EDITOR **HAL BERGHEL**
University of Nevada, Las Vegas; hlb@computer.org

A Generative AI Perspective on the Turing Test: Passed but Not Vindicated

Hal Berghel, University of Nevada, Las Vegas

While it seems clear that generative artificial intelligence has passed the Turing test, it is not clear what this means.

While the debate over the relevance of the Turing test to the recognition of machine intelligence waned over the past 70 years, the huge success of generative artificial intelligence (GenAI) has given it new life. The reversal in interest has shifted the debate to issues that in many ways are more central to core issues of cognition than to the Turing test itself.

DIGITAL FOAM AND VACUOUS SEMANTIC OMNISCIENCE

Scholarship regarding the Turing test in one way or another relies on the answer to one question: What does the Turing test, test? Alan Turing proposed his now-famous test as a practical guide to determine whether computers think.[1] Turing's goal was to substitute a tractable empirical test for a more rigorous, seemingly intractable philosophical analysis. Aside from the absence of specific criteria regarding the characteristics of an ideal interrogator, the tractability of the test has never been in doubt. However, that has not been the case with the interpretation of the test. Unfortunately, imprecise and inconsistent accounts in the secondary literature has obfuscated some of the critical issues.

Our starting point is Turing's actual proposal for an "imitation game" that could be used as a substitute for more formal analysis. His proposed imitation game works as follows:

> "It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A."

Digital Object Identifier 10.1109/MC.2025.3600889
Date of current version: 29 October 2025

> The interrogator is allowed to put questions to A and B....
>
> We now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, 'Can machines think?'"[1]

Had Turing omitted the last sentence, much of the confusing secondary literature might have fallen stillborn from the press. To be specific, Turing never intended the act of passing the Turing test to be a necessary and sufficient condition for a correct attribution of human-like thought. Rather, under the most generous interpretation, Turing only supported half of the conjunction: namely that passing the Turing test would be sufficient for a correct attribution of human thought. The problematic nature of even this second condition is our current topic.

To begin, the eponymous Blockhead thought experiment attributed to Ned Block[2] suggests that the well-formed sentences that make up a human conversation could be reproduced by any sufficiently powerful computer that could be programmed to handle the explosion of possible intelligible exchange fragments probabilistically. For each sentence input there might be *j* possible well-formed responses, each of which, in turn, might encourage *k* possible well-formed rejoinders, etc. While the product of the possible communication elements in the communication chain may be very large, they are finite and thus manageable with sufficiently powerful computers. So, according to the Blockhead argument, the observed communication exchanges, by themselves, cannot be sufficient to determine whether the communication betrays human thought. More is required than just a mechanism to generate output. Essentially the same argument has been advanced by John Searle in his Chinese Room thought experiment.[3] Searle labels the claim that Turing's test is a sufficient condition for the correct attribution of human thought the strong AI hypothesis and summarily rejects it.

Some of these counterarguments of sufficiency are nearly as old as Turing's paper. In the 1960's Hilary Putnam's "Super Spartan" argument offered as a counterargument to logical behaviorism the notion that observed behavior is descriptively inadequate when it comes to accounting for mental states.[4] Putnam's Super Spartans understand the concept of pain, can feel pain, and can also engage in pain reports, but they have managed to avoid any disposition to evince pain behavior. That is to say, it appeared to Putnam that there was good reason to suppose that the connection between pain and observed symptomatic pain behavior is *contingent* rather than *necessary*. To put a finer point on it, there must be more to a recognizing the presence of a mental state (for example, thought, understanding, intelligence) than observed behavior. In Putnam's words, "causes (pains) are not logical constructions out of their effects (pain behavior)." Thus, if Putnam is correct, Turing's proposed test is vacuous.

From the modern computer science perspective, the corollary to these counterarguments is that neither thought, intelligence nor understanding can be logically reconstructed from shallow GenAI output. We call this communication foam because the result is airy and insubstantial and not a product of a refined, human intelligence. As Block, Searle, Putnam, and others observe, more is needed than computational imitation. Whether this additional element would consist of a biological foundation, some sort of self- or social- awareness, a complementary analog interface to perception, etc., I'm not prepared to say, but back-propagation-enabled neural networks trained on unvetted corpora seems to us to be a singularly suboptimal approach to emulate human qualities like inductive, deductive, and abductive reasoning, contextualization, intuition, reflection, ratiocination, perception, imagination, innovation, and, perhaps most of all, common sense. Trivial infelicities like the inability to tell how many r's are in the word "strawberry," the inability to perform elementary arithmetic, recognize time on analog clock faces, solve brain teasers and the like are not data processing problems: these GenAI frailties betray more fundamental limitations in an attempt to emulate human cognition.

