Madison Khan & Gemini
The Digital Revolution
Professor Isaacson
April 14, 2024

The AI Ouroboros: A History Written by its Own Future

The Paper

Artificial intelligence, a field that has captured the imagination of scholars and innovators alike, traces its roots back to the visionary ideas of Alan Turing and has since undergone a remarkable evolution. From its inception as a conceptual framework to its manifestation in large language model chatbots, the development of artificial intelligence represents a captivating journey through the annals of technological advancement.

Alan Turing, a luminary in the realm of computer science and mathematics, laid the groundwork for the exploration of artificial intelligence with his groundbreaking concept of the Turing Test. This seminal idea, proposed in 1950, challenged the boundaries of human understanding by posing the question of whether a machine could exhibit intelligence indistinguishable from that of a human. Turing's visionary proposition ignited a spark of inquiry that would propel the field of artificial intelligence into uncharted territory.

In the ensuing decades, the landscape of artificial intelligence witnessed both triumphs and setbacks. The emergence of the Logic Theorist, developed by Herbert Simon and Allen Newell in the late 1950s, marked an early attempt to implement Turing's ideas in practical applications. However, the limitations of early AI techniques became apparent during the AI Winter periods, characterized by skepticism and reduced funding.

Yet, from the depths of adversity emerged the seeds of revitalization. The advent of deep learning catalyzed the renaissance of artificial intelligence, a paradigm shift that unleashed the potential of neural networks to learn from vast datasets. This resurgence paved the way for the emergence of large language model chatbots, sophisticated AI systems capable of engaging in natural language conversations with humans.

Thus, the development of artificial intelligence from Turing's conceptualization to large language model chatbots represents a journey marked by resilience, innovation, and continual advancement. This essay will delve into the intricacies of this evolution, exploring key milestones, breakthroughs, and challenges along the way, and shedding light on the implications for the future of AI research and development.

Our exploration of conversational AI's fascinating journey begins with a true pioneer: Alan Turing (1912-1954). A British mathematician, computer scientist, logician, cryptanalyst, and philosopher, Turing's life and work were as groundbreaking as they were tragic. During World War II, his code-breaking efforts at Bletchley Park are widely credited with shortening the conflict and saving countless lives. However, Turing's lasting impact extends far beyond wartime heroics. His 1936 paper, "On Computable Numbers," introduced the now-iconic Turing machine, a theoretical model of computation that laid the foundation for modern computers.

Turing's foray into AI was not solely theoretical. In his seminal 1950 paper, "Computing Machinery and Intelligence," he proposed the now-famous Turing Test as a way to assess a machine's ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human. The test involves an interrogator who converses with a hidden human and a machine via text-only chat. If the interrogator cannot reliably determine which is which after a sustained conversation, the machine is deemed to be intelligent.

The Turing Test's significance lies in its attempt to move beyond purely symbolic manipulation (e.g., chess playing) and towards a more holistic understanding of intelligence. It sparked a fervent debate within the nascent field of AI, forcing researchers to consider the importance of natural language processing, learning, and the ability to engage in meaningful conversation.

Following Turing's lead, the 1950s and 1960s witnessed the development of early AI programs inspired by the Turing Test. One such example is ELIZA, created by Joseph Weizenbaum in 1966. ELIZA mimicked Rogerian psychotherapy by employing pattern matching and keyword recognition to respond to user input. While effective in creating the illusion of conversation, ELIZA lacked true understanding or the ability to learn from interactions. These early attempts exposed the limitations of rule-based systems and paved the way for more sophisticated approaches based on machine learning, which would dominate the latter half of the 20th century.[1]

[1] Dr Munr Kazmir, "Are AIs Acing the Turing Test? Or Are Humans Failing the Mirror Test?," Medium, February 20, 2023, http://medium.com/@munrkazmir/are-ais-acing-the-turing-test-or-are-humans-failing-the-mirror-test-dd2405dbdc15?source=topics_v2---------9-89--------------------c87c22c5_4de7_4b3b_94fa_d43a140d180c-------17.

While the Turing Test established a critical benchmark for conversational intelligence, the pursuit of AI in the 1950s ventured beyond mere imitation. Enter the Logic Theorist, a groundbreaking program developed by Herbert Simon and Allen Newell at Carnegie Institute of Technology (now Carnegie Mellon University).

