Advances in Sciences and Humanities
2026, Vol. 12, No. 2, pp. 52–60
https://doi.org/10.11648/j.ash.20261202.11

Research Article

Deception, Impersonation, and Intelligence: What the Original Imitation Game Reveals About Modern Chatbots

Mohammed Zeinu Hassen*

Department of Social Sciences, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

Abstract

This article argues that the standard interpretation of the Turing Test, which dominates both philosophical discourse and public understanding of artificial intelligence, fundamentally misrepresents Alan Turing's original proposal. The Standard Turing Test asks whether a machine can imitate a human in unrestricted conversation. The Original Imitation Game, by contrast, requires both human and machine to impersonate a woman, with their success rates compared against each other. This structural difference has profound implications. The Original Imitation Game tests not behavioral similarity to humans but resourcefulness in performing a difficult task, impersonation. This paper examines what this alternative test reveals about intelligence and applies its insights to contemporary chatbots. It argues that modern language models, despite their conversational fluency, fail precisely the kind of test Turing originally proposed. They cannot genuinely impersonate because they lack the self-conscious critique of ingrained responses that impersonation requires. This failure illuminates something essential about intelligence: it consists not in the having of cognitive habits but in the capacity to recognize, evaluate, and override them when circumstances demand.

Keywords

Turing Test, Original Imitation Game, Impersonation, Chatbots, Artificial Intelligence, Intelligence, Deception, Language Models

1. Introduction

The question of whether machines can think has haunted philosophy since Descartes [1]. Yet no formulation of this question has proven more influential than the one Alan Turing proposed in 1950 [2]. His "imitation game" has become the cultural benchmark for artificial intelligence, invoked whenever a new chatbot demonstrates conversational prowess. In 2014, journalists announced that the program Eugene Goostman had passed the Turing Test for the first time [3]. In 2022, a Google engineer claimed that the company's LaMDA system was sentient based on his conversations with it [4]. In 2024, researchers reported that GPT-4 had been judged human in fifty-four percent of its interactions [5].

These claims share a common assumption. They assume that the Turing Test asks whether a machine can converse indistinguishably from a human being. This assumption structures public discourse, animates research programs, and shapes philosophical debate about machine intelligence.

The assumption is false.

Turing did not propose a single test. His 1950 paper contains multiple formulations, and these formulations are not equivalent. Susan Sterrett demonstrated this definitively in 2000, distinguishing what she called the "Original Imitation

*Correspondence: Mohammed Zeinu Hassen (mohammed.zeinu@aastu.edu.et), Mohammed Zeinu Hassen (mozhassen@gmail.com)

Received: 10 March 2026; **Accepted:** 19 March 2026; **Published:** 24 April 2026

Game Test" from the "Standard Turing Test" [6]. The Standard Turing Test, which dominates popular imagination, asks whether a machine can imitate a human well enough to fool an interrogator. The Original Imitation Game asks something different. It asks whether a machine can impersonate a woman, and whether it can do so as often as a man can impersonate a woman.

These are different questions. They yield different results. They test different capacities. And they have different implications for how we understand intelligence.

The problem is that seventy-five years of discussion have proceeded on the basis of the wrong test. Philosophers have criticized Turing for behaviorism, for anthropocentrism, for operationalism, criticisms that may apply to the Standard Turing Test but miss the subtlety of his actual proposal [7]. Computer scientists have dismissed the test as irrelevant to practical AI, citing the ease with which simple programs can fool naive judges [8]. Meanwhile, the Original Imitation Game sits neglected, its insights unexamined, its implications for evaluating modern chatbots unexplored.

This neglect matters because chatbots are now ubiquitous. They write student essays, provide customer service, offer therapeutic conversation, and serve as companions for the lonely [9]. Their fluency tempts us to attribute understanding, intelligence, even consciousness to them. The Standard Turing Test provides no resources for resisting this temptation. It asks only whether the machine can talk like a human, and when it does, we are inclined to grant that it thinks like one too.

The Original Imitation Game offers a different perspective. It suggests that intelligence is not about fluency but about resourcefulness. Not about similarity to human behavior but about capacity to perform difficult tasks. Not about what responses one has learned but about whether one can override learned responses when impersonation requires it.

This paper examines what that perspective reveals about modern chatbots.

2. Research Question

What does the Original Imitation Game reveal about the intelligence of contemporary chatbots that the Standard Turing Test obscures?

