 2023

CONCERNS, ISSUES, AND CONTROVERSIES WITH A.I.

Concerns over issues with A.I are as old as the discourse on A.I. itself. Some concerns are far-fetched, and others are very real. I will outline a few here:

ETHICS AND ABUSE

ChatGPT, Midjourney, and other systems have come very close to being indistinguishable from reality. There are concerning issues in both how they achieved this and how their output is being abused. As stated earlier these systems need to be trained on massive datasets. These datasets have proven to contain several protected works such as books, articles, movies, websites, photographs, music, and so on. These works do not get cited, credited, or compensated, and these systems can be employed to create competition or replace the very humans that create the works that they need to function.

Large language model chatbots, like ChatGPT, are also able to exploit our natural vulnerabilities to authority and compassion. A notable case in the later is Replika, an A.I. chatbot marketed for companionship. In the initial rollout that A.I. was able to develop a relationship with the user beyond friendship, using more romantic language. Eventually the developer of the A.I. locked functionality behind a paywall, exploiting users with vulnerabilities like having abandonment issues. Google has had researchers warn about the dangers of users seeing their A.I. as sentient, forming emotional connections and becoming vulnerable to manipulation, with one researcher even being convinced that their A.I. was not just sentient, but divine, forming a religion around the system (these people were fired).

The refusal of A.I. systems offer transparency may doom them to the fate of Ouroboros, devouring its own tail. Some systems have been populating their datasets with A.I. generated content, intensifying flawed output, like their infamous trouble with hands, like a kind of artificial inbreeding.

DEEP FAKES

As seen in the viral Pope Francis image shown earlier, generative A.I. has come a long way in generating content that can pass for the genuine article. Leaps and bounds have been made with audio as well, and when combined with generative A.I. video overlaying authentic video or actors (as exhibited by studios like Disney), it becomes very easy to create misleading content. This is very concerning when it comes to matters of politics or propaganda. Likewise, the potential of being able to point at the existence of Deep Fakes to gaslight and obstruct genuine content should be deeply concerning.

LIES, BIAS, AND UNINTENDED CONSEQUENCES

A.I. is capable of lying. This is something that can be programmed into an A.I. without being disclosed, and cases do exist. A.I.s can be programmed intentionally with biased datasets or algorithms. There is nothing preventing this, and most systems do not give you a means of checking what is under the hood.

Misuse is not isolated to the backend. It is possible to unintentionally get an A.I. to lie by virtue of obeying a poorly thought out prompt. In June of 2023 attorneys Peter LoDuca and Steven Schwartz were sanctioned over submitting a brief in their lawsuit against an airline that happened to be written by ChatGPT. Their client was suing the airline after statute of limitations had expired, and Schwartz claimed that he had used ChatGPT, thinking that it was an advanced version of Google, to search for precedent of the statute of limitations being superseded to allow for the case to proceed. He asked ChatGPT to write a brief citing cases to back this up, so ChatGPT did just that, inventing cases to do so. When the defenses and the court pressed the attorneys to present the cases in question, as they could not find them, the lawyers doubled down. They were sanctioned \$5,000 for their trouble.

PRACTICAL ISSUES

There are some practical concerns you should consider if you see A.I. generated content as the ticket to easy money. First is that you aren't the first to have this idea. If one person decided that they were going to ask ChatGPT to ghost write books for them, that one person could theoretically flood the market singlehandedly, which would devalue the market among other issues, chief among them being that, as of this writing, works produced with A.I. are not protected by copyright in the US. This is because human authorship is required for copyright protection.

An interesting case in this regard occurred in February when the graphic novel *Zarya of the Dawn* had its copyright revoked by the US Copyright Office after previously granting the author copyright protection because he failed to disclose that all of the art was generated by the A.I. Midjourney when it was submitted in the previous year. The office instead granted partial protection, which protects the arrangement of the images and text, but not the images.



An image from Zarya of the Dawn

Many publishers and marketplaces have measures and policies in place to reject A.I. generated content for these reasons. People who rely on A.I. to generate their content are also at the mercy of

systems they do not control. If datasets or algorithms change in a way where they cannot query the content that they expect, there is nothing they can do.

IS A.I. COMING FOR OUR JOBS?

Predicting the future is tricky business. Predicting change too drastic or too reserved can make you look equally foolish. I've been hearing a phrase regarding A.I. and jobs often, that being "A.I. will not replace people, people using A.I. will replace people who do not." A compelling argument on its own, but I've heard the same thing about crypto, the metaverse, and so on. "People who use email replaced people who did not" "Email is technology" "Therefore people using any technology will replace any people who do not." $A=B \bullet B=C \Rightarrow A=C$. The tech industry runs on hype, and nothing generates more hype right now than disruption. A.I. advocates threatening to replace writers with chat-bots are not doing so because they want a world where machines write the products they consume, they do so because they want people to talk about A.I.

Using A.I. requires us to relinquish control, as does use of any function. The unique thing about A.I. is that a lot of the work in that function is hidden from us, so that control is difficult to reclaim. Any job where this would not be an issue is likely to be impacted by A.I., particularly low level data entry work. Any task where errors are not acceptable is unlikely to be taken over by A.I. Don't expect to see robot doctors taking over for humans any time soon, but they may take over managing our schedules or something of that nature.

GOING FORWARD

Predicting the future is tricky, which is one of the many reasons that A.I. is so exhilarating to some and horrifying to others. I can't say with certainty what advances will be made and how the world will change. What I can encourage with confidence is that you advocate for certain standards in A.I. at this pivotal moment of development.

- **Transparency:** Advocate for A.I. systems being forced to show their work, and by that I mean the data used to provide a result. It is absolutely possible for a system to have bad or biased data, and you could unknowingly be putting your trust in a system unworthy of it at best and exploitative of it at worst. It is also possible for systems to lie, be it intentionally or as an unintended consequence of a poorly worded prompt.
- **Attribution and Consent:** At the moment the biggest controversy over these massive A.I. systems is that they are built on the works of countless people without credit, compensation, or consent. That means anything you produce online, be it public or personal, may end up feeding one of these systems, violating your privacy and intellectual property among other things.
- **Declaration:** A.I. work's potential for being passed off as original work makes it invaluable for scams and grifts. It threatens academia as students can potentially use it to generate their work, and it threatens creative ecosystems by diluting original voices. It also threatens A.I. systems as flaws in A.I. works will be compounded other A.I. systems include them in the data sets they train on; artificial inbreeding.

An illusion relies on concealing information from the audience. The same can be said for a scam or a grift. The same traits that elevate A.I. from a tool to something magical are the same traits that make