Has AI Succeeded?
Large Language Models and the Turing Test

William J. Rapaport

Department of Computer Science and Engineering,
Department of Philosophy, Department of Linguistics,
and Center for Cognitive Science
University at Buffalo, The State University of New York,
Buffalo, NY 14260-2500

rapaport@buffalo.edu
http://www.cse.buffalo.edu/~rapaport/

January 17, 2026

Abstract

Passing the Turing Test is often taken to be a sign of intelligence. Some people believe that large language models (LLMs) have passed the Turing Test. Some of those people infer that such LLMs show signs of intelligence or thinking. Others say that such LLMs, despite passing, are not intelligent. And still others believe that such LLMs have not passed the Turing Test. This essay attempts to sort these issues out, arguing that, even though LLMs may *not* be intelligent, Turing was right: What intelligence "really is" is less important than whether we *take* an LLM to be intelligent. Passing or not passing a Turing Test is irrelevant to the social and moral issues involved with taking the output of LLMs at face value.

Keywords: Artificial Intelligence, large language models, Turing Test

"I meant, were these things … well, intelligent? Could they talk?"
"Aye. They could talk. They were intelligent, for-bye,[1] which is not always the same thing."
— C.S. Lewis, *That Hideous Strength* (1946, Ch. 9, §3, p. 190–191)

Nevertheless[2] I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.
— Alan M. Turing (1950, §6, p. 442)

1 Introduction

What will happen when we accept a computer as having passed a Turing test?
— William J. Rapaport (2000, §9, p. 487)

Passing the Turing Test[3] is often taken to be a sign of intelligence. Some people believe that large language models (LLMs)[4] have passed the Turing Test. Some of those people infer that such LLMs show signs of intelligence or thinking (Bubeck et al., 2023; Cappelen and Dever, 2025). Others say that such LLMs, despite passing, are not intelligent (as John Searle's (1980) Chinese Room Argument holds). And still others believe that such LLMs have not passed the Turing Test. This essay attempts to sort these issues out, arguing that, even though LLMs may *not* be intelligent, Turing was right: What intelligence "really is" is less important than whether we *take* an LLM to be intelligent. Passing or not passing a Turing Test is

[1]Scottish for "besides".

[2]I.e., despite 'Can machines think?' being "too meaningless to deserve discussion".

[3]I capitalize 'Test', because it is not clear that the Turing Test is a "test". However, although Turing introduced it as a 'game', he also referred to it as a 'test' (Turing, 1950, §6, p. 446).

[4]As I understand the recent literature, the terminology for these language-processing systems is not yet fixed. Following Bender and Koller (2020, p. 5185, col. 1), let's say that a *language model* is any neural-network deep-learning system based on "transformer" technology (Vaswani et al. 2017; see also Levinstein 2023a) "trained only on the task of string prediction, whether it operates over characters, words or sentences … ". (Although language models so defined may be "models" of language (cf. Kalai et al. 2025, §3, p. 5), this use of 'model' must be distinguished from the notion of a model of the *world*, as discussed in §7, point 14, below.) Then a *large* language model is one based on (very) large amounts of training data (https://openai.com/research/better-language-models). Other terms sometimes used to generalize over these are 'foundation model' and 'generative AI' (Bommasani et al., 2022; Ronge et al., 2025). Shanahan (2024) distinguishes between (1) LLMs as "generative mathematical models of the statistical distribution of tokens in the vast public corpus of human-generated text" (p. 70, col. 1) and (2) "the systems in which they are embedded" (col. 3), such as ChatGPT. This is an important distinction, but I will usually ignore it here (as does Shanahan occasionally; see p. 68). For convenience, I will call them all 'LLMs'.

irrelevant to the social and moral issues involved with taking the output of LLMs at face value.

2 A Tale of Two AIs

> ... it was the age of wisdom, it was the age of foolishness ...
> — Charles Dickens, *A Tale of Two Cities* (1859, Book I, Ch. 1, p. 3)

2.1 AI: General Educated Opinion

Historically, AI is the branch of computer science that deals with the computability of cognition.[5] Its goal is to explore the extent to which cognitive processes such as language, reasoning, planning, acting, perceiving, etc., are computable in the sense of Turing Machine[6] computability (Barr 1983; Rapaport 1998; Rapaport 2023, Ch. 18; Dietrich et al. 2024a). This is perhaps the original notion of AI.[7] A better term might have been 'computational cognition' (Rapaport 1995; Rapaport 2023, §18.2.2; Poole et al. 1998 calls it 'computational intelligence').

There have been two main paradigms of AI research: (1) "Good Old-Fashioned Artificial Intelligence" ("GOFAI"; Haugeland 1985, p. 112), i.e., logical, symbolic, and knowledge-based techniques, and (2) connectionist or neural-network techniques. The former dominated AI's early years; the latter has dominated more recently. Both, however, have always been and still are being pursued.[8]

AI, thus understood, has not yet succeeded: There are many important features of intelligence that AI systems don't yet exhibit. And although you can converse

[5] I take computer science to be "the scientific ... study of what problems can be solved, what tasks can be accomplished, and what features of the world can be understood computationally, i.e., using a language [equivalent to that of a Turing Machine] ..., and then to provide algorithms to show how this can be done efficiently, practically, physically, and ethically" (Rapaport, 2023, §20.5, p. 460).

[6] As with 'Turing Test', I capitalize 'Machine' because Turing Machines are not (physical) machines (Rapaport, 2023, §1.4.1, p. 44, "Terminological Digression"). Turing Machine computation includes heuristic computation: A heuristic for problem P is an algorithm for a different (but related) problem Q, where the solution to Q is "good enough" or "satisfices" (Simon, 1996) as a solution to P (Rapaport 1998, p. 406; cf. Romanycia and Pelletier 1985; Oommen and Rueda 2005). Other notions of computation relax some of the constraints of Turing Machines or add other features, such as neural computation (Anderson and Piccinini, 2024), hypercomputation, or trial-and-error computation (Rapaport, 2023, Ch. 11). For a useful overview, see Maley and Shagrir 2026. I will not discuss these here.

[7] "The study [of AI] is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it" (McCarthy et al., 1955). But AI, as "machine intelligence", is at least ten years older (Turing, 1947, 1948, 1950, 1951a,b; Turing et al., 1952). See Copeland 2023.

[8] Both (1) and (2) are reminiscent of Kahneman's (2011) "slow" and "fast" thinking.

with an LLM, its responses to your questions and solutions to your problems are not trustworthy.

However, there is no scientific (as opposed to practical) reason to believe that AI cannot succeed (Rapaport, 2025b): Many essential aspects of cognition *have* been shown to be computable (see any introductory AI textbook); so, AI has partially succeeded. It will fail to completely succeed if and only if there is some essential aspect of cognition that can be shown to be non-computable.[9] But there is no generally accepted logical or non-logical argument to that effect: There are no logical ones (à la the Halting Problem), and there is no general agreement about non-logical ones (such as Landgrebe and Smith's (2023) arguments that no currently known mathematical technique can work).[10] Thus, there is no scientific reason to believe that cognition cannot eventually be shown to be computable.[11]

"General educated opinion" has not yet "altered" to declare that the Turing Test has been passed (Mitchell, 2024).[12]

2.2 AI: The Use of Words

To the general public and even many researchers, however, AI is the study of neural-network deep-learning systems in general, and LLMs in particular, as well as their commercial development and deployment.

One goal of AI, thus understood, is to develop computers that can identify ("learn") statistical patterns in large amounts of data and to communicate with users in natural language. This goal can be seen as trying to develop an "intelligent" agent we can talk to—a robot or softbot of the sort found in science-fiction stories. Other AI systems, not concerned with understanding language, can win at Go, solve protein-folding problems, drive cars, diagnose and treat medical problems, and so on.

[9] With the possible exception of phenomenal consciousness (qualia). See Block 2025; Schneider et al. 2025; Schwitzgebel 2025; Seth 2025 for discussion.

[10] For analysis of their arguments, see Rapaport 2024.

[11] On *practical* impossibility, see Marcus 2024c; Rapaport 2025b. (And on the practical impossibility—better: unlikelihood—of a *musical* Turing Test, see Neely 2024; thanks to Michael I. Rapaport for the pointer.)

[12] There is one caveat: The notion of cognition in general and even the specific cognitive processes listed above are vague (cf. Pylyshyn 1980, p. 119, col. 1; Dietrich et al. 2024b). Gunderson (1985, p. 180) noted that 'Can machines think?' is not a yes-no question. Neither is 'Is cognition computable?', because 'cognition' is vague (in addition to the different meanings of 'computable'). And vague things can only be understood computationally by making them precise. On the other hand, that is one of the advantages of computational cognitive science: It forces researchers to be maximally precise down to the finest detail. (In any case, "The point of science is not the holy grail but the quest—the searching and the asking" (Gleick, 2021, p. 36).)

To a first approximation, an AI (in the guise of an LLM) does "super-autocomplete" (Chomsky et al. 2023; cf. Marcus 2022),

> always fundamentally trying to ... produce a "reasonable continuation" of whatever text it's got so far, where by "reasonable" we mean "what one might expect someone to write after seeing what people have written on billions of webpages, etc." ... So at any given point, it's got a certain amount of text—and its goal is to come up with an appropriate choice for the next token to add. (Wolfram, 2023, pp. 1, 47)[13]

AI, thus understood, has succeeded: AI systems have made great accomplishments, and AI researchers have even won Nobel prizes (see, e.g., Ronge et al. 2025, p. 16). The goal of having a computer that you can talk to, that can answer your questions, and that can solve your problems, has been realized. AI systems are in daily use by a wide number and variety of specialists and ordinary people. Every day, ordinary people talk to LLMs, and use them to answer questions, to get advice, and to even just be companions (Cappelen and Dever, 2025).

Science fiction has become reality; the Turing Test has been passed (Jones and Bergen 2025; but cf. Marcus 2025a): "The use of words" has "altered" (Montero, 2025).

3 Turing Tests and Intelligence

> What does a bowl of alphabet soup know?
> — *Ziggy* cartoon (17 September 2002), https://tinyurl.com/ziggy20020917

What is the relation between the ontological issue of the nature of intelligence and cognition, on the one hand, and the epistemological issues of how to detect or identify them (Gonçalves, 2021, pp. 23, 61)? Is the Turing Test a test of intelligence?

The vast literature on the Test—as well as the occurrence of the word in the phrases 'artificial intelligence' and 'machine intelligence'—takes passing the Turing Test as a sign of intelligence, usually followed by a caveat that 'intelligence' itself is in need of definition. Are thinking and intelligence the same thing? Turing (1950) proposed the computer version of his Imitation Game to deal with the question whether machines could *think*, yet he titled his paper "Computing Machinery

[13]For excellent overviews of *how* LLMs work, see Gubelmann 2022, §2, and Levinstein 2023a. An alternative view of *what* LLMs do is this Facebook post by Rachel Meredith Kadel (6 April 2023): "When you enter a text into it [an LLM], you're asking 'What would a response to this sound like?' "; you're not asking what a truthful response would be (although you may think that you are) (https://tinyurl.com/rmkadel20230406).

and *Intelligence*". This suggests that he considered thinking and intelligence to be more or less synonymous (if equally vague).[14] Accordingly, I will use 'Intelligent', with a capital 'I', to cover all such vague terms as 'think', 'intelligent', 'understand', 'intellect', etc.[15]

As a formal test, the Turing Test is at the very least underspecified. It is better understood as a way to consider how an entity might be judged as, or taken to be, Intelligent. It is a description of how we will or should react to apparent Intelligence—not only Intelligence that is produced primarily computationally, but also non-human Intelligence more generally (including that of other animals and even extra-terrestrials).[16]

The important point is that the Turing Test *replaces* questions about Intelligence:

> Let us fix our attention on one particular digital computer *C*.[[17]] Is it true that ... with an appropriate programme, *C* can be made to play satisfactorily the part of A [the man] in the imitation game, the part of B [the woman] being taken by a man [i.e., a human]? (Turing, 1950, §5, p. 442)

This is consistent with how the general public treats LLMs: People don't explicitly ask if the LLM that they are interacting with is Intelligent. They treat them as "satisfactorily playing the part" of a human expert.