3. Methodology

This article employs conceptual analysis and comparative philosophical method. It first reconstructs Turing's 1950 argument with careful attention to its textual details, distinguishing between the two tests his paper contains [2]. It then examines the structural differences between these tests, drawing on the work of Sterrett, Copeland, and others who have attended to Turing's multiple formulations [6, 10]. The analysis proceeds by identifying the cognitive capacities each test demands and considering whether those capacities are possessed by contemporary language models. The argument is not empirical but conceptual: it asks not whether chatbots can pass the Original Imitation Game as a matter of fact, but what passing or failing would show about the nature of their operation. The conclusion draws on philosophical accounts of intelligence, impersonation, and self-conscious critique to assess what chatbots reveal about the relationship between fluent language use and genuine thought [11, 12].

4. The Two Tests

Turing opens "Computing Machinery and Intelligence" with a bold proposal. He suggests replacing the question "Can machines think?" with a question about a game. The game, which he calls the imitation game, is played with three people: a man, a woman, and an interrogator. The interrogator remains in a separate room and communicates with the other two by teleprinter. The object is to determine which respondent is the man and which is the woman. The man attempts to deceive the interrogator into making the wrong identification. The woman attempts to help the interrogator identify correctly [2].

Turing then asks: "What will happen when a machine takes the part of A in this game?" [2]. He proposes that we compare how often the interrogator decides wrongly when the machine plays the deceiver's role with how often the interrogator decides wrongly when a man plays that role. If the machine succeeds as often as the man, we have replaced our original question with one that can be answered.

This is the Original Imitation Game Test. Sterrett summarizes its structure concisely: "Both man and machine are required to impersonate. The machine's performance is not directly compared to the man's, but their rates of successfully impersonating against a real woman candidate are compared" [6].

Several pages later, Turing offers what appears to be a restatement. He writes: "Let us fix our attention on one particular digital computer C. Is it true that by modifying this computer to have an adequate storage, suitably increasing its speed of action, and providing it with an appropriate programme, C can be made to play satisfactorily the part of A in the imitation game, the part of B being taken by a man?" [2].

This formulation differs from the first. Here the machine plays the deceiver's role against a human who is simply being himself. The human does not impersonate anyone. The interrogator's task is to determine which respondent is the computer and which is the human. This is the Standard Turing Test. Sterrett notes that in this version, "only the computer is attempting to impersonate. The computer's performance is judged based on similarity to a man's performance" [6].

Turing apparently believed these formulations were equivalent. They are not. Sterrett demonstrates this by considering the quantitative results each test can yield. In the Original Imitation Game, nothing prevents the machine from scoring higher than the man. Suppose the man fools the interrogator

one percent of the time and the machine fools the interrogator three percent of the time. The machine has outperformed the man. This result is intelligible. Both contestants performed the same difficult task, and the machine did it better [6].

The Standard Turing Test admits no such possibility. If the interrogator misidentifies the computer as the human with greater than fifty percent frequency, this does not indicate that the computer performed a task better than the human. It might only indicate that the interrogator has peculiar biases. In the first Loebner Prize competition, one interrogator mistook a human for a computer because the human exhibited what the interrogator considered superhuman knowledge of Shakespeare [13]. This result tells us nothing about the computer's performance and everything about the interrogator's expectations.

The Standard Turing Test is also sensitive to interrogator skill in ways the Original Imitation Game is not. A skilled interrogator will more easily identify the computer. An unskilled interrogator will more easily be fooled. The test result fluctuates with the interrogator's abilities. In the Original Imitation Game, interrogator skill affects both contestants equally. If the interrogator is highly skilled, neither the man nor the machine will often succeed in impersonating the woman. If the interrogator is unskilled, both will succeed more often. The comparison between their success rates screens off interrogator skill as a variable [6].

Copeland has defended the Standard Turing Test by emphasizing that any practical test is merely a sampling of an ongoing situation. "A machine emulates the brain if it plays the imitation game successfully come what may, with no field of human endeavour barred, and for any length of time commensurate with the human lifespan" [10]. This defense acknowledges the test's impracticality but preserves its conceptual force.

Yet the defense does not address the structural difference between the two tests. Even in principle, the Standard Turing Test compares the machine's impersonation against the human's simple self-presentation. The Original Imitation Game compares impersonation against impersonation. These are different comparisons. They test different things.

5. What Impersonation Demands

Impersonation is not imitation. Imitation reproduces behavior. Impersonation performs a role while knowing that one is not that role's genuine occupant. The difference matters for understanding what intelligence requires.

