Responses to the Turing Test from 2000 to 2010

Cassius Reed, Abrahm Geissinger

The Enigma of Alan Turing

Middlebury College

Alan Turing's 1950 paper "Computing Machinery and Intelligence" introduced the controversial idea of the Turing Test: a test which could be applied to many sorts of machines to determine whether or not they possessed intelligence. The test is simple; if a computer could out perform a human in the imitation game- a game in which a computer and a human attempt to convince a human interviewer of their humanness- the machine would be considered intelligent. Over the 75 years since the publication of Turing's paper, many philosophers, mathematicians, and computer scientists have published their own responses. In the decade following the turn of

the century, the authors of these responses explored many new objections, agreements, and shortcomings.

The first article published after the turn of the century was in some ways the most comprehensive. Selmer Bringsjord, Clarke Caporale, & Ron Noel's article: *Animals, Zombanimals, and the Total Turing Test: The Essence of Artificial Intelligence,* proposed the idea of the philosopher's zombie and built a thorough philosophical argument against the Turing Test. Jose Hernandez-Orallo in his 2000 article *Beyond the Turing Test* argues the test lacks specificity and proposes tests with more comprehensive evaluations. Mark Halpern's *The Trouble with the Turing Test* (published 6 years later) has similar critiques. Ari N. Schulman provides final objections to the Turing Test in: *Why Minds Are Not Like Computers.* (2009) Schulman argues human consciousness and understanding cannot be replicated by computational systems, challenging the AI-as-mind analogy.

Hugh Loebner's *The Turing Test* (2006) and Stuart M. Shieber's *The Turing Test as an Interactive Proof* (2007) attempts to provide arguments for the Turing Test's legitimacy. Loebner discusses the practical side of the Turing Test—especially in regards to the Loebner Competition for realistic AI- Shieber reframes the Turing Test as an interactive proof process, suggesting that intelligence could be demonstrated statistically rather than logically.

The early 2000s were a time of tremendous technological change, and it is unsurprising that there would be additional discussion of AI and the Turing test. However, it is first useful to have a sense of technological developments from 2000-2009. Broadband internet became a true reality in 2000, allowing widespread internet access. Social sites, blogs, and video sharing platforms became possible, providing a way for far more people to access and use the internet in their daily lives. Apple's iPhone and Google's Android brought smartphones to the masses, and

these quickly became ubiquitous. Laptop computers also started to replace desktops as components shrunk while becoming more powerful.

Perceptions of AI began to shift after 2000, but AI was still generally seen as a thing of the future. Pop culture, especially in the form of movies, molded public opinion around AI. Movies like *I-Robot*, *Terminator 3*, and sequels to *The Matrix*, continued the idea that AI could potentially be dangerous, especially if it attained consciousness. Certain technological advancements gradually introduced the public to early concepts of AI, with systems such as search and spam filters, speech recognition, and game AI such as chess robots. However, these advances were mainly seen as clever programming and not as true advancement in AI.

Mark Halpern and Christopher Ariza reflect the culture of skepticism growing throughout the early 2000s in their articles: *The Trouble with the Turing Test*, and *The Interrogator as Critic: The Turing Test and the Evaluation of Generative Music Systems* respectively. Halpern asserts the Turing Test is an inherently flawed methodology for determining machine intelligence. He claims that when successful, machine intelligence is merely a product of prior human ingenuity. For example, an intelligent response on behalf of a machine is simply the result of intelligent programming on the part of the human. His critique is largely against the philosophical underpinning of the test. He argues that possessing the ability to supply good answers is not equivalent to possessing intelligence. Halpern furthers his critique of the Turing Test with criticism of the Loebner competition (A competition for creating the most anthropomorphic artificial intelligence). Halpern draws from examples in which judges misidentified humans as machines as evidence for the insufficient nature of the Turing Test.

Christopher Ariza's paper denounces the use of the Turing Test on generative music systems. He claims that since music lacks natural language discourse, and since lingual discourse

is a key element of the Turing Test, the use of the Turing Test on generative music systems is a misapplication. While Ariza's argument against the Turing Test is largely denouncing incorrect applications, he claims that it can often lead to overestimation of machines. Ariza provides as close to a neutral perspective as can be found in this time period.

Throughout 2000-2009, there were very few papers that actually backed the Turing Test in its entirety. Arguments for it focused mainly on whether it was technically possible, and not on the philosophical question of whether it actually measures intelligence. Hugh Loebner, the sponsor of the eponymous Loebner competition, a contest for who could make the most human-like AI, responded to Halpern's aforementioned article with an argument supporting the Turing Test. His response to Halpern's article was centered on refuting Halpern's claims regarding the execution of the competition. The counterargument was that Halpern shouldn't have used the 1991 competition as a case study as it was rife with errors when compared to Turing's framework. The most notable restriction was that the competition limited dialogue to one area of conversation. In an effort to correct these criticisms and improve the credibility of the contest, competitions after 2004 were corrected to conform to Turing's core "Imitation Game" structure. This requires a paired comparison in which a judge compares exactly two entities (one human, one computer). He also clarifies that Turing's prediction about a 70 percent chance of correct identification after five minutes by the year 2000 was a rough benchmark, not a be-all end-all.