Turing proposed this machine version of his Imitation Game because he felt that defining " 'machine' and 'think' ... to reflect ... the normal use of the words ... is dangerous" (Turing, 1950, §1, p. 433). Turing's explicit reason for this is that "a statistical survey" of how the words are used "is absurd". But an implicit reason may be that, at least up until 1950 or so, 'machines can think' was an oxy-

[14]Curiously, however, besides the title of Turing's essay, the words 'intelligence' or 'intelligent' appear in only two other places: once referring to the "intelligence" that a *human programmer* might need to improve a program's speed (Turing, 1950, §7, p. 456), and once to contrast "the completely *disciplined* behavior involved in computation" (my italics) with (undisciplined?) "intelligent behavior" (§7, p. 459). As for other terms, Turing talks about "understanding" in only two places: once when discussing the similarity of the Imitation Game to *viva voce* (i.e., oral) examinations of "understand[ing] something" (§6(4), p. 446), and once when discussing teaching the computer "to understand and speak English" (§7, p. 460). (We'll come back to the *viva voce* version in §4, below.) His most frequently used terms are 'intellect' and 'intellectual capacities'.

[15]Buckner (2024, §2.4) adds 'rationality' to this list. For now, I exclude phenomenal consciousness.

[16]Rapaport 2025a, §3.5. But see Schwitzgebel and Pober 2024 for a contrasting view.

[17]It is interesting to observe that, whereas in Turing 1936 'computer' referred only to humans, here he distinguishes between 'digital computers' and 'human computers' (§4, p. 436), characterizing the former using a highly simplified description of a Turing Machine. So, by 1950, the "use of the word" 'computer' had "altered" to mean a genus with at least two species.

moron.[18] Note that to say this is to embrace a form of (Cartesian) mind-body dualism: On one hand, there are physical machines; on the other, there are thinking things. Turing rejects this dualism, claiming that a certain kind of machine—a digital computer—might be able "to give a good showing in the [imitation] game" (Turing, 1950, §3, p. 436). On the other hand, Turing seems to reintroduce a dualism, noting that the game "has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a [hu]man" (Turing, 1950, §2, p. 434). But this is more likely the anti-dualistic claim that (some) bodies *can* exhibit mental qualities.

So, for Turing, his Test was not a test of Intelligence ("general educated opinion"); rather, it was a way to make a judgment about Intelligence ("the use of words").

4 A Tale of Two Turing Tests

> ... it was the epoch of belief, it was the epoch of incredulity ...
> — Charles Dickens, *A Tale of Two Cities* (1859, Book I, Ch. 1, p. 3)

> [I]f I let myself believe anything on insufficient evidence, ... [t]he danger to society is not merely that it should believe wrong things ... but that it should become credulous, and lose the habit of testing things and inquiring into them ....
> —William K. Clifford (1877, p. 294)

If an LLM is *taken as* Intelligent (i.e., is *judged* to have passed a Turing Test), *is* it Intelligent? That, of course, is precisely the question that Turing "believe[d] to be too meaningless to deserve discussion" (Turing, 1950, §6, p. 442). Turing would

[18] At least one of the *Oxford English Dictionary*'s definitions of 'machine' almost explicitly rules out thinking. See its definitions of 'machine' at http://www.oed.com/view/Entry/111850, senses IV.6.b and V.8.b (a sense in which a machine could be a *person* who acts *unthinkingly*), and especially of its cognate 'mechanical' at http://www.oed.com/view/Entry/115544, sense A.II.7 ("acting or performed without thought"). Similarly, the 1958 Merriam-Webster *Second Unabridged Dictionary*'s fifth definition of 'machine' specifically says that machines are "unintelligent" (i.e., by definition) (Neilson et al., 1958, p. 144, col. 3). See also Sieg 2008, p. 527, fn. 1; p. 574; and Gonçalves 2024. Wittgenstein (1934, p. 47) said that "the sentence, 'A machine thinks (perceives, wishes)', seems somehow nonsensical. It is as though we had asked 'Has the number 3 a colour?' "; but cf. Proudfoot 2024. For a discussion contemporaneous with Turing 1950, see Mays 1952, especially this passage: "The O.E.D. definition does bring out one thing at least, a machine is usually thought of as something which does not possess a private life of its own, it does not indulge in reverie when at its task, it lacks consciousness, intelligence and will" (p. 149). Although 'machines think' may have been an oxymoron ca. 1950, people have been trying to get machines to think since at least the 18th century (Lepore, 2024).

not say that LLMs *can* "think" or *be* Intelligent, only that they would pass a Turing Test, period, full stop.

He went on to say that eventually we would *call* their output 'thinking' (or 'Intelligence'): "the use of words ... will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted" (§6, p. 442). As Lakoff and Johnson (1980) showed us, words can alter by metaphorical extension (Rapaport 2000; Rapaport 2023, Ch. 18; Montero 2025). A species-referring term (e.g., 'computer' or 'fly' or 'think') can be "promoted" to being a genus-referring term. Or it may be that eventually we will simply no longer *care* whether an AI is "really" Intelligent; we'll just become habituated to treating them so: Smartphones aren't really smart (Davis, 2017, p. 157). Turing also went one step further: "general educated opinion" will *also* "alter"; i.e., we will change our understanding of both machines and thinking.[19]

Have either of those two things happened? It's safe to say that most people no longer think that "machines can think" is an oxymoron, i.e., that computers can't think *simply because* they are machines. The issues now are: (A) Have LLMs passed the Turing Test? (B) If they have passed, have they passed in the "right" (e.g., cognitive) way? And, of course, we cannot help but ask (C) *Are* they Intelligent?

The impressive linguistic (and other) abilities of LLMs engender two kinds of reactions from users, who may be thought of as (perhaps unwitting) judges of a Turing Test. Some are "credulous", treating them as (if) intelligent (i.e., holding that LLMs *have* passed Turing Tests); others are "incredulous", holding that LLMs have *not* passed Turing Tests, lacking important features of intelligence.

Question (A) is ill-formed: There is no such thing as "the" Turing Test. There are only various Turing-like tests, some of which follow the rules of the Imitation Game (such as they are: e.g., one judge and two participants, five-minute time limit, etc.), and some of which don't (cf. Moor 2001, p. 90). So a better question is whether any current LLM has passed a Turing-like test. Even better: Has any such system consistently passed numerous such tests as determined by numerous judges?

Turing suggested that the imitation "game **(with the [human] player B omitted)** is frequently used in practice under the name of *viva voce* to discover whether some one really understands something or has 'learnt it parrot fashion' " (Turing, 1950, §6, p. 446, my boldface). (Think of dissertation defenses!) Note that the kinds of Turing Tests that LLMs participate in are typically of the *viva voce* variety. The *viva voce* version has been criticized as lacking a control: the woman in the Imitation Game; the human in the computer version. But the *viva voce* version

[19]Davidson's (1978) claim that words have their "literal" meaning in metaphors is relevant here.

has an implicit control.[20] To see why, consider two kinds of Turing Test, or two ways in which an entity (computer or human) can "take a Turing Test". Daniel C. Dennett (2023a, p. 270) calls them "friendly conversation" and "aggressive probing":

Friendly Conversation: Suppose that the judge is an ordinary person, that the test is a simple (perhaps one-off) interaction, and that the judge does not distinguish the interaction from that of a human, either because they cannot (or merely do not) do so, or else because they are not interested in doing so. I call such a judge "credulous". In that case, other ordinary people are an implicit control group, and the entity being judged has passed and has been taken by that judge to be Intelligent.

Aggressive Probing: But suppose that the judge is an AI expert, that the test is a sophisticated experiment, and that the judge *can* distinguish the interaction from that of a human (either because of linguistic differences[21] or because the entity is lacking in one or more of the features that experts consider to be part of Intelligence; §7, below). I call such a judge "incredulous". In that case, an "ideally" Intelligent human—i.e., one who exhibits most or all of those features—is the implicit control, and the entity being judged has not passed.[22]

Note that judges need not have to be aware that they are conducting a Turing Test: A time-traveler from the distant past conversing with an LLM would almost certainly take it as Intelligent. The general public is put in the position of being a judge (whether they realize it or not) when they use Google, whose first response to a search is now often an "AI overview", i.e., an LLM-generated response. If a user accepts the AI overview as the answer to their question, they have implicitly accepted the system's reply as being on a par with the reply that a (knowledgeable) human would have given. They have believed something (Google's AI response) at face value, without questioning it or inquiring after the source of the information. Being uncritical in this way is not necessarily a bad thing; many of us do it all the time.[23] But LLMs are not (yet) completely trustworthy, and therein lies a problem, to which we'll return (§8).

Turing himself suggested that passing only required "an average interrogator" (Turing, 1950, §6, p. 442). Indeed, two years later, he went further and said that "A considerable proportion of a jury, *who should not be expert about machines*, must

[20]For more on the notion of a control in the Turing Test, see Gonçalves 2023.

[21]Argamon et al. 2003; Kabir et al. 2023; Pagan et al. 2025.

[22]For some observations on what I am calling credulous and incredulous judging, see Cappelen and Dever 2025, Chs. 4–5.

[23]Roose 2025 recites a list of ways that LLMs can be helpful "in my day-to-day life". But do we *need* LLMs for this? Or would a (more trustworthy?) Google search do just as well? But see my comments about *Wikipedia* in §8.

be taken in by the pretence" (Turing et al. 1952, p. 495, my italics; cf. Proudfoot 2013, p. 398). However, that does not necessarily mean that Turing required his judge to be credulous, because I am more concerned with the judge's "expertise" on—i.e., views about the nature of—*Intelligence*, not necessarily on *computers*. Turing Tests judged credulously are, perhaps, too easily passed! Bisk et al. (2020, p. 8723, col. 2) claim that "Turing was careful to show how easily a naïve tester could be tricked", which suggests that for the Turing Test to be valid, the judge *should* be an expert, despite Turing's preferences.

Incredulous judges include those who are more concerned with the failures of current LLMs than with their successes. LLMs have so far failed to be generally accepted by incredulous judges as being Intelligent (but see Cappelen and Dever 2025). But they have clearly been accepted by the credulous general public.

5 Responding to the Output

The winner of the Imitation Game is determined by the "interrogator", whom I am calling a "judge". The judge's determination is not the result of an objective, quantitative amassing of points. Rather, it is based on the judge's subjective and qualitative[24] *response* to the system's *output*. (I use the term 'output' of LLMs to avoid the somewhat theory-laden term 'behavior'.) Let's consider these in turn.

5.1 Response

> The extent to which we regard something as behaving in an intelligent manner is determined as much by our own state of mind and training as by the properties of the object under consideration. . . .
> —Alan M. Turing (1948, p. 431)[25]

> In the end, it's the judges' choice: Do you want a traditional gumbo or one with a modern twist?
> — Contestant on the TV cooking competition *Beat Bobby Flay*, ca. July 2025

Is the winner of a TV cooking competition the "best chef"? As the above-quoted chef said, in the end, it's the judge's choice: The judgments on a cooking competition are dependent on the judges' responses to the dish, not (only) on whether

[24] Might the judge be a referee or umpire whose determination is objectively based on whether the player is following the rules? After all, incredulous judges might have a set of criteria for Intelligence that they use as guidance. But determining whether a system's output satisfies those criteria can still be subjective.

[25] This passage is continued in §5.2, below.

the dish is (objectively) good (if there is such a thing). The winner is judged only in comparison to the other competitors and only in part on their cooking ability. Another part of the decision depends on who the judges are and on their particular tastes: A blue-cheese dish judged by a blue-cheese-hating judge might be judged poorly even if other judges love it. Losing chefs are not necessarily "bad" chefs; it's just that their dishes were not as tasty according to the judges as the dishes of the winner.