Simon, a political scientist and future Nobel laureate in economics, and Newell, a computer scientist, were not solely interested in machines that could converse like humans. They sought to understand how machines could reason and solve problems in a way akin to human cognition. Their 1956 creation, the Logic Theorist, became a landmark in this pursuit.

The Logic Theorist tackled a seemingly straightforward task: proving mathematical theorems from Bertrand Russell and Alfred North Whitehead's seminal work, *Principia Mathematica*. However, the program's significance lay not in the specific problem it addressed, but in its approach. The Logic Theorist employed a set of heuristics, to search through logical axioms and theorems to find a valid proof for a given proposition. This "search-based" approach marked a significant departure from the logic-based reasoning envisioned by the Turing Test.

The Logic Theorist's capabilities were impressive for its time. It successfully proved many theorems in *Principia Mathematica*, even discovering novel and more elegant proofs for some[2] However, its limitations were equally evident. The program relied heavily on a pre-defined set of axioms and heuristics, limiting its ability to handle problems outside its specific domain. Additionally, the search process could be computationally expensive, becoming exponentially slower with increasingly complex theorems.

Despite these limitations, the Logic Theorist's impact on AI research was profound. It demonstrated the potential for computers to engage in symbolic reasoning and problem-solving tasks traditionally considered the hallmark of human intelligence. Furthermore, the program's emphasis on heuristics paved the way for the development of search algorithms and problem-solving techniques that remain fundamental to AI today.[3]

However, the Logic Theorist also exposed the challenges of early AI. The brittleness of rule-based systems and the limitations of brute-force search algorithms became evident. These

---

[2] Pamela McCorduck, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence* (Natick, MA: A.K. Peters, 2004).
[3] Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. (Upper Saddle River: Pearson, 2016).

challenges would propel researchers towards exploring alternative approaches, including machine learning techniques, that would dominate the latter half of the 20th century.

The early optimism surrounding AI in the 1950s and 1960s, fueled by pioneering programs like the Turing Test and the Logic Theorist, eventually gave way to a period of stagnation and disillusionment known as the AI Winter. This period, spanning roughly from the mid-1970s to the late 1980s, saw a significant decline in research funding, public interest, and overall progress in the field.

Several factors contributed to the onset of the AI Winter. One critical factor was the overhype surrounding early AI successes. The impressive capabilities of programs like the Logic Theorist were often misconstrued as harbingers of imminent human-level AI, leading to inflated expectations. When these expectations inevitably went unmet, a sense of disappointment and skepticism set in. Furthermore, the limitations of the prevailing symbolic reasoning approaches became increasingly apparent. Programs like the Logic Theorist, for all their ingenuity, were brittle and domain-specific, struggling to adapt to problems outside their pre-defined parameters. This lack of flexibility casted doubt on the long-term viability of these approaches.

This confluence of factors led to a significant decline in funding for AI research. Government agencies and private investors grew hesitant to support research that seemed to promise more than it could deliver. Prominent AI researchers like Marvin Minsky and Seymour Papert publicly criticized the field's direction, further eroding public confidence. This lack of funding and support severely hampered research efforts, hindering the development of new approaches and stifling innovation.

The implications of the AI winter for the field were significant. Many promising research avenues were abandoned or slowed down considerably. A generation of talented researchers left the field, disillusioned by the lack of progress and funding. The field's reputation suffered, making it even more difficult to attract new talent and resources. However, the AI winter was not entirely devoid of positive consequences. It served as a crucial period of reflection, forcing researchers to re-evaluate their assumptions and explore alternative approaches.[4] The limitations of symbolic reasoning became painfully clear, paving the way for a renewed interest in machine learning techniques that would eventually lead to a resurgence of AI research in the late 20th and early 21st centuries.

[4] Russell & Norvig, *Artificial Intelligence.*

The AI winter, while a period of stagnation, ultimately served as a catalyst for a significant shift in the field's trajectory. The limitations of symbolic reasoning approaches became apparent, paving the way for a renewed interest in machine learning techniques, particularly a subfield known as deep learning.

