Sarah Hall and Alexis Zhai

Professor Olinick

Breaking the Code

21 November 2025

2020-2025 Responses to the Turing Test

Introduction:

Between 2020 and 2025, the rapid evolution of AI transformed the Turing Test from theory into reality, with models like GPT-4 and Gemini effectively playing the "Imitation Game". However, this technological achievement has led to complex philosophical debates about the future of AI and the nature of intelligence itself. This report evaluates the contemporary status of the Turing Test and presents some of the major responses/criticisms to Turing's original thesis.

Background Information:

Between 2020 and 2025, global events and technological breakthroughs created the perfect conditions for artificial intelligence to move rapidly into everyday life. The COVID-19 pandemic drew all attention to digital platforms, with schools, workplaces, and social life abruptly moving online. With this shift, demand for continued digital advancements grew exponentially, and the global reliance on computers paved the way for artificial intelligence to become woven into mainstream culture. During this same period, OpenAI's release of GPT-3 in 2020, followed by the launch of ChatGPT in 2022, was the foundation of what is referred to as the "AI Spark." The popularity of large language models such as ChatGPT has rapidly changed the way that machines are used, now being tasked with tasks previously referred to as deeply human, such as writing, brainstorming, art, and programming. However, with the advancement

of artificial intelligence came widespread anxiety around how AI was to be used, with fears around fake images, videos, altered audio clips, or the spread of misinformation becoming foreseeable problems with expanding machine capabilities. By the mid-2020s, AI development had transformed into an arms race among major companies, all seeking dominance in the highly marketable field of continued artificial intelligence innovation. As the AI exploded into a tangible part of human culture, questions surrounding the ethics and safety behind these computer systems became the topic of mainstream concern. Between 2020 and 2025, AI transformed from a future possibility to having a strong prevalence in everyday life, renewing debate around the Turing test and the desire to answer Turing's question: "Can machines think?"

Background Information About the Turing Test:

It is also worth noting that by 2025, the concept of the Turing Test shifted from a theoretical goal to a realized milestone. Two studies we found from this period highlight this process. The first study by Drew Turney in 2024 [1] focuses on the conversational imitation game. Researchers organized an experiment involving 500 human participants who engaged in five-minute conversations with four respondents: a human, the ELIZA (which is not an LLM), GPT-3.5, and GPT-4. The research shows that GPT-4 was judged to be human 54% of the time, while real human respondents were identified 67% of the time. This indicates that in such an experiment, GPT-4 can already deceive human interrogators and is not far from becoming as good as humans in playing the imitation game. The second study by Mei et al. in 2024 [2], titled *A Turing test of whether AI chatbots are behaviorally similar to humans*, tried to measure AI through a "Behavioral Turing Test". The research focused not only on conversation, but action and decision-making process. They analyzed how AI solved behavioral economic games like the Trust Game and generated a "Big Five" personality survey of GPT-3.5 and GPT-4. Results

showed that GPT-4's personality traits were already very close to, and statistically indistinguishable from, human average data. As a result, we can see that after a leap from GPT-3.5 to GPT-4, AI is now very close to the point where they are functionally indistinguishable from humans.

Positive Comments on the Turing Test:

Several scholars have defended the value of the Turing test, placing high value on the test as a framework for philosophical debates about machine intelligence. For example, García in his 2024 paper *Thought experiments in the Jefferson-Turing controversy: A Kuhnian perspective* reconstructs the arguments of Jefferson and Turing, identifying the dispute as a powerful identification of several crucial "scenarios of controversy," which are built upon contrasting relevant experiences [3]. García argues that the test functions best as a thought experiment rather than purely as a question of machinery intelligence, highlighting how many debates surrounding artificial intelligence are grounded not in a factual difference in understanding but from fundamentally different conceptual bases. By reconstructing the arguments between Turing and Jefferson, García demonstrates how the Turing test continues to be a useful tool in examining the question of machine intelligence. Similarly, Traynor's 2021 paper *Beyond the limits of imagination: abductive inferences from imagined phenomena* highlights the philosophical benefit of the Turing test, emphasizing how the imitation game is useful in moving debates about machine intelligence beyond purely conceptual disputes, rather focusing on basing the arguments on empirical evidence. Traynor sees this change as a way to move away from questioning solely what intelligence is, and rather questioning what behaviors and evidence from machines would lead one to infer machine thought. In his paper, Traynor explains how "We can see Turing's test as attempting to break free from conceptual arguments, by showing that an alternative concept of