And similarly for Intelligence: In the three-person Turing Test (one judge plus two candidates), the candidates are judged in relation to each other. A candidate passes a Turing Test based on the judge's *response* to its output and as compared to the other's output, not (only) on whether the candidate's output is (objectively) Intelligent (if there is such a thing) (Pettit, 1991; Watt, 1996; Proudfoot, 2005, 2013, 2017).[26] Just as there is no independent gold standard of culinary ability, there is no independent gold standard of Intelligence (other than the comparison to human Intelligence, itself a vague notion). Nevertheless, just as dishes must be edible, appeal to the eye, be tasty, etc., there are features of Intelligence that may be necessary, as we will see in §7.

Thus, Intelligence is in the eye of the beholding judge; it is not necessarily something internal to the system being tested. Consider the ant who traces an outline in the sand that an observer takes to be Winston Churchill (Putnam, 1981, p. 1). An LLM is like the ant, its output like the sand tracing, and the Turing Test judge like the ant's observer. It is the judge's interpretation of the output that matters (cf. Crane 2024, pp. 7, 10). Thinking that an LLM is Intelligent tells us more about *us* than it does about either the LLM or Intelligence: "[T]he failure of experts to distinguish between imitations and the real thing should not be taken as much more than a statement of the competence of the experts" (Schank, nd, p. 14).

Consider an actual conversation with the Eliza computer program (Weizenbaum, 1966) in which the user pretended to be Hamlet.[27] The interaction was predictable Eliza: repetition, pattern matching, canned responses. But towards the end of the conversation, the Eliza program repeated a comment from earlier in the conversation that, when placed in conjunction with the conversation at the end, caused the user to remark "How interesting! I never thought of that!". A random conjunction of two sentences caused the user to think about the play *Hamlet* in a new way. But, given the nature of Eliza, that the user was impressed said more

[26]This "response-dependence" view is related both to Chalmers's (2025) "quasi-interpretivism" and to the first-person point of view. For the first-person perspective in (GOF)AI and for an earlier argument in favor of the prominence of the judge's response, see Rapaport 2000, §6.1; Rapaport 2023, §18.8.4; and the citations at https://tinyurl.com/rapaport2023-1884. Cf. Cappelen and Dever 2025, p. 88. Also see footnote 35, below.

[27]https://cse.buffalo.edu/~rapaport/hamlet.script.html

about the user than about Eliza.

5.2 Output

> [Stanley Cavell] also turns away from the craftsman's idea that you can know if the automaton is human (or not) by looking inside. On Cavell's account, everything important to know about the automaton, or for that matter a human, happens outside as well as inside.
> —Marie Theresa O'Connor (2024, §3, p. 11)

That a system passes (i.e., is judged to be Intelligent) matters, of course. But question (B)—*how* it passes—is irrelevant. The judge responds only to output, not internal processing, which the judge is assumed to have no access to, or knowledge of.

We don't say whether a *human* is Intelligent based on their internal processing—on how their neurons fire.[28] We make that judgment based on (our responses to) the human's output (cf. Cappelen and Dever 2025, p. 39). After all, it's possible for both a neuron-firing human and an algorithm-executing computer to exhibit cognitive abilities. (This is the fundamental suggestion of functionalism.) Moreover, focusing on the output, "no matter whether the implementation substrate is a brain or a neural network . . . is the standard procedure in linguistics, where data consisting primarily of acceptability judgments is used to postulate underlying linguistic competence" (Futrell and Mahowald, 2025, p. 6).

Yes, the judge could be an AI researcher who has access to the system's program, but such access, if not explicitly prohibited by Turing, is at least not assumed by him:[29]

> . . . If we are able to explain and predict its behaviour or if there seems to be little underlying plan, we have little temptation to imagine intelligence. With the same object therefore it is possible that one man would consider it as intelligent and another would not; the second man would have found out the rules of its behaviour. (Turing, 1948, p. 431)

We take things as they appear to us, and then wonder if they "really" are as they appear (cf. Weizenbaum 1962). But what if appearance is all that we have to go on? (Cf. Kant!) Proudfoot's response-dependence gives priority to appearance.[30]

[28]And much else besides, but for simplicity I will lump all of the electrochemical behavior in the brain together as "neuron firings".

[29]What follows is the rest of the passage quoted at the beginning of §5.

[30]So do the Eliza effect and the Intentional Stance; see §5.3.

What if we had access to the "reality"? In the case of LLMs, that would mean access to the statistical processing. One possibility is that we would find nothing but "donkey work" (Turing et al., 1952, p. 500). Would that matter?

I'm certainly not saying that internal processing doesn't exist; neither cognition nor computation is magic (Rapaport, 2023, §3.16.6). Nor am I saying that internal processing doesn't matter at all. (I'm a cognitive scientist, not a behaviorist.) Although AI conceived as computational cognition requires that the processes be at least algorithmic, Turing placed no limitations on the nature of the algorithms: "We should wish to permit every kind of engineering technique to be used in our machines" (Turing, 1950, §3, p. 435). So both symbolic GOFAI and neural-network deep-learning techniques should be allowed, and both might even be required, as in hybrid (or "neurosymbolic") systems combining GOFAI with neural networks (Pylyshyn 1980, §6; Fabiano et al. 2025; Marcus 2025b,c). (I assume that most LLMs use only very advanced versions of the latter, and none—or very little—of the former.) The issues of donkey work and stochastic parroting (Bender et al., 2021) are—to continue the non-human animal metaphors—red herrings.[31]

If we find Intelligent structures in the code (e.g., GOFAI), would that matter? Probably not: If we determine that an LLM is Intelligent based on appearance, then any Intelligence in the code is irrelevant (except that it entitles us to say that reality matches the appearance). So the underlying "reality" is irrelevant unless it fails to produce output judged to be Intelligent. The nature of the internal processing only matters if the LLM fails on the criteria for Intelligence to be discussed in §7. In that case, the donkey work is not sufficient. If the LLM passes on all criteria by donkey work, then the donkey work is sufficient. All that matters for the internal processing is that it produce cognition—more precisely, that it produce output incredulously judgeable as cognition.

What I *am* saying is that the *nature* of the internal processing doesn't matter in order for the judge to make a determination about Intelligence based solely on output: "The bet was, we could build a distributed net whose text interpretation reasonably mimicked an arrested human's. All we contracted to was product. We didn't promise to duplicate anything under the hood" (Powers, 1995, p. 275).

5.3 Problems with Responding Only to Output

> The Previous [inhabitants] had many other customs that were inexplicable, none more so than their propensity to intermingle fact with fiction, which

[31] Piccinini 2026, p. 1, argues against the view that "psychology … is *autonomous* from neuroscience." My view here is *not* that output Intelligence is "autonomous" from internal processing. But my view is consistent with viewing the *determination* of Intelligence as "autonomous" from *knowledge* of the internal processing.

made it very hard to figure out what had happened and what hadn't.
—Jasper Fforde, *Shades of Grey* (2009, pp. 7–8)

It is human nature to find meaning in chance.
—Eric Schwitzgebel (2017).

Responding to the linguistic output alone is not unproblematic.

Although grammatical language can express true propositions and provide useful information, it can also be purely fictional, false, or just nonsense. Language itself is neutral between fictional and non-fictional uses, and there is normally no way to tell just by looking at the language which of them it is (Rapaport, 2025a, §3.7). In LLMs, this manifests in the unmarked presence of "hallucinations" (better: confabulations) and the lack of concern with truth.

There is another problem with a focus on the output,

> ... another reason that a Turing test will eventually be passed. It is less interesting from a computational point of view, more so from a sociological point of view. It is simply that—to return to the ... Internet dog[[32]]—... *humans tend to treat other entities with which they interact as if they were human.* (Rapaport, 2000, §9, p. 486, italics added)

We almost cannot help but attribute causality, motive, etc., to inanimate objects. Consider animated cartoons and, especially, our almost universal and automatic—perhaps cognitively impenetrable—reaction to Heider and Simmel's (1944) classic movie of triangles and circles that we cannot help but see as cognitive agents (Luccioni and Marcus, 2023). It is difficult, if not impossible, to describe or react to them "neutrally".

This phenomenon—related to pareidolia and anthropomorphism[33]—has been called the "Eliza Effect", having famously been seen when some people reacted to Eliza as if it were a real psychotherapist and not merely a computer program. If, for

[32] https://tinyurl.com/steiner1993. See also https://tinyurl.com/hafeez-cartoon.

[33] On pareidolia, see https://en.wikipedia.org/wiki/Pareidolia and Floridi 2025. Andrews and Huss (2014) distinguish between anthropomorphism—the (possibly mistaken) attribution "of a human property to a non-human animal" (p. 2)—and "anthropectomy"—the (possibly mistaken) *failure* to attribute a human property to a non-human animal (p. 7). Both notions are highly relevant in the present context if the "human property" in question is Intelligence and if we replace 'non-human animal' with 'AI system'. Coghlan (2024, p. 3) argues against anthropomorphism, understood as "the claim that people tend to form the ... false belief" "that some machines really or truly have phenomenally experienced feelings". He cites as evidence that we don't form that belief in the case of puppets or fictional characters. (On dolls—hence also puppets—see O'Connor 2024.) But, in those cases, we know (or antecedently believe) that they have no inner mental life, whereas in the Eliza Effect we don't. Curiously, Coghlan nowhere mentions Dennett's Intentional Stance. On all of this, see also Epley et al. 2007; Astobiza 2024; Buckner 2024, §2.5.

example, having a "purpose" or a "communicative intent" is necessary for language use (cf. Bender and Koller 2020, p. 5187, col. 1), then the Eliza Effect would be that the audience either *assumes* by default or abductively (hence defeasibly) *infers* that the producer has such a purpose or intent. In the Eliza Effect, the human participants unintentionally determine—or unthinkingly take it to be the case—that what they have been interacting with exhibits Intelligence.[34]

To fall prey to the Eliza Effect is to respond to a system as if it were Intelligent. But such a response is what Dennett (1971, p. 87) called the Intentional Stance,[35] which has its benefits, too:

> Attributing human characteristics and motivations to nonhuman agents increases the ability to make sense of an agent's actions, reduces the uncertainty associated with an agent, and increases confidence in predictions of this agent in the future. (Epley et al., 2007, p. 866, col. 1)[36]

Problems arise only when taking the Intentional Stance might lead us to trust "counterfeit people", something Dennett (2023b) admonished us not to do.

6 A Tale of Two Intelligences

> Some aspects of their [LLMs'] behavior appear to be intelligent, but if it's not human intelligence, what is the nature of their intelligence?
> —Terrence J. Sejnowski (2023, p. 311)

> They [LLMs] do things that are very much like thinking. You could say they *are* thinking, just in a somewhat alien way.
> — Douglas Hofstadter, quoted in Somers 2025, p. 30, col. 3

Let's turn to question (C). The Intelligence that was intended to be perceived and responded to in Turing's Test was *human* Intelligence. But what is perceived and responded to in the output of LLMs is—if Intelligence at all—a different kind of Intelligence. This is not surprising: Many different kinds of Intelligence are seen in non-human animals (Yong, 2022), and there is a range of Intelligences even among humans (Gardner, 1983; Sternberg, 1985). Has the AI of §2.2 succeeded

[34]On the related notion that "comprehension automatically implies belief" and that such (temporary) belief is distinct from, and precedes, understanding, see Pennycook et al. 2012. Related to these ideas are "the mechanisms of a psychic's con" (Bjarnason, 2023).

[35]I will assume the reader's familiarity with Dennett's theory of the Intentional, Design, and Physical Stances. Dennett 1990 is relevant here as well for its stance on first-person interpretation, not unlike response dependence.

[36]Epley et al. 2007, p. 872, col. 1, explicitly links this to Dennett.

in producing Intelligence, just not human Intelligence? (And if it's not human Intelligence, should we trust it?)

Consider two more-or-less parallel situations:

1. Human cognition is nothing but (the result of) neuron firings.
   ($\therefore$?) Cognitive behavior is epiphenomenal.

