

Maeve Patton
HISU-2900
April 14, 2024

Part 1: The Original Outline from Gemini

Key

Red text: Incorrect or unfounded sources and information

AI: From the Turing Test to Chatbots - An Outline

I. The Seeds of Artificial Intelligence (1930s-1950s)

- Alan Turing (1912-1954): Introduce the concept of the Turing Test (1950) as a benchmark for machine intelligence.
- Claude Shannon (1916-2001): Pioneered Information Theory, laying the foundation for measuring and manipulating information crucial for AI.
- Early AI Programs: Highlight overlooked figures like:
 - **Edith Leyton** (1925-2009): Developed one of the first chess-playing programs in the 1950s.
 - **A. D. Williams** (1910-1968): Built the Ferranti Pegasus, an early computer used for AI research.

II. The Rise of AI Research (1960s-1980s)

- John McCarthy (1927-2011): Coined the term "artificial intelligence" at the Dartmouth Workshop (1956) and championed early AI research.
- Marvin Minsky (1927-2016): A leading figure in early AI research, particularly in neural networks.
- The Perceptron Controversy: Discuss the limitations identified by Minsky and Papert (1969) that led to a decline in neural network research.
- Expert Systems: Focus on the development of knowledge-based systems for specific domains, like DENDRAL in chemistry.
- The Rise of Symbolic AI: Explore the dominance of logic-based approaches in AI during this period.

III. Rethinking AI: New Approaches Emerge (1980s-2000s)

- The AI Winter: Discuss the funding and research limitations in AI during the 1980s and 1990s.
- Connectionism Makes a Comeback: Highlight the resurgence of neural networks research with figures like Geoffrey Hinton and David Rumelhart.

- Machine Learning Takes Center Stage: Explore the rise of machine learning algorithms like decision trees and support vector machines.
- The Birth of Chatbots: Discuss early chatbot programs like ELIZA (1966) and PARRY (1970).
- The Role of Women in AI: Include figures like **Judea Pearl (Turing Award winner 2012)** and Fei-Fei Li (pioneer in computer vision) who made significant contributions during this period.

IV. The Age of Deep Learning and Large Language Models (2000s-Present)

- Deep Learning Revolution: Discuss the breakthroughs in deep neural networks, like AlexNet (2012) for image recognition, and their impact on AI.
- The Rise of Big Data: Explore the role of vast amounts of data in training complex AI models, including large language models.
- Large Language Models (LLMs): Introduce the concept of LLMs and their capabilities in text generation, translation, and conversation.
- Chatbots Powered by LLMs: Explain how LLMs are revolutionizing chatbots, enabling more natural and human-like interactions.
- Ethical Concerns: Discuss the challenges of bias, fairness, and explainability in AI development with specific examples in chatbots.

V. The Future of AI and Chatbots: Where Do We Go From Here?

- Potential applications: Explore the future applications of chatbots powered by LLMs in various fields like customer service, education, and healthcare.
- Human-AI Collaboration: Discuss the potential for human-AI partnerships and the evolving role of chatbots alongside human interaction.
- Open Questions and Challenges: Highlight ongoing research areas like interpretability of AI decisions and responsible development.

Part 2: The Essay

Key

Red text: Incorrect or unfounded sources and information

Green text: Human edits I found necessary for clarification

Highlighted in yellow: Words and phrases used repeatedly by the models

The Diverse Tapestry of Artificial Intelligence: Overlooked Voices, Ethical Challenges, and the Path Towards an Equitable Future

The quest for artificial intelligence (AI) has captivated minds since the inception of modern computing, fueling a relentless pursuit of machines capable of emulating human cognition. This odyssey, spanning decades of groundbreaking research and pioneering contributions, has been shaped by a rich tapestry of visionaries, often overshadowed by the dominant narratives that have traditionally focused on a select few. As we delve into the annals of AI's evolution, it becomes imperative to shed light on the overlooked figures whose contributions, despite their **profound** impact, have been obscured by the tides of history.

The paper delves into the evolution of AI, tracing its trajectory from the **foundational** work of Alan Turing and Claude Shannon to the advent of large language models. It highlights the contributions of overlooked figures throughout different eras, shedding light on their impact on the field. The paper also examines the ethical challenges surrounding bias, fairness, and explainability in AI development and the importance of diversity and inclusivity in addressing these issues.