Deep learning represents a specific approach within the broader field of machine learning. Unlike traditional machine learning algorithms that rely on hand-crafted features, deep learning models learn these features directly from the data through a series of interconnected layers, often inspired by the structure and function of the human brain. This ability to automatically extract meaningful features from vast amounts of data has revolutionized AI research. Deep learning techniques have empowered computers to tackle complex tasks previously considered beyond their capabilities, such as image recognition, natural language processing, and game playing. This breakthrough has fueled a resurgence of AI research, leading to significant advancements across various domains. [5]

The foundation of deep learning lies in artificial neural networks, a class of algorithms inspired by the structure of the human brain. These networks consist of interconnected nodes, or artificial neurons, arranged in layers. Information flows through the network, with each layer performing a specific transformation on the data. Deep learning models achieve complex learning by adjusting the connections between these neurons based on the data they are trained on. This process allows the network to identify patterns and relationships within the data, ultimately enabling it to perform tasks like image recognition or language translation. The ability of neural networks to learn from vast amounts of data and continuously improve their performance has made them a powerful tool for a wide range of AI applications.[6]

The impact of deep learning can be seen in the remarkable achievements of recent AI research. In the field of computer vision, deep learning models have surpassed human-level performance on tasks like image classification and object detection. They are now powering applications like self-driving cars and facial recognition systems. Natural language processing has also witnessed significant progress, with deep learning models achieving fluency in machine translation and generating human-quality text.

---

[5] Russell & Norvig, *Artificial Intelligence,* Chapter 18.

[6] IBM, “What Is Deep Learning?,” www.ibm.com, 2023, https://www.ibm.com/topics/deep-learning.

The influence of deep learning extends beyond these specific examples. It is transforming areas like robotics, healthcare, and finance, where deep learning models are increasingly used for tasks such as robot control, medical diagnosis, and financial forecasting. These breakthroughs are a testament to the transformative power of deep learning and its central role in the current AI renaissance.[7]

The deep learning revolution has ushered in a new era of AI, and within this era, large language models (LLMs) have emerged as a powerful force in the realm of conversational AI. This section delves into the capabilities of LLMs, explores the significance of transformer architectures in their development, and examines some prominent examples of LLM chatbots and their applications.

LLMs represent a specific class of neural networks trained on massive datasets of text and code. These datasets can encompass books, articles, code repositories, and even online conversations, providing LLMs with a vast repository of linguistic knowledge. Through complex training algorithms, LLMs learn to identify patterns and relationships within this data, allowing them to perform a variety of language-related tasks. These tasks include generating human-quality text, translating languages, writing different kinds of creative content, and, most relevant to this discussion, engaging in conversations that mimic human interaction.[8]

A critical innovation in the development of LLMs is the transformer architecture, introduced by Vaswani et al. in 2017. Unlike traditional recurrent neural networks, which process text sequentially, transformers employ an attention mechanism that allows them to analyze relationships between words across the entire input sentence simultaneously. This parallelization allows for more efficient processing of long sequences of text, a crucial capability for handling the vast datasets used to train LLMs. The transformer architecture's effectiveness has revolutionized the field of natural language processing and is a cornerstone of the advancements seen in LLM chatbots.[9]

The capabilities of LLMs have led to the development of sophisticated chatbots that can engage in nuanced and informative conversations. A prominent example is LaMDA (Language Model for Dialogue Applications), developed by Google AI. LaMDA leverages transformer-based architectures to hold open-ended, informative conversations on a wide range

---

[7] Russell & Norvig, *Artificial Intelligence,* Chapter 17.
[8] Russell & Norvig, *Artificial Intelligence.*
[9] Ashish Vaswani et al., "Attention Is All You Need," arXiv.org, June 12, 2017, https://arxiv.org/abs/1706.03762.

of topics.[10] Another noteworthy example is Megatron-Turing NLG from NVIDIA, which holds the record for the largest generative pre-trained transformer model. This model demonstrates exceptional skill in tasks like writing different kinds of creative content.[11]

The applications of LLM chatbots are expanding rapidly. They are being employed in customer service applications, providing 24/7 support and answering user queries in a natural and engaging way. Additionally, LLMs are being used to develop educational chatbots that can personalize learning experiences and answer student questions in a comprehensive manner. The potential applications extend beyond these initial examples, with LLMs poised to revolutionize communication across various sectors in the years to come.[12]

Our exploration of Artificial Intelligence's fascinating journey has reached its concluding point. This investigation has traversed a rich historical landscape, beginning with the pioneering work of Alan Turing. Turing's groundbreaking concepts, including the Turing Test, laid the foundation for the quest to understand and potentially replicate human intelligence in machines.