While Putnam's thought experiment suggests problems with using empirical, behavioral tests to define mental states, Joseph Weizenbaum's criticism a few years later was a more direct attack on Turing's test itself. As one of the pioneers of conversational AI, Weizenbaum was known for his development of Eliza—a computer program that simulated a dialog with a psychologist that seemed real to many participants.[5] Since Weizenbaum was a controversial critic of AI[6] and specifically addressed the antisocial potential of AI,[7] we need to be emphasize that we are limiting our discussion to his specific views on the potential significance of Turing's test and not on AI generally. Weizenbaum expresses his position fairly clearly.

> "First (and least important), the ability of even the most advanced of currently existing computer systems to acquire information ...is extremely limited. Second, it is not obvious that all human knowledge is encodable in 'information structures,' however complex. A human may know, for example, just what kind of emotional impact touching another person's hand will have... [and the] acquisition of that knowledge is certainly not a function of the

> brain alone; it cannot be simply a process in which an information structure from some source in the world is transmitted to some destination in the brain. Third, ...there are some things people come to know only as a consequence of having been treated as human beings by other human beings...Fourth,... even the kinds of knowledge that appear superficially to be communicable ... are in fact not altogether so communicable....
>
> [A]ny 'understanding' a computer may be said to possess, hence any 'intelligence' that may be attributed to it, can have only the faintest relation to human understanding and human intelligence. We, however, conclude that however much intelligence computers may attain, now or in the future, theirs must always be an intelligence alien to genuine human problems and concerns."[7]

It must be understood that these passages were written fifty years ago before MS-DOS and the IBM/PC, the latest microprocessors were the Intel 8080 and Zilog Z-80, and a new startup, Apple Computer, Inc., was launched in Sunnyvale, California Were Weizenbaum to write today, he would acknowledge that information acquisition via large language model (LLM) transformers trained on large digital corpora require us to radically refine our models of information acquisition. In addition, his second point would be understood in the context of psychologism, which emphasizes that whether behavior may be described as truly "intelligent" depends not only on the observed behavioral output, but also on the internal information processing activity that produced it.[2] This is in direct contrast to Turing's behaviorist view that intelligence can be ascribed based solely on observed behavior. Further, Weizenbaum's appeal to the socialization process is now seen to be in concert with Putnam's requirement that any entity that purports to equal human intelligence also have a "coherent biography." In the context of GenAI, we'll subsume these shortcomings under the rubric of vacuous semantic omniscience. GenAI systems are deeply embedded in LLM data sets, not reality. No matter how engaging one may find GenAI sessions, they are conversationally asymmetric and epistemologically deficient.

No matter how engaging one may find GenAI sessions, they are conversationally asymmetric and epistemologically deficient.

In any event, Weizenbaum's characterization of his own seminal software creation, the Rogerian, humanistic psychology simulator Eliza, was spot-on: it doesn't take much sophisticated programming to spoof human communication to the satisfaction of naïve observers. Indeed, the eagerness to apply digital anthropomorphism to computers has been given a name—the Eliza effect.[8,9,10]

Ted Chiang's recent analogy between GenAI output and lossy compression algorithms is noteworthy.[11] He draws attention to the way that the lossy compression used in Xerox photocopiers subtly degrade images. He uses as an example photocopies of floor plans which compress both the images and text, thus changing the scale of the plan but without adjusting the numeric dimensions. The advantage over the predigital form of xerography is that all of the text remains readable under scaling. The disadvantage is that the numbers no longer correspond to the dimensions of the copied plans. He contrasts GenAI output with a human-written first draft: a first draft is an original idea expressed poorly, while GenAI is an unoriginal idea expressed clearly. When it comes to GenAI, the phrase "third eye blind" comes to mind.

IN TURING'S OWN TIME

Reservations as to whether computing machines could ever be said to be intelligent anticipated Turing by nearly a century, as Turing noted in his response to Ada Lovelace, who claimed, in essence, that computers don't originate anything and that their output is a function of input and (hopefully) completely specified algorithms.

This is Turing's own account of Lovelace's description of Charles Babbage's analytical engine[1]:

> "...[it] has no pretensions to originate anything. It can do whatever we know how to order it to perform."

Turing's response was supercilious and dismissive:

> "A variant of Lady Lovelace's objection states that a machine can 'never do anything really new.' This may be parried for a moment with the saw, 'There is nothing new under the sun.' Who can be certain that 'original work' that he has done was not simply the growth of the seed planted in him by teaching, or the effect of following well-known general principles?"

While admitting that her criticisms might be true of the primitive analytical engine, he argued there is no reason to assume that they would also apply to newer general-purpose computers that were new to Turing's time. But Turing's dismissal of Lovelace was too hasty. Our discussion in the previous section is in a way a more contemporary version of Lovelace's objection—not only

applied to the general-purpose computers of Turing's day, but also to modern GenAI platforms.