2. LLM "cognition" (as seen in its output) is nothing but (the result of) statistical prediction.
   ($\therefore$?) The "cognitive" output of an LLM is epiphenomenal (besides being untrustworthy).

And consider two possible reactions to each:[37]

1a. The only thing that has to be studied is the neuron firings.

1b. The cognitive behavior also has to be studied on its own, because everyone in their normal, day-to-day activities pays attention only to that output.

2a. The only thing that has to be studied is the statistical prediction.

2b. The LLM output also has to be studied on its own, because Turing Test judges (credulous or not) pay attention only to that output.

Stich and Ravenscroft (1994, pp. 14–15) distinguish between "external" folk psychology as judged by observers and "internal" folk psychology that "is part of the mechanism subserving" those features.[38] The Churchlands, famously, deny the importance of external folk psychology and hence the usefulness of any kind of internal folk psychology. As Colin McGinn (1999) puts it, their "position is that folk psychology was cobbled together in an earlier, pre-scientific age, as a speculative theory of what causes people's behavior, and it is high time to examine it critically with a view to finding a more streamlined theory of our inner workings." By analogy, we should ignore (b)—an LLM's "psychology"—in favor of (a) its statistical machine-learning processing.

However, if an entity is to be taken as Intelligent, it is (b) that matters (be it an LLM, extra-terrestrial, non-human animal, or human). What counts are the output and its interpretation by us, not the underlying processes (be they neuron firings, statistical machine learning, or donkey work of any kind).

Surely, to whatever extent we humans have a folk psychology that the Intentional Stance capitalizes on (even if it is not a full-blown scientific theory), LLMs also have their own:

[37]For a humorous take on these, see "Elementary Physics Paths", https://xkcd.com/2933/

[38]Proudfoot 2005 makes a similar distinction between "internalist" and "externalist" interpretations of the Turing Test.

> Kartik Talamadupula … said that when talking about measuring the ability of an LLM, the bar is always about predicting the next token (or word). "Behavior like 'answering questions' or 'logical reasoning' or any of the other things that are ascribed to LLMs are just human interpretations of this token completion behavior" … (Gregory, 2026, pp. 14–15)

In our ordinary, everyday dealings with an LLM, we *have* to treat it via its (folk) psychology, i.e., from an Intentional Stance (cf. Cappelen and Dever 2025, pp. 68, 75).[39] We can't very well manipulate its statistical algorithms in ordinary interactions with it, any more than we can manipulate a human interlocutor's neuron firings.

Recently, there have been several such "psychological" studies of LLMs,[40] and Cappelen and Dever (2025) advocate them. These are high-level, behavioristic studies of the output of LLMs, comparing it to the psychology, hence high-level mental—as opposed to neurological—behavior of humans. Psychological studies may be the only way to understand what a neural network does: Although we can examine the internal program of a GOFAI system; we can't for a neural network (or at least not as easily). The focus of such psychological studies is on the cognitive capabilities of statistical processing.

However, there is a problem with dealing with LLMs only "psychologically": One aspect of Turing's presentation of the Turing Test is that you should be able to ask the system why it said something:

> Interrogator: In the first line of your sonnet which reads 'Shall I compare thee to a summer's day', would not 'a spring day' do as well or better?
> Witness: It wouldn't scan. (Turing, 1950, §6(4), p. 446)

Turing's point is that we would accept that response as more than mere "artificial signaling" (Turing 1950, §6(4), p. 446, quoting Jefferson 1949, p. 1110, col. 1). But should we? Luccioni and Marcus (2023) observe that "It's also silly to ask these sorts of models for questions about themselves; … they don't actually 'know'; they are just generating different word strings on different occasions, with no guarantee of anything."

Lance Fortnow (2023) actually tried this with ChatGPT, in the context of the "black box" problem for neural networks ("seek[ing] … algorithms that can explain themselves"):

[39] If the psychology of an LLM differs from that of a human, as Sejnowski 2023, Cappelen and Dever 2025, and the quotation from Hofstadter suggest, does that contradict the Turing Test?

[40] E.g., Trott et al. 2023; Han et al. 2024; Suzgun et al. 2024; Cong and Rayz 2025; Han and McClelland 2025; Nadler et al. 2025

> You can speculate why Google Translate uses male pronouns but you can't ask it why. But you can ask ChatGPT.
>
>> I used female pronouns in the translation because Vera T. Sós is a female name. Vera is a feminine given name, and since the context of the original text is about her life and accomplishments, it is appropriate to use female pronouns in the English translation to maintain clarity and accurately convey information about her.

In the comments section of his blog, I asked Fortnow, "How do you know that ChatGPT's 'explanation' is really that, rather than merely more statistical language parroting?". He replied that "There is some circular reasoning here and the best you could hope for is an overly simplified view of its reasoning. When I have been asking ChatGPT to explain its actions, it does seem to be making a good effort at it." So Fortnow seems to side with Turing.[41]

But is that kind of "higher-level explanation" really what is wanted (especially in the black-box problem)? Or is it "merely" an after-the-fact rationalization—a confabulation—as in the "explanations" that some split-brain patients give for their odd behavior?[42] It's certainly possible that such a rationalization is, in fact, a good high-level summary output of the low-level, internal statistical processing. Still, are such explanations trustworthy? Molfese et al. (2025) offer evidence that they are not, and Bommasani et al. (2022, p. 126) note that "It is important to be discerning of the difference between the ability of a model to create plausible-sounding explanations and providing true insights into its behavior." Cappelen and Dever (2025, p. 102) observe that "It's possible that ... [an LLM] says these things when asked to justify its answer, but that the things said don't play a reason-giving role in its reaching its answer". The same can be said about us! Perhaps all of our own explanations of our behavior are similarly illusions or rationalizations. So relying solely on psychology is not the answer, either.

To ask what an LLM does and to answer that it predicts the next word (as Shanahan (2024) recommends) is like asking what I'm doing now and answering "firing neurons". That's true, but I'm also writing an essay (*by* firing neurons).[43] We need to do "psychological" or Intentional-Stance analyses of LLMs *in addition to* statistical, "neural", or Physical-and-Design-Stance analyses. Although an

[41] See also Molfese et al. 2025. Johnson-Laird and Ragni 2023 offers a possible way around this problem. Cf. Kolata 2024: "The computer may not be able to give a clear explanation of its decision pathway, but does that matter if it gets the diagnosis right?". But contrast Marcus 2025d (citing Kambhampati et al. 2025): "The chains of thoughts that LLMs produce don't always correspond to what they actually do." See also Beger et al. 2025, pp. 4–5.

[42] On split-brain confabulations, see Gazzaniga 1989, esp. p. 951; Wolman 2012, esp. p. 262, col. 2. On confabulation in general, see Hirstein 2009.

[43] Cf. Rapaport 2023, §16.9, esp. p. 379.

LLM's explanations of its behavior may not be the real reasons it behaves as it does (where the real reasons are its statistical predictions), is this any different from our own psychological explanations of our behavior (where our real reasons are neural)?

7 Requirements for Human Intelligence

7.1 How to Judge Incredulously

> "You think the bet [that an AI system could pass an English literature master's exam] was about the *machine*?" ...
> "It wasn't about teaching a machine to read?"...
> "No."
> "*It was about teaching a human to tell.*"
> —Richard Powers, *Galatea 2.2: A Novel* (1995, pp. 317–318, my italics)[44]

In Powers's novel, a deep-learning neural-network system is trained to pass an English-literature master's exam. Its prescient conversations with its programmer were—at least in 1995—pure speculation (as were those in Turing 1950 and Hofstadter 2016). But now almost anyone can have a similar conversation with an LLM. Yet the differences between these fictional interactions and real ones are instructive.

Incredulous judges might be impressed by LLMs' linguistic abilities, but still be skeptical. (Cf. Sejnowski's (2023) "Parable of the Talking Dog" who can speak but makes things up.) What features must the output of an LLM exhibit to be incredulously judged Intelligent? What must be added to an LLM's capabilities to make it a full natural-language *understanding* program as well as a computational theory of general intelligence?[45] Quite a few things, some of which LLMs already do, but many that they don't:[46]

[44] Cf. Rathi et al. 2025!

[45] I.e., what should be in the Chinese Room's rule book? On LLMs and the Chinese Room, see Rapaport 2025a, §5.3. See Rapaport 2025a, §2.4, for a discussion of where LLMs lie on the spectrum of things studied by computational linguistics.

[46] Some caveats to the following list of things that AIs still lack: Dreyfus (1992) didn't think that AI would ever succeed, and famously cited several abilities that he thought that GOFAIs lacked, some of which they eventually came to have. I think that AI *can* succeed, but—Dreyfus-like—I am about to offer several abilities that I think that LLMs lack. Unlike Dreyfus, however, I am not arguing that AIs will never be able to do these things. Rather, these should be understood as research goals. The list is based on, and extends, the list in Rapaport 1995, §2.1.2, pp. 50–51. (Cf. Rapaport 2023, §18.10.) It is not intended to be exhaustive. For related lists, see Marcus and Davis 2019; Marcus 2020, 2024d,e; Hendrycks et al. 2025; Tellex and Watkins 2026. The items are in no particular order, are not necessarily independent of each other, need not all be present, and need not be present to any

1. An AI system must take *discourse* as input, not isolated sentences. LLMs clearly satisfy this. What is not so clear is whether they can deal with the compositional structure of individual sentences: LLMs' token-prediction seems to be primarily linear, not structured (Marcus and Davis, 2019, p. 87).

2. The system must understand *ungrammatical* input and recover from *misinterpretations*. LLMs seem to be able to do this, although not always satisfactorily.[47]

3. It must be able to *ask* questions, not just answer them. And it should also be able to ask questions about itself (Powers, 1995, pp. 229, 268). Some LLMs seem to be able to ask questions (Rothman, 2025, p. 32, col. 1), but it is not clear if they are real questions prompted by the LLM's real curiosity (if it has that at all!) or just a (polite?), statistically likely continuation of a conversation.

4. It must be able to *understand plans*, especially the speech-act plans of users. LLMs seem to be able to do this, though again not always satisfactorily.

5. It must be able to *make plans* (Michie 1971, pp. 101–103; Kambhampati 2023) and to *think ahead* (Newport, 2024), both in general and for natural-language *generation*, in particular in order to ask and answer questions and to initiate or end conversations. (I'm talking about plans for the system itself to carry out, not plans that it devises for a user to follow.) To do this, it must have intentions (Bender and Koller's "purpose" or "communicative intent";

"maximal" degree. Moreover, LLMs' abilities seem to advance on an almost daily basis, so they may already be able to do some of the listed things that I say they cannot do (cf. Marcus 2025f). The list can still serve as a guide to what an AI should be able to do.

[47]For example:

> I asked [ChatGPT] to find anagrams for the word 'threads.' It came up with nonsense words that didn't even have the same letters as the original. And it got "defensive" when I pointed out its mistakes. [It] list[ed] the anagrams I found after it told me it couldn't find any anagrams. It then told me I was correct and proceeded to define the words I had listed. Given sufficient memory for a database of English words, in 1980 I could have programmed my pre-IBM CPM desktop to correctly find the anagrams. . . . "Using all seven letters only once", it comes up with 'struther', which has eight letters, is not a word, contains no a and no d, and, when I point out that there is no u in 'threads', in perfect English apologizes and says, OK, you're correct, in that case the anagram is—struther. It didn't even learn to eliminate its last incorrect response. It took self control for me not to type in "are you on drugs?" And yet, if you tell it to write a six page paper explaining the origins of the Russian revolution, it's difficult to tell that a human didn't write it. (Michael Seymour, personal email, 2 January 2024)

For similar examples, see von Hippel 2023; Marcus 2024b.

recall §5, above). One aspect of the Eliza Effect is that the audience assumes that the producer has such a purpose or intent. But LLMs don't have such intentions. There is no evidence that LLMs can make such plans; they are reactive only.