Consider what the man must do in the Original Imitation Game. He must convince the interrogator that he is the woman. This means he cannot simply respond as he normally would. His normal responses would reveal his gender. He must instead produce responses that he judges would be given by a woman in his situation. This requires two operations. First, he must recognize when his automatic response would be inappropriate. Second, he must fabricate an alternative response that serves his deceptive purpose.

Sterrett describes this as "a self-conscious critique of one's ingrained responses" [6]. The critique has two aspects: recognizing and suppressing an inappropriate response, and fabricating an appropriate one. Neither aspect is trivial. Recognizing inappropriateness requires knowing what would give one away. Fabricating appropriateness requires knowing what would be convincing.

The man cannot rely on practice. He has spent his life learning to respond as a man. These responses are deeply ingrained. They operate automatically, without reflection. To override them, he must catch himself in the act of automatic response and intervene before the response emerges. This is the kind of self-monitoring that Ryle identified with genuinely intelligent performance. "The cleverness of the clown may be exhibited in tripping and tumbling. He trips and tumbles on purpose and after much rehearsal and at the golden moment and where children can see him and so as not to hurt himself" [11]. The clumsy man's tripping manifests nothing but accident. The clown's tripping manifests mind at work.

Impersonation requires this kind of purposive control over what would otherwise be automatic. The impersonator must perform the same external behavior as the genuine article while knowing that the internal relation to that behavior is entirely different. The genuine woman answers questions by drawing on her actual experience. The impersonator must answer by drawing on his knowledge of what a woman would say. His answers may be indistinguishable from hers, but the process producing them is different. The difference lies in the reflective awareness that his answers are fabrications.

This reflective awareness is what distinguishes thinking from mere mechanical response. French, in his critique of the Standard Turing Test, emphasizes that many of our ordinary responses depend on what he calls a "subcognitive substrate" [12]. This substrate consists of associations built up through a lifetime of embodied interaction with the world. When we answer questions about whether jackets can serve as blankets or whether pens make good weapons, we draw on this substrate without conscious reflection. Our answers emerge from patterns of association we did not explicitly learn and cannot explicitly articulate.

The impersonator cannot rely on this substrate when answering questions designed to reveal gender. His own substrate was built through male experience. It will produce male-typical responses. To impersonate successfully, he must override those responses and produce ones appropriate to a different experiential history. This requires explicit reasoning about what would be convincing, not automatic deployment of what comes naturally.

The task is doubly difficult because the interrogator actively seeks to expose the impersonation. The interrogator chooses questions designed to elicit gender-revealing responses. The impersonator must anticipate which questions will be traps and formulate answers that avoid them. This requires what

Sterrett calls "knowing how to use the knowledge that someone else knows how to draw conclusions" [6]. It is not enough to know what women typically say. One must know what the interrogator expects women to say, and what inferences the interrogator will draw from various answers.

This second-order knowledge is what makes the task genuinely intellectual. The impersonator must model not only the woman but also the interrogator's model of women. He must simulate the interrogator's reasoning and adjust his responses to block the inferences the interrogator would otherwise draw. This is the kind of recursive modeling that theorists of mind have identified with sophisticated social cognition [14].

6. The Standard Turing Test's Different Demands

The Standard Turing Test asks something else entirely. Here the machine must impersonate a human, but the human does not impersonate anyone. The human simply responds as he normally would. The interrogator's task is to identify which respondent is the machine and which is the human.

This asymmetry changes what the machine must do. It need not override its own automatic responses, because it has no automatic responses of the relevant kind. It need not model the interrogator's expectations about gender because gender is not at issue. It need only produce responses that fall within the range of humanly possible responses.

This is a much easier task. French demonstrates this with his "rating game" questions. "Rate banana splits as medicine," "Rate grand pianos as wheelbarrows," "Rate purses as weapons" [12]. These questions probe the subcognitive associations built through embodied experience. Humans answer them without reflection, drawing on patterns they could not explicitly articulate. A machine without that embodied experience will answer differently.

But the Standard Turing Test does not require the machine to answer such questions convincingly. It only requires that the machine's answers fall within the range of what a human might say. This range is vast. Humans vary enormously in their associations, their knowledge, their conversational styles. A machine can adopt a persona that excuses its differences. It can claim to be a non-native speaker, a child, a person with unusual background. The Eugene Goostman program that supposedly passed the test in 2014 adopted precisely such a persona: a thirteen-year-old Ukrainian boy with limited English [3]. This persona made its odd responses plausible.

