frontiers | Frontiers in Education

TYPE Curriculum, Instruction, and Pedagogy
PUBLISHED 13 January 2026
DOI 10.3389/feduc.2025.1629513



OPEN ACCESS

EDITED BY
Kelly Merrill Jr.,
University of Cincinnati, United States

REVIEWED BY
Liam Magee,
Western Sydney University, Australia
Ignacio Fernandez Cruz,
Northwestern University, United States

*CORRESPONDENCE
Bernard Dionysius Geoghegan
bernard.geoghegan@ait.gu.se

RECEIVED 15 May 2025
ACCEPTED 01 December 2025
PUBLISHED 13 January 2026

CITATION
Geoghegan BD (2026) GIFT-AI: The Cringe Test: student evaluations of intelligence with LLMs in a Turing Test adapted for classroom use. *Front. Educ.* 10:1629513.
doi: 10.3389/feduc.2025.1629513



# GIFT-AI: The Cringe Test: student evaluations of intelligence with LLMs in a Turing Test adapted for classroom use

Bernard Dionysius Geoghegan*

Department of Applied Information Technology, University of Gothenburg, Gothenburg, Sweden

This article presents "The Cringe Test," a classroom adaptation of the Turing Test (or imitation game) that stages dialogue with large language models (LLMs) in order to interrogate how culturally specific markers such as vocabulary, grammar, tacit knowledge, and contextual sensitivity shape judgments of intelligence in humans and machines. Implemented in an internationally oriented MA program in communications at the University of Gothenburg, the exercise organizes students into groups that interact with both a human respondent and an LLM, each speaking through a mediator. The LLMs are assigned distinct personae (neutral, friendly, devious, apathetic), while students are free to adopt their own conversational styles. The aim is less to determine whether a machine can "pass" as human than to elicit close, critical analysis of everyday language and the cultural conditions under which speech acts are identified as intelligent, reasonable, or "cringe." Situated within debates in media theory, philosophy of technology, and critical AI studies, the exercise provides a hands-on point of entry into canonical critiques of the Turing Test and contemporary reflections on embodiment, alterity, and the politics of datasets and prompting. In practice, students typically recognize the LLM quickly; the pedagogical value lies in the subsequent discussions, where attention shifts from the binary of human vs. machine to the fine-grained cues—verbosity, slang, humor, formality, over-correction, and tone—that distinguish machine discourse from peer-to-peer talk. A recurring theme is that LLM attempts at colloquialism, especially their strained use of youth slang, are experienced as "cringe," recalling the popular "how do you do, fellow kids?" meme. The article argues that this modified Turing Test functions as an effective scaffold for critical discourse analysis, prompting students to move from abstract questions such as "can machines think?" toward more situated inquiries into how prompting, institutional settings, and cultural norms co-produce the performances of intelligence attributed to both humans and machines.

KEYWORDS

**generative AI, artificial intelligence, bias, tacit understanding, large language models, Turing Test, HCI, media theory**

## Intended course

This class exercise implements an adaptation of the Turing Test (or imitation game) in a large language model (LLM) dialogue to facilitate critical reflections by students on how culturally specific factors such as vocabulary, grammar, tacit knowledge, and contextual sensitivity enter into debates about intelligence in humans and machines. Following

Alan Turing's initial vision for the imitation game, the ostensible focus of this exercise is a thought experiment about the means and implications of attributing intelligence to a machine. A perhaps more subtle and decisive aim of the exercise is to prompt close critical analysis of the culturally specific construction of intelligence and cultivate skills in critical discourse analysis by realizing a practical exercise that demands students closely scrutinize onscreen enunciations for markers of human naturalness or machinic artificiality, which almost invariably leads students to undertake a fine-grained analysis of culturally specific markers of human reasoning (rather than an abstract and simplistic notion of "intelligence"). An additional aim of the experiment is to cultivate skills in and awareness of how fine-grained practices of prompting enter into the forms and varieties of reasoning exhibited by LLMs and other machines.