machines is involved in the best explanation of certain possible data" [4]. His understanding of the test paints the imitation game as a tool for investigating what could count as evidence of thinking.

Another scholar, Ivie writing in 2022 in *Metaphor: Key to Critical and Creative Thinking* builds of this positive understanding of the Turing test, focusing more on the conceptual and creative value of Turings approach to understanding machine thought. Ivie views the Turing test as a creative depiction that helped lead to the conceptual conditions for the development of AI, seeing them as more than just a prediction of what AI could become. According to Ivie, "The idea of the Turing machine made a bridge between the logical and the physical worlds, thought and action, which crossed conventional boundaries" [5] suggesting that Turing's descriptions of the machine in his proposed imitation game allowed researchers to think about thought computationally, aiding in the development of AI. This understanding of the Turing test highlights Ivie's take on the test as maintaining value as a creative catalyst for machine development and interest in artificial intelligence.

Finally, in 2025, Mühlhoff in What AI Are We Talking About? emphasises the sociological importance of the Turing test. Arguing that "the practice of projecting intelligence onto machines was profoundly shaped by the notion of the Turing test" [6]. Rather than serving as a black and white benchmark that determines if a machine is intellegent, Mühlhoff sees the test as a cultural lens used by many to view machine behavior. This means that the Turing test not only dictacts how individual researchers examine systems, but also how ordinary uses of programs such as ChatGPT understand them.

Together, although each with unique individual interpretations, these scholars between 2020 and 2025 all underscore the importance of the Turing test far beyond determining if a

machine is capable of fooling a human. Rather, the test has morphed into a conceptual tool of understanding and questioning the way that researchers, philosophers, and the general public understand machine intelligence and behavior. Instead of viewing the test as a pass or fail benchmark, the test continues to shape the intellectual interpretation of AI, sparking debates on what it means to think, and the question of artificial thought.

Negative Comments on the Turing Test:

On the other hand, the Turing Test received a number of negative comments in the 2020s. We will first introduce a fundamental critique that lies in the test's reliance on behaviorism – the assumption that external performance is sufficient to prove internal experience. In *The Measurement Problem of Consciousness* [7], Browning and Veit argue that behavioral similarity does not imply mental similarity. While functions like pain response, navigation, or decision-making that we evolved can be achieved in artificial systems, they are achieved entirely differently (or non-consciously): "Although a particular function may be realized by a robot that has no consciousness, in an evolved animal, it is through consciousness that this function is realized." This suggests that complex behavior achieved through evolution (which definitely involves consciousness) may be achieved much more simply and "nonconsciously" in artificially designed systems. Although AI might exhibit the same behavior, it can only do so in a non-conscious way through mechanical processes. This sort of algorithmic mimicry, according to the authors, is not human thought.

Valiunas' & Beck's Papers: the Unpredictable Nature of Thinking

Having considered negative comments on the Turing Test, we will now examine three major objections in detail. The first objection is about the unpredictable nature of human

thinking, highlighted in Valiunas' 2020 article, *Turing and the Uncomputable* [8] and Beck's 2020 paper, *Do We Want Dystopia?* [9].

To begin with, Valiunas' paper states that Turing "boldly suggested that appearance was as good as actuality when it came to identifying the machine's behavior: If it appeared to be thinking, then we may as well say it is thinking." The author believes that Turing's argument can be transformed into this: the human mind and the machine operate on the same principles of logic. Consequently, the psychological mind could properly be described in terms of Turing machines because they both lie on the same level of description of the world.