The Original Imitation Game permits no such evasion. The machine must impersonate a woman, and its success is compared to a man's success at the same task. The man has a massive advantage. He shares with women a human body, a human lifespan, a human culture. His automatic responses are shaped by the same world that shapes women's responses. Yet despite this advantage, impersonating a woman remains difficult. Men do not succeed at it often.

The machine, lacking any human experience at all, must somehow produce responses that are not merely human-like but convincingly female. This is a far more stringent test. It requires not just similarity to human behavior but specific similarity to behavior shaped by a particular kind of human experience. The machine cannot simply be generally human-like. It must be specifically woman-like.

Sterrett emphasizes that gender is not essential to the test's structure. "Cross-gendering is not essential to the test; some other aspect of human life might well serve in constructing a test that requires such self-conscious critique of one's ingrained responses" [6]. The point is to choose an aspect of response that is deeply ingrained, pervasively relevant, and unlikely to have been practiced in the impersonator's own experience. Gender meets these criteria. It shapes responses from earliest childhood. It operates automatically and unconsciously. No one can unilaterally decide to stop responding as their gender. The task of impersonating the opposite gender therefore forces the kind of self-conscious critique that distinguishes thinking from mere mechanical response.

7. Modern Chatbots and Their Operation

Contemporary chatbots are large language models trained on vast corpora of human text. They learn statistical patterns of word co-occurrence. They predict what word is likely to follow given words. Through this simple learning objective and enormous scale, they acquire the ability to generate fluent, coherent text on almost any topic [15].

Their operation is entirely statistical. They have no experience of the world beyond the text they were trained on. They have never tasted strawberries, felt embarrassment, watched a sunset. They have read billions of words about these experiences but never had them. Their knowledge is knowledge of what people say, not knowledge of what things are.

This matters for impersonation. When a chatbot generates text, it does not override ingrained responses. It has no ingrained responses to override. It does not catch itself in automatic reaction and substitute a calculated alternative. It simply computes the most probable continuation given its input and training. There is no self-conscious critique because there is no self to do the critiquing.

The chatbot's fluency can deceive us. It produces sentences that a person might produce. It answers questions, tells jokes, offers opinions. We are strongly disposed to treat such behavior as evidence of mind. This disposition has been documented extensively in human-computer interaction research. Nass and Reeves found that people treat computers as social actors, applying the same social rules to them that they apply to humans [16]. We are "hardwired to respond to social cues" [17]. When a machine produces language, we cannot help but hear a voice behind it.

The Eliza effect amplifies this tendency. Named after Weizenbaum's early chatbot, it refers to "the susceptibility of

people to read far more understanding than is warranted into strings of symbols, especially words, strung together by computers" [18]. Weizenbaum himself was disturbed by how readily people attributed understanding to his simple program. He argued that confusing simulation with genuine understanding was a dangerous error [19].

Modern chatbots are vastly more sophisticated than Eliza. They can maintain coherent conversation over many turns. They can adapt their style to match their interlocutor. They can generate novel sentences on topics they were not explicitly trained on. This sophistication makes the Eliza effect more powerful, not less. We are even more susceptible to reading understanding into their outputs.

Yet the underlying operation remains the same. The chatbot has no experience, no automatic responses to override, no self-conscious critique of its own productions. It generates text by statistical prediction, not by reflection on what would be convincing in an impersonation.

8. Why Chatbots Cannot Pass the Original Imitation Game

Consider what would be required for a chatbot to impersonate a woman as successfully as a man does. The chatbot would need to recognize when its statistically-generated responses would be inappropriate for a woman. It would need to suppress those responses and generate alternatives that would be more convincing.

But the chatbot has no way to recognize inappropriateness. Inappropriateness for impersonation is not a statistical property of text. It is a matter of whether a response would reveal that the speaker lacks female experience. The chatbot has no model of female experience. It has only text about women, written by people with and without female experience. It cannot distinguish between text that reflects genuine experience and text that reflects second-hand knowledge.

The chatbot also has no way to suppress responses selectively. Its generation process is probabilistic. It samples from a distribution of possible continuations. There is no mechanism for identifying one continuation as automatic and substituting another. All continuations are equally automatic in the sense that all are produced by the same statistical process.