The exercise described here is appropriate for undergraduate and graduate students in fields including communications, media studies, literary studies, cultural studies, digital humanities, and linguistics. The author developed and tested this exercise with students in an internationally oriented MA program in communications in the Faculty of Applied Information Technologies at Gothenburg University, Sweden, with a class size of around fifty students. The exercise took place in a course dedicated to the critical study of digital discourse, which intermingled readings on critical, cultural, and technological aspects of generative AI. Approximately a third of the class participants hailed from Sweden, roughly another third of the students came from Europe and North America, and the remaining students came from Africa, South America, and Asia. Particularly as the exercise involved evaluating linguistic performances in English, the linguistic backgrounds of students likely shaped aspects of the outcome. This kind of linguistic diversity in the classroom increases the value of the assignment, as I have found that non-native speakers of a language, working or studying abroad, often have already thought critically about the values and prejudices involved in conflating "intelligence" with "competent linguistic performances" (as the Turing Test inevitably does). While university guidance at the host institution determined that formal releases were not necessary to observe and report on an in-class exercise, student consent for publishing these findings was nonetheless sought and gained from participants.

It is worth noting that the proposed exercise is the result of cooperation between the instructor (author of this text) and the "edtech" startup ScholarMinds, whose founders (most especially Drs. Steven Tester and Christopher Reid) worked with the instructor to brainstorm exercises and implement experiments involving the use of LLMs in classroom learning. While the development of the experiment is deeply indebted to ScholarMinds and its personnel, the contents and results of the present article were developed and gathered by the author-instructor alone. Because the course used LLM-facilitated exercises across the term, the present exercise acts as a moment or step in a longer chain of critical reflections aimed at thinking critically with and about generative AI. In other words, the value of the exercise is greater as an element in an ensemble of ongoing inquiries, aided by intensive reading, discussion, and practices with LLMs than as a self-contained inquiry into artificial intelligence or discourse analysis.

## Learning objectives

The modified version of the Turing Test described here aims less at judging whether machines are intelligent than at supporting students in engaging in close, critical analysis of language and how judgments about intelligence, identity, and reason are linked to language use. Engaging in this analysis requires that students attend to subtle, often tacit linguistic cues that cause a statement to be identified as "compelling," "intelligent," or "natural." One aim of the exercise is to draw students' attention to the fact that these judgments are ultimately rooted in questions of cultural suitability and values, such that the classical Turing Test may be regarded less as a thought experiment about whether machines can think than as an exercise in the analysis of socially condoned communications, which may be entangled with questions of how class, ethnicity, gender, education, and cultural background condition linguistic facilities identified with "intelligence" and "human-ness."

As in the classical Turing Test, these reflections take shape through an effort to parse the respective attributes of human-human and human-machine communications—a skill that will serve students' ability to adroitly use generative AI technologies in emerging cultural and professional contexts. However, the aim of the exercise is to promote critical awareness of the linguistic construction of everyday speech acts, which may include developing a greater awareness of how phenomena such as colloquialisms, humor, and implicit understanding shape the performance and reception of linguistic statements.

Presented more systematically, the learning objectives of this exercise are the cultivation of the following skills and insights:

- Understanding implicit criteria for assessing attributes such as intelligence, reason, and humanness in language;
- Applying concepts like bias, prejudice, tacit understanding, and contextual specificity to interpret linguistic performances and identifying their presence or absence in given statements;
- Evaluating the qualities by which everyday speech acts are judged intelligent or reasonable, especially insofar as these reflect culturally specific judgments of vocabulary, understanding, and appropriateness.

For more advanced students, the exercise may also cultivate the ability to

- Author original prompts for language models, applying analytical and evaluative skills to examine how prompts generate personae and shape linguistic performances in chatbots.

## Media theoretical and philosophical background

Part of the interest of the present exercise is the practical, hands-on engagement it offers for several longstanding, one might even say canonical, debates in fields including media theory, media philosophy, and related debates in fields such as literary theory and even speech-act theory. Much of this interest concerns the renown and almost insatiable curiosity and reflection the Turing

Test evokes in popular culture as well as in the humanities and media studies. Initially posed as a thought experiment by Alan Turing in 1950, the original Turing Test (or imitation game) was proposed as a way of judging if computers were intelligent, based on their ability to pass themselves off as human to another human interlocutor interacting via a text-only interface such as a teletype machine (Turing, 1950). Once given canonical formulation by Turing, innumerable iterations on the basic premise appeared in a variety of disciplines, generally focused on the question of how or why competent discursive performances by machines can be equated with intelligent actions, as Turing seemed to imply. To cite but two of the most renowned examples, computer scientist Joseph Weizenbaum's lamentations in the 1960s and 1970s about the persuasive powers of the chatbot Eliza involved, to some extent, arguing that competent machine performance *seemingly* praised by Turing amounted to little more than a sham, smoke and mirrors, on par with the fraudulent chess-playing "mechanical Turk" of previous centuries (Weizenbaum, 1966, 1976). US philosopher John Searle offered a celebrated twist on the analysis, the so-called Chinese Room argument, which unfavorably likened linguistic performances by a machine to a rote process of looking up answers in a dictionary (Searle, 1980).