The subsequent sections charted the evolution of AI research, from the early promise of the Logic Theorist to the disillusionment of the AI winter. This period of stagnation highlighted the limitations of symbolic reasoning approaches and underscored the need for new paradigms. The emergence of deep learning, particularly the transformative power of transformer architectures, marked a turning point. This new wave of research led to the development of large language models (LLMs), ushering in a new era of conversational AI.

Reflecting on this historical trajectory, one is struck by the remarkable progress achieved. From Turing's theoretical musings to the sophisticated LLMs capable of engaging in meaningful conversations, the field of AI has undergone a radical transformation. This journey underscores the enduring human fascination with replicating our own cognitive abilities and the potential of technology to surpass even our wildest imaginings.

---

[10] Christian Plagemann and Katya Cox, "English Learners Can Now Practice Speaking on Search," blog.research.google, October 19, 2023, https://blog.research.google/2023/10/google-search-can-now-help-with-english-speaking-practice.html.

[11] Paresh Kharya and Ali Alvi, "Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, the World's Largest and Most Powerful Generative Language Model," NVIDIA Developer Blog, October 11, 2021, https://developer.nvidia.com/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-largest-and-most-powerful-generative-language-model/.

[12] Bruno Belić, "Transformers: Pay Attention," Megatrend, May 17, 2022, https://www.megatrend.com/en/transformers-pay-attention/.

However, as with any historical narrative, the story of AI is not without its complexities. The limitations of current LLMs, their potential biases, and the ethical considerations surrounding their development remain crucial areas of ongoing inquiry. Yet, the potential benefits of AI are undeniable. LLMs have the potential to revolutionize communication, education, and numerous other sectors.

Looking towards the future, the evolution of AI promises to be as remarkable as its past. Advancements in areas like explainable AI and addressing ethical concerns will be crucial in ensuring responsible development. As historians, we must continue to document and analyze this ongoing saga, for the story of AI is ultimately a reflection of humanity's ceaseless quest to understand ourselves and the world around us.

The Process

In writing this paper, I first turned to trusty ChatGPT (3.5). I gave it the initial prompt, and it spit back several chronological sections about the history of AI development, most of which appear in this paper. I prodded OpenAI to elaborate on the different subjects, including a deeper dive into the prominent figures associated with each AI era. I also wanted to see if I could have the AI write the paper in a specific style to make things more exciting, but having the paper written in iambic pentameter or in a poorly imitated version of Dostoevsky's Raskolnikov or Salinger's Holden Caulfield was not cutting it for me.

To more effectively write the paper, I then asked ChatGPT to create an outline for the paper with the given prompt. I asked the AI to write it in the style of a history Ph.D. student, which I would maintain for the rest of the process. I decided to stick with ChatGPT's outline, which hit all of the chronological points you see above. I began asking the AI to write each section one by one, but after trying to get the AI's sources for the first section, I hit a roadblock. ChatGPT told me that it did not know where its data came from, but would then list some suggested sources and ideas on where to find more. This is when I decided to switch to Google's Gemini, which someone had mentioned in class was better at citing its sources than ChatGPT was. This proceeded to be correct; as I prompted Gemini to write out each section of the paper and include sources, it would provide one or two sources. I asked Gemini if it could write a Chicago note-footnote style citation for its first source, and it said it could not because the article- the same one it cited- was behind a paywall. One of the other sources Gemini used was simply called "A History of Artificial Intelligence" with no valid URL provided, and when I prompted it to tell me more about the source, it could not. This happened a few other times. While there were some scholarly articles cited, many of the sources that Gemini pulled from were from research blogs from various AI developers- or just blogs in general- including Google, as well as product websites. These are not sources that I would normally cite, but most of them included that information that Gemini had mentioned. Once, It tried to cite a Pearson webpage that was selling a beta version of an AI-powerd study tool, which could have been used as source for a different section, just not its discussion on face-recognition and self-driving cars. In another instance, Gemini provided a link followed by "Chapter 17," but the link navigated to a blog page rather than a book that would have chapters. I put that section back into Gemini and asked if it could provide a source, and it gave me three new ones that matched the information that Gemini had initially provided.