Shieber defends the TT as a sufficient criterion for intelligence by reinterpreting it as a statistical, rather than deductive, mechanism.
He argues that the TT should be viewed as an interactive proof (a concept borrowed from theoretical computer science), not a classical deductive proof.

Despite these two positive responses, overall responses to the Turing Test over the course of the decade were critical. In his 2000 article, Hernandez-Orallo finds the Turing Test to be useless outside of philosophical exercises. He claims that the Turing Test provides no objective definition of intelligence. He claims that the test is rather an anthropomorphical test and measures little more than humanness. Hernandez-Orallo suggests moving past the Turing Test to more comprehensive evaluation systems that measure specific cognitive abilities not imitation.

Schullman ends the decade with *Why Minds Are Not Like Computers (*2009*)*, which builds off of Hernandez-Orallo's objections to the Turing Test be anthropomorphical. Schullman argues that the test is measuring humanness—which (Schullman argues) is impossible for machines to replicate. He argues that the assumption that the mind's complexity is reducible to computational concepts is based on an inherent misunderstanding of minds and computers. He asserts the Turing Test is based on the idea of the mind as reducible to inputs (senses), processing (the mind), and outputs (speech movement etc.) Schullman objects to the external input-output view of the mind as it omits internal, first person experience, which he claims is a key part of intelligence. Schullman also notes the shift in the goals of AI. He notes that early efforts aimed towards replicating reasoning and the thought process, but after successive failures, AI goals have regressed to replicating the brain neuron by neuron without understanding intelligence at all.

Nine years earlier, Bringsjord et al. objected on the same basis through the analogy of the philosopher's zombie or the Zombanimal. The philosopher's zombie is a theoretically possible human being, in which every neuron has been replaced by a chip which performs the exact same function. This would create a "person" who acted identically to their real counterpart, but had no understanding of the processes they performed. Bringsjord et al. ultimately make the argument

that the Turing Test does not fail to test for intelligence, but is rather insufficient to test for personhood. They argue that personhood requires certain properties, such as free will, creativity, phenomenal consciousness, and robust abstract reasoning, which are exceptionally hard to reduce to circuits and algorithms using current AI techniques.

This shift towards evaluating AI on personhood, consciousness, or emotional intelligence, as opposed to pure intelligence, continued throughout the decade and became more prevalent in the following 15 years. Ultimately, many of the articles most critical of the Turing Test express a deep fear that one day, machines will truly be able to outsmart humans. To avoid this frightening concept many authors shift the goal posts in an effort to maintain a clear, definable difference between humans and machines.

Over the course of the decade following the turn of the century, articles written in response to the Turing Test expressed many different grievances and agreements. Most arguments against the Turing Test were based on philosophical objections. Many authors argued the test lacked specificity and tested for imitation rather than intelligence. Attempts to keep the Turing Test relevant focused on its technical simplicity and its accuracy in comparison with humans. The attitude to the Turing Test in the 2000s highlighted the growing cultural fear of artificial intelligence, and the sudden and large developments in tech. Many articles encapsulate this fear of AI encroaching on or surpassing human intelligence. Authors shifted the standard of successful AI from simple intelligence to emotional intelligence, and independent creative thought. As AI continued to improve, people have continued to shift the standard further from simple intelligence in a scrambling effort to identify and maintain a clear difference between humans and machines.

Works Cited

Ariza, Christopher. "The Interrogator as Critic: The Turing Test and the Evaluation of

Generative Music Systems." *Computer Music Journal* 33, no. 2 (2009): 48–70. The MIT Press.

http://www.jstor.org/stable/40301027.

Bringsjord, Selmer, Clarke Caporale, and Ron Noel. "Animals, Zombanimals, and the Total

Turing Test: The Essence of Artificial Intelligence." *Journal of Logic, Language, and

Information* 9, no. 4 (2000): 397–418. Springer. http://www.jstor.org/stable/40180234.

Halpern, Mark. "The Trouble with the Turing Test." *The New Atlantis* 11 (2006): 42–63.

Center for the Study of Technology and Society. http://www.jstor.org/stable/43152219.

Hernandez-Orallo, Jose. "Beyond the Turing Test." *Journal of Logic, Language, and

Information* 9, no. 4 (2000): 447–466. Springer. http://www.jstor.org/stable/40180237.

Loebner, Hugh. "The Turing Test." *The New Atlantis* 12 (2006): 5–7. Center for the Study of

Technology and Society. http://www.jstor.org/stable/43152236.

Schieber, Stuart M. "The Turing Test as Interactive Proof." *Noûs* 41, no. 4 (2007): 686–713.

Wiley. http://www.jstor.org/stable/4494555.

Schulman, Ari N. "Why Minds Are Not Like Computers." *The New Atlantis* 23 (2009): 46–68.

Center for the Study of Technology and Society. http://www.jstor.org/stable/43152940.