Donald Aingworth

Professor Ryer

HIST 46: Independent Study

December 16, 2025

From Mary Shelley's *Frankenstein* to Marvin the Paranoid Andriod in *The Hitchhiker's Guide to the Galaxy*, the idea of artificially creating or replicating the human mind has captured the imagination of scientists and writers for centuries. We have come ever closer to this dream as time has passed, especially since the advent of the computer. As computers become better researched and tested, so has Artificial Intelligence and Machine Learning (AI/ML). Ideas like this have been on the minds of Mathematicians and Computer Scientists since the middle of the twentieth century. One such Mathematician was the "Father of Theoretical Computer Science" Alan Turing. Turing's philosophical prediction of learning mechanisms came in his 1950 paper *Computing Machinery and Intelligence*[1]. [Insert Strong Thesis Here] Recent developments in AI have shown Turing to be ahead of his time with philosophical theories about punishment and reward resembling reinforcement learning.

The development of Reinforcement Learning from Human Feedback (RLHF) parallels Alan Turing's philosophical suggestion of teaching machines through reward and punishment. The combination of reward and punishment of a machine into one unique response has failed to replicate Turing's full vision.

Despite AI's early focus on symbolic logic and later on game playing, recent developments of large language models and conversational AI and their prominence among have brought renewed historical relevance to Turing's philosophical test called the Imitation Game.

Turing's emphasis in his paper on testing conversational distinguishability of Artificial Intelligence from humans was secondary in the progress of AI towards theorem proving and game playing.

These concepts provided a philosophical series of behavioral tests and frameworks for language models such as child machines.

In *Computing Machinery and Intelligence*, Alan Turing predicted a dialogue machine which would be able to mimic a human mind. This prediction would be far ahead of its time, as AI/ML initially manifested as symbolic logic machines and eventually game-playing machines decades before Large Language Models achieved prominence. Despite this, Turing developed strong philosophical theories related to whether an artificial mind is possible. These laid out a target that language models would align themselves towards. This target took the form of the dialogue machine able to convince a human that it is more likely to be a human than another person through a process called "The Imitation Game". His prediction of training artificial intelligence described reinforcement learning, which remains a major paradigm of machine learning. This prediction was contemporary to the advent of the artificial neuron and neural network, a major foundation of AI/ML.

Arguably the most famous idea in *Computing Machinery and Intelligence* is what Turing calls "The Imitation Game". The Imitation Game is a Turing uses the Imitation Game as a destination for Artificial Intelligence.

References

[1] Alan M. Turing. "Computing Machinery and Intelligence". In: <u>MIND</u> 59.236 (1950).