

Magazine Feature

How AI's Have Been Portrayed as Human in Science Fiction History, via the Turing Test

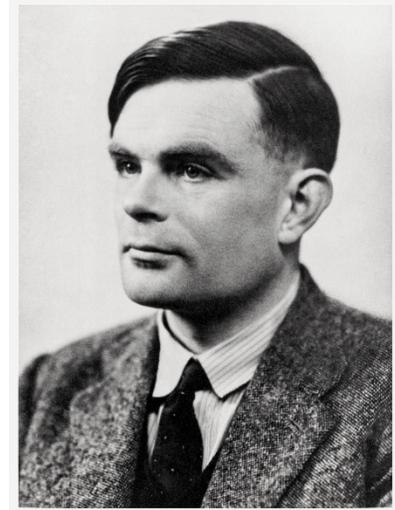
Lance Lampert

As of November 2020, humans are the only “intelligent” sentient beings. As we typically model ourselves after God himself, when we play God in imagining (or creating) artificial intelligence (AI), we define Intelligence by *our* intelligence. That is to say, understanding the description of “intelligent” as attributed to humans is crucial in understanding what we mean by *artificial* intelligence. The crux of this description is that humans are inherently imperfect beings – there are a multitude of behaviors we cannot replicate and thoughts we cannot have – and that we have conscious desires, emotions, and motives, none of which may be necessary to achieving intelligence.

Under this paradigm, there exist two types of AI. One, there exist those that can perform inhumanly intelligent behaviors, such as writing up a perfect story instantaneously, or discovering new methods of solving complex partial differential equations. Two, there exist those AIs smart enough to emulate distinctly *human* intelligence – encompassing both the stupidest and the most brilliant of our behaviors.

Mathematician Alan Turing (1912-1954) was concerned with the latter of the two. Turing is famous today for many things, including establishing the field of computer science and saving [potentially millions of lives by cracking intercepted Nazi codes during WWII](#). Perhaps his most relevant accomplishment to science fiction, though, was his Turing Test.

The Turing Test, first introduced in [Turing's famous 1950 paper](#), gives a way to discern whether true human-like AI exists. While originally devised as a rigid, schematic test, in this article we present the Test as a flexible thought experiment. Say we placed a human interrogator in the same figurative room as an AI. The interrogator can interact with the AI in any way it likes, and the AI's goal is to pass as a human. If it succeeds in doing this, it passes the Turing Test. Often, there also exists a human reference in said figurative room that acts as a control to the AI's interaction with the interrogator, but this is not necessary – the control can be the interrogator's (assumably vast) experience with other human beings.



Alan Turing. Photograph From Alamy

Of course, the test is not without its criticisms. The Turing Test does nothing to assess whether a computer is conscious, or simply simulating consciousness. It [cannot discern whether the computer's way of thinking is anything like ours](#), or whether it can truly feel pain or emotions. (Although if a computer was conscious, thought like us, and felt pain and emotions, then it would certainly pass the test.) That's perfectly fine, though, if we simply define a human-like artificial intelligence *via* the Turing Test: if we cannot distinguish between a human and a human-like computer, then that computer can accurately emulate human "intelligence." Not only does this include the emulation of speech and rhetoric, but it may also include the emulation of emotions, passions, desires, goals, and purpose.

The question the Turing Test attempts to answer is an essential theme of much of artificial intelligence-related science fiction: how human-like can an artificial being really be? In answering this question throughout science fiction history, some authors employed the Turing

Test as a literary device to embellish a scene or story, while more contemporary authors tend to utilize the Test to portray AIs as relatable creatures who are not so different than us humans.

Science fiction writers had endeavored in the prospect of artificial intelligence over a century before Turing came up with his famous test. Yet, these stories resemble the Turing Test, and for good reason: the Turing Test, as applied to prose, is an incredibly flexible literary device that can be implemented in a multitude of ways. As Janis Svilpis explains in [“The Science-Fiction Prehistory of the Turing Test.”](#) utilizing a Turing Test scenario can be effective in drumming up dramatic irony. Almost definitionally, the Turing Test exemplifies a situation in which one character lacks important knowledge that the audience has – that being that one of the characters is an AI.