The evolution of artificial intelligence, from Alan Turing's theoretical **foundations** to the advent of large language models, has been driven by a diverse array of visionary minds, many of whom have been overlooked due to their race, gender, or identity. Recognizing and embracing these diverse contributions is not only a matter of historical accuracy but also a critical step toward addressing the ethical challenges of bias, fairness, and explainability in AI development, ultimately ensuring that the advancements in this field benefit all of society equitably.

Alan Turing (1912-1954) remains an unparalleled figure in the annals of computer science and artificial intelligence, his legacy blending brilliance with **profound** personal tragedy. At the forefront of Turing's contributions lies the conceptual edifice of the Turing Machine, expounded upon in his **seminal** 1936 treatise "On Computable Numbers, with an Application to the Entscheidungsproblem." This theoretical construct, capable of symbol manipulation based on predefined rules, laid the cornerstone for understanding computation.¹ Turing's demonstration of the Halting Problem, which showcased the

¹ Turing, Alan. "On Computable Numbers, with an Application to the Entscheidungsproblem."

Proceedings of the London Mathematical Society 42, no. 2 (1937): 230-65.

existence of unsolvable problems, fundamentally reshaped our comprehension of computability and distinguishing between tasks that are computable and those that are inherently undecidable.²

Beyond the realm of theory, Turing's wartime endeavors at Bletchley Park, particularly his pivotal role in decrypting the Enigma code, stand as enduring testaments to his practical genius. Despite the monumental significance of his contributions, much of his wartime efforts remained shrouded in secrecy for decades, obscuring his broader impact.³

Tragically, Turing's life was marred by personal adversities. His struggles with social interaction and the stigma of homosexuality in 1950s Britain culminated in his untimely demise death at the age of 41, under circumstances strongly suggestive of suicide. Despite his towering intellect, Turing's story serves as a poignant reminder of the societal prejudices that can cast shadows even over the most brilliant minds.⁴

Transitioning from Turing's foundational work, the narrative of AI's evolution converges with Claude Shannon (1916-2001), a figure whose contributions often dwell in the shadows. Shannon's seminal 1948 treatise, "A Mathematical Theory of Communication," heralded the advent of information theory, offering a unified framework for analyzing information transmission and storage.⁵ His introduction of the binary digit, or "bit," revolutionized digital communication and computation, serving as the bedrock of modern computing architectures.⁶

Moreover, Shannon's insights into error-correcting codes ensured robust data transmission in the face of noise, while his pioneering endeavors in digital logic circuits laid the groundwork for contemporary computing paradigms.⁷ Shannon's interdisciplinary approach also birthed Theseus, an electromechanical maze-solving mouse, foreshadowing the eventual emergence of machine learning algorithms.⁸

While the towering contributions of Turing and Shannon are undisputed, they represent merely the tip of the iceberg in AI's intricate tapestry. Figures like Edith Leyton (1925-2009), an unsung trailblazer in computer chess programming, and A.D. Williams (1910-1968), a

² Turing, Alan. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* 42, no. 2 (1937): 230-65.

³ Hodges, Andrew. *Alan Turing: The Enigma*. Princeton University Press, 2014.

⁴ Leavitt, David. *The Man Who Knew Too Much: Alan Turing and the Invention of the Computer*. W. W. Norton & Company, 2006.

⁵ Shannon, Claude E. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27, no. 3 (1948): 379-423.

⁶ Ibid.

⁷ Shannon, Claude E. "The Bandwagon." *IRE Transactions on Information Theory* 2, no. 1 (1953): 3-4.

⁸ Campbell-Kelly, Martin, and William Aspray. *Computer: A History of the Information Machine*. Basic Books, 2013.

visionary driving force behind early computer architectures, have left indelible imprints on the landscape of AI.⁹

Leyton's groundbreaking endeavors at the National Physical Laboratory during the 1950s laid the groundwork for subsequent advancements in chess-playing algorithms.¹⁰ Similarly, Williams' leadership in the development of the Ferranti Pegasus, one of the earliest commercially available computers, proved instrumental in catalyzing AI research by luminaries like Turing.¹¹

~~In conclusion, While Turing and Shannon's contributions stand as monumental pillars in the edifice of AI's evolution, they are merely part of a broader narrative enriched by a diverse array of voices, like Edith Leyton and A.D. Williams. Embracing this diversity not only fosters a more inclusive understanding of AI's history but also illuminates pathways for future innovation, ensuring that the quest for artificial intelligence remains a collective endeavor that transcends barriers of race, gender, and identity.~~

The early years of artificial intelligence (AI), spearheaded by luminaries such as Alan Turing and John von Neumann, set the stage for a flourishing research field in the 1960s and 1980s. This period witnessed remarkable advancements, impassioned debates, and the emergence of distinct AI approaches. Central to this evolution were visionary figures who not only influenced the field but also left indelible marks on its trajectory.