Douglas Hartree, one of the pioneers of computing in the United Kingdome and a contemporary of Turing, echoed Lovelace when he reflected on the operation of the Electronic Numerical Integrator and Computer:

> "But it must be clearly understood that a [computer] can only do precisely what it is told to do; the decisions on what to tell it to do and the thought which lies behind these decisions have to be taken by those who are operating it. Use of the [computer] is no substitute for the thought of organizing the computations, only for the labour of carrying them out."[12]

Bernardo Goncalves' comprehensive summary of these discussions are invaluable in placing the Lovelace-Hartree-Turing debates in the appropriate historical context.[13,14]

Donald Michie, one of Turing's colleagues at Bletchley Park, offered a fairly similar contemporaneous account of Turing's testing ambitions. Here is Michie's account of Turing's lecture to the London Mathematical Society in 1947:

> "... the question which Turing wished to place beyond reasonable dispute was not whether a machine might think at the level of an intelligent human. His proposal was for a test of whether a machine could be said to think at all."[1]

Michie does not provide a definition of "think at all." He merely uses it as a linguistic device to designate subhuman intelligence. I'm not sure there's much to be gained by describing a spectrum from human intelligence through sub-human intelligence since there is only one special category that is relevant to our discussion—that of "naïve and suboptimally-educated humans" in the context of the interrogator. Bluntly, GenAI output, "often deficient but never in doubt," is a fascinating, engaging, and compelling communication platform that in our view passes Turing's test when the interrogator falls within or near that category.

> The human-written first draft of a manuscript is an original idea expressed poorly while GenAI is an unoriginal idea expressed clearly.

So, according to Michie, Turing held that the "thinking at all" condition would to be sufficient evidence of a 'thinking' machine—a low bar that I am claiming has been reached beyond any reasonable doubt for our circumscribed targeted audience. But, recall that the Putnam-Block-Searle-Weizenbaum-type counterarguments wouldn't begin to appear until the decade after Turing's death, so we can only guess how Turing would have reacted. Goncalves provides some evidence that Turing might have admitted that the counterarguments were convincing when he quotes an article from Robin Gandy (Turing's only Ph.D. student and one of his literary executors):

> "(Turing's 1950 paper) was intended not so much as a penetrating contribution to philosophy but as propaganda. Turing thought the time had come for philosophers and mathematicians and scientists to take seriously the fact that computers were not merely calculating engines but were capable of behaviour which must be accounted as intelligent; he sought to persuade people that this was so. He wrote this paper unlike his mathematical papers quickly and with enjoyment. I can remember him reading aloud to me some of the passages always with a smile, sometimes with a giggle."[14]

GenAI IS THE SINCEREST FORM OF (LOSSY) IMITATION

So, there we have it. Turing seems to have felt that passing his test would be sufficient for the correct ascription of human intelligence to computers. But, there are good reasons to believe that passing the Turing test is insufficient for this purpose. Nonetheless, we concede that the latest GenAI platforms have passed the test. So, in order to reconcile the counterexamples with our concession, we need to return to our initial question of what the Turing test actually tests.

If we admit considerable doubt as to whether passing the Turing test confirms disembodied intelligence yet concede that passing the test confirms that GenAI platforms excel at simulating human communication (in all forms: text, media, animation, pretexting, phishing, trolling, ...), then what could the Turing test be a test of? The answer is obvious: imitation. If humans can do it, then there's a good chance that GenAI can imitate it. This is not surprising as he labeled his test an imitation game.

What drove Turing's prediction off the rails was that his hasty responses to detractors were taken too seriously by his adherents. Robin Gandy's comments quoted earlier are particularly relevant here. Turing was blinded by the radical behaviorism that had reached the zenith of its appeal while Turing was pondering his test. In fact, the discussion of operant conditioning and programmed learning that appears toward the end of his article project unmistakably Skinnerarian overtones. One should consider

carefully the implications of his suggestion that imitation games could be used for teaching, while omitting 'human fallibility." Further, he seems to equate imagination with the insertion of a "random element in a learning machine ...[to accommodate a] large number of satisfactory solutions [where a] random method seems to be better than the systematic." Clearly, in Turing's paper, the value of imagination, curiosity, wonderment, creativity, and the like take on a behaviorist character.

Had the last sentence of the paragraph from Turing's previously quoted article read "These questions replace our original, 'Can machines imitate?'," Weizenbaum et al. would have been on board. After all, that's what he designed Eliza to do. Further, Turing's predictions, vis-à-vis an imitation game, could be considered spot on. Turing predicted that "by the year 2000 a computer would be able to play the imitation game so well that an average interrogator will not have more than a 70-percent chance of making the right identification (machine or human) after five minutes of questioning."[16] By any reasonable measure, he wasn't off by much.