6. More generally, an AI system must be an agent that can *act*.[48] In particular, it must be an agent that can act (or even communicate) without being prompted to (or asked a question); i.e., it must have initiative and motivations. Moreover, even if LLMs

   > can consciously decide on a plan, and even if we agree they can then devise a plan, these LLMs must be able to determine that it is moving forward in its plan and that means it must be able to determine what is now (or what has become) true based on the actions it is taking. (Saba, 2023)

   This is also something that they cannot (yet) do.

7. Even more generally, an AI system must be capable of *decision-making* (Luccioni and Marcus, 2023). And all of this requires being goal-oriented (Jordan, 2019).

8. Part of decision making is the ability to *make inferences* and *revise beliefs* (Kambhampati, 2023; Mitchell et al., 2023; Marcus, 2024a). Although LLMs seem to be able to make inferences, they do not do so on the basis of a reasoning module, and so they can just as easily "infer" a falsehood as a truth (Berglund et al., 2024; Mirzadeh et al., 2024). And although they can seem to revise their beliefs (when they say things like, "I misunderstood you"; Włodarczyk 2023), they do not do so on the basis of a belief revision module.[49] Belief revision (as well as other aspects of understanding) requires an awareness of what the system is saying and why it says it (see point 10).

9. An AI system must be able to *remember* what it is told, what it has said, what it has learned, and what it formerly believed in cases where it has changed its beliefs. LLMs lack anything that we could call a belief (Levinstein 2023b; Schulte 2023, p. 46), and it is not at all clear that current LLMs can remember what they are told (but see Metz 2024; Roose 2024b). It also has to be able to recall information from previous conversations. Some LLMs might be

[48]Thanks to Johan Lammens for emphasizing this to me. "Agency is what distinguishes us from machines" (Gleick, 2024, p. 30).

[49]E.g., Martins and Shapiro 1988.

capable of this (OpenAI, 2024), but it can come with a cost to the user (Hill and Freedman, 2025).

10. Related to this is the need for the system to be *aware* of what it is saying, in the sense of feedback from its own output. (This is not necessarily related to consciousness, though it may be a sort of "higher-order consciousness".) The "psychological" level (see §6, above) must influence its own processing at the Design or Physical Level (even if that processing is mere donkey work), either by conscious awareness or by unconscious influence. Instinct and unconscious experience can give us information about the world we live in, but *thinking about* the instinct or experience—making it conscious—allows us to talk and theorize about it, and to consciously *use* that information (Rapaport, 2025b, §2.5, p. 21). This is needed for Intelligence; LLMs lack it.

11. To have and use a reasoning module, an AI system must *care about truth* (Witt, 2025). An AI system needs (a) the ability (and the desire) to justify what it says (Kriegel, 2024, p. 466), (b) knowledge of the difference between fiction and non-fiction, and (c) the intention to write or speak truthfully or accurately (as well as being honest about when it is intentionally writing or speaking fictionally; Chomsky et al. 2023). In short, it must be trustworthy. (See §8, below.)

12. Language learning, linguistic negotiation (Rapaport, 2003), and the ability to have a real conversation (along the lines of the fictional ones by Turing, Powers, and Hofstadter cited earlier) require an AI system to be able to *interact* with its interlocutors, to exhibit joint attention, and to construct a model of the user's beliefs (Bender and Koller 2020, 5190; Bisk et al. 2020, p. 8722, col. 2). There is no evidence that LLMs can do any of these.

13. Related to this, an AI system must be capable of *experiential* learning, especially if there are "things that *cannot* be learned (about language) by merely reading large bodies of text data" (Sahlgren and Carlsson 2021, §2.3; Nadler et al. 2025; cf. Jackson's (1986) Mary, and Cappelen and Dever 2025, p. 85). After all, compare learning French in artificial situations in school with using it in real-life situations in France: "Language learning continues for a speaker's whole lifetime: the grammar of human languages evolves, and humans flexibly adapt to novel linguistic situations ..." (Bommasani et al., 2022, p. 26). Note, however, that Bubeck et al. (2023, §§5.2.1, 5.2.2) suggest that GPT-4 is capable of learning from current interactions.

14. One of the most all-compassing characterizations of Intelligence is that of McCarthy and Hayes (1969, pp. 465–466):

> … an entity is intelligent if it has an adequate model of the world (including the intellectual world of mathematics, understanding of its own goals and other mental processes), if it is clever enough to answer a wide variety of questions on the basis of this model, if it can get additional information from the external world when required, and can perform such tasks in the external world as its goals demand and its physical abilities permit.

It is not at all clear that LLMs satisfy any aspect of this definition (cf. Hutson 2023). At best, LLMs seem to be "clever enough to answer a wide variety of questions", but not on the basis of any "model of the world". And so it must have *background knowledge*, including "world knowledge" and "commonsense knowledge". By this, I mean that it must have a knowledge base of true statements ("beliefs") about the world and about language, not just the ability to statistically predict words that form sentences that look to us as if they constitute such knowledge. Although LLMs appear to have some world or background knowledge (though not common sense),[50] they only have what they can probabilistically predict from their training. Such knowledge should not merely be a list of unrelated propositions. The knowledge must be organized as a *model* of (relevant aspects of) the world (Thorpe 1989; Marcus 2025c,e; Han and McClelland 2025, p. 3799). They do not have an explicit knowledge representation system (Brachman and Levesque, 2004) or a "world model" (Marcus, 2025b,c,e).[51] Even if LLMs automatically "induce" models in some fashion (Loo et al., 2026), it's not clear that they *use* those models.

15. An AI system must also be able to *learn* about the world and about language. Current LLMs "learn" statistical correlations, but cannot "learn" new things (OpenAI, 2023, p. 18):[52]

> Consciousness depends on a brain's ability to maintain a *constantly updated* conception of itself as a distinct entity interacting with a model of the external world. The layers of neural networks that make up systems like ChatGPT, however, are static: once they're trained, they never change. ChatGPT maintains no persistent state, no model of its sur-

[50]In the sense of Marcus and Davis 2019; Brachman and Levesque 2022.

[51]To have a model, and to deal with the symbol-grounding problems, a cognitive entity must have internal representatives of external objects as part of its knowledge representation system (Rapaport, 2025a, §5.3.2).

[52]Bringsjord et al. (2018) argue that machine-learning systems don't "learn" anything at all. See also Rothman 2024. For an opposing viewpoint, see Futrell and Mahowald 2025, p. 9 and §3.2.

> roundings that it modifies with new information, no memory of past conversations. (Newport, 2023, my italics)

To the extent that LLMs are taken to be Intelligent, the way that they become so differs considerably from the way that humans do.[53]

16. The system also needs to be embedded in a social context (Rapaport 2025a, §5.3.4; cf. Humphrey 1976; Bisk et al. 2020). LLMs are not (yet) so embedded: "Text generated by an LM is not grounded in communicative intent, any model of the world, or any model of the reader's state of mind. It can't have been, because the training data never included sharing thoughts with a listener, nor does the machine have the ability to do that" (Bender et al., 2021, p. 616, col. 2). Here, Bender et al. make two important points. The first is the statement of what LLMs lack. The second is an interesting explanation of that lack, interesting because it suggests that if an LLM's training could include sharing thoughts with an interlocutor, it could have the things that it now lacks. This is essentially part of the learning that Turing's "child" machine would have (Turing, 1950, §7, pp. 455–456).

17. An AI system needs to *understand causality and uncertainty* (Jordan, 2019).

18. And, of great importance, any biases that such systems show are embedded in the data on which they are trained (OpenAI, 2023, p. 7). Intelligence requires the computer *to be "aware" of such biases* and to be able to overcome them without external prompts. And the difficulty of doing this is one of the chief dangers of (current) LLMs. (Though, to be fair, it's a problem for humans, too.)

7.2 Discussion

So I think we can be confident that, despite passing a credulously judged Turing Test, LLMs lack many of the features required for Intelligence (or for passing an incredulously judged Turing Test).

I am not saying that an AI couldn't be Intelligent, only that current LLMs are not. However, Peter J. Denning believes that many of these things are unattainable:

> Have we become so mesmerized by LLMs we do not see the rest of what we do in language? … We build relationships. We take care of each other. We recognize and navigate our moods. We build and exercise power. We

[53] For discussion of *how* humans and LLMs learn or become Intelligent, see Rapaport 2025a, §6, point 11.

> make commitments and follow through with them. We build organizations and societies. We create traditions and histories. We take responsibility for actions. We build trust. We cultivate wisdom. We love. We imagine what has never been imagined before. We smell the flowers and celebrate with our loved ones. None of these is statistical. (Denning, 2023, p. 27, col. 1)

Of course, it's not clear that these are necessary conditions for Intelligence. But note that even if "none of these is *statistical*", some or all of them might yet be *computable*. Yet Denning thinks that they are not even that:

> An analogy familiar to computer scientists is the gap between Turing machine-computable functions and all functions: the machines are a countable infinity, the functions are an uncountable infinity. There are not enough LLMs to handle all the functions visible in human interactions. (Denning 2023, p. 27, col. 2; cf. Denning and Lewis 2019)

But, in the absence of a mathematical proof of non-computability, whether this is the case is an empirical question (Rapaport, 2025b). As with the objections of Landgrebe and Smith (2023), even if AI may never precisely match human Intelligence, it may heuristically come sufficiently close; i.e., its Intelligence may only differ from ours no more than, say, yours differs from mine.

Because judges are basically forced to treat the systems from the Intentional Stance and are prone to Eliza Effects, credulous judges (mostly, but not exclusively, non-experts) are overly quick to give a passing score to AI systems whose output is indistinguishable from humans. The difficulty in identifying the output as coming from an AI system is exactly what Turing predicted when he said that the use of words would alter. If we judges cannot distinguish an AI system's output from that of a human, we will inevitably treat the two kinds of Intelligent systems alike, leading to various social, ethical, and legal problems. Incredulous judges (experts) will require higher standards.

Even if "the use of words" may be altering among the general public, "educated opinion" is, at the very least, divided. The upshot—currently—is that, given both the general untrustworthiness of much current LLM output and the lack of the essential features listed above, judges should be urged to err on the side of incredulity. (The same, of course, holds true for our responses to our fellow humans! We should always be somewhat skeptical; we should always emulate the proverbial toddler who constantly asks "Why?".)

8 Risks

> To sum up: it is wrong always, everywhere, and for any one, to believe any-

> thing upon insufficient evidence.
> —William K. Clifford (1877, p. 295)

> Devices that use heuristics to create the *illusion* of intelligence present a risk we should not accept.
> —David Lorge Parnas (2017, p. 5, col. 3, my italics); cf. Halpern 2023

> How convincing does the illusion of understanding have to be before you stop calling it an illusion?
> — James Somers (2025, pp. 28–29)

LLMs have excellent language abilities: a seeming ability to understand and generate fluent language. They also seem to be able to answer questions, solve problems, provide information, etc. The first ability makes us think that they might be Intelligent and leads us to trust the second ability, which provides their real service. That is where they become risky.

As many stories about the wishes granted by genies make clear, just as you should be careful about what you wish for, you should also be careful about what you ask an LLM. Whenever you deal with one, you are in the position of a Turing Test judge. But if you're not trying to find an LLM's weakness that reveals itself as a mere computer, then you need to ask it questions in a way that will raise the probability of getting a reasonable answer. (There is now a small industry offering such advice; e.g., Saygin et al. 2000; Chen 2023; Shastri 2025. But cf. Witt 2025!)

Many commentators have pointed to the alleged dangers of artificial general intelligence and the Singularity (see, e.g., Eden et al. 2012). But there is a more serious—because more pressing—problem: our current willingness to credulously accept Turing Test-passing entities as Intelligent. Turing was prescient: Whether or not LLMs get closer to, or achieve, artificial general Intelligence, they are already accepted by many people as being Intelligent, simply on the basis of casual interactions with them. And as LLMs get better, more people will accept them, flaws and all.