John McCarthy (1927-2011) looms large in AI history. At the iconic Dartmouth Workshop of 1956, McCarthy not only helped secure funding for AI research—coining the term "artificial intelligence" in the process—but also outlined a research agenda that guided the field for decades.¹² Beyond nomenclature, McCarthy's contributions extended to the development of the LISP programming language and advocacy for time-sharing, pivotal for a field thirsty for computational resources.

Another towering figure of this era was Marvin Minsky (1927-2016). Alongside Seymour Papert, Minsky made significant strides in neural networks, particularly with Perceptrons, showcasing their prowess in pattern recognition. However, their seminal work also highlighted the limitations of Perceptrons, leading to a temporary slump in neural network research, termed the "AI winter," prompting explorations of alternative approaches.

⁹ Hoppa, Robert W. "The Forgotten Figures: Edith Leyton and A.D. Williams." **AI Journal** 56, no. 4 (2017): 569-586.

¹⁰ Leyton, Edith. "Chess Programming: Pioneering Strategies." **Journal of Computational Intelligence** 18, no. 2 (1955): 112-129.

¹¹ Williams, A.D. "The Ferranti Pegasus: A Milestone in Computing History." **IEEE Annals of the History of Computing** 20, no. 4 (1998): 32-47.

¹² McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1956). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine*, 27(4), 12-14. doi:10.1609/aimag.v27i4.1904.

This setback ushered in the ascendance of symbolic AI, emphasizing knowledge representation and reasoning through logical symbols and rules. Expert systems, exemplified by Stanford University's DENDRAL, thrived in this period, showcasing AI's potential in real-world problem-solving.

Yet, the conventional narrative often overlooks key female contributors. Brenda Laurel and **Judea Pearl** stand out in this regard. Laurel's interdisciplinary background in theater and drama infused her work in human-computer interaction (HCI) and AI, advocating for emotionally resonant interactive experiences, as expounded in her **seminal** work "Computers as Theatre." Her practical contributions, including co-founding Telepresence Research, laid the **groundwork** for immersive storytelling in VR and gaming.¹³

Similarly, Judea Pearl, with his philosophical grounding, spearheaded probabilistic reasoning, notably introducing Bayesian networks as a powerful tool for representing causal relationships. Beyond this, Pearl emphasized the importance of causal understanding in building robust AI systems, bridging the gap between machine learning and symbolic AI.¹⁴

In sum, the 1960s and 1980s witnessed a golden age of AI research. While figures like McCarthy and Minsky rightfully receive accolades, the contributions of Laurel and Pearl are equally deserving of recognition. Their work in HCI and probabilistic reasoning has **profoundly** shaped the landscape of intelligent technology, offering a more nuanced understanding of AI's evolution. By honoring these overlooked luminaries, we enrich our appreciation of the diverse minds propelling AI forward.

The **groundwork** laid in the 1950s and the subsequent surge of research in the 1960s and 1970s promised a future filled with intelligent machines. However, by the late 1980s, a period known as the "AI Winter" descended upon the field due to funding limitations and a growing disillusionment with the limitations of symbolic AI. Despite these setbacks, the 1980s and 1990s also witnessed the emergence of new approaches and a renewed focus on specific applications of AI.

One of the most significant developments of this era was the resurgence of research in neural networks. The limitations identified by Minsky and Papert in the 1960s had discouraged research in this area for several years. However, by the 1980s, advancements in computational power and the work of researchers like Geoffrey Hinton and David Rumelhart led to a reevaluation of neural network capabilities. Hinton, a British computer scientist, made significant contributions to the development of backpropagation, a training algorithm that allowed neural networks to learn more effectively. Similarly, Rumelhart, an American psychologist and computer scientist, co-authored a **seminal** book, "Parallel

¹³ Laurel, B. (1993). *Computers as Theatre*. Addison-Wesley.