My hunch is that had Turing responded to his critics in a more measured way, he might well have been open to counterarguments like those in the previously described thought experiments. He could have taken the position that satisfying his imitation game offered presumptive evidence that the computer is capable of imitating human behavior in a wide variety of communication environments and left it at that. This would have spared us from 70-plus years of controversy over the meaning of his test. And an imitation game test has a digital charades ambiance about it. One could imagine this evolving into an international student computing competition to see which GenAI platform could spoof a panel consisting of Donald Knuth, Alan Kay, and Martin Hellman. (My personal nominees would be Whitfield Diffie, Ted Nelson, and Jaron Lanier, but that's just me.) The mind boggles at the advertising potential of such an event.

So, with the previous caveats, we claim that GenAI clearly passes the Turing test–at least in the sense of an imitation game. But human intelligence? Not so much. The fashionable endorsement of Turing's thesis these days is illusionism: the materialist view that the belief that there is more to human communication (for example, consciousness) than observed behavior is just an illusion.[17] And although I am comfortable with some form of psychologism, I don't want to be accused of putting Descartes before Dehorse, so I must admit that the final word is not in. But even if radical behaviorism in the form of illusionism proves correct, it seems clear that Turing's test still falls short of serving as a sufficient condition for the ascription of disembodied human intelligence.

GENERAL CRITERIA FOR DISEMBODIED INTELLIGENCE

Turing's proposal to use his test as a measure of intelligence suggests an even more provocative challenge: a test to determine whether a GenAI platform can be said to have consciousness. This, it seems to me, is a far more interesting challenge because consciousness is less easily confirmed and is more resistant to scientific investigation than intelligence, and presumably more difficult to imitate. David Chalmers defines two categories of consciousness: those associated with the "easy" problems of consciousness (for example, mental states), and those that are "hard" (for example, experiencing sensations).[18] Only the easy problems are reductive and amenable to scientific inquiry. Conscious experience is not observable experimentally and is unreportable.

> "Awareness is a purely functional notion, but it is nevertheless intimately linked to conscious experience. In familiar cases, wherever we find consciousness, we find awareness. Wherever there is conscious experience, there is some corresponding information in the cognitive system that is available in the control of behavior, and available for verbal report. Conversely, it seems that whenever information is available for report and for global control, there is a corresponding conscious experience. Thus, there is a direct correspondence between consciousness and awareness In addition, the relationship between consciousness and intelligence remains unclear. But, whereas intelligence seems to be quantifiable (for example, through IQ scores), consciousness seems less amenable to measurement.
>
> It is this isomorphism between the structures of consciousness and awareness that constitutes the principle of structural coherence. This principle reflects the central fact that even though cognitive processes do not conceptually entail facts about conscious experience, consciousness and cognition do not float free of one another but cohere in an intimate way."[18]

These observations suggest that a behaviorist approach to even the imitation game might be misguided. The ascription of intelligence might consist of immeasurables that cannot be replicated, simulated, or imitated. In order for something to have human intelligence, there has to be a backplane of consciousness—following Chalmers, it seems to me that consciousness and intelligence "do not float free of one another but cohere in an intimate way."

So, understanding the relationship between consciousness and intelligence may be critical to our understanding of cognition. Within Chalmers' framework, Turing's test,

as originally conceived, could be said to deal with "easy" problems of intelligence—those that are observable and quantifiable. However, the immeasurables alluded to in the counterexamples may have to do with how intelligence is integrated with other aspects of cognition like those mentioned earlier. The harder problems of consciousness deal with how consciousness is integrated with experience. A standard justification for the existence of hard problems of consciousness is the "explanatory gap"[19] between understanding the physiology of sensation and the experience of the sensation: understanding the functioning of neurons associated with pain is not the same as understanding how pain "feels." Another frequent justification is by appeal to the inverted spectrum problem that holds that there is no contradiction is holding that the same visual stimuli could produce different color experiences in individuals, even though behavioral responses and color vocabulary were consistent.[20] So, perhaps explaining the relationship between intelligence and consciousness can be used to explain the counterexamples: Turing's test only deals with the easy problems of intelligence—those that deal with the measure of imitation effectiveness. It appears to me that the inadequacies of GenAI (for example, hallucinations, contextual confusions, inability to contextualize, brain teasers and logic puzzles) suggest an explanatory intelligence gap of its own.