Despite their hallucinations and confabulations, "millions of people *do* trust A.I. models, and their outputs are being given prominent real estate on . . . Google, . . . Facebook . . . , even in basic Microsoft Office applications" (Roose, 2024a). It is now difficult to avoid Google's "AI Overview", because it's often the first response you get when you ask a question. I recently did a Google search for the actor David Alan Grier. Google's top reply was an "AI overview" that told me that "David Alan Grier is an American actor, comedian, and associate professor of science and technology policy at George Washington University". I was rather impressed until I checked further and discovered that there are *two* David Alan Griers: an actor *and* a professor (Grier, 2005). Only in small print at the end of the "overview" was there a caveat: "Generative AI is experimental".

Two recent developments, however, make this both better and worse: The good news is that some of these AI overviews now come with links to their sources. I try to avoid the AI overviews, but when I do read one, I quickly continue to read the links to double check the accuracy of the AI result. The bad news is that often that is the only way to see those sources: The AI overviews are the only responses.

But things may change. When *Wikipedia* first came out, there was great controversy over its trustworthiness. After all, its articles were not only not necessarily written by "experts", but were editable by anyone, at any time. Readers did not know who the writers were or whether the writing was vetted in anyway. But as we became more used to *Wikipedia*, we became more trusting of it, because people who were knowledgeable about the articles would read them and realize that they were generally on the right track. *Wikipedia* has come to be a generally accepted source (even if it sometimes needs to be read with caution, as we should do with any text).

Eventually, the same thing may happen to LLMs as the processing improves. It may turn out that current techniques of processing will never be trustworthy; there may always be hallucinations. But there may be improvements (e.g., use of GOFAI algorithms) that will make things more trustworthy. This is all consistent with the response-dependent approach.

Indeed, users might not only be accepting of LLMs, they might "become unable or unwilling to distinguish artificial systems from human systems" (Schwitzgebel et al., 2023, p. 2). And therein lies the danger, because such entities currently *don't* pass "aggressive"—incredulously judged—Turing Tests, yet we are all too willing to see Intelligence in their behavior. We come to trust and rely on them, when we should be more critical and less accepting of them.

The source of this danger—and it is a danger, especially given that most interactions with them will be credulous ones—has three interacting sources: (1) the Eliza Effect, (2) LLMs' ignorance of truth and their (current) lack of trustworthiness, and (3) the fact that "The basic tool for the manipulation of reality is the manipulation of words" (Dick, 1978).[54]

Dennett, too, was prescient: Over 40 years ago, he wrote:

> The problem of overestimation of cognitive prowess, of comprehension, of intelligence, is not, then, just a philosophical problem, but a real social prob-

[54]Cf. Arendt 1974; Berkowitz 2024: "If everybody always lies to you, the consequence is not that you believe the lies, but rather that nobody believes anything any longer. ... And a people that no longer can believe anything cannot make up its mind. It is deprived not only of its capacity to act but also of its capacity to think and to judge. And with such a people you can then do what you please."

lem, and we should alert ourselves to it, and take steps to avert it. (Dennett, 1985, p. 140)

Sadly, we did not. We underestimated how soon it would happen.

If the output of an LLM is indistinguishable from that of a human, we need to be (reasonably) skeptical and to think critically about that output. Of course, the same goes for output generated by humans: Any trust we put into what we read or hear someone say must be similarly tempered. A consequence of the Turing Test and the apparent ability of LLMs to pass them is that it is more important to treat all output skeptically and critically than it is to know who (or what) generated the output or how it was generated. We should always be at least slightly skeptical (Rapaport, 2023, §2.4.4). And we need not only to be *able* to think critically, but to *actually* think critically, in order to overcome paralyzing skepticism (Graham and Metaxas, 2003; Singer, 2023; Waxman, 2024).[55]

The scary part of all this is that the credulous general public—not to mention some apparently overly credulous experts who may be more interested in the hype of AI—are not being as critical as they should be, especially given the lack of current LLMs' abilities to pass incredulously judged Turing Tests.

It may or may not be the case that LLMs as currently implemented will achieve artificial general intelligence. Perhaps it will require GOFAI to do so. But suppose that, in the limiting case, an LLM-based system does so. We would *still* need to treat its output critically and not with blind acceptance of its Intelligence:

> Generative AI is a probabilistic system, not a deterministic one; it returns likelihoods, not truth. When the stakes are real, skilled human agents have to remain accountable for the call—noticing when the model has drifted from reality, and treating its output as a hypothesis to test, not an answer to obey. It's an emergent skill, and a critical one. The future of expertise will depend not just on how good our tools are but on how well we think alongside them. (Appiah, 2025)

This is not limited to LLMs (or any future version of AI): We need to treat LLM output in precisely the same way that we should treat the output of other humans. All the harms of LLM-based AIs are also possible harms of *human* linguistic interaction, but humans can override them (even if they don't). Similarly, even if

[55] William Perry's (1970; 1981) "Scheme of Intellectual and Ethical Development" can shed some light on these issues. On his scheme, "Dualists" uncritically believe what "Authorities" tell them, before transitioning to "Multiplism", believing that all opinions are equally good. Both of these positions seem to describe many credulous users of LLMs. It is "Contextual Relativists"—who try to understand things relative to their contexts—who have a chance to become incredulous users. For discussion of Perry, see Rapaport 2018; Rapaport 2023, §2.6, pp. 22–23; and https://tinyurl.com/phics-perry.

LLMs learn from uncurated data "exactly" as humans do, LLMs can, theoretically, be trained on (positively) curated data: Both LLMs and humans need to be better educated. In the words of Oscar Hammerstein II, "You've got to be carefully taught".[56]

9 Conclusion

Thesis:

AI has succeeded. The general (credulous) public treats LLMs as Intelligent. The Turing Test has been passed. The use of words has altered.

Antithesis:

AI has not yet succeeded. There are many necessary components of Intelligence that LLMs do not exhibit. The Turing Test has not been passed. General (incredulous) educated opinion has not yet altered.

Synthesis:

I have told a tale of two AIs, two Turing Tests, and two Intelligences. The Turing Test was never about what Intelligence *is*. It was always about how we would *react* to a computer that seemed to be Intelligent. Whether LLMs "really" are Intelligent or not is irrelevant to how we have to learn how to deal with them. They are here, and we must learn to treat them with a critical eye (cf. Hsu 2025). The two AIs are at opposite ends of a spectrum. AI success is possible, but may only come when they approach each other, when primarily statistical LLMs incorporate GOFAI techniques, and when we accept them *skeptically* as (possibly alien) Intelligences.

Goat: "Looks like artificial intelligence is really getting better."
Rat: "What can we do about regular intelligence?"
Goat: "I think that's stuck where it is."
Rat: "Scientists need to prioritize."
— "Pearls before Swine" comic strip, 17 June 2023[57]

Acknowledgements

This is a fairly complete revision of an unpublished essay that has circulated online (Rapaport, 2025a). Thanks to John Richards and two anonymous referees for comments on that version.

[56] https://tinyurl.com/2p95zp56
[57] https://www.gocomics.com/pearlsbeforeswine/2023/06/17

References

Anderson, N. G. and Piccinini, G. (2024). *The Physical Signature of Computation: A Robust Mapping Account*. Oxford University Press, Oxford.

Andrews, K. and Huss, B. (2014). Anthropomorphism, anthropectomy, and the null hypothesis. *Biology and Philosophy*, 29:711–729. Page refs. to https://tinyurl.com/andrews-huss2014.

Appiah, K. A. (2025). The age of de-skilling. *The Atlantic*. 26 October, https://tinyurl.com/appiah2025.

Arendt, H. (1974). Hannah arendt: From an interview. *New York Review of Books*, 25(16). 26 October 1978.

Argamon, S., Koppel, M., et al. (2003). Gender, genre, and writing style in formal written texts. *Text & Talk*, 23(3):321–346.

Astobiza, A. M. (2024). Do people believe that machines have minds and free will? Empirical evidence on mind perception and autonomy in machines. *AI and Ethics*, 4:1175–1183.

Barr, A. (1983). Artificial intelligence: Cognition as computation. In Machlup, F. and Mansfield, U., editors, *The Study of Information*. Wiley, New York. Page ref. to https://tinyurl.com/barr1983.

Beger, C., Yi, R., et al. (2025). Do AI models perform human-like abstract reasoning across modalities? https://arxiv.org/abs/2510.02125.

Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proc. 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623.

Bender, E. M. and Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proc. 58th Annual Meeting Assn. Comp. Ling.*, pages 5185–5198.

Berglund, L., Tong, M., et al. (2024). The reversal curse: LLMs trained on 'A is B' fail to learn 'B is A'. https://arxiv.org/abs/2309.12288.

Berkowitz, R. (2024). On fake Hannah Arendt quotations. *Amor Mundi*. https://tinyurl.com/berkowitz2024.

Bisk, Y., Holtzman, A., et al. (2020). Experience grounds language. *Proc. 2020 Conf. Empirical Methods Nat. Lang. Processing*, pages 8718–8735. https://tinyurl.com/bisketal2020.

Bjarnason, B. (2023). The LLMentalist effect: How chat-based large language models replicate the mechanisms of a psychic's con. *Out of the Software Crisis*. https://softwarecrisis.dev/letters/llmentalist/.

Block, N. (2025). Can only meat machines be conscious? *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2025.08.009.

Bommasani, R., Hudson, D. A., et al. (2022). On the opportunities and risks of foundation models. https://arxiv.org/abs/2108.07258v3.

Brachman, R. J. and Levesque, H. J. (2004). *Knowledge Representation and Reasoning*. Morgan Kaufmann, San Francisco.

Brachman, R. J. and Levesque, H. J. (2022). *Machines Like Us: Toward AI with Common Sense*. MIT Press, Cambridge, MA.

Bringsjord, S., Govindarajulu, N. S., et al. (2018). Do machine-learning machines learn? In Müller, V., editor, *Philosophy and Theory of Artificial Intelligence 2017*, pages 136–157. Springer, Cham, Switzerland.

Bubeck, S., Chandrasekaran, V., et al. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. https://arxiv.org/abs/2303.12712.

Buckner, C. J. (2024). *From Deep Learning to Rational Machines*. Oxford University Press.

Cappelen, H. and Dever, J. (2025). Going whole hog: A philosophical defense of AI cognition. https://arxiv.org/abs/2504.13988.

Chalmers, D. J. (2025). What we talk to when we talk to language models. https://philarchive.org/rec/CHAWWT-8.

Chen, B. X. (2023). Get the best from ChatGPT with these golden prompts. *New York Times*. 25 May.

Chomsky, N. et al. (2023). The false promise of ChatGPT. *New York Times*. 8 Mar.

Clifford, W. K. (1877). The ethics of belief. *Contemporary Review*, 29:289–309.

Coghlan, S. (2024). Anthropomorphizing machines: Reality or popular myth? *Minds and Machines*, 34(25).

Cong, Y. and Rayz, J. (2025). Language models demonstrate the good-enough processing seen in humans. In Barner, D. et al., editors, *Proc. 47th Annual Conf. Cog. Sci. Soc.*, pages 4642–4649.

Copeland, B. J. (2023). Early AI in Britain: Turing et al. *IEEE Annals of the History of Computing*, 45:19–31.

Crane, T. (2024). On the explanation of intentionality. *Australasian Philosophical Review*, 8:5–19.

Davidson, D. (1978). What metaphors mean. In *Inquiries into Truth and Interpretation*, pages 245–264. Oxford University Press, 1986, Oxford.

Davis, M. D. (2017). Universality is ubiquitous. In Floyd, J. and Bokulich, A., editors, *Philosophical Explorations of the Legacy of Alan Turing*, pages 153–158. Springer.

Dennett, D. C. (1971). Intentional systems. *Journal of Philosophy*, 68:87–106.

Dennett, D. C. (1985). Can machines think? In Shafto, M., editor, *How We Know*, pages 121–145. Harper & Row.

Dennett, D. C. (1990). The interpretation of texts, people and other artifacts. *Philosophy and Phenomenological Research*, 50(Supplement):177–194.

Dennett, D. C. (2023a). *I've Been Thinking*. W.W. Norton, New York.

