John Doe
Professor Smith
HIST 21: History of Computers
December 11, 2025

A History of Artificial Intelligence, with a Focus on the Imitation Game

According to a 2024 survey by the Digital Education Council, 86% of students claim to use Artificial Intelligence in their studies[1], raising the question of how successful this AI is relative to its earliest benchmarks. An early test theorized for Artificial Intelligence came from the "Father of Theoretical Computer Science" Alan M. Turing in his 1950 paper *Computing Machinery and Intelligence*[7]. For this test, Turing predicted a machine able to engage in conversation with people. Despite AI's early focus on symbolic logic and later on game playing, recent developments of large language models and conversational AI have brought renewed historical relevance to Turing's philosophical test called the Imitation Game.

In order to understand the impact of Turing's philosophical theories, familiarity with the Imitation Game is necessary. Turing's Imitation Game consists of a machine capable of replicating human conversation attempting to convince a person it is more likely to be human than a real person is. Conversation about a mechanical mind prior to this was largely philosophical about whether an artificial brain is possible and the existence of a mind outside the brain[2]. Turing effectively dismissed these questions by claiming that a more appropriate question is whether the machine is able to convince others that it is human.

Despite Turing's emphasis in his paper on testing conversational distinguishability of Artificial Intelligence from humans, early AI focused on symbolic logic and proving theorems. Turing's early work led to the development of symbolic logic machines. Early and modern computers are largely based on Turing Machines, originally outlined in the 1937 paper *On computable numbers, with an application to the Entscheidungsproblem*[8]. Allen Newell and Herbert Simon's 1956 report *The logic theory machine–A complex information processing system*[6] and the associated program Logic Theorist was an early example of AI that operated on symbolic logic. Newell and Simon described the goal of Logic Theorist to be "a program for constructing chains of theorems, not at random but in response to cues that make discovery of cues possible within a reasonable computing time"(Newell and Simon, p.5). These innovations were revolutionary, but they strayed from Turing's vision of conversational AI.

The game-playing machines also played a prominent role in the development of AI decades before the advancements required for conversational AI Turing initially described. An example of this is found in chess. Chess programs have existed since Turing and Champerowne created the Turochamp algorithm in 1950, which was able to play a game of chess[3]. Over the next half century, chess programs advanced until the IBM supercomputer Deep Blue won a six-game chess match against world chess champion Garry Kasparov in 1996[9]. These had similar concepts to the Turing Machine, in that the programs sought to play the same games that humans did. However, they diverged from Turing when the computers' goal was to move beyond the skill capacity of humans.

Conversational AI also has a long history, despite a slower start. An early conversational AI was ELIZA, developed between 1964 and 1967 by Computer Scientist Joseph Weizenbaum. ELIZA used pattern matching to simulate a Rogerian Psychoanalyst[5]. Rogerian Psychoanalysis focuses on a patient's own opinions and thoughts, with an emphasis on empathy and unconditional positive regard[10]. This and the limitations of the computer did limit the use of ELIZA, which failed in taasks such as genuine conversation and non-psychiatric use[5]. These limitations prevented ELIZA from success in the Imitation Game. However, advancements in machine learning would allow for advancements in conversational AI.

One such advancement in machine learning was the model of Reinforcement Learning from Human Feedback (RLHF), which uses human opinion to improve its model and better resemble

ideal human response. RLHF takes an intial language model capable of responding to a wide range of prompts. The language model is then presented to a (human) reviewer who gives the model a prompt. Once the model and one or more similar models generate responses to the prompt, the responses are then shown to the reviewer, who chooses the best one[4]. This is most easily done with two models, but studies have trained models using up to four response options[11]. A reward model then turns this selection into a single number that acts as a reward for each model, teling it what works well and what works poorly. The model in turn improves itself based on this response[4]. This parallels the idea of a machine trained by reward and punishment, introduced by Turing in *Computing Machinery and Intelligence*[7]. RLHF allows AI to improve its similarity to human conversation through direct response by humans. This assists in building a congruence between conversational AI and human conversation, granting it better likelihood at success in the Imitation Game.

Artificial Intelligence has a long history in Computer Science, from its introduction working with symbolic logic to learning and playing chess. Ever since Alan Turing introduced the Imitation Game, the question of whether a conversational AI can convince others it is human has been on the mind of Computer Scientists and philosophers. Recent developments have allowed for conversational AI to better resemble human language. With the rise of use of AI in society, the question is extant of whether it can replace humans or will only ever asymptote them. The answer to this question is found through Alan Turing's Imitation Game, which rises in relevance as Artificial Intelligence advances.

References

[1] Digital Education Council. AI or Not AI: What Students Want. 2024. URL: https://www.digitaleducationcouncil.com/post/digital-education-council-global-ai-student-survey-2024.

[2] Rene Descartes. Meditations on First Philosophy. With Selections from the Objections and Replies. Trans. by John Cottingham. 2nd. Cambridge Texts in the History of Philosophy. Cambridge University Press, 1996.

[3] Andrew Hodges. Alan Turing: The Enigma. Simon & Schuster, 1983.

[4] Nathan Lambert et al. "Illustrating Reinforcement Learning from Human Feedback (RLHF)". In: Hugging Face Blog (2022). https://huggingface.co/blog/rlhf.

[5] Pamela McCulloch. Machines Who Think. A.K. Peters, Ltd., 2004.

[6] A. Newell and H. Simon. "The logic theory machine–A complex information processing system". In: IRE Transactions on Information Theory 2.3 (1956), pp. 61–79. DOI: 10.1109/TIT.1956.1056797.

[7] Alan M. Turing. "Computing Machinery and Intelligence". In: MIND 59.236 (1950).

[8] Alan M. Turing. "On computable numbers, with an application to the Entscheidungsproblem". In: Proceedings of the London Mathematical Society 42.2 (1937).

[9] Bruce Weber. "Swift and Slashing, Computer Topples Kasparov". In: The New York Times (June 12, 1997).

[10] Lucy Yao and Rian Kabir. "Person-Centered Therapy (Rogerian Therapy) [Updated 2023 Feb 9]". In: StatPearls [Internet]. StatPearls Publishing, 2025. URL: https://arxiv.org/abs/2005.14165.

[11] Daniel M. Ziegler et al. Fine-Tuning Language Models from Human Preferences. 2020. arXiv: 1909.08593 [cs.CL]. URL: https://arxiv.org/abs/1909.08593.