Valiunas, however, believes human minds and Turing machines are fundamentally dissimilar: "The mind of man and that of the Turing machine are essentially different in kind – not just in the material conditions of their existence but in the way they function, their logical structures." He believes that it is an essential property of the mechanical systems which we have called "discrete state machines" that the phenomenon – unpredictable overwhelming effect from small initial errors – does not occur, "He (Turing) is describing a (universal) machine whose mind operates with a perfection that the human mind will never possess — yet that is also constrained by necessity as the human mind is not." As a result, human minds cannot be described by Turing machines alone, which are fundamentally different.

This theme is also mentioned in Beck's paper, in which the author states: "A machine can be taught when to bend a rule only by supplying it with more rules. This is not unpredictability, nor is it thought, nor, most importantly, does it result in compassion or love." A simulation, no matter how persuasive, is fundamentally limited compared to human thought and experience. Beck further provides a quote that highlights this difference: "What makes a diamond is how it came about, not what it looks like beneath the loupe." As a result, both authors conclude that the

Turing Test is fundamentally flawed because it fails to consider the internal and unique quality of human thought: unpredictability.

Hasselberger's Paper: Thinking is More Than Calculation

The second major objection to the Turing Test is that it falsely reduces thinking to mere calculations. This is illustrated in Hasselberger's 2021 paper, *Can Machines Have Common Sense?* [10].

The paper is Hasselberger's review of Erik J. Larson's book about artificial intelligence. In the book, Larson challenges the optimism he called "AI myth", which thinks the discovery of human-level AI is inevitable. Larson believes that although nowadays AI is good at deduction and induction (deep thinking), it cannot perform any kind of "abduction", which means insightful hypothesizing that seeks the best explanation of some particular event or phenomenon. According to Larson, abduction is "a leap to a previously unforeseen explanation" and it is something AI cannot do.

Haseelberger, in his paper, then argues that the Turing Test has some kind of "intelligence error": the idea that human intelligence can be reduced, without remainder, to calculation and problem-solving. Hasselberger believes that intelligence involves more than logical inference. True intelligence should include "anthropo-centric" concepts like love, trust, betrayal, hope, guilt, etc. The belief that intelligence can be understood in terms of only computation or calculation is incomplete. Hasselberger fears that focusing on the Turing Test may cause humanity to become dependent solely on algorithm-like decision-making.

The paper also mentions that currently, machines pass the Turing test mainly through "trickery and evasion", such as repeating the content of the person's statements in the form of a

question, changing the subject, being evasive instead of flexibly responsive, etc. However, a real Turing Test should require the computer to succeed through genuine comprehension of natural language. Hence, the Turing Test is fundamentally flawed because it fails to measure the abductive and emotional reasoning that constitutes human intelligence.

The Turing Trap Paper:

The last objection is about the concept of the Turing Trap. It is illustrated in Brynjolfsson's 2022 paper, *The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence* [11], and it might be the most direct attack on Turing's theories.

The paper discusses the profound economic and political consequences of focusing AI development on "mimicking human capabilities", a pursuit often guided by the goal set by the Turing Test. The author states that "Alan Turing proposed a test of whether a machine was intelligent: could a machine imitate a human so well that its answers to questions were indistinguishable from a human's? Ever since, creating intelligence that matches human intelligence has implicitly or explicitly been the goal of thousands of researchers, engineers, and entrepreneurs." As a result, building more advanced human-like AI (HLAI) became the primary goal of modern science.

However, the author believes that this might lead us into a "Turing Trap": the danger posed by an excessive focus on deploying HLAI. While HLAIs offer benefits, such as productivity and increased leisure, they can lead to machines becoming better substitutes for human labor, causing unemployment, unfair opportunities distribution, and increased social inequality.