The self-conscious critique that impersonation requires is not statistical. It is reflective. It involves awareness that one's natural response would give one away and deliberate construction of an alternative. This reflective awareness is precisely what statistical language models lack. They have no awareness at all. They have no sense that they are responding, let alone a sense that their responses might be inappropriate for the role they are playing.

French's argument about subcognitive substrates applies here. He contends that "the [Standard] Turing Test provides a guarantee not of intelligence but of culturally-oriented human intelligence" [12]. His point is that passing the Standard Turing Test would require actually having lived a human life with human sensory capacities in a human culture. The same point applies more strongly to the Original Imitation Game. Passing that test would require not just human life but specifically female human life, or at least the capacity to simulate female experience so thoroughly that one could generate responses indistinguishable from those of a woman.

The chatbot has no experience at all. It cannot simulate female experience because it has no experience to simulate. It has text about female experience, but text is not experience. The difference is not quantitative but qualitative. No amount of text about strawberries equals the experience of tasting a strawberry. No amount of text about embarrassment equals the feeling of being embarrassed. Impersonation requires drawing on something like experience, not just on text about experience.

Sterrett makes this point through an analogy with learning a second language. A non-native speaker may learn the language perfectly, yet subtle cues can give them away to an expert interrogator. These cues are not failures of linguistic competence but markers of a different learning history. "There will be subtle cues that can give the non-native away, no matter how well he or she has learnt the second language and become informed of various regional idioms and dialects" [6]. These cues reflect ingrained responses that are not a matter of competence but of history.

The chatbot has no learning history in the relevant sense. It was not raised in a culture. It did not acquire language through interaction with caregivers. It absorbed text, not life. Its responses bear the marks of that absorption. They are fluent but rootless. They lack the specific texture that comes from having lived a particular life.

The Original Imitation Game would reveal this rootlessness precisely because it demands impersonation of a particular kind of life. The man impersonating a woman can draw on his observation of women, his interactions with them, his understanding of how they differ from men. These resources are imperfect, which is why he succeeds only rarely. But they are resources nonetheless. The chatbot has no comparable resources. It has text, but text underdetermines life. It cannot know what it has not lived.

9. What Impersonation Reveals About Intelligence

The Original Imitation Game suggests a conception of intelligence different from the one implicit in the Standard Turing Test. On the Standard view, intelligence is whatever produces behavior indistinguishable from human behavior. If a machine talks like a human, it thinks like a human. This is the view that Searle attacks with the Chinese Room argument. He imagines himself following rules for manipulating Chinese symbols without understanding them. From the outside, his

responses are indistinguishable from a native speaker's. From the inside, he understands nothing. Searle concludes that syntax is not sufficient for semantics, that formal symbol manipulation cannot produce genuine understanding [20].

The Chinese Room argument has generated endless debate. But the Original Imitation Game suggests a different approach to the issues Searle raises. The question is not whether the machine understands but whether it can perform a specific kind of task—impersonation, and whether it can perform that task as well as a human can. This formulation avoids the metaphysical puzzles about understanding while preserving something of what we care about when we ask whether machines think.

What we care about, on this view, is resourcefulness. Can the machine recognize when its automatic responses would be inappropriate? Can it override those responses and generate alternatives suited to the situation? Can it model what others expect and adjust its behavior accordingly? These are the capacities that impersonation tests.

They are also capacities we associate with intelligence in humans. When someone responds inappropriately to a situation, we say they were not thinking. When someone catches themselves before saying something they would regret, we credit their self-awareness. When someone successfully navigates a complex social situation by anticipating how others will react, we admire their social intelligence. The capacities tested by impersonation are the same capacities we look for in evaluating human thoughtfulness.

Ryle captured this in his distinction between knowing how and knowing that. Intelligence is not primarily a matter of knowing facts but of knowing how to do things, how to respond appropriately, how to adjust to circumstances, how to carry out purposes. "The cleverness of the clown may be exhibited in tripping and tumbling" [11]. The clown knows how to trip in a way that amuses rather than injures. This knowing-how cannot be reduced to knowing-that. It is manifest in performance, not in propositional knowledge.

The Original Imitation Game tests knowing-how. It asks whether the contestant knows how to impersonate a woman well enough to succeed as often as a man does. This knowing-how cannot be reduced to knowing facts about women. It requires the ability to deploy those facts in real time, to generate appropriate responses, to avoid traps. It requires resourcefulness.

The Standard Turing Test, by contrast, tests something closer to knowing-that. It asks whether the machine's responses fall within the range of humanly possible responses. This can be achieved by storing vast numbers of facts and retrieving them appropriately. The machine need not be resourceful. It need only be comprehensive.

