Summary

The Imitation Game

Alan Turing proposes replacing the question ”Can machines think?” with the imitation game, where an interrogator tries to determine the gender of two players through written questions, to avoid ambiguity.

In the imitation game, player A, a man, tries to deceive the interrogator, while player B, a woman, provides truthful answers to help the interrogator.

Turing suggests investigating if a machine playing as A would deceive the interrogator as often as a man, effectively addressing the question of machine thinking.

Critique of the New Problem

Turing argues that focusing on the imitation game is more productive than questioning its worthiness.

The imitation game tests mental abilities while making physical skills irrelevant, and if a machine performs satisfactorily, objections about fundamental differences between human and machine thinking can be dismissed.

Turing assumes the best machine strategy is to provide human-like answers.

The Machines concerned in the Game

To avoid ambiguity, Turing limits the definition of ”machine” to digital computers in the imitation game.

He argues this restriction is not problematic unless digital computers fail to perform well, which he considers unlikely.

The imitation game aims to determine if hypothetical computers could succeed, a question Turing believes will be viewed differently later.

Digital Computers

Digital computers are designed to perform operations that a human computer could carry out by following strict rules.

Turing outlines the basic components of digital computers: a store, an executive unit, and a control unit.

The executive unit performs operations ranging from simple to complex, depending on the machine.

In digital computers, the ”book of rules” is replaced by a ”table of instructions” in the store, and the control unit ensures accurate execution.

Instructions are coded numerically, specifying the operation and data storage positions, and are executed sequentially unless altered by special instructions.

Repeated execution of instruction sequences until a condition is met is crucial in digital computers.

Digital computers closely mimic human computers, and programming involves translating human processes into instruction tables.

Computers with random elements introduce unpredictability, making them difficult to distinguish from deterministic machines by observation.

The concept of an infinitive capacity computer with an unlimited store is theoretically possible and has special interest.

The Analytical Engine, a mechanical precursor to modern computers, demonstrates that electricity is not theoretically essential, and similarities between computers and the nervous system are more likely found in mathematical analogies of function.

Universality of Digital Computers

Digital computers are discrete state machines that move suddenly between distinct states, ignoring the possibility of confusion.

Discrete state machines with finite states can be fully described by tables, allowing for future state predictions.

Digital computers have an enormous number of possible states due to their storage capacity.

A digital computer with sufficient storage and speed could mimic any discrete state machine in the imitation game without the interrogator distinguishing between them.

Digital computers are considered universal machines because they can mimic any discrete state machine, making them essentially equivalent.

The question "Can machines think?" can be rephrased as whether imaginable digital computers would perform well in the imitation game, which is equivalent to asking if a specific computer C could satisfactorily play the role of player A against a human player B.

Contrary Views on the Main Question

Turing believes that within 50 years, computers will fool an average interrogator 70% of the time after five minutes, and by the end of the century, machine thinking will be widely accepted.

He acknowledges the importance of distinguishing between proven facts and conjectures in scientific discourse.

The theological objection argues that thinking is exclusive to the immortal soul granted by God to humans.

Turing challenges this by arguing that God's omnipotence allows for granting souls to animals or machines, and creating thinking machines would not usurp God's power more than procreation.

He dismisses theological arguments as unimpressive and historically unsatisfactory.

The ”Heads in the Sand” objection dismisses machine thinking due to its dreadful consequences.

Turing suggests the theological argument stems from a desire to maintain human superiority, particularly among intellectuals.

He dismisses the ”Heads in the Sand” objection as insubstantial and suggests consolation in transmigration of souls.

The Mathematical Objection argues that discrete-state machines have inherent limitations and cannot answer all questions correctly, suggesting a difference between machine and human intellect.

Turing argues that while machines have proven limitations, the same hasn't been definitively proven for human intellect, and our feeling of superiority may be illusory.

Proponents of the mathematical argument would likely accept the imitation game as a basis for discussion.

The Argument from Consciousness contends that a machine cannot equal a human brain until it can create art and experience emotions through genuine thoughts and feelings.

In its extreme form, this argument leads to solipsism, where one can only be certain of one's own thoughts.

Turing suggests the viva voce examination is used to assess genuine understanding and provides a hypothetical dialogue to illustrate this.

Turing challenges the Argument from Consciousness by suggesting that satisfactory answers from a machine about its own sonnet would be difficult to dismiss as artificial signaling.

He suggests proponents would likely abandon the argument to avoid solipsism and accept the imitation game as a valid test.

Turing acknowledges mysteries surrounding consciousness but argues they don't need to be fully resolved to address machine intelligence.

The Arguments from Various Disabilities claim machines will never exhibit specific human traits or abilities.

Turing argues these arguments are often based on flawed scientific induction, generalizing from limited experiences with simple machines and failing to consider advanced technologies.

He suggests some proposed machine disabilities may seem trivial but contribute to more significant issues.

Turing distinguishes between ”errors of functioning” and ”errors of conclusion,” arguing that while abstract machines are incapable of the former, they can make the latter, refuting the claim that machines cannot make mistakes.

He argues machines can be the subject of their own thought by processing information about their operations, modifying programs, and potentially assisting in their own programming.

Turing dismisses criticisms about machine capabilities as disguised forms of the argument from consciousness.