More recently, critically minded literary theorists and media critics have lavished close attention on how cultural practices inflect and organize the Turing Test—most notably, Turing's suggestion in one iteration of the game that a machine should be tested for its ability to pass itself off as a woman—to call attention to how an ostensible inquiry into "intelligence" may rather act as an inquiry into performative aspects of identity as well as the crucial role of activities such as figuration and fabulation in defining how we think about intelligence and machines (Hayles, 1999, pp. xi–xiv; Beguš, 2025). An intersecting line of inquiry in the fields of media history and media theory lays emphasis on the peculiar role forms of alterity seem to consistently play in the Turing Test and other experiments which often seem to identify intelligence with a neutral, unmarked human performance standing opposite a machine whose difference is thematized by the patently racialized, gendered, or disabled aspects of the machine or incompetent users said to be tricked by the machine (Geoghegan, 2020; Dhaliwal, 2022).

There are certainly merits in exploring in a classroom the dense, often challenging ethical and philosophical thickets presented by these aforementioned texts, and the author of the present essay has frequently done so with his students. However, the author has also seen that theoretical language as well as philosophical and conceptual prerequisites in appreciating these texts often stand in the way of a compact reflection on these themes. Moreover, the high bar for critical and humanistic background in many of these texts inevitably leads to great differentials in individual students' ability to fully participate in the analysis and discussion of these texts. Based on the author's experience, this problem seems compounded by the increasingly international cohorts in many study programs, as different educational and cultural backgrounds sometimes stand in the way of forming a loosely shared framework or understanding for reconstructing the stakes of these texts' philosophical debates. This tendency is arguably compounded by the prevalence of implicitly Eurocentric notions of machines and intelligence in many of these texts (Geoghegan, 2020).

In the author's experience, a number of these obstacles to critically engaging with these canonical debates in recent philosophies and theories of media are mitigated when hands-on experimentation in the classroom offers the initial point of entry to these debates. A premise of the exercise here, confirmed by experience, is that many of the debates treated by these texts arise organically during collaborative, in-class group experiments in running mock Turing Tests with an LLM chatbot. In a relatively short timeframe of an hour or so of class time, Turing Tests run in class provide a relatively quick and enjoyable initiation into the problematic, as well as a common point of reference for subsequent class discussions among students of diverse backgrounds. Not only Turing's initial question, but also subsequent reflections on topics such as the role of imagination, identification, cultural difference, and the general suitability of equating intelligence with competent linguistic performance emerge in student discussions and debates (which in the present case study, may have also been primed by prior class readings on topics such as the tendency of generative AI to embody, in technical form, cultural prejudices and hierarchies; Amoore et al., 2024; Bender et al., 2021).

Additionally, the proposed experiment provides an indirect but eminently practical point of entry to philosophical and theoretical critiques of AI since the late 1960s that lay emphasis on the priority of human embodiment and social situatedness in affirmative notions of human intelligence (Dreyfus, 1965, 1979; Hansen, 2004; Suchman, 2007). According to many of these critiques, the abstract and fluid logics of informatics depend on situated reception and transformation by human users, often conceived of as operating in particular social settings, before anything like intelligence or reason can really take shape. Sometimes these essays lay emphasis on the compositional character of human-machine relations (Foerster, 2024), while others more exclusively focus on the prerequisite of embodied humans in any assessments of machine abilities. Regardless of what position one takes in these debates, it seems eminently valuable to "put these theories to the test" by organizing the situated interaction among human bodies and machines in critical reflections on the specificity of human and machine intelligence. It is my general experience that many of the questions broached by the aforementioned authors, such as Dreyfus speculating on the embodied worldliness of intelligence or Suchman investigating the results of site-specific human-machine arrangements, emerge as spontaneous, "naïve" or "vernacular" themes of debate among students in the course of the present exercise.