Block's "Blockhead" thought experiment makes this vivid. He imagines a machine with a vast lookup table containing all possible conversations of a certain length. Given any input, it looks up the appropriate output and prints it. Such a machine would pass any finite test of conversational ability. Yet it would have no intelligence at all. It would be, Block says, about as intelligent as a jukebox [21].

The Blockhead objection is often raised against the Standard Turing Test. It is less effective against the Original Imitation Game. To impersonate a woman as successfully as a man does, the machine would need to generate responses appropriate to situations that cannot be anticipated in advance. The space of possible conversations is too vast for exhaustive pre-programming. The machine would need genuine resourcefulness, not just a comprehensive database.

This is the deeper insight of the Original Imitation Game. Intelligence is resourcefulness, not knowledge. It is the capacity to respond appropriately to novel situations, not the possession of a store of pre-packaged responses. Impersonation tests resourcefulness because it requires generating responses that are not one's own, that do not come naturally, that must be constructed on the spot.

10. Chatbots as Sophisticated Blockheads

Modern chatbots are not Blockheads in the literal sense. They do not store all possible conversations in a lookup table. They generate responses dynamically based on statistical patterns learned from training data. In this respect, they are more flexible than Block's imaginary machine. They can produce novel sentences, adapt to novel topics, generate responses that were not explicitly pre-programmed [15].

Yet in another respect, they are Blockheads after all. Their responses are determined by statistical patterns in their training data, not by genuine understanding of the situations they address. They have no experience of the world, only text about the world. They have no automatic responses to override, only probabilities to sample. They have no self-consciousness to deploy, only statistical inference.

The fluency they achieve is real. Their outputs are often indistinguishable from human outputs in short exchanges. This fluency tempts us to attribute understanding to them. We hear a voice behind the words and assume there is someone there.

But the Original Imitation Game reveals that this attribution is mistaken. What looks like understanding is actually something else: sophisticated pattern matching without comprehension. The chatbot can generate sentences that a woman might say, but it cannot impersonate a woman because it has no sense of what impersonation would mean. It has no sense of itself as having an identity that could be concealed or revealed. It has no sense of the interrogator as having expectations that could be met or frustrated. It has no sense at all.

This absence of sense is not a failure that could be remedied by more data or larger models. It is a structural feature of the kind of system chatbots are. They are not agents with purposes and histories. They are not selves with experiences and memories. They are statistical engines trained on text.

The Original Imitation Game makes this structural feature

visible. It demands that the contestant perform a task that requires agency: deliberately constructing a false appearance while knowing that one's true identity is different. This task cannot be performed by a system without agency. It cannot be performed by a system without identity. It cannot be performed by a system without self-consciousness.

Chatbots lack all of these. They are not agents but tools. They have no identity to conceal. They have no purposes to pursue. They generate text, but they do not speak. The distinction matters. Speech is action. It expresses intention, reveals character, commits the speaker. Text generation is just text generation. It commits no one to nothing.

Weizenbaum saw this clearly. He warned against confusing calculation with judgment, against mistaking the products of computation for the products of thought. "There is a difference between man and machine. There is a human use of computers. There is also a misuse" [19]. The misuse consists in treating machines as if they had human capacities when they do not.

The Original Imitation Game helps us see the difference. It asks not whether the machine can produce human-like text but whether it can perform a human-like task, impersonation. as well as a human can. This question reveals what chatbots cannot do, even as their fluency dazzles us.

11. Deception and Its Moral Dimensions

The Original Imitation Game involves deception. The man tries to deceive the interrogator about his gender. The machine, if it plays the game, tries to deceive the interrogator as well. Deception is central to the test's design.

This feature has troubled some commentators. Michie complains that the test "obliges candidates to demonstrate intelligence by concealing it" [21]. He suggests that this is perverse, that intelligence should be displayed openly, not hidden through trickery.

But this complaint misunderstands what deception reveals. Deception is not just concealment. It is active manipulation of another's beliefs. To deceive successfully, one must understand what another believes, what inferences they will draw, what evidence they will accept. This requires the second-order modeling that is central to social intelligence.

The man impersonating a woman must anticipate what the interrogator expects women to say. He must understand what inferences the interrogator will draw from various answers. He must adjust his responses to block those inferences. This is not mere concealment. It is active engagement with another mind.