Dennett, D. C. (2023b). The problem with counterfeit people. *The Atlantic*. 16 May.

Denning, P. J. (2023). The smallness of large language models. *Commun. ACM*, 66(9):24–27.

Denning, P. J. and Lewis, T. G. (2019). Intelligence may not be computable. *Amer. Sci.*, 107(6):346–349.

Dick, P. K. (1978). How to build a universe that doesn't fall apart two days later. In Sutin, L., editor, *The Shifting Realities of Philip K. Dick*, pages 259–280. Pantheon, 1995, New York.

Dickens, C. (1859). *A Tale of Two Cities*. Washington Square Press, 1963, New York.

Dietrich, E. et al. (2024a). How much like us do we want AIs to be? *Techné*, 28(2):137–168.

Dietrich, E., Fields, C., Sullins, J. P., van Heuveln, B., and Zebrowski, R. (2024b). How much like us do we want AIs to be? *Techné: Research in Philosophy and Technology*, 28(2):137–168.

Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT Press, Cambridge, MA.

Eden, A. H., Moor, J. H., Søraker, J. H., and Steinhart, E., editors (2012). *Singularity Hypotheses: A Scientific and Philosophical Assessment*. Springer, Berlin.

Epley, N., Waytz, A., and Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4):864–886.

Fabiano, F., Ganapini, M. B., et al. (2025). Thinking fast and slow in human and machine intelligence. *Communications of the ACM*, 68(8):72–79.

Fforde, J. (2009). *Shades of Grey: The Road to High Saffron*. Penguin, New York.

Floridi, L. (2025). AI and semantic pareidolia: When we see consciousness where there is none. *SSRN*. https://dx.doi.org/10.2139/ssrn.5309682.

Fortnow, L. (2023). Complexity and explainable AI. *Computational Complexity (blog)*. https://tinyurl.com/fortnow20230410.

Futrell, R. and Mahowald, K. (2025). How lingustics learned to stop worrying and love the language models. https://arxiv.org/abs/2501.17047.

Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. Basic Books, 2011, New York.

Gazzaniga, M. S. (1989). Organization of the human brain. *Science*, 245(4921):947–952.

Gleick, J. (2021). Eclipsed by fame. *New York Review of Books*, 68(7):34–36.

Gleick, J. (2024). The fate of free will. *New York Review of Books*, 71(1):27–30.

Gonçalves, B. (2021). Machines will think: Structure and interpretation of Alan Turing's imitation game. https://tinyurl.com/goncalves2021. PhD diss., Univ. São Paulo.

Gonçalves, B. (2023). The Turing test is a thought experiment. *Minds and Machines*, 33:1–33.

Gonçalves, B. (2024). *The Turing Test Argument*. Routledge, New York.

Graham, L. and Metaxas, P. T. (2003). 'Of course it's true; I saw it on the internet!'. *Communications of the ACM*, 46(5):71–75.

Gregory, J. G. (2026). Learning how learning works. *Communications of the ACM*, 69(2):14–16.

Grier, D. A. (2005). *When Computers Were Human*. Princeton University Press, Princeton, NJ.

Gubelmann, R. (2022). A loosely Wittgensteinian conception of the linguistic understanding of large language models like BERT, GPT-3, and ChatGPT. *Grazer Philosophische Studien*, 99:485–523.

Gunderson, K. (1985). *Mentality and Machines, 2nd Ed.* Univ. Minnesota Press.

Halpern, S. (2023). What we still don't know about how A.I. is trained. *The New Yorker*. 28 Mar.

Han, S. J. and McClelland, J. L. (2025). Humans learn proactively in ways that language models don't. In Barner, D. et al., editors, *Proc. 47th Ann. Conf. Cog. Sci. Soc.*, pages 3795–3801.

Han, S. J., Ransom, K. J., et al. (2024). Inductive reasoning in humans and large language models. *Cognitive Systems Research*, 83(101155).

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. MIT Press, Cambridge, MA.

Heider, F. and Simmel, M. (1944). An experimental study of apparent behavior. *Amer. J. Psych.*, 57(2):243–259.

Hendrycks, D., Song, D., et al. (2025). A definition of AGI. https://arxiv.org/abs/2510.18212.

Hill, K. and Freedman, D. (2025). Chatbots can go into a delusional spiral. Here's how it happens. *New York Times*. https://tinyurl.com/hill-freedman2025.

Hirstein, W. (2009). Introduction: What is confabulation. In Hirstein, W., editor, *Confabulation: Views from Neuroscience, Psychiatry, Psychology and Philosophy*, pages 1–12. Oxford University Press.

Hofstadter, D. R. (2016). Dull Rigid Human meets Ace Mechanical Translater. In Cooper, S. B. and Hodges, A., editors, *The Once and Future Turing*, pages 117–128. Cambridge University Press.

Hsu, H. (2025). What happens after A.I. destroys college writing? *The New Yorker*. https://www.newyorker.com/magazine/2025/07/07/the-end-of-the-english-paper.

Humphrey, N. (1976). The social function of intellect. In Bateson, P. and Hinde, R., editors, *Growing Points in Ethology*, pages 303–317. Cambridge University Press.

Hutson, M. (2023). Can we stop runaway AI? *The New Yorker*. https://tinyurl.com/hutson20230516.

Jackson, F. (1986). What Mary didn't know. *Journal of Philosophy*, 83:291–295.

Jefferson, G. (1949). The mind of mechanical man. *Brit. Med. J.*, 1(4616):1105–1110.

Johnson-Laird, P. N. and Ragni, M. (2023). What should replace the Turing test? *Intelligent Computing*, 2. Article 0064.

Jones, C. R. and Bergen, B. K. (2025). Large language models pass the Turing test. https://arxiv.org/abs/2503.23674.

Jordan, M. I. (2019). Artificial Intelligence—The revolution hasn't happened yet. *Harvard Data Science Review*, 1(1).

Kabir, S., Udo-Imeh, D. N., et al. (2023). Who answers it better? An in-depth analysis of ChatGPT and Stack Overflow answers to software engineering questions. https://tinyurl.com/kabiretal2023.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Strauss and Giroux, New York.

Kalai, A. T. et al. (2025). Why language models hallucinate. https://arxiv.org/abs/2509.04664.

Kambhampati, S. (2023). Can LLMs really reason and plan? *Blog@CACM*. https://tinyurl.com/kambhampati2023.

Kambhampati, S., Stechly, K., et al. (2025). Stop anthropomorphizing intermediate tokens as reasoning/thinking traces! https://arxiv.org/abs/2504.09762.

Kolata, G. (2024). A.I. chatbots defeated doctors at diagnosing illness. *New York Times*. https://tinyurl.com/kolata20241117.

Kriegel, U. (2024). Knowledge-by-acquaintance first. *Phil. & Phen. Res.*, 109:458–477.

Lakoff, G. and Johnson, M. (1980). Conceptual metaphor in everyday language. *J. Phil.*, 77(8):453–486.

Landgrebe, J. and Smith, B. (2023). *Why Machines Will Never Rule the World*. Routledge, New York.

Lepore, J. (2024). The chit-chatbot. *The New Yorker*, pages 12–16. 7 October.

Levinstein, B. (2023a). A conceptual guide to transformers: Part I. *Substack*. https://tinyurl.com/levinstein2023PartI.

Levinstein, B. (2023b). What's going on under the hood of LLMs. *Substack*. https://tinyurl.com/levinstein2023PartV.

Lewis, C. S. (1946). *That Hideous Strength*. Macmillan, 1965, New York.

Loo, A., Pavlick, E., and Feiman, R. (2026). LLMs model how humans induce logically structured rules. *Journal of Memory and Language*, 146(104675).

Luccioni, S. and Marcus, G. (2023). Stop treating AI models like people. *Marcus on AI*. https://tinyurl.com/luccioni-marcus20230417.

Maley, C. J. and Shagrir, O. (2026). The mind-brain is a computer, but what is (neural) computation? In Piccinini, G., editor, *Neurocognitive Foundations of Mind*, pages 144–163. Routledge, New York.

Marcus, G. (2020). The next decade in AI: Four steps towards robust artificial intelligence. https://arxiv.org/abs/2002.06177.

Marcus, G. (2022). The dark risk of large language models. *Wired*. https://tinyurl.com/marcus20221229.

Marcus, G. (2024a). CONFIRMED: LLMs have indeed reached a point of diminishing returns. *Marcus on AI*. https://tinyurl.com/marcus20241109.

Marcus, G. (2024b). Evidence that LLMs are reaching a point of diminishing returns—and what that might mean. *Marcus on AI*. https://tinyurl.com/marcus20241413.

Marcus, G. (2024c). Seven reasons why the world should say No to Sam Altman. *Marcus on AI*. https://tinyurl.com/marcus20240210.

Marcus, G. (2024d). Superhuman AGI is not nigh. *Marcus on AI*. https://tinyurl.com/marcus20240412.

Marcus, G. (2024e). There must be some misunderstanding. *Marcus on AI*. https://tinyurl.com/marcus20240208.

Marcus, G. (2025a). AI has (sort of) passed the Turing Test; here's why that hardly matters. *Marcus on AI*. https://tinyurl.com/marcus20250402.

Marcus, G. (2025b). The fever dream of imminent 'superintelligence' is finally breaking. *New York Times*. https://tinyurl.com/marcus20250903.

Marcus, G. (2025c). Generative AI's cripplig and widespread failure to induce robust models of the world. *Marcus on AI*. https://tinyurl.com/marcus20250628.

Marcus, G. (2025d). A knockout blow for LLMs? *Marcus on AI*. https://tinyurl.com/marcus20250607.

Marcus, G. (2025e). LLMs are not like you and me—and never will be. *Marcus on AI*. https://tinyurl.com/marcus20250812.

Marcus, G. (2025f). Six (or seven) predictions for AI 2026 from a Generative AI realist. *Marcus on AI*. 20 Dec., https://tinyurl.com/marcus20251220.

Marcus, G. and Davis, E. (2019). *Rebooting AI*. Vintage Books, New York.

Martins, J. P. and Shapiro, S. C. (1988). A model for belief revision. *Artif. Intell.*, 35(1):25–79.

Mays, W. (1952). Can machines think? *Philosophy*, 27(101):148–162.

McCarthy, J. et al. (1955). A proposal for the Dartmouth Summer Research Project on Artificial Intelligence. https://tinyurl.com/mccarthy-etal1955.

McCarthy, J. and Hayes, P. J. (1969). Some philosophical problems from the standpoint of Artificial Intelligence. In Meltzer, B. and Michie, D., editors, *Mach. Intel. 4*. Edinburgh Univ. Press.

McGinn, C. (1999). Can we ever understand consciousness? *New York Review of Books*, 46(10):44–48.

Metz, C. (2024). OpenAI gives ChatGPT a better 'memory'. *New York Times*. https://tinyurl.com/cmetz20240213.

Michie, D. (1971). Formation and execution of plans by machine. In Findler, N. and Meltzer, B., editors, *Artificial Intelligence and Heuristic Programming*, pages 101–124. American Elsevier, New York.

Mirzadeh, I., Alizadeh, K., et al. (2024). GSM-Symbolic: Understanding the limitations of mathematical reasoning in large language models. *[Apple] Machine Learning Research*. https://tinyurl.com/mirzadeh-etal2024.

Mitchell, M. (2024). The Turing Test and our shifting conceptions of intelligence. *Science*, 385(6710).

Mitchell, M. et al. (2023). Comparing humans, GPT-4, and GPT-4V on abstraction and reasoning tasks. https://arxiv.org/pdf/2311.09247.

Molfese, F. M., Moroni, L., et al. (2025). ReTraceQA: Evaluating reasoning traces of small language models in commonsense question answering. https://arxiv.org/abs/2510.09351. 10 Oct.

Montero, B. G. (2025). What is it like to be a computer? *New York Times*, page A19. 10 November.

Moor, J. H. (2001). The status and future of the turing test. *Minds and Machines*, 11:77–93.