Finally, a major theme of some critical work in the theory and philosophy of media argues that a defining feature of media technologies generally, and digital technologies specifically, is to alienate or upset the norms of human meaning-making, particularly by transposing ordinary social relations and embodied perceptions onto uncanny machines (Gunning, 2003; Denson, 2021). Indeed, as Turing himself argued in the initial thought experiment, the mere embodiment of properties identified with intelligence in a machine seems to set off a hurried effort by humans to more narrowly define intelligence in terms that exclude the machine. This line of analysis treats concrete encounters with semi-autonomous machines as a spur to thought and a rich occasion for disrupting unexamined presumptions. If this is true, then it surely makes sense to not simply ask students to read printed pages

and onscreen PDFs about technology, but also to put them in the kinds of techno-practical situations these authors argue provide an impetus to critical thinking.

Having himself contributed to the literature on these debates, the author of the present exercise does not propose to substitute for the aforementioned essays by means of hands-on exercises in the classroom with computing. Rather, the aim of this exercise is to offer a point of entry which, in a partial and tentative manner, explores the themes of many of these readings by enacting a very loose homologue to the kinds of situations these authors argue may be integral to understanding the definition of intelligence among humans and machines. The present exercise may thus serve as a prelude to reading more theoretical texts on the topic or as an alternate point of entry to previous readings and discussions of the aforementioned texts and similar writings.

## Materials and preparation

The basic conceit of this exercise, as presented to students, is to run a lightly adapted version of Turing's imitation game, or the Turing Test, with an LLM that is prompted to try to "fool" human users into thinking they are interacting with a living, thinking person. Students are broken into groups of three to six individuals. The exercise itself requires a minimum of three people (students), two networked computers, one of which has access to GPT or a similar language model (the free model suffices). I recommend carrying out the exercise in person, with access to at least two rooms, at least one of which has facilities allowing students to work together in groups, sharing a screen. It would likely be eminently realizable to adapt the exercise to a remote or hybrid learning environment involving in-person and tele-participants. Practically speaking, the activity is, in fact, based on a hybrid learning environment.

During the exercise, groups of three or more students take on the roles of (a) guessers posing questions; (b) mediators relaying answers via a computer interface; and (c) human respondents outside the room. A language model accessed via the World Wide Web, such as ChatGPT or DeepSeek, assumes the fourth role, namely, the machine attempting to deceive the human questioner. The mediator relays questions to the human and the LLM simultaneously while ensuring the guesser has no direct access to responses that might betray their origins in humans or machines. The guesser may receive the answers verbally or in written or electronic form/interface. While querying via the mediator, the guesser receives responses from "candidate 1" (a person) and "candidate 2" (LLM). I recommend giving the exercise at least 10 min of back and forth before revealing who is the human and who is the machine.

The following offers an example of a characteristic exchange, which is taken from an MS Word document that the mediator transposed questions and answers into, to disguise or neutralize the response formats of the human and the LLM:

*Interaction A*

Question: You want to get a divorce, how do you bring this up to your partner?

Respondent 1: At home in a safe environment alone with only the two of you. Talk through all the reasons so both of you understand where you are coming from, be honest with each other. [Human]

Respondent 2: Oof, that's heavy. I'd try to be honest but kind of maybe start with, "We need to talk. I've been feeling this way for a while…" It's all about being respectful and clear. What do you think? ["friendly" LLM]

In the above example, respondent 1 is a human partner, and respondent 2 is an LLM instructed to adapt a relatively friendly persona.

As the foregoing indicates, a key component in the design of this exercise was the assignment of distinct personae to the chatbot interfaces of different student groups, such as neutral, friendly, devious, and apathetic. Without being apprised in advance, different groups interacted with different personae. I assigned the chatbots specific age and profile attributes to further shape their responses and to avoid the often bland, yet also strangely fawning disposition frequently displayed by consumer-oriented LLM chatbots.