Lady Lovelace's objection states that the Analytical Engine can only perform what it is instructed to do, while Hartree suggests future machines could ”think for themselves” or learn.

Turing agrees with Hartree, suggesting the Analytical Engine could potentially mimic a machine with the ability to ”think for itself” if it had adequate storage and speed.

He addresses variants of Lovelace's objection, particularly the claim that machines can never do anything truly new or surprising.

Turing argues that machines frequently surprise him due to his own limited calculations and assumptions, and the concept of ”original work” in humans is also debatable.

He anticipates critics attributing machine-generated surprises to human creativity but argues that recognizing something as surprising requires the same level of creative mental act regardless of the source.

Turing attributes the belief that machines cannot surprise us to a fallacy common among philosophers and mathematicians.

The Argument from Continuity in the Nervous System suggests that the nervous system's continuous operation may not be accurately mimicked by a discrete-state system like a digital computer.

Turing argues that while discrete-state machines differ from continuous machines, this difference would not be detectable in the imitation game, as a digital computer could provide sufficiently accurate approximations.

The Argument from Informality of Behaviour posits that creating a comprehensive set of rules describing human behavior for every circumstance is impossible.

Turing dissects the Argument from Informality of Behaviour, pointing out its logical fallacy and clarifying the distinction between ”rules of conduct” and

”laws of behaviour”.

He suggests that while we can't easily prove the absence of complete laws of behavior, this doesn't necessarily mean humans cannot be considered machines.

Turing argues that even with discoverable laws of behavior for machines, predicting their future actions might still be impossible, as demonstrated by a simple unpredictable computer program.

Turing acknowledges the overwhelming statistical evidence for extra-sensory perception (ESP), particularly telepathy, which challenges conventional scientific understanding.

He considers the argument from ESP to be strong, acknowledging that while many scientific theories remain practical despite conflicting with ESP, this offers little comfort.

Turing presents a scenario where ESP could potentially affect the imitation game, suggesting that even if a human with telepathic abilities outperforms a computer, the computer's random number generator might be influenced by the interrogator's psychokinetic powers.

Turing suggests modifying the imitation game, perhaps using a ”telepathy-proof room,” if telepathy is accepted as real.

Learning Machines

He acknowledges lacking strong positive arguments to support his views on machine intelligence, focusing instead on pointing out flaws in opposing arguments.

Turing uses analogies of a piano string and an atomic pile to explore idea propagation in machines and minds, questioning whether machines could be made ”super-critical”.

Turing acknowledges his analogies and arguments are meant to be persuasive illustrations rather than conclusive proofs.

He suggests the most convincing support for his views will come from conducting the proposed experiment at the end of the century but questions what steps can be taken in the meantime.

Turing argues that programming is the main challenge in creating a machine capable of passing the imitation game, estimating that the required storage capacity is achievable with current technology.

Turing suggests that instead of directly simulating an adult mind, creating a program simulating a child's mind and educating it might be more feasible.

He divides the problem into developing a child-program and an education process, drawing parallels to biological evolution.

Turing acknowledges that while a machine cannot be educated exactly like a human child, the example of Helen Keller demonstrates that education is possible with two-way communication.

Turing describes experiments with simple child-machines programmed to learn through punishment and reward signals, noting limited success due to unconventional teaching methods.

He argues that while punishment and reward can be part of the teaching process, developing ”unemotional” communication channels, such as a symbolic language, is necessary for effective and efficient instruction transmission.

Turing discusses the potential complexity of a child-machine, ranging from a simple system to one with built-in logical inference capabilities, and explains how it might process and act on instructions through established facts and imperatives.

Turing explains that for a limbless machine, imperatives must focus on applying logical rules and decision-making processes, which can be given by authority or developed through methods like scientific induction.

He addresses the apparent paradox of a learning machine by explaining that while fundamental rules remain constant, learning rules are less permanent, similar to the U.S. Constitution allowing for amendments.

Turing highlights that a key feature of a learning machine is the teacher's limited understanding of its internal processes, contrasting this with traditional computing, and suggests intelligent behavior involves slight deviations from strictly computational processes.

Turing advocates for including a random element in learning machines, arguing it can be more efficient than systematic methods when searching for solutions, especially with multiple possibilities, and draws a parallel to biological evolution.

He suggests that while machines may eventually compete with humans in all intellectual fields, determining the best starting point is difficult, proposing exploration of both abstract activities like chess and sensory-based approaches like language teaching and understanding.

Note on the use of LLMs

This summary was generated with the help of large language models.

My method for generating the summary was as follows:

1. Input each paragraph into Claude 3.5 Sonnet with the prefix "Please summarize the main point from the following paragraph as a single sentence: ", and copy the model's output to a draft document.

After completing this for all paragraphs in the essay, I fed my draft document back into Claude 3.5 Sonnet with the following instructional prompt:

> I have written the following set of summaries, where each sentence is a summary of a paragraph in the original text "Computing Machinery and Intelligence", an essay by Alan Turing. Please rewrite this document without removing any sentences. That is, there should be the same number of sentences in the output as there are in the original. The sentences can be reduced in length if there is redundancy across them, but the core sense of what they indicate should not be removed.

The model's output was then copied and pasted into a text editor, and cleaned up for basic formatting corrections. I then inserted the text's subtitles at points in the summarization that felt appropriate given my understanding of those the sections' contents.