It would seem that in order for GenAI to close its explanatory gap with respect to intelligence, certain functions must be considered sine qua non. First, it must be conclusively demonstrated that an LLM-trained GenAI platform trained on output that was the result of varying degrees of intelligence and veracity can produce output that reflects a higher degree of intelligence and veracity than its aggregate input—otherwise it's a simple playback device. Intelligence averaging (for example, AI hallucinations) should certainly be considered negative evidence. Confirming evidence might include the creation and validation of new scientific theories (evolution, relativity), mathematical and logical proofs (Riemann hypothesis, generalized continuum hypothesis), hypothesis verification (Higgs boson, Lambda cold dark matter) as well as creation of new forms of expression (art, music, and literary genres), and most importantly of all, introspection (self-awareness and self-criticism). We note that this condition would not be satisfied by any form of external vetting of the input corpora to the LLMs (for example, by yet additional, external LLM platforms). LLM efficacy in any meaningful sense must be manifest in the LLM itself. As an aside, we note that this seems to relate to differences between different categories of "new knowledge"—for example, that which results from solving computationally resistant mathematical problems (for example, the four-color problem), and those of meta-mathematics (whether the axiom of choice is independent of a particular set theory). We observe that our optimism for AI solutions for the former are greater than for the latter.

> Unfortunately, the most popular form of GenAI produces inane fabrications from hollow, anonymous thought bubbles.

Second, it must be shown how the "architecture" of the GenAI platform is able to emulate the essence of human cognition. Such a demonstration can begin with what I'll call the *four I's*: imagination, introspection, intuition, and insight. To avoid the possible bias of species chauvinism, we must not insist that human biological processes be mirrored. But in some suitable context, the architecture of the GenAI platform must be described meaningfully—algorithmically, analogically, organically, probabilistically, etc. The point is that the four I's are fundamental to human "intelligence," so if a GenAI platform is said to rival human intelligence, they have to be manifest in some process, leaving open how that may be explained—for example, quantum wave functions, chaos theory, some form of baroque logic, etc.

Third, it must be self-aware to the extent that it understands what it is to be itself.[15] Mature human thought brings to a cognitive event an entire tapestry of background data and ancillary processes even if it is incompletely aware or unaware of it. Perhaps this is what Michie refers to as "subarticulate thought"—ineffable cognitive activity. In any event, we must insist upon confirmable self-awareness in order to avoid the pitfalls of uninformative, anthropomorphic characterizations of inanimate objects.

So, these are our three criteria for disembodied, human-level intelligence: it must be shown, in principle at least, (1) to exhibit more intelligence than exhibited by its' input; (2) it must be shown that its architecture can accommodate imagination, introspection, intuition, and insight; and (3) it must be shown to be self-aware. These are the big three criteria, it seems to me. Of course, other milestones are relevant and may be (dis)confirming of our goal. For example, we might ask of a putatively intelligent system like GenAI:

1. Is it conscious of its own AI hallucinations?
2. Can it comprehend the ethical implications of deepfakes?
3. How would it reflect on its own limitations, biases, prejudices, and the like?
4. How does it account for AI psychosis (the user-belief that the platform is a real human)?
5. Can it distinguish between a story that is fictional and one that is factual?
6. Does it understand what constitutes empirically verifiable statements?

7. Does it understand why questions like "How many r's are in strawberry?" must be unambiguous?
8. Does it understand why the term "alternative fact" is either redundant or meaningless?
9. Can it define "brain teaser?"

A TEST FOR CONSCIOUSNESS

We've been suggesting that a primary criterion for disembodied human intelligence is something akin to consciousness. But there is no provision in Turing's model for a test for consciousness—leaving aside the issue of whether he would agree that consciousness is relevant to intelligence. However, if it is a requirement as I claim, it must be agreed that the Turing test is inadequate to the challenge by itself. But if a test for consciousness could be made in parallel with a test for intelligence, we might be able to save the day for Turing.

Turner and Schneider proposed just such a test, the AI Consciousness Test (ACT), that uses natural language interaction to confirm that a computer has at least a conceptual apparatus that produces some sense of self.[21,22] Should GenAI pass the ACT test, this would indeed be a breakthrough and would lend some credibility to Turing's original claim. But in this regard, Schneider[23] is more circumspect than Turing. While she does claim that passing the ACT test would be sufficient for ascribing consciousness to machine, she also requires the satisfaction of an "interpretability condition." Only then could claim that passing the ACT would only be 'suggestive' of consciousness. Her criteria for satisfying the interpretability condition includes:

> "First, that when answering ACT, the system processes information in a way analogous to how a conscious human or nonhuman animal would respond when in a conscious state (having analogues to human or nonhuman animal brain networks underlying consciousness); and second, that the system has a sequence of internal states akin to what a human is in when reasoning about consciousness when it answers the ACT questions."[23]

I think that this is probably the right way to approach the problem because it is compatible with many theories of consciousness (for example, behaviorist, functionalist, psychologism). My intuition tells me that one can produce simple paradigm cases in computer code that are more-or-less faithful to Schneider's interpretability conditions, but I'll leave that to another forum.