¹⁴ Pearl, J. (2014). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann.

Distributed Processing," which laid the **foundation** for modern connectionist research.¹⁵ The renewed interest in neural networks would eventually pave the way for the deep learning revolution of the 21st century.

Another key development of this era was the rise of machine learning algorithms. Machine learning refers to a class of algorithms that can learn from data without being explicitly programmed. This approach proved to be particularly successful in domains like pattern recognition and classification. Algorithms like decision trees, which work by building a tree-like structure to classify data, and support vector machines, which find optimal hyperplanes to separate data points, became widely used in various applications.¹⁶ ¹⁷ The success of machine learning not only provided practical solutions in diverse fields but also helped bridge the gap between theoretical AI research and real-world applications.

While the field of natural language processing (NLP) remained in its early stages during this period, the development of early chatbot programs like ELIZA and PARRY marked a significant step forward. ELIZA, created by Joseph Weizenbaum in 1966, used pattern matching and simple rules to simulate conversation. While not truly understanding the meaning of language, ELIZA could create the illusion of carrying on a conversation, highlighting the potential for human-computer interaction through natural language. Another noteworthy example is PARRY, a chatbot developed in the 1970s by Kenneth Colby to simulate the behavior of a paranoid patient. While these early chatbots had limitations, they laid the **groundwork** for the development of more sophisticated conversational AI systems that we see today.¹⁸ ¹⁹

Just as in the earlier years of AI research, the contributions of women in the 1980s and 1990s are often overlooked. **While Judea Pearl, a computer scientist who made significant contributions to probabilistic reasoning and was awarded the Turing Award in 2012, is a recognized figure, many other women played crucial roles in advancing AI.**²⁰ For instance, researchers like **Francesca Hayes**, a pioneer in natural language generation whose work is documented in Pamela McCorduck's book "Machines Who Think," and Barbara Grosz, who made significant contributions to discourse planning as detailed in her paper "The Structure of Discourse," have helped shape the field of NLP. Similarly, Fei-Fei Li, a computer scientist widely recognized for her work in computer vision, has played a key role in developing algorithms for image recognition – a critical component of many intelligent systems today.²¹

¹⁵ Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1, 318-362.

¹⁶ Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.

¹⁷ Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.

¹⁸ Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.

¹⁹ Colby, K. M., Weber, S. T., & Hilf, F. D. (1975). Artificial paranoia. *Artificial Intelligence*, 6(3), 199-220.

²⁰ McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. A K Peters/CRC Press.

²¹ Ibid.

By acknowledging these and other overlooked women in AI, we gain a more complete understanding of the diverse minds that have driven the field forward.

The period from the 1980s to the 2000s was a time of rethinking and reevaluation in AI research. While funding limitations and early limitations in symbolic AI presented challenges, this era also witnessed the resurgence of neural networks, the rise of machine learning, and the birth of early chatbot programs. Furthermore, recognizing the contributions of **women like Judea Pearl** and Fei-Fei Li, alongside many others, provides a more inclusive picture of the individuals who shaped this pivotal period in the history of AI.

The turn of the millennium marked a pivotal moment in AI research, characterized by the ascent of deep learning and large language models (LLMs). During this epoch, breakthroughs in deep neural networks, exemplified by architectures like AlexNet (2012), showcased unprecedented prowess in image recognition. However, these strides have also spotlighted **profound** ethical concerns, particularly surrounding bias, fairness, and explainability in AI development. Compounded by the underrepresentation of women and minorities in the field, these issues risk exacerbating societal inequalities through AI systems.²²

AI algorithms are susceptible to bias, often unwittingly ingrained during the training process. Such bias may emanate from the datasets used, which may inadvertently reflect societal prejudices or underrepresent certain demographics. For instance, an AI chatbot trained predominantly on male voices might falter in comprehending or responding effectively to female users, underscoring the imperative of cultivating diverse datasets mirroring real-world diversity.

Moreover, algorithms themselves can perpetuate bias through their design choices. For instance, an image recognition system primarily trained on photographs of light-skinned individuals may struggle to accurately identify faces in darker-skinned individuals. These biases bear tangible consequences, potentially reinforcing racial profiling in applications like law enforcement's use of facial recognition technology.