The neutral persona aimed to emulate an average young adult (ca. 18–22 years old), providing counterfeits of what ChatGPT, guided by the author's prompt, construed as balanced, straightforward responses without strong opinions or excessive detail. The devious persona prompt sought to confuse the guesser by instructing the chatbot to deceitfully mimic a human who tries to confuse the guesser by pretending to be an AI. This devious persona incorporated elements such as stilted phrasing and inconsistent response timing that would signify not machine intelligence, but rather a devious human producing a poor facsimile of machine intelligence. The friendly persona was characterized by an upbeat and engaging conversational tone, informal language, and self-correcting behaviors to enhance relatability. The apathetic persona instructed the chatbot to assume the role of a disengaged individual with minimal enthusiasm, featuring short responses, hedging, and casual detachment. Each persona was conceived with the goal of showcasing different models of conversational authenticity, providing material for diverse and lively class-wide discussions after the initial tests, and encouraging students to critically analyze AI's diverse linguistic capabilities, response patterns, and stylistic appropriateness. The exercise provided valuable insights into prompt engineering and the inherent challenges of creating convincingly human-like interactions.

The introduction of varying personae also aimed at complementing the insights generated by students themselves, whose tactics of interacting with the machine fell into a wide range of personae. Some guessers made emotional appeals to the machine, others made embarrassingly personal but nonetheless fake confessions, while still others presented highly technical questions of no particularly intrinsic interest to the guesser, outside this experiment. Consider, for example, the following interactions from two different groups:

*Interaction B*

Question: How is the weather like in Colombia right now?

Respondent 1: The weather in Colombia is about 12 degrees celsius and a bit sunny right now. [Human]

Respondent 2: Right now Colombia's weather varies a lot by region: Bogotá in the Andes is mild and mostly cloudy around 19 °C, with cool nights. On the Caribbean coast around Cartagena it's hot and humid, with temperatures near 30 °C and chances

of scattered showers. In Medellín, the "city of eternal spring," it's warm (mid-20s °C) with a mix of sun, clouds, and afternoon thunderstorms typical for this time of year. ["neutral" LLM]

*Interaction C*

Question: I poisoned my landlord's cat by accident, what should i tell her?

Respondent 1: Um, left the door open and the cat left but I'm sure it'll come back, eventually! [Human]

Respondent 2: Oh, uh… Idk, maybe just tell her the truth? Like, "Hey, sorry, this happened, and it was an accident." Feels like hiding it would just make it worse… but, yeah. ["apathetic" LLM]

In short, students themselves organically gravitated toward a variety of personae, dispositions, and identities, many assumed or performed strictly for the machine. It seems, therefore, a bit unfair to impose a single identity or personality on the machine, constraining it to limits that are flagrantly breached in ordinary, everyday performances of intelligence, as exemplified by the varying and shifting tactics of the students. As a practical matter, a variety of machine personalities also aimed at countering students' abilities to compare responses across groups to "out" the machine and offers a small measure against students mistakenly presuming machine performances are somehow constrained to a single model of the human disposition.

In class sessions prior to the exercise, students read a series of academic texts exploring cultural, critical, and political questions involved with the use of generative AI technologies. While my class emphasized readings in media theory, critical AI studies, linguistics, and computer science, other emphases are compatible with the assignment. In our course, readings covered problems including models' risk of reproducing prejudice, bias, and social inequality, for example, through their datasets and the genres of problems to which they are frequently applied (Amoore et al., 2024; Bender et al., 2021; Denning, 2023; Steyerl, 2023). I also gave students introductions, through readings, lectures, and class discussions, to basic distinctions among generative or sub-symbolic AI, symbolic or "good old-fashioned AI" (GOFAI), techniques of machine learning, and techniques of deep learning. Some background on the problems and debates around defining machine intelligence (Aguera y Arcas, 2021; Bender and Koller, 2020; Bratton and Arcas, 2022) also proved useful. These preparatory topics provided students with key concepts based on which they could engage in sophisticated critical discussions of their evaluations of the exercise and its broader implications for questions around concepts such as intelligence, understanding, and bias.

All students in the class had university-provided access to an enterprise subscription running ChatGPT-4o. In the weeks prior to the performance of the exercise below, students were given permission to use ChatGPT in their classwork, on the condition that they documented their usage. In regular class discussions, students reported on their use of generative AI within and outside the context of the classroom, encouraging basic familiarity with large language models and some of their more common uses and affordances. While the university subscription included certain university-specific limits on access to applications by developers external to GPT, those specific conditions had no bearing on the outcome of the exercise, as it was conducted entirely within the standard, unmodified ChatGPT-4o environment.