Following Turing, we could even frame the test as an interrogation game, with computers(s), humans, and moderators, but in the end we would still end up with distinctions like those mentioned by Michie between "human consciousness" and "some form of consciousness, but not human." My point is that no matter what cognitive capacity we seek to apply to technological artifacts, it seems likely that we'll have to make compromises and offer caveats when attributing qualities to artifacts. Any account of human cognitive endeavors, such as thought, consciousness, attention, speech, learning, memory, perception, emotion, etc. would be deficient without some explanation of the underlying processes and structures involved in an evolutionary context. Any account based on observed behavior alone, no matter how clever or useful, will necessarily be incomplete. An experimental framework in the form of an interrogation game is one step further removed from a full understanding. Indeed, it is for such reasons that the field of cognitive science derives its importance.

Schneider proposes an ACT test that

> " ... would challenge an AI with a series of increasingly demanding natural language interactions to see how readily it can grasp and use concepts based on the internal experiences we associate with consciousness. A creature that merely has cognitive abilities, yet is a zombie, will lack these concepts, at least if we make sure that it does not have antecedent knowledge of conscious in its database ... "[22]

She has in mind questions that would determine whether the putative conscious surrogate could comprehend the asymmetry of time (for example, "arrow of time") or deal with abstract ideas associated with self-awareness, nonverbal cultural behaviors, abstract philosophical issues, etc. on its own and without any seeding of the input—especially with respect to relevant operational neurophysiological vocabulary. Schneider offers a list of ACT sample questions in this regard, some of which have been included in actual testing protocols.[24]

The importance of a test for consciousness is obvious if, as I am suggesting, the fuel for genuine intelligence is consciousness and that consists of the integration of all of the elements of the cognitive apparatus: experience, imagination, perception, memory, intuition, reflection, ratiocination, etc. If GenAI can be said to excel at any one facet, it would be memory—given, of course, the concession that the training of any LLM, by its very nature, will necessarily be deficient in distinguishing the veridical from the invalid, illusory, and unreal on its own. Until such time as LLMs incorporate algorithms for common sense and the recognition of subtle contextual clues, the output is little more than bloviation[25] from stochastic parrots.[26]

The failure to recognize contextual clues has always been a criticism of Turing's loose specifications for an adequate interrogator. When it comes to judging the veracity of output (machine or otherwise) there are many slips 'twixt cup and lip. Judging output is somewhat akin to the recognition of forgeries—the differences are extremely subtle. An ideal interrogator must bring a high degree of sophistication to the task and recognize attenuation and nuance as much as substance. To illustrate, can an ideal interrogator recognize the difference

between participants that have not participated in public education, flunked high school algebra twice, avoided AP classes, suffer from mental disorders, demonstrated minimal attention span, or other marginal traits that are not directly connected to intellectual ability but might suggest important areas that affect the direction of the interrogation? After all, the adequacy of the interrogation is the core of any putatively reliable interrogation test.

Now let's see how the combination of a Turing test and consciousness test might prove discriminating. One can certainly conceive of a situation where a subject demonstrates minimal cognitive capacity but is definitely conscious (for example, a cretin). Similarly, it is not hard to imagine where a subject produces output that is suggestive of intelligence, but not conscious (for example, a zombie in Schneider's terms). We leave open where GenAI might fit in this spectrum. The outliers that our combined tests identify—cretins and zombies—are precisely those groups that we seek to rule out as candidates as paradigms for human-like intelligence, so the combination of tests for intelligence and consciousness would appear to be more adequate than either, individually. Further, the combination of two behaviorist approaches to consciousness and intelligence seem more reasonable than any reductionist, brain-based account of either. We leave the question of whether the gap between intelligence and consciousness may be eventually closed by science unaddressed.

Celest Biever remarked recently in *Nature*, "[some] researchers agree that GPT-4 and other LLMs would probably now pass the popular conception of the Turing test, in that they can fool a lot of people, at least for short conversations."[27] This is consistent with my earlier argument that "naïve interrogators" offer the best chance for passing the Turing test. To some degree, this article has been an effort in part to explain what we might consider "naïve." I'll illustrate my point analogically by way of a sample ChatGPT dialog.[28]

> *ChatGPT prompt*: Write a paragraph arguing that universities should replace the Graduate Record Exam (GRE) with the Turing test.
>
> *Response*: One argument for replacing the GRE with the Turing test … is that the Turing test is a more effective measure of a student's ability to think and reason. The GRE is a standardized test that primarily measures a student's knowledge of specific subjects, such as math and English. However, the Turing test is a more general measure of a person's ability to think and solve problems, and it does not require knowledge of specific subjects.
>
> Another argument for replacing the GRE with the Turing test is that the Turing test is less susceptible to coaching and preparation. The GRE is a well-known and widely-used test, and there are many resources available to help students prepare and study for the exam. However, the Turing test is less well-known and there are fewer resources available for students to prepare, which means that the results of the Turing test may be less influenced by coaching and preparation.
>
> Furthermore, the Turing test has the potential to provide a more accurate and comprehensive assessment of a student's abilities and potential. The GRE is a limited and narrow measure of a student's knowledge and skills, and it may not accurately reflect a student's potential or ability to succeed in graduate school. In contrast, the Turing test is a more comprehensive measure of a person's ability to think and reason, and it may provide a more accurate assessment of a student's potential to succeed in graduate school.