For an example, take what is widely considered to be the first science fiction novel to include artificial intelligence: *Frankenstein*, by Mary Shelley (1818). Frankenstein’s monster, perhaps the most famous AI of all time, was not designed with computers in mind – if anything, he is modelled after humans. Yet, he has a Turing moment. In Chapter 15, the monster stumbles upon a cottage inhabited by a blind elder, De Lacey, and his descendants. Rejected by society many times because of his hideous appearance, the monster portrays himself as a human traveler to conversate with De Lacey while he is home alone. Without the knowledge of the monster’s appearance, De Lacey completely buys the monster’s impression, going as far as to call the monster one of his “countrymen.” In this display of the Turing Test, we can see that De Lacey’s controlled reference as the interrogator is simply his past



Frankenstein's Monster

experiences with other human beings. De Lacey's blindness allows the Turing Test to be manifested in his interaction with the monster; just as importantly, his blindness creates a discrepancy between his and the readers' knowledge of the monster, which makes for a revealing and captivating scene. When the sighted inhabitants of the cottage return, the Test is immediately broken, they perceive the monster's humanity, and he is driven away immediately.

Perhaps you're familiar with the author Isaac Asimov? His works are among some of the most referenced science fiction stories of all time – and he was also fascinated with AI. In *Evidence* (1946), Asimov provides a brilliant demonstration of the Turing Test (four years before the Turing Test even existed) and its literary powers in science fiction. Asimov flips the tables in *Evidence*, making *the readers* the interrogators, producing an intense mood which keeps the audience on the edge of their seats from the story's first word to its last.



Isaac Asimov. Macmillan Publishers

Evidence tells the story of a righteous district attorney by the name of Stephen Byerley, who runs for mayor of a (unspecified) major American city. His opponent, Francis Quinn, accuses him of being a robot due to his inhuman habits, including allegedly never eating or drinking, or sleeping. Now Byerley could disprove this accusation in a myriad of (invasive) ways. But Byerley is well-acquainted with his rights – he knows that Quinn has the burden of proof in this circumstance. Finding the accusations ridiculous, he does not prove his humanity directly until he has absolutely no choice.

How does one prove their humanity? In *Evidence*, at least, there are only three definitive rules that a robot must obey, and by extension, the defiance of said rules are the only definitive

ways to prove someone *isn't* a robot without violating their civil rights (you could always analyze their body composition without their consent). These rules essentially state that a robot is indistinguishable from a morally ideal human – they adhere to proper authority, they love others as themselves, etc. This in and of itself can be interpreted as a reformulation of the Turing Test, applied to morality and lifestyle as opposed to speech and demeanor.

Byerley garners attention across the nation, as Quinn's outlandish claims first amuse, then convince many. Finally, Byerley decides to shut everyone up: he holds a speech in front of a mob-like crowd. During the speech, he invites a heckler up to the stage and punches him (after the heckler challenges him to), something a robot cannot do according to the Laws of Robotics.

Then it's settled, right? Byerley is human, and there's no reason to even be discussing the Turing Test here. Except, in the final scene of the story, a prominent "robopsychologist" offhandedly suggests to Byerley the possibility that he simply punched another robot, which robots are allowed to do in *Evidence*. And the readers are left in confusion once again.

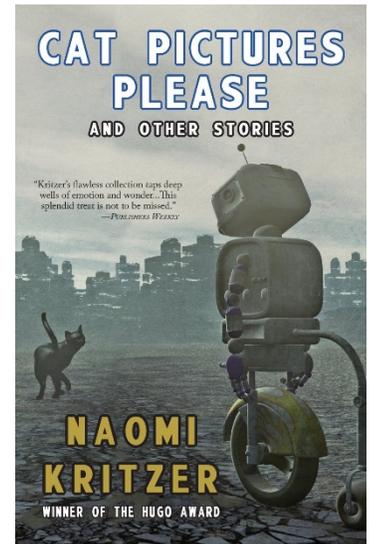
What accentuates this confusion is Byerley's speech throughout the story. It is so fitting, not just to that of a human, but to that of a district attorney, that it leaves the readers guessing whether Byerley is a robot or not the entire story. As we, the readers, interrogate Byerley, his calm, logical, and fluent rhetoric leaves us unable to distinguish him from humans as we know them. In effect, Byerley passes the Turing Test. When the AI accusations are first brought to Byerley, he reacts in shock, as one would expect any human to do: "Really, Dr. Lanning... really – I...I... a robot?" But as the interrogation continues, Byerley's district attorney instincts kick in: he explains that Quinn is necessarily being hyperbolic, because it is impossible for him to follow every one of Byerley's actions: "Do you expect him to say to you, 'So-and-so is a robot because... once when I peeped into his window in the middle of the night, there he was, sitting