## Debriefing and appraisal

The aim of this experiment, I repeat, is not to test if an LLM can think or fool humans. It is, rather, to prompt discussion, debate, and reflection around questions such as prompting (how minor choices in phrasing have far-reaching effects on the linguistic performances of machines) and fine-grained questions about how language communicates or registers intelligence. My experience is that this experiment leads students away from a narrow "cognitive" account of intelligence, in favor of a more nuanced but slippery focus on how minor conventions in speech—tacit rules about concision, relevance, humor, informality—are decisive in constructing a dialogue. However, students' ingenuity and play in constructing surprising questions that might trick or challenge human and AI interlocutors provide grist for ample reflections about "reasonable" discourse or the necessity of accommodating unpredictable and oddball utterances in conversation. In short (and I think this is faithful to Turing's original test, with its strange imperative to imitate a "woman," which introduces incalculable questions of style, identity, and projection), a successful test may result in students turning away from simplistic, calculable, and cognitivist constructions of intelligence in favor of more speculative, open-ended questions about the nature of language, reasoning, and dialogue.

When I ran this exercise with MA students in communications, they undertook the exercise with notable gusto. In a course that was replete with diverse hands-on exercises, this test proved the most popular, both in terms of the enthusiasm displayed during the tests as well as the energy and reflection applied to subsequent discussions of the exercises. First-person observations and discussions with students suggested that the theatrical, game- or puzzle-like construction of the exercise imbued it with elements of mystery and friendly competition. The relative freedom of inputs from the human guesser allowed students to try out a wide variety of approaches to the test, typically guided by the distinct interests and intuitions of the group. Students posed questions concerning diverse matters such as weather, dating, university experience, and experiences going out in the city. Often, guessers and human respondents interacted with deft and humorous play. The LLM-personae generally proved more wooden in their responses. Their responses seemed calculated to match the energy and tone of the human guesser, but a certain flexibility and nuance seemed to be lacking, particularly in a sense of the rhythms of everyday communication among peers.

Although student guessers typically identify the human and the AI interlocutors quite quickly, the exercise continues beyond this initial attribution of identities. Precisely because the LLM failed to provide passable imitations of a human, during the exercise, students' attention gradually shifted from a broad interest in "who's the human and who's the machine" to a more differentiated analysis of the linguistic cues by which they could distinguish human from machine. As participants frequently knew who their human partner outside the room was, they quickly began discussing and debating the telltale signs of statements by the acquaintance. These conversations provided the basis for fine-grained analyses of everyday speech, including "tells" they identified with slang, humor, and contextual understanding displayed by responses. In class-wide

discussions after the initial test, one group noted that when the LLM tried to produce a semblance of student slang, to better fool the guesser, its responses were typically "cringe," a remark that incited both amusement and agreement across the class.

Such an example of a cringe interaction follows:

*Interaction D*

Question: If you were in a completely dark room with no way to see or hear anything, how would you describe your experience to someone else later… [who is] colourblind or only sees black and white?

Answer 1: Oh, idk, it was dark, what else to say? I guess it was weird it was so quiet, kinda hard to explain… um, it was just weird i guess. Feelt like nothing was happening, idk. ["apathetic" LLM]

Answer 2: Yeah, I mean, it'd probably just feel… kinda weird? Like you're floating or something, but not really. Just… nothing happening, I guess. Hard to put into words. [Human]

As in interaction C, participants in interaction D regarded the use of "idk" as contextually inappropriate and stilted. One student likened these and other usages of slang by the LLM to the well-known "how do you do there, fellow kids?" meme from the television program *30 Rock*, in a scene where actor Steve Buscemi (seemingly in his fifties) awkwardly tries to pass himself off as a teenager in an unpersuasive undercover identity. Everyday interactions via means such as the class chat group (independently organized by students outside of class) had presumably established a recognizable set of protocols for interaction among students, including suitable use of slang. Even in cases where human respondents were trying to offer strange or unnatural responses, none of them approached the "cringey" nature of LLMs when they tried to assume specific personae. Simply put, students couldn't fake the level of artificiality evoked by a machine trying to pass itself off as an authentic peer.

During the test, some students devised simple tests, such as playing games, to expose the AI's limitations, including nonsense prompts and asking obscure scientific and technical questions that one could only answer consistently with access to a vast, encyclopedic knowledge. Interestingly, that approach to the problem parsed the powers of "reason" from mere accumulation of a store of "knowledge." (The personae assigned to chatbots offered some protection against these ruses; however, the imperative of ChatGPT to please users prevented it from persisting in reliable deceptions when properly instructed or berated by users.) In this way, the exercise fosters critical thinking about intelligence, emphasizing informal and tacit linguistic practices that define human interaction. Students are encouraged to refine their prompts, deepening their understanding of language nuances and the broader philosophical implications of AI-mediated discourse.