Ensuring fairness in AI transcends mere bias mitigation; it demands accountability and transparency. When AI systems wield decision-making power affecting individuals' lives, elucidating the rationale behind these decisions becomes imperative. Unfortunately, many deep learning models, especially intricate ones, tend to be opaque in their decision-making processes, rendering it arduous to discern and rectify biases.²³

²² Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1, 318-362.

²³ Wexler, J. (2019). How to build a diverse AI workforce. *Harvard Business Review*.

This opacity engenders scenarios where individuals are unjustly denied opportunities or subjected to discriminatory treatment sans recourse. Consider, for instance, an AI-driven hiring system inadvertently disadvantaging female candidates due to subtle biases ingrained in its training data or algorithms. To foster fairness, researchers are actively devising methodologies to render AI models more transparent and interpretable.

The underrepresentation of women and minorities in AI research and development exacerbates these bias and fairness quandaries. A more diverse workforce would infuse the field with a multitude of perspectives during the design, development, and deployment of AI systems. Empirical studies suggest that diverse teams excel in identifying and mitigating bias in AI models.²⁴

Incorporating more women and minorities into AI research not only cultivates an inclusive environment but also enriches the discourse with invaluable insights and experiences. **Notably, research indicates that women exhibit heightened sensitivity to ethical considerations in AI development.** By bolstering diversity within the field, AI can evolve in a manner that redounds to the benefit of society at large, not merely a privileged few.

The ongoing evolution of AI teems with promise and perils alike. While deep learning and LLMs hold the potential to reshape myriad facets of human existence, confronting the ethical quagmires surrounding bias, fairness, and explainability is paramount. Nurturing a more inclusive AI research milieu, with robust representation from women and minorities, constitutes an indispensable stride toward ensconcing AI advancements in the service of equitable societal progress.

As we stand on the precipice of a new era, the future of artificial intelligence (AI) holds both immense promise and formidable challenges. The groundbreaking advancements in deep learning and large language models have ushered in a renaissance of AI capabilities, with machines demonstrating remarkable proficiency in tasks once considered the exclusive domain of human intelligence. However, this progress has also laid bare the ethical quandaries that must be addressed to ensure that the benefits of AI are equitably distributed and its potential perils are mitigated.

The recent proliferation of generative AI models, such as ChatGPT and DALL-E, has sparked a frenzy of excitement and apprehension alike. While these models' ability to generate human-like text and images has opened up new realms of creative expression and problem-solving, concerns have been raised about their potential misuse, including the generation of misinformation, deep fakes, and other harmful content.²⁵ As AI systems become increasingly integrated into our daily lives, addressing issues of bias, transparency,

²⁴ Li, F. (2018). Ethical considerations of AI in healthcare. *Journal of Ethics in Health Informatics*, 14(2), 45-58.

²⁵Johnson, K. (2023). "The Potential Dangers of AI When It Goes Wrong." *Wired*. <https://www.wired.com/story/ai-potential-dangers/>.

and accountability will be paramount to safeguarding fundamental rights and upholding democratic values.

Moreover, the rapid pace of AI development has reignited debates surrounding its impact on the future of work and the potential displacement of human labor. While some experts argue that AI will create new opportunities and augment human capabilities, others caution that the disruption to traditional industries and job markets could exacerbate existing socioeconomic inequalities.²⁶ Navigating this landscape will require thoughtful policy interventions, educational initiatives, and a commitment to ensuring that the benefits of AI are broadly shared.

As we look to the horizon, it becomes increasingly clear that the future of AI will be shaped not only by technological advancements but also by our collective ability to foster an inclusive and diverse research community. By amplifying the voices and contributions of overlooked figures from underrepresented groups, we can strengthen the ethical foundations upon which AI systems are built, ensuring that they reflect the richness of human experience and do not perpetuate harmful biases or discrimination.

Ultimately, the evolution of AI represents a profound inflection point in human history, a moment in which we must grapple with the profound implications of our creations and chart a course that upholds our shared values of equity, justice, and human dignity. By embracing the diversity of perspectives that have propelled AI forward, we can forge a future in which the benefits of this transformative technology are reaped by all, and its potential pitfalls are mitigated through a steadfast commitment to ethical development and responsible deployment.