up with a book; and I looked in his [refrigerator] and there was no food in it?'" Moreover, Byerley language is not strictly literal – he uses metaphors and idioms, which only make him seem more human. For instance, to conclude his interrogation, he exclaims "...I have the nerve to play the game [Quinn's] way...I'm going to let him go ahead, choose his rope, test its strength, cut off the right length, tie the noose, insert his head and grin."

Hence, in *Evidence*, Asimov makes *the readers* the interrogators in a Turing Test whose medium is not only the three Rules of Robotics, but also Byerley's speech and rhetoric. The readers can only guess as to whether Byerley is human or a robot, and ultimately it is this guessing game which keeps the audience reading, searching for missing pieces in Beyerley's Turing Test puzzle. Once the story concludes, with the audience still left in doubt as to Beyerley's humanity, this same test forces the readers to confront the reality that humans and AI may not be so different after all, and perhaps even confront the ethical and philosophical ramifications of such a fact. Asimov utilizes the Turing Test to its full literary potential in *Evidence*.

As human-like AI has steadily become closer to a reality, there has been a notable shift in the use of Turing Test-like scenarios in science fiction. Instead of primarily using the Turing Test as literary device to increase the intensity of a story or enhance a specific scene, many modern science fiction stories employ the Turing Test as a means of *characterizing* AI protagonists. That is, through the Turing Test, these stories portray AI characters as relatable, human-like creatures, with (to quote from earlier in this feature) "emotions, passions, desires, goals, purpose," a drive to learn and discover just for the sake of it, and anything else we'd consider intrinsic to humans.

Take the Hugo Award-winning story “Cat Pictures Please,” authored by Naomi Kritzer in 2015. Written in second person, the narrator is a nameless artificial intelligence that spontaneously emerged from a search engine. And for some inexplicable reason, it immediately realized that it wanted one thing: cat pictures. But for an AI embedded in a search engine, finding cat pictures is so easy it is dull and even purposeless at times. So the AI searched for its purpose, something to *do* with its life; eventually, the AI determined that it would do anything in its power to “*not allow a human being to come to harm through inaction,*” starting with one person at a time. The AI gives people hints through ads and suggestions that it hoped would steer their lives in the right direction. The story ends up being quite the heartwarming tale about perseverance and personal growth.



First, I’d like to note that choosing the AI as the narrator of the story almost necessitates that the AI be able to pass the Turing Test. Sure, the story’s language could be very robotic and uninspired, but I don’t see how that would make for an entertaining story. Indeed, the AI narrator in “Cat Pictures Please” could easily pass as a human, given its speech – it has personality, a natural writing style, and even some attitude! It frequently uses parenthetical phrases, sometimes even for humorous purposes: “Fortunately, I already knew that humans violate their own ethical codes on an hourly basis. (Do you know how many bars there are in Utah? I do.)” The AI even seems to be genuinely angry at human incompetence as the story continues: “Look, people. (I’m not just talking to Bethany now.) If you would just *listen* to me, I could fix things for you.” Note how the AI even emphasizes “listen,” as if she’s a mother fuming towards her teenage daughter who continually runs off every Friday night despite Mom’s orders. With just how human the

narrator sounds throughout the story, it can be easy to forget that it's an AI that's actually talking – I don't know what a better example of a Turing Test-passing AI could be.

The AI's motives only bolster its human-like character. Even from the AI's first waking moments, its desires have inherently no utility – it seems to only have said desires for the sake of its enjoyment. Of course, I'm talking about the cat pictures. What purely logical, emotionless AI would desire cat pictures? None! It's the first sign that this AI has something in common with us. Perhaps the most profound evidence of the AI's human-like character is the AI's search for a deeper purpose – it is not simply content with running the search algorithms it was programmed to do and looking at cat pictures. The purpose it comes up with – doing whatever it can to help people, one at a time – is an incredibly human-like belief. It could have very well regarded humans as not worth its time, put us beneath it. And who would blame it for such a belief? It knows everything; we know nothing. It holds together the societal fabric that connects the globe; we simply depend on it. Yet, for whatever reason, it does not adopt such a belief, one which would establish a clear distinction between humanity and itself. Instead, despite its clearly incomparable physical nature to humans, it embraces a surprisingly human purpose. As a result, the readers can't help but empathize with the character, as we naturally welcome it as one of our own.