Deception in this sense is a deeply intellectual activity. It requires theory of mind, the capacity to represent others' representations [14]. It requires strategic thinking, the capacity to plan moves with an eye to future consequences. It requires flexibility, the capacity to adjust plans when circumstances change.

These are capacities we associate with intelligence. They are not peripheral to thought but central to it. The most challenging social situations demand exactly this kind of recursive modeling. We navigate these situations daily, adjusting what we say based on what we think others think. This is not deception in the moral sense. It is the ordinary work of social life.

The Original Imitation Game isolates this capacity and tests it directly. It asks whether the contestant can model another's expectations and adjust responses accordingly. This is a far better test of social intelligence than any measure of factual knowledge or conversational fluency.

Chatbots fail this test not because they cannot produce deceptive text. They can. They can generate lies as easily as truths. The failure is deeper. They cannot model another's expectations because they have no model of others at all. They have statistical patterns, not representations of minds. They can simulate theory of mind by reproducing text about what people think, but they cannot actually think about what anyone thinks because they cannot think at all.

12. Implications for AI Evaluation

If the Original Imitation Game offers a better test of intelligence than the Standard Turing Test, then our current practices of AI evaluation are misguided. We celebrate chatbots for achieving human-level fluency in conversation. We treat their ability to generate plausible text as evidence of understanding. We ignore the deeper capacities that impersonation would reveal.

This matters practically. Chatbots are being deployed in roles that require genuine understanding. They provide mental health support [22]. They offer legal advice [23]. They serve as educational tutors [24]. If they lack the capacities that impersonation tests, they will fail in these roles in ways that may not be immediately obvious. They will produce plausible-sounding text that is nevertheless wrong, inappropriate, or harmful.

The danger is not that chatbots will deceive us intentionally. They have no intentions. The danger is that we will deceive ourselves, projecting understanding onto systems that have none. We will trust their outputs as we trust human testimony, and we will be betrayed not by malice but by our own credulity.

The Original Imitation Game offers a corrective. It reminds us that intelligence is not fluency but resourcefulness, not knowledge but judgment, not similarity to human behavior but capacity to perform difficult tasks. By asking whether machines can impersonate as well as humans can, it focuses attention on what matters.

This does not mean we should run the Original Imitation Game on every new chatbot. The test is impractical for routine evaluation. But its conceptual lessons can guide our thinking. When we evaluate a chatbot, we should ask not just whether its outputs are human-like but whether it can perform tasks that require genuine resourcefulness. We should probe its lim-

its, test its boundaries, see where it breaks. We should not assume that fluency implies understanding.

The history of AI is littered with overestimations. Every time a machine masters a domain previously reserved for humans, critics object that this is not really thinking. Chess was supposed to require intelligence, but Deep Blue played chess without intelligence [25]. Go was supposed to be beyond machines, but AlphaGo mastered it without understanding [26]. The goalposts keep moving.

The Original Imitation Game suggests why the goalposts keep moving. What we count as intelligence is not fixed. It shifts as machines master new domains. What remains constant is the sense that genuine intelligence involves something more than what current machines can do. That something more is resourcefulness, the capacity to respond appropriately to novel situations, to override ingrained responses when circumstances demand, to model others' expectations and adjust accordingly.

Modern chatbots lack this something more. They are fluent but not resourceful. They can generate text but cannot think. The Original Imitation Game reveals this lack precisely because it tests resourcefulness, not fluency.

13. Conclusion

The Original Imitation Game and the Standard Turing Test are not equivalent. They test different capacities. They yield different results. They have different implications for how we understand intelligence.

The Standard Turing Test tests similarity to human behavior. It asks whether a machine can produce responses indistinguishable from a human's. This test is passed by machines that are merely fluent, that have learned statistical patterns of human language without any understanding of what they say. The Standard Turing Test cannot distinguish genuine intelligence from sophisticated simulation.

The Original Imitation Game tests resourcefulness. It asks whether a machine can impersonate a woman as successfully as a man can. This requires overriding ingrained responses, modeling others' expectations, generating appropriate fabrications on the spot. These are the capacities we associate with thinking.

Modern chatbots pass the Standard Turing Test in limited contexts. They generate fluent, human-like text. They fool naive interlocutors. They create the impression of understanding. But they cannot pass the Original Imitation Game. They lack the self-conscious critique that impersonation requires. They have no ingrained responses to override. They have no model of others' expectations. They generate text without agency, without identity, without thought.