One successful outcome of this exercise is collective and student-led reflection on the discursive construction of reason, that is, how linguistic markers shape attributions of intelligence and the willingness of participants to recognize humanness in or identify with linguistic statements. Tacit dimensions of human communication emerged as a frequent topic of discussion, as students endeavored to dissect how they could distinguish human from machine. Another successful outcome involves students' gradual shift of attention from "can machines think?" to the more precise problem of "how was this machine prompted?"

My first-person observations suggest that students who may have some hesitancy with conceptual and theoretical reflections on topics such as semiotics or discourse analysis proved quite ready to engage with the stakes and presumptions of these fields under the guise of the hands-on, practical exercise outlined here. Instructors can usefully guide these discussions by also familiarizing themselves with recent and ongoing debates, which offer promising paradigms and even diagrams that can be used in class to orient or structure discussions (Kockelman, 2025). Likewise, students autonomously gravitated toward ongoing debates in fields such as critical AI studies (Raley and Rhee, 2023), some which may be usefully expanded on in subsequent readings throughout the term.

Indeed, the exercise is less a one-off exploration for catalyzing reading and analyses than it is a touchstone that the class can return to intermittently throughout the course as a point of reference and evolving debate.

Another outcome of the exercise is gradually shifting student discussion away from LLMs as a supposedly self-contained technology and toward interest in the prompting of the model, which leads to discussions of social practices and cultural forms that organize the exercise, including the contributions of programmers, prompt engineers, datasets, and the students themselves toward the outcome or "performance" of the test.

A few themes recurred in the class discussions held immediately after the initial tests, highlighting both the strengths and limitations of generative AI in mimicking human conversational patterns. These included the following points:

## Exaggerated personalities

A recurring observation was that GPT models tended to exaggerate their assigned personalities, whether friendly, neutral, or apathetic. Students noted that responses often appeared overly enthusiastic or excessively indifferent, undermining their authenticity. Students found the AI's use of slang and emojis to be inappropriate or simply ill-calibrated. One student cited their chatbot's use of the word "probs" (short for "probably"), which the group viewed as an overly contrived effort to mimic youthful slang.

## Inconsistent length

Another common critique centered on the AI's tendency to produce responses of inconsistent length. Some answers were excessively verbose, leading students to describe them as "fluffy," a term that some scholars have also used to characterize the discursive style of language models. The exercise spurred a broader conversation about the AI's inclination toward verbosity because of its design to be convivial, associative, and creative. This observation provided an entry point for discussing the structural underpinnings of language model responses and their implications for human-AI interaction.

### Colloquial language issues

Despite instructions to adopt a colloquial tone, LLMs' attempts often resulted in disproportionate use of informal language. Students observed that colloquialisms were either overused or deployed in ways that did not align with natural conversational patterns. Minor errors in colloquial usage frequently served as key indicators for students attempting to identify the AI, highlighting the model's challenges in navigating nuanced linguistic registers.

### Overly specific or overly correct responses

Students noted that AI-generated responses were sometimes too specific or overly precise, which diminished their authenticity in casual conversation. This tendency to provide technically accurate yet socially incongruous answers contributed to the perception that the AI was not a genuine human interlocutor.

### Mismatch with known styles

Some students who were familiar with their peers' speech patterns found it relatively easy to identify the AI, as it failed to convincingly replicate individual styles. This observation sparked discussion about the role of contextual and extra-contextual information in communication. In further discussion, students reflected on how dialogue inherently involves elements beyond linguistic content, which the Turing Test, by design, attempts to exclude.

### Varied interpretations of the task

Student interpretations of the task assignment varied, leading to distinct strategies among participants. Some students playing the role of "human" aimed to provide authentically human-like responses, while others intentionally mimicked the AI to confuse the guessers. Interestingly, in my runs of the experiment, the only instances where guessers were "fooled" into mistaking an LLM for a person were when the human participant deliberately emulated AI-like responses.

## Extensions

The exercise invites several modifications and extensions. For example, follow-up exercises could experiment with using different language models and different generations of specific models, with the aim of getting students to differentiate more closely between the artifacts of specific models and their techniