

Donald Aingworth
Professor Ryer
HIST 46: Independent Study
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 conversational AI is relative to its earliest benchmarks. An early test theorized for Artificial Intelligence came from 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 in 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 pitted against a human in an effort to convince the third party that they are the human and the other is the machine. Conversation about a mechanical mind prior to this had a heavy philosophical emphasis centered around dualism, the question of the existence of a mind outside the brain[2]. Turing's Imitation Game reframed the question at hand by claiming that a more appropriate question is whether the machine is able to convince others that it is human.

Despite the emphasis in *Computing Machinery and Intelligence* 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 trace their conceptual origins to the Turing Machine, 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 in their ability to teach reasoning to a machine, but they strayed from conversational AI. Symbolic logic machines found proofs using axioms and logic. However, they were incredibly specialized to logic but faltered in other areas like conversation. When playing the Imitation Game, a topic the machine has low capacity in due to an extreme and unbalanced focus on logic using its creation would be a breaking point.

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 programs sought to play the same games that humans did. Their innovation led to significant progress in optimization programs, which many fields of AI use. However, like symbolic logic machines, their specialized nature lacked the ability to engage with humans beyond playing the game they were programmed for, and they would face a severe uphill battle in the Imitation Game.

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. This and the physical constraints of the computer did limit the

use of ELIZA, which failed in tasks such as non-psychiatric use[5]. These 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 initial language model capable of responding to a wide range of prompts. It and one or more similar models generate responses to the prompt, from which the reviewer chooses the best one. A reward model then turns this selection into a single number that acts as a reward for each model, telling it what works well and what works poorly. The model in turn improves itself based on this response[4]. This parallels a thought experiment introduced by Turing in *Computing Machinery and Intelligence* of a machine trained by reward and punishment[7]. RLHF allows AI to improve its similarity to human conversation through direct response by humans. Using human response to set benchmarks allows the AI to test its resemblance to human conversation by asking its eventual reviewer directly. This assists in building a congruence between conversational AI and human conversation, granting it better chances 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, there remains a question of whether it can replace humans or will only ever approach but never reach them. Alan Turing's Imitation Game is a long-standing benchmark for AI that should be strongly considered. As Artificial Intelligence advances from symbolic logic to playing games to pattern matching and language models that use RLHF, we come closer and closer to a machine that can sufficiently replicate human communication and potentially beat various opponents in the Imitation Game.

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).