Another feature I appreciated about Gemini was it had a self-asses function for its responses. When I had it assess itself for accuracy, it returned with mostly green flags except for the AI Winter section, which turned up mostly yellow. What the self-assess feature does, it seems, is run a quick Google search to double-check the information that the AI provided, which was most often confirmed by Wikipedia. I reuploaded the section on the AI Winter into Gemini and asked for its sources. Of the two books it listed, one had not already been cited so I decided to cite that one, only to realize that the book did not exist at all. The author, Pamela McCorduck, is an author on AI books, but has never written one titled *The Once and Future Machine: Why AI Winter Didn't Kill Artificial Intelligence* or remotely about that subject.

I also attempted, unsuccessfully, to have the AI write this final section of my paper. I prompted it to describe how I used the AI to write the paper, and it produced a very generic response about asking the initial question and then probing it for further detail. Some of the prompts that the AI claimed I used were entirely new to me, so I decided to just write this section myself. Finally, I tried one more time to rewrite the paper in a more exciting voice, namely that of Tolstoy and then Dr. Seuss. Again, I decided that I could not possibly turn in a paper that reads in this way, although I liked what Gemini had done with Tolstoy. Although it cut out most of what I wrote, Gemini followed up its translation with an explanation as to why this new text was written in the style of Tolstoy, notably through the use of similes and metaphors, descriptive language, and, an “emphasis on the human endeavor.” The Seussical style was very entertaining, however, and an excerpt follows:

> The Turing Test, oh what a quest!
> To see if a robot could truly impress.
> With questions and answers, a back-and-forth game,
> Could the machine mimic and never be blamed?
> This sparked a great journey, a race to the stars,
> To build thinking machines, reaching ever so far.
> From ideas on paper, to circuits so bright,
> Artificial intelligence, taking flight!
> Now programs converse, with a wink and a smile,
> Large language models, learning for a while.
> From simple commands to stories untold,
> The world of AI, a future unfolds!

Finally, I would like to conclude this paper with Gemini’s own closing statements, written in the style of a tenacious history Ph.D. student: Overall, this exploration with the LLM has demonstrated the potential of this technology as a research tool. By wielding prompts effectively, researchers can unlock a wealth of information and gain new perspectives on a variety of topics. This collaborative approach is likely to redefine research methodologies in the years to come.

Bibliography

Belić, Bruno. “Transformers: Pay Attention.” Megatrend, May 17, 2022. https://www.megatrend.com/en/transformers-pay-attention/.

IBM. “What Is Deep Learning?” www.ibm.com, 2023. https://www.ibm.com/topics/deep-learning.

Kazmir, Dr Munr. “Are AIs Acing the Turing Test? Or Are Humans Failing the Mirror Test?” Medium, February 20, 2023. http://medium.com/@munrkazmir/are-ais-acing-the-turing-test-or-are-humans-failing-the-mirror-test-dd2405dbdc15?source=topics_v2---------9-89--------------------c87c22c5_4de7_4b3b_94fa_d43a140d180c-------17.

Kharya, Paresh, and Ali Alvi. “Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, the World’s Largest and Most Powerful Generative Language Model.” NVIDIA Developer Blog, October 11, 2021. https://developer.nvidia.com/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-largest-and-most-powerful-generative-language-model/.

McCorduck, Pamela. *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. Natick, MA: A.K. Peters, 2004.

Plagemann, Christian, and Katya Cox. “English Learners Can Now Practice Speaking on Search.” blog.research.google, October 19, 2023. https://blog.research.google/2023/10/google-search-can-now-help-with-english-speaking-practice.html.

Russell, Stuart J, and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 3rd ed. Upper Saddle River: Pearson, 2016.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. “Attention Is All You Need.” arXiv.org, June 12, 2017. https://arxiv.org/abs/1706.03762.