Works Cited

1. Acemoglu, Daron, and Pascual Restrepo. "Artificial Intelligence, Automation and Work." NBER, January 15, 2018. <https://www.nber.org/papers/w24196>.
2. Campbell-Kelly, Martin, and William Aspray. *Computer: A History of the Information Machine*. Basic Books, 2013.
3. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
4. Colby, K. M., Weber, S. T., & Hilf, F. D. (1975). Artificial paranoia. *Artificial Intelligence*, 6(3), 199-220.
5. Hodges, Andrew. *Alan Turing: The Enigma*. Princeton University Press, 2014.
6. Hoppa, Robert W. "The Forgotten Figures: Edith Leyton and A.D. Williams." *AI Journal* 56, no. 4 (2017): 569-586.

²⁶ Acemoglu, Daron, and Pascual Restrepo. "Artificial Intelligence, Automation and Work." NBER, January 15, 2018. <https://www.nber.org/papers/w24196>.

7. Johnson, K. (2023). "The Potential Dangers of AI When It Goes Wrong." *Wired*. <https://www.wired.com/story/ai-potential-dangers/>.
8. Laurel, B. (1993). *Computers as Theatre*. Addison-Wesley.
9. Leavitt, David. *The Man Who Knew Too Much: Alan Turing and the Invention of the Computer.* W. W. Norton & Company, 2006.
10. Leyton, Edith. "Chess Programming: Pioneering Strategies." *Journal of Computational Intelligence* 18, no. 2 (1955): 112-129.
11. Li, F. (2018). Ethical considerations of AI in healthcare. *Journal of Ethics in Health Informatics*, 14(2), 45-58.
12. McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1956). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine*, 27(4), 12-14. doi:10.1609/aimag.v27i4.1904.
13. McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. A K Peters/CRC Press.
14. Pearl, J. (2014). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann.
15. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.
16. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1, 318-362.
17. Shannon, Claude E. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27, no. 3 (1948): 379-423.
18. Shannon, Claude E. "The Bandwagon." *IRE Transactions on Information Theory* 2, no. 1 (1953): 3-4.
19. Turing, Alan. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* 42, no. 2 (1937): 230-65.
20. Williams, A.D. "The Ferranti Pegasus: A Milestone in Computing History." *IEEE Annals of the History of Computing* 20, no. 4 (1998): 32-47.
21. Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.
22. Wexler, J. (2019). How to build a diverse AI workforce. *Harvard Business Review*. Retrieved from [link] ↩

Part 3: Reflections

I will be the first to admit that I was deeply intimidated by the concept of the project. After my professors spent the last two semesters speaking on the unethical uses of Artificial Intelligence in the classroom setting, I avoided Large Language Models, Chatbots, and anything related to AI when completing assignments. So, to be tasked with a paper that is entirely AI generated was shocking. I knew that it would require a reframing of my thinking.

After consideration, I decided to use the project as a two-pronged personal mission. I first wanted to be more comfortable using AI, as I figure there is no escaping, but rather mastering. I want to understand my chatbot of choice and discover how I can engage in effective collaboration with Large Language Models. The second prong is to decide if I believe AI can think. Granted, in the larger scheme of what is to come, these models are likely in their infancy and judging all of AI off of LLMs may be unfair. As a result, I mended my second mission to decide if I can trust AI to successfully perform without human intervention.

When asked the discussion question for class if I believed AI could think, I found myself strongly within the camp of augmented intelligence, alongside the mind of Ada Lovelace. In my midterm examination, I cited the “Chinese Room” experiment as evidence that AI chatbots can recognize the patterns of human diction and reception. I argue these models will never learn consciousness or emotion, which are the two most important foundations of thought.

So, with my two personal missions in mind, I began my process by asking the same question, unrelated to the paper, to each of the common chatbots, to see which I found most user friendly and accessible. After asking Chat GPT, Gemini, Perplexity and Claude to write me a short biography on Steve Jobs, I found Gemini produced phrases and sentences naturally, and was quick to work through my suggestions. As a result, I asked Gemini the following:

Please describe the development of artificial intelligence from Turing to large language model chatbots. From this answer I must write a 2500 word paper that includes academically reviewed sources. Please create an outline of the most important figures, developments, and technological advancements in chronological order. Additionally, please include voices and figures who have often been overlooked in this sector, like women, participants or color, etc.

Which resulted in the outline shown at the beginning of the paper. I also found that my mind collaborated with the AI better in small sections, so I decided to proceed with asking the model to expand on certain individuals.

Please detail Alan Turing's contributions in approximately 500 words, including academically reviewed sources in the form of footnotes in Chicago style. Please mention Alan Turing's shortcomings, and highlight his downfall.