To explain his argument, the author first differentiates the idea of “automation” from the idea of “augmentation”. Automation, according to the paper, is when AI “replicates” existing human capabilities. In this case, machines will become better substitutes for human labor, causing workers to lose economic and political bargaining power, leading to the Turing Trap. Conversely, augmentation is when AI focuses on “augmenting” human capabilities, enabling us to do things we never could before. In this way, humans and machines will become complements rather than substitutes.

The paper argues that automation eventually will reduce the marginal value of workers. Gains will disproportionately go to owners, entrepreneurs, and architects, resulting in greater wealth and political power concentration. The resulting risk is humankind being “trapped in an equilibrium in which those without power have no way to improve their outcomes.” Hence, we should reverse the excess incentives for automation. According to the paper, "A good start would be to replace the Turing Test, and the mindset it embodies, with a new set of practical benchmarks that steer progress toward AI-powered systems that exceed anything that could be done by humans alone." Based on these ideas, the Turing Test is argued to be economically risky, as it directs technological progress toward the displacement of human labor rather than the expansion of human potential.

Conclusion:

In conclusion, works in the 2020s defined the limits of Turing’s vision. Technically, the imitation game has been won, and AI LLMs like GPT-4 and Gemini are already close to being functionally indistinguishable from humans in many ways. However, some believe that the successful replication of function and behavior is not enough, as algorithmic mimicry cannot produce the unpredictable, anthropo-centric, conscious nature of human thought. Furthermore,

Brynjolfsson's warning of the "Turing Trap" [11] presents a potential risk of prioritizing labor automation over human augmentation. Based on these ideas, it is safe to conclude that passing the Turing Test does not necessarily mean the arrival of true mechanical intelligence. Rather than viewing it as the finish line, we shall see it as a gateway to a new era of scientific exploration and technological evolution.

Bibliography:

1. Turner, B. (2024, July 25). GPT-4 has passed the Turing test, researchers claim. Live Science. https://www.livescience.com/technology/artificial-intelligence/gpt-4-has-passed-the-turing-test-researchers-claim

2. Mei, Q., Xie, Y., Yuan, W., & Jackson, M. O. (2024). A Turing test of whether AI chatbots are behaviorally similar to humans. Proceedings of the National Academy of Sciences, 121(9), e2313925121. https://doi.org/10.1073/pnas.2313925121

3. Traynor, M. (2021). Beyond the limits of imagination: abductive inferences from imagined phenomena. Synthese, 199(5/6), 14293–14315. https://www.jstor.org/stable/48691642

4. García, P. (2024). Thought experiments in the Jefferson-Turing controversy: A Kuhnian perspective. Theoria: An International Journal for Theory, History and Foundations of Science, 39(1), 43–65. https://doi.org/10.1387/theoria.24674

5. Ivie, S. D. (2022). Metaphor: Key to Critical and Creative Thinking. Journal of Thought, 56(1/2), 55–76. https://www.jstor.org/stable/27184756

6. Mühlhoff, R. (2025). What AI Are We Talking About? In The Ethics of AI: Power, Critique, Responsibility (1st ed., pp. 27–37). Bristol University Press. https://doi.org/10.2307/jj.20433222.7

7. Browning, H., & Veit, W. (2020). The Measurement Problem of Consciousness. Philosophical Topics, 48(1), 85–108. https://www.jstor.org/stable/48628587

8. Valiunas, A. (2020). Turing and the Uncomputable. The New Atlantis, 61, 44–75. https://www.jstor.org/stable/26898500

9. Beck, S. (2020). Do We Want Dystopia? [Review of The Passengers; Zed; Machines Like Me, by J. Marrs, J. Kavenna, & I. McEwan]. The New Atlantis, 61, 87–96. https://www.jstor.org/stable/26898502

10. Hasselberger, W. (2021). Can Machines Have Common Sense? [Review of The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do, by E. J. Larson]. The New Atlantis, 65, 94–109. https://www.jstor.org/stable/27020762

11. Brynjolfsson, E. (2022). The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence. Daedalus, 151(2), 272–287. https://www.jstor.org/stable/48662041