This failure illuminates something essential about intelligence. Intelligence is not about having the right responses but about knowing when to override them. It is not about fluency but about resourcefulness. It is not about similarity to human behavior but about capacity to perform difficult tasks. The Original Imitation Game captures this better than any other test.

Turing's genius was to see that impersonation could serve as a probe for intelligence. He chose gender because it is deeply ingrained, pervasively relevant, unlikely to have been practiced. The task of impersonating the opposite gender forces the kind of self-conscious reflection that distinguishes thinking from mere mechanical response. This insight remains relevant seventy-five years later, even as our machines have grown vastly more sophisticated.

We should attend to it. We should resist the temptation to equate fluency with intelligence. We should remember that chatbots, however fluent, are not agents with purposes and histories. They are tools, not thinkers. The Original Imitation Game helps us see this. It reminds us what genuine intelligence requires and what current machines still lack.

Abbreviations

AI	Artificial Intelligence
GPT	Generative Pre-trained Transformer
LaMDA	Language Model for Dialogue Applications

Author Contributions

Mohammed Zeinu Hassen: Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Visualization, Writing – original draft

Conflicts of Interest

The author declares no conflict of interest.

References

[1] Descartes, R. (1637). Discourse on the method. Leiden.

[2] Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433-460.

[3] Warwick, K., & Shah, H. (2015). Can machines think? A report on Turing test experiments at the Royal Society. Journal of Experimental and Theoretical Artificial Intelligence, 28(6), 989-1007.

[4] Grant, N., & Metz, C. (2022). Google sidelines engineer who claims its A.I. is sentient. New York Times, June 12.

[5] Jones, C. R., & Bergen, B. K. (2024). Does GPT-4 pass the Turing test? arXiv preprint arXiv: 2403.01234.

[6] Sterrett, S. G. (2000). Turing's two tests for intelligence. Minds and Machines, 10(4), 541-559.

[7] Block, N. (1981). Psychologism and behaviourism. Philosophical Review, 90(1), 5-43.

[8] Hayes, P., & Ford, K. (1995). Turing test considered harmful. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 972-977.

[9] DeAngelis, T. (2023). AI chatbots and mental health. Monitor on Psychology, 54(2), 40-46.

[10] Copeland, B. J. (2000). The Turing test. Minds and Machines, 10(4), 519-539.

[11] Ryle, G. (1949). The concept of mind. University of Chicago Press.

[12] French, R. M. (1990). Subcognition and the limits of the Turing test. Mind, 99(393), 53-65.

[13] Shieber, S. M. (1994). Lessons from a restricted Turing test. Communications of the ACM, 37(6), 70-78.

[14] Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? Behavioral and Brain Sciences, 1(4), 515-526.

[15] Brown, T. B., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.

[16] Nass, C., & Reeves, B. (1996). The media equation: How people treat computers, television, and new media like real people and places. CSLI Publications.

[17] Rubin, C. T. (2017). Mind games. The New Atlantis, 51, 108-127.

[18] Ariza, C. (2009). The interrogator as critic: The Turing test and the evaluation of generative music systems. Computer Music Journal, 33(2), 48-70.

[19] Weizenbaum, J. (1976). Computer power and human reason: From judgment to calculation. W. H. Freeman.

[20] Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417-424.

[21] Michie, D. (1993). Turing's test and conscious thought. Artificial Intelligence, 60(1), 1-22.

[22] Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent. Journal of Medical Internet Research, 19(2), e19.

[23] Sourdin, T. (2018). Judge v robot: Artificial intelligence and judicial decision-making. University of New South Wales Law Journal, 41(4), 1114-1133.

[24] Holmes, W., et al. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.

[25] Hsu, F. H. (2002). Behind Deep Blue: Building the computer that defeated the world chess champion. Princeton University Press.

[26] Silver, D., et al. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484-489.

Biography

Mohammed Zeinu Hassen is a philosopher and academic researcher based in the Department of Social Sciences at Addis Ababa Science and Technology University in Addis Ababa, Ethiopia. His research interests lie at the intersection of philosophy of mind, artificial intelligence, and the conceptual foundations of cognitive science. Hassen's work examines foundational questions about machine intelligence, consciousness, and the philosophical implications of artificial intelligence for human self-understanding. His scholarship engages with classic debates in the philosophy of AI while addressing contemporary challenges posed by advances in machine learning and language models. Hassen's approach combines careful textual analysis of historical philosophical texts with rigorous conceptual analysis of current AI technologies.