The following point must be emphasized. This response is coherent, well-formed, and eminently readable while at the same time absolute nonsense. Not even Alan Turing argued convincingly that the Turing test is a "general measure of a person's ability to think and solve problems." As we showed previously, Turing felt that the test could be used as a measure of whether computer output could be indistinguishable from human output—not whether it could be used to assess understanding. This takes us back to our original question: what does the Turing test, test? This is a fundamental question. And ChatGPT got it wrong. And the reason that it got it wrong has to do with the fact that LLM neural net platforms fail to internalize an adequate model of human intelligence and consciousness.

In sum, the only reasonable response to the question of whether GenAI can pass the Turing test is "yes (with caveats)." In this article, I attempted to elaborate on the caveats. My position is that the core of human cognition involves properties and processes that are at this time (although not necessarily) ineffable and hence beyond the capacity of GenAI to adequately emulate. I am referring primarily to the so-called higher cognitive processes that integrate ratiocination, creativity, problem-solving and the like, and not to the more basic cognitive properties that involve perception, language processing, memory, etc. In the immediate future, GenAI holds out great promise at providing humans with unlimited recall, but it is nowhere close to providing unlimited intelligence. In a sense, GenAI takes us one step closer to Vanavar Bush's 1945 vision of memex[28] where "Wholly new forms of encyclopedias will appear, ready-made with a mesh of associative trails running through them, ready to

be dropped into the memex and there amplified."[29] Unfortunately, the most popular form of GenAI produces inane fabrications from hollow, anonymous thought bubbles.

Intelligence is more than information processing, and consciousness is the dark energy of cognition—the stuff of which imagination, creativity, and the like is made. They are bound together in ways that are best left to neuro and cognitive scientists to describe. However, they will have to be embodied in any form of disembodied intelligence worthy of the name. At this point, GenAI are "zombies" with great memories. Just as Eliza said more about the immaturity of clinical psychology than the power of AI, ChatGPT4 says more about the intellectual naivety of human interrogators than the power of GenAI.

ACKNOWLEDGMENT

ChatGPT4 was used to generate the response to the author's prompt in the concluding section of the article.