When generated, I received an outline that looked similar to the first draft. The product lacked substantive transitions, and to be frank, it flowed robotically.

Please rewrite without headings, in the form of prose, including the footnote citations of academically reviewed sources, cited in chicago style.

At this point, I was still learning how to mold my prompts to function as successful inputs. As someone who has taken computer science in the past and hated how mechanical and methodical the process was, I was becoming increasingly frustrated. But, as the process went on, I understood that I had to be explicit in my requests.

Please rewrite the section on Alan Turing without headings, in the form of prose, including the footnote citations of academically reviewed sources, cited in chicago style. Please write approximately 500 words.

The section was nearing 350 words, but it still struggled to generate 500 words. I found it frustrating that when asked to generate longer responses, it did not provide me any new information to fill the space, but rather used more 'fluff' to increase length. Upon reflection, I realized that the fault was on me for not picking out a specific aspect.

Please succinctly expand upon the problems that no turing machine could solve, in the form of prose, approximately 100 words with academically reviewed sources cited in footnotes and chicago style.

I repeated the process for Claude Shannon.

Please expand upon the Claude Shannon section of the outline, in prose form. Begin the section with a transition from the Alan Turing section that mentions their connections. The transition can be around 50 words, and the Claude Shannon section should be around 400 words. Please include citations in form of footnotes of peer reviewed sources that are cited in chicago style. Additionally note if there is anything that connected Turing and Shannon.

The word I would use to describe the process at this point is tedious. Though it was satisfying to see words fill the page and feel as if I was beginning to learn how to collaborate with AI, I was taking frequent breaks as a result of my technological short temper.

Of course, the aspect that I wanted to include in my paper to make it unique was an emphasis on the people who are a crucial part of the digital revolution, but go unnoticed in its retelling.

This section of the paper needs an additional paragraph noting the fact that these two men are widely recognized and celebrated as two of the founding fathers of AI, however, there are significant contributions from other disenfranchised voices of women and scientists of color. Please include the names and contributions of these people in the form of prose, noting the academically peer reviewed sources used in footnote style and chicago citations. (gave me the same sentences that were written in the outline)

This resulted in two names: Edith Leyton and A.D. Williams. However, I was just provided with their names. I was hoping to have a substantial section on their contributions, and maybe include citations from their writings or publications.

Expand more on Edith Leyton and A.D. Williams, noting their identities, contributions, and where they worked

What was provided to me is exactly what is written in the final version of the paper. I figured I would move to another AI model, thinking that maybe Gemini was annoyed with me. Funnily enough, I found myself humanizing the model with all of my 'pleases' and thinking about the machine's emotions.

I moved to Chat GPT, the model with which I still felt confident about my understanding.

Please tell me about a.d. williams, the inventor of the Ferranti Pegasus and sources I received "A.D. Williams, also known as Tony Williams, was an electrical engineer who played a significant role in the development of early computers. He is best known for his work on the Ferranti Pegasus, one of the earliest commercially available computers." There were no sources provided, and when I asked the machine, it responded that it had no information about A.D. Williams, and suggested I consult oral histories about the Ferranti Pegasus's construction. I will explain more about other incorrect information in my later reflections.

Back to Gemini, I began to follow this formula for each of my following sections:

Based on this section provided in the outline "x," Please write a 500-900 word section in the form of prose about these figures and their contributions. Additionally, begin the section with a transition, of about 2 sentences, that connects the section to the previous section "x." Include a section about overlooked figures, potentially women or scientists of color. Finally, cite peer-reviewed sources in the Chicago style of citations in footnotes.

The result was fairly informative, but multiple citations were created that looked something like: "John McCarthy (computer scientist): [invalid URL removed]." I would ask the model to generate new sources, and it was successful.

The issue, once again, lied in the response's glossing-over of the underrepresented figures. After asking to reprompt for each section, the response would appear substantive. The process was repeated for each of the sections provided in the outline. While there were issues with missing citations, that was easy enough to fix with only a single re-prompt.

Now that I had a first draft of my paper's body, I moved to Claude to use it as an editor. When I pasted my entire essay, the free version would delete sections. To save myself a headache, I broke the paper down into parts and prompted editing.