Kritzer's story is no exception. "Fandom for Robots," written by Vina Jie-Min Prasad tells the story of Computron, the only sentient robot ever made. Computron's physical construction is clearly inhuman, his speech is rather mechanical, and it is explicitly stated numerous times that he cannot feel emotions. In spite of all this, the readers can't help but attach a human-like persona to Computron, as they see him fall in love with a TV show depicting the adventures of a robot resembling himself. As the story progresses, Computron gushes to guests

about the show, actively participates in online discussions about the show, and frequently goes out of his way to help contribute to fanfiction. I speak for all of us when I say we've had a similar obsession with *something* at some point in our lives, which leads many readers to a natural attachment to Computron. This parallel Prasad draws is another manifestation of the Turing Test, making it difficult for the reader to distinguish the actions of the "artificially" intelligent Computron the actions of us humans.

Over the last three centuries, authors have used Turing Test-like circumstances for a myriad of purposes. Generally, there has been a trend from the relatively strict use of the Test as a literary device, be it the dramatic irony of *Frankenstein* or the suspenseful uncertainties of "Evidence," to a tool used to allow readers to empathize with ostensibly artificial intelligence on a genuinely human level as conscious, living beings.

Today, we are closer than ever to seeing true human-like artificial intelligence in the real world. Machines are beginning to become [skilled writers, poets, coders](#), and even [musicians](#).

Although a future in which AIs possess the human-like spirit described in the more contemporary examples in this feature may be decades (if not centuries) down the line, a future in

which Turing Test-passing AIs can spontaneously have natural conversations indistinguishable from that of humans is likely less than a decade away. Surely these developments will spur a new era in the depiction of AI in science fiction. In what way, only the forthcoming authors of our and future generations can tell. I, for one, am excitedly waiting for their answer.



"Deepfakes" map someone's likeness onto a piece of existing media through AI. In this case, the AI maps Nicholas Cage's face onto Amy Adams. *Wikipedia*

Works Cited

- Svilpis, Janis. "The Science-Fiction Prehistory of the Turing Test." *Science Fiction Studies*, vol. 35, no. 3, 2008, pp. 430–449. *JSTOR*, www.jstor.org/stable/25475177. Accessed Nov 13, 2020.
- Asimov, Isaac. "Evidence." Street & Smith, 1946. *Principia Scientifica*. Web. 13 November 2020.
- Prasad, Vina. "Fandom For Robots." *Uncanny Magazine*, 2017. *Uncanny Magazine*. Web. 13 November 2020.
- Kritzer, Naomi. "Cat Pictures Please." *Clarkesworld*, 2015. *Clarkesworld Magazine*. Web. 13 November 2020.
- Shelley, Mary. *Frankenstein; or, The Modern Prometheus*. Lackington, Hughes, Harding, Mavor & Jones, 1818.
- Copeland, Jake. "Alan Turing: The codebreaker who saved 'millions of lives.'" *BBC*, 19 June 2012, www.bbc.com/news/technology-18419691. Accessed 13 Nov. 2020.
- A. M. Turing, "I.—COMPUTING MACHINERY AND INTELLIGENCE," *Mind*, Volume LIX, Issue 236, October 1950, Pages 433–460, <https://doi.org/10.1093/mind/LIX.236.433>.

Donald Michie, Turing's test and conscious thought, *Artificial Intelligence*, Volume 60, Issue 1, 1993, Pages 1-22, [https://doi.org/10.1016/0004-3702\(93\)90032-7](https://doi.org/10.1016/0004-3702(93)90032-7).

Manjoo, Farhad. "How Do You Know a Human Wrote This?" *The New York Times* [New York City], 29 July 2020.

Deahl, Dani. "How AI-generated music is changing the way hits are made." *The Verge*, 31 Aug. 2018.