Nadler, E. O. et al. (2025). Statistical or embodied? Comparing colorseeing, colorblind, painters, and large language models in their processing of color metaphors. *Cognitive Science*, 49(e70083).

Neely, A. (2024). Why AI is doomed to fail the musical Turing Test. YouTube video. Viewed 8 May 2024 at https://tinyurl.com/neely2024.

Neilson, W. A., Knott, T. A., and Carhart, P. W., editors (1958). *Webster's New International Dictionary of the English Language, 2nd Edition, Unabridged*. G.& C. Merriam, Springfield, MA.

Newport, C. (2023). What kind of mind does ChatGPT have? *The New Yorker*. https://tinyurl.com/newport20230413.

Newport, C. (2024). Can an A.I. make plans? *The New Yorker*. tinyurl.com/newport20240315.

O'Connor, M. T. (2024). In the craftsman's garden: AI, Alan Turing, and Stanley Cavell. *Minds and Machines*, 34(22).

Oommen, B. J. and Rueda, L. G. (2005). A formal analysis of why heuristic functions work. *Artificial Intelligence*, 164(1-2):1–22.

OpenAI (2023). GPT-4 system card. https://tinyurl.com/openai20230323.

OpenAI (2024). Memory and new controls for ChatGPT. https://tinyurl.com/openai20240213.

Pagan, N., Törnberg, P., et al. (2025). Computational Turing test reveals systematic differences between human and AI language. https://arxiv.org/abs/2511.04195.

Parnas, D. L. (2017). The real risks of artificial intelligence. *Communications of the ACM*, 60(10):27–31.

Pennycook, G., Cheyne, J. A., et al. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, 123:335–346.

Perry, Jr., W. G. (1970). *Forms of Intellectual and Ethical Development in the College Years: A Scheme*. Holt, Rinehart and Winston, New York.

Perry, Jr., W. G. (1981). Cognitive and ethical growth. In Chickering, A. et al., editors, *The Modern American College*, pages 76–116. Jossey-Bass, San Francisco.

Pettit, P. (1991). Realism and response-dependence. *Mind*, 100(4):587–626.

Piccinini, G. (2026). New foundations for the philosophy of mind and the mind sciences. In Piccinini, G., editor, *Neurocognitive Foundations of Mind*, pages 1–17. Routledge, New York.

Poole, D., Mackworth, A., and Goebel, R. (1998). *Computational Intelligence: A Logical Approach*. Oxford University Press, New York.

Powers, R. (1995). *Galatea 2.2: A Novel*. Farrar Straus & Giroux, New York.

Proudfoot, D. (2005). A new interpretation of the Turing test. *The Rutherford Journal*, 1.

Proudfoot, D. (2013). Rethinking Turing's test. *Journa of Philosophy*, 110(7):391–411.

Proudfoot, D. (2017). The Turing test—from every angle. In Copeland, B. J. et al., editors, *The Turing Guide*, pages 287–300. Oxford Univ. Press.

Proudfoot, D. (2024). Wittgenstein and Turing on AI. In Helliwell, A. C. et al., editors, *Wittgenstein and Artificial Intelligence, Vol. 1*, pages 17–37. Anthem Press, London.

Putnam, H. (1981). *Reason, Truth, and History*. Cambridge University Press, Cambridge, UK.

Pylyshyn, Z. W. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, 3:111–169.

Rapaport, W. J. (1995). Understanding understanding: Syntactic semantics and computational cognition. In Tomberlin, J. E., editor, *AI, Connectionism, and Philosophical Psychology (Philosophical Perspectives, Vol. 9)*, pages 49–88. Ridgeview, Atascadero, CA.

Rapaport, W. J. (1998). How minds can be computational systems. *J. Exp. Theor. Artif. Intell.*, 10:403–419.

Rapaport, W. J. (2000). How to pass a Turing test. *J. Logic, Lang. Info.*, 9(4):467–490.

Rapaport, W. J. (2003). What did you mean by that? Misunderstanding, negotiation, and syntactic semantics. *Minds and Machines*, 13(3):397–427.

Rapaport, W. J. (2018). William Perry's scheme of intellectual and ethical development. https://cse.buffalo.edu/~rapaport/perry-positions.html.

Rapaport, W. J. (2023). *Philosophy of Computer Science*. Wiley-Blackwell.

Rapaport, W. J. (2024). Is artificial general intelligence impossible? *Cosmos + Taxis*, 12(5+6):5–22.

Rapaport, W. J. (2025a). Large language models and the Turing Test: The 'use of words' vs. 'general educated opinion'. https://cse.buffalo.edu/~rapaport/Papers/llmtt.pdf.

Rapaport, W. J. (2025b). Yes, AI *can* match human intelligence. https://cse.buffalo.edu/~rapaport/Papers/aidebate.pdf. 30 July draft of book chapter in progress.

Rathi, I., Bergen, B. K., and Jones, C. R. (2025). Judging the judges: Displacing and inverting the Turing test to investigate the interrogator. In Barner, D. et al., editors, *Proc. 47th Annual Conf. Cog. Sci. Soc.*, pages 6016–6023.

Romanycia, M. H. and Pelletier, F. J. (1985). What is a heuristic? *Computational Intelligence*, 1(2):47–58.

Ronge, R., Maier, M., and Rathgeber, B. (2025). Towards a definition of *Generative Artificial Intelligence*. *Philosophy & Technology*, 38(31).

Roose, K. (2024a). How do you change a chatbot's mind? *New York Times*. https://tinyurl.com/roose20240830.

Roose, K. (2024b). Meet my A.I. friends. *New York Times*. https://tinyurl.com/roose20240509.

Roose, K. (2025). How to AI. In Jackson, L., editor, *The Morning*. New York Times. https://tinyurl.com/roose20250604.

Rothman, J. (2024). What does it really mean to learn? *The New Yorker*. https://tinyurl.com/rothman20240827.

Rothman, J. (2025). After the algorithm. *The New Yorker*, pages 31–39. 1 & 8 September.

Saba, W. S. (2023). How did scientists succumb to Aunt Edna? The dangers of a superintelligent AI is fiction. *Blog@CACM*. https://tinyurl.com/wssaba2023.

Sahlgren, M. and Carlsson, F. (2021). The singleton fallacy: Why current critiques of language models miss the point. *Frontiers Artif. Intell.*, 4(682578).

Saygin, A. P., Cicekli, I., and Akman, V. (2000). Turing test: 50 years later. *Minds and Machines*, 10:463–518.

Schank, R. C. (n.d.). Speculations on understanding, mechanical and human. Unpublished(?) talk at SUNY Buffalo(?). Copy available upon request from rapaport@buffalo.edu.

Schneider, S. et al. (2025). Is AI conscious? A primer on the myths and confusions driving the debate. https://philpapers.org/archive/SCHIAC-22.pdf.

Schulte, P. (2023). *Mental Content*. Cambridge Univ. Press, Cambridge, UK.

Schwitzgebel, E. (2017). The Turing machines of Babel. *Apex Magazine*, (98). https://www.apexbookcompany.com/a/blog/apex-magazine/post/the-turing-machines-of-babel.

Schwitzgebel, E. (2025). Two dimensionalism about other minds, and its implications for brain organoids and robots. *The Splintered Mind*. https://tinyurl.com/schwitzgebel20251022.

Schwitzgebel, E. et al. (2023). Creating a large language model of a philosopher. *Mind & Language*, 39(2):237–259.

Schwitzgebel, E. and Pober, J. (2024). The Copernican argument for alien consciousness; The mimicry argument against robot consciousness. https://arxiv.org/abs/2412.00008.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3:417–457.

Sejnowski, T. J. (2023). Large language models and the reverse Turing test. *Neural Computation*, 35(3):309–342.

Seth, A. K. (2025). Conscious artificial intelligence and biological naturalism. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X25000032.

Shanahan, M. (2024). Talking about large language models. *Commun. ACM*, 67(2):68–79.

Shastri, S. (2025). The art and discipline of prompt engineering. *Blog@CACM*. https://cacm.acm.org/blogcacm/the-art-and-discipline-of-prompt-engineering/.

Sieg, W. (2008). On computability. In Irvine, A., editor, *Philosophy of Mathematics*, pages 525–621. Elsevier, Oxford.

Simon, H. A. (1996). *The Sciences of the Artificial, Third Edition*. MIT Press, Cambridge, MA.

Singer, N. (2023). At this school, computer science class now includes critiquing chatbots. *New York Times*. 6 Feb., https://tinyurl.com/nsinger20230206.

Somers, J. (2025). *The New Yorker*, pages 28–33. 10 November.

Sternberg, R. J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence*. Cambridge University Press, Cambridge, UK.

Stich, S. and Ravenscroft, I. (1994). What is folk psychology? *Cognition*, 50(1–3):447–468. Page ref. to https://tinyurl.com/stich-ravenscroft1994.

Suzgun, M., Gur, T., et al. (2024). Belief in the machine: Investigating epistemological blind spots of language models. https://arxiv.org/abs/2410.21195.

Tellex, S. and Watkins, D. (2026). Elephants don't write sonnets: The Grounded Turing Test for embodied AI. In Konidaris, G., editor, *Designing an Intelligence*. MIT Press, Cambridge, MA.

Thorpe, C. (1989). *Computing Reviews*. Cover quotation, dated 22 August 1989.

Trott, S., Jones, C., et al. (2023). Do large language models know what humans know? *Cog. Sci.*, 47(e13309).

Turing, A. M. (1936). On computable numbers. *Proc. London Math. Soc., Ser. 2,*, 42:230–265.

Turing, A. M. (1947). Lecture to the London Mathematical Society on 20 February 1947. In Copeland, B. J., editor, *The Essential Turing*, pages 378–394. Oxford University Press (2004).

Turing, A. M. (1948). Intelligent machinery. In Copeland, B. J., editor, *The Essential Turing*, pages 410–432. Oxford University Press (2004), Oxford.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236):433–460.

Turing, A. M. (1951a). Can digital computers think? In Copeland, B. J., editor, *The Essential Turing*, pages 476–486. Oxford University Press (2004).

Turing, A. M. (1951b). Intelligent machinery, a heretical theory. *Philosophia Mathematica*, 4(3):256–260.

Turing, A. M. et al. (1952). Can automatic calculating machines be said to think? In Copeland, B. J., editor, *The Essential Turing*, pages 487–506. Oxford Univ. Press, 2004.

Vaswani, A., Shazeer, N., et al. (2017). Attention is all you need. In *31st Conf. Neural Info. Processing Systems (NIPS 2017).* https://tinyurl.com/vaswani-etal2017.

von Hippel, P. T. (2023). ChatGPT is not ready to teach geometry (yet). *Education Next*. 4 Jan.

Watt, S. (1996). Naive psychology and the inverted Turing test. *Psycoloquy*, 7(14). https://www.cogsci.ecs.soton.ac.uk/cgi/psyc/newpsy?7.14.

Waxman, O. B. (2024). AI in the classroom. In Vered, A. and Weintraub, P., editors, *Artificial Intelligence: A New Age of Possibilities*, pages 44–47. Time [Magazine].

Weizenbaum, J. (1962). How to make a computer *Appear* intelligent. *Datamation*, 8(2):24–26.

Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9(1):36–45.

Witt, S. (2025). The A.I. prompt that could end the world. *New York Times*. https://tinyurl.com/witt20251010.

Wittgenstein, L. (1933–1934). *The Blue and Brown Books*. Basil Blackwell, 1964, Oxford.

Włodarczyk, M. (2023). #ChatGPT_dialogues. https://tinyurl.com/wlodarczyk2023.

Wolfram, S. (2023). *What Is ChatGPT Doing... and Why Does It Work?* Wolfram Media. wolfr.am/SW-ChatGPT and wolfr.am/ChatGPT-WA.

Wolman, D. (2012). A tale of two halves. *Nature*, 483:260–263.

Yong, E. (2022). *An Immense World: How Animal Senses Reveal the Hidden Realms Around Us*. Random House, New York.