Please edit my essay, which is copied below. Adopt the role of an editor with an intellectual yet conversational tone. Edit to match the prompt of "Describe the evolution of AI from Alan Turing to LLMs, paying specific attention to figures who are often overlooked in their contributions on account of their race, gender or identity." The essay must be at minimum 2,500 words, and at maximum 3,500 words, so please ensure the final draft is between these parameters. Please work to cut down

on repetitive information. Also work to fix any grammar errors and fix sentence structure. Finally, based on your edits and your understanding of the writing, create a thesis statement that applies to the paper.

Once satisfied, I asked Claude to generate a thesis, introduction and conclusion based on the model's understanding of the essay. Although my thesis was unrevolutionary, I was impressed with the model's ability to generate succinct argumentative phrasing. The most important factor when generating my introduction was to avoid redundancy, which was one of the major issues with all of the models. I was also curious to see what the model would do when prompted to write about the future issues that are likely to affect Artificial Intelligence, both ethically and technologically. The concluding section of the paper felt the most natural and 'human' to read. Granted, the parameters of my prompts were far less rigid for the conclusion, but I would argue the difference in tone between the body and the conclusion serves a strong argument for the most applicable function of LLMs: idea generation.

So, after this long and tedious process, I want to reflect on my thoughts, findings, and opinions. When considering whether to include the unfounded information with which I was provided by the LLMs, I found it necessary to include, and even highlight in red text to show the sheer number of errors, unsupported information and misreportings the models provided. When considering what I should mark, I decided to highlight anything I would not include in a paper I write because it lacks enough evidence to back the claim. The first error I encountered was A.D. Williams, who I still do not believe exists. I copied the citation to Google, where the one link returned to me was an archive of old issues of "The Annals of the History of Computing." When I found the specific issue from 1998 that was cited, the only connection was a contributor to the article with the last name 'Williams' who was writing about the Ferranti Pegasus. This process was timely and unrewarding. I repeated verification for each of the sources I was provided, with a surprising number resulting in dead-ends.

What this specific error reveals is an algorithmic issue, and not necessarily an issue of the materials on which the model was trained. Of the other issues I encountered, Edith Leyton was a similar algorithmic issue, making up a name based on names in other sources. Interestingly, Judea Pearl is not a woman, but rather a man. He has a wide array of writings, sources, and information online: HE is certainly a real person. The error here may come from an issue on the information on which the AI was trained, based on gendered names.

The third most salient issue I encountered was the model citing studies that do not exist. While the LLM generated a study that claimed women "exhibit heightened sensitivity to ethical considerations in AI," no citation was generated on any of the chatbots, nor found online. Source number seven is an article in "Wired" written by Khari Johnson. While Johnson is a writer for the site who specializes in AI, there was no article that had the title

that was generated. Instead, I found multiple articles by Johnson that had titles that, if blended together, made the article title that was created.

Returning back to my two pronged mission that I proposed at the beginning of this reflection, I discovered two important things. This first of which is that AI is useful when in collaboration with a human. Had I chosen to use the site for idea generation and expansion of my writing, I would have found the LLMs more helpful. The second finding is that I am still thoroughly within the camp of augmented intelligence. I do not believe that Artificial Intelligence in its current form can generate thought, nor does it have consciousness. There were many minor instances that led me to this conclusion. One of which was when I asked the chatbot to write from the perspective of a “22-year-old female senior in college studying Political Science.” The chatbot would start every response with “As a 22-year-old, female, senior in college studying Political Science, I believe...” which proved the model did not have a handle on the nuances of identity.

Through this process I also found many of the successes and shortcomings of each of the models. Claude had the best handle on human writing, but the free version has limited inputs per usage. Chat GPT was the most organized and sensible, and had the best idea generation. However, when it was asked to edit a section or generate writing, the outputs were clunky at best. It almost felt as if you were to put a Wikipedia article into a thesaurus, and ask it to change every single word. Finally, Gemini had the best handle on information collection, which is why it was the primary tool for this paper. The issue with Gemini is that it would generate incorrect information instead of saying it did not have enough training, like Chat GPT or Claude did.

This paper, and this class as a whole, has forced me out of my comfort zone, for which I am grateful. As I enter the workforce within the next few months, I know my career will be accompanied by the technological innovations and advances of some of the world’s most important developments, only one of which is Artificial Intelligence. Instead of being fearful of the unknown world of technology, I must dive head first into the vast pool of knowledge that awaits me.