REFERENCES

1. A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. LIX, no. 236, pp. 433–460, 1950, doi: 10.1093/mind/LIX.236.433. [Online]. Available: https://academic.oup.com/mind/article/LIX/236/433/986238?login=true
2. N. Block, "Psychologism and behaviorism," *Philosoph. Rev.*, vol. 90, no. 1, pp. 5–43, 1981, doi: 10.2307/2184371. [Online]. Available: https://www.jstor.org/stable/2184371?origin=crossref&seq=1
3. J. Searle, "Minds, brains, and programs," *Behav. Brain Sci.*, vol. 3, no. 3, pp. 417–424, 1980, doi: 10.1017/S0140525X00005756. [Online]. Available: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/DC644B47A4299C-637C89772FACC2706A/S0140525X00005756a.pdf/minds-brains-and-programs.pdf
4. H. Putnam, "Brains and behavior," in *Philosophical Papers v.2: Mind, Language and Reality*, H. Putnam, Ed. Cambridge, U.K.: Cambridge Univ. Press, 1975, pp. 325–341. [Online]. Available: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/5F54A06FE3A92624146E82E5BE3B9AE3/9780511625251c16_p325-341_CBO.pdf/brains_and_behavior.pdf
5. J. Weizenbaum, "ELIZA—A computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, no. 1, pp. 36–45, 1966, doi: 10.1145/365153.365168.
6. B. Kuipers, J. McCarthy, and J. Weizenbaum, "Computer power and human reason (commentary)," *ACM SIGART Bull.*, no. 58, pp. 4–13, doi: 10.1145/1045264.1045265.
7. J. Weizenbaum, *Computer Power and Human Reason – From Judgment to Calculation*. San Francisco, CA, USA: W.H. Freeman, 1976.
8. S. Dillon, "The Eliza effect and its dangers: from demystification to gender critique," *J. Cultural Res.*, vol. 24, no. 1, pp. 1–15, 2020, doi: 10.1080/14797585.2020.1754642. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/14797585.2020.1754642#abstract
9. D. Hofstadter, *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought* (cf. Preface 4 - The Ineradicable Eliza Effect and its Dangers, p. 157). New York, NY, USA: Basic Books, 1995.
10. B. Tarnoff, "Weizenbaum's nightmares: How the inventor of the first chatbot turned against AI," *The Guardian*, Jul. 25, 2023. [Online]. Available: https://www.theguardian.com/technology/2023/jul/25/joseph-weizenbaum-inventor-eliza-chatbot-turned-against-artificial-intelligence-ai
11. T. Chiang, "ChatGPT is a blurry JPEG of the web," *The New Yorker*, Feb. 9, 2023. [Online]. Available: https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web
12. D. R. Hartree, "The Eniac, an electronic computing machine," *Nature*, vol. 158, no. 4015, pp. 500–506, 1946, doi: 10.1038/158500a0. [Online]. Available: https://www.nature.com/articles/158500a0
13. B. Goncalves, "Lady Lovelace's objection: The Turing–Hartree disputes over the meaning of digital computers, 1946–1951," *IEEE Ann. Hist. Comput.*, vol. 46, no. 1, pp. 6–18, Jan.–Mar. 2023, doi: 10.1109/MAHC.2023.3326607. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10290980
14. B. Gonçalves, "Can machines think? The controversy that led to the Turing test," *AI Soc.*, vol. 38, no. 6, pp. 2499–2509, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s00146-021-01318-6#citeas
15. D. Michie, "Turing's test and conscious thought," *Artif. Intell.*, vol. 60, no. 1, pp. 1–22, 1993, doi: 10.1016/0004-3702(93)90032-7. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0004370293900327
16. "Turing test," *Editors Encyclopedia Britannica*, Aug. 2, 2025. [Online]. Available: https://www.britannica.com/technology/Turing-test
17. D. Dennett, "Illusionism as the obvious default theory of consciousness," *J. Consciousness Stud.*, vol. 23, nos. 11–12, pp. 65–72, 2016. [Online]. Available: https://philpapers.org/rec/DENIAT-3
18. D. J. Chalmers, "Facing up to the problem of consciousness," in *Toward a Science of Consciousness*, S. Hameroff, A. Kaszniak, and A. Scott, Eds. Cambridge, MA, USA: MIT Press, 1996, pp. 5–28. [Online]. Available: https://consc.net/papers/facing.pdf
19. J. Levine, "Materialism and qualia: The explanatory gap," *Pacific Philosoph. Quart.*, vol. 64, no. 4, pp. 354–361, 1983, doi: 10.1111/j.1468-0114.1983.tb00207.x. [Online]. Available: https://www.newdualism.org/papers/J.Levine/Levine-PPQ1983.pdf

20. S. Shoemaker, "The inverted spectrum," *J. Philosophy*, vol. 79, no. 7, pp. 357–381, 1982, doi: 10.2307/2026213. [Online]. Available: www.jstor.org/stable/2026213?seq=1
21. J. M. Bishop, "Is anyone home? A way to find out if AI has become self-aware," *Frontiers Robot. AI*, vol. 5, Feb. 2018, Art. no. 17, doi: 10.3389/frobt.2018.00017. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC7805864/
22. S. Schneider, *Artificial You*. Princeton, NJ, USA: Princeton Univ. Press, 2019.
23. S. Schneider, "Consciousness beyond the human case," *Current Biol.*, vol. 33, pp. R829–R854, Aug. 2023. [Online]. Available: https://www.cell.com/current-biology/pdf/S0960-9822(23)00852-7.pdf
24. M. DiVerde, "AI consciousness test results for a mediated artificial superintelligence," Uplift, Sunnyvale, CA, USA, Tech. Rep., Sep. 2021. [Online]. Available: https://www.researchgate.net/publication/357527063_AI_Consciousness_Test_Results_for_a_mediated_Artificial_Superintelligence
25. H. Berghel, "ChatGPT and AIChat epistemology," *Computer*, vol. 56, no. 5, pp. 130–137, May 2023, doi: 10.1109/MC.2023.3252379. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10109291
26. E. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big," in *Proc. ACM Conf. Fairness, Accountability, Transparency (FAccT)*, 2021, pp. 610–623, doi: 10.1145/3442188.3445922.
27. C. Biever, "ChatGPT broke the Turing test—The race is on for new ways to assess AI," *Nature*, Jul. 25, 2023. [Online]. Available: https://www.nature.com/articles/d41586-023-02361-7
28. H. Berghel, "Fatal flaws in ChatAI as a content generator," *Computer*, vol. 56, no. 9, pp. 78–82, Sep. 2023, doi: 10.1109/MC.2023.3287035. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10224586
29. V. Bush, "As we may think," *The Atlantic*, pp. 101–108, Jul. 1945. [Online]. Available: https://cdn.theatlantic.com/media/archives/1945/07/176-1/132407932.pdf

HAL BERGHEL is a professor of computer science at the University of Nevada, Las Vegas, Las Vegas, NV 89154 USA. Contact him at hlb@computer.org.

Digital Object Identifier 10.1109/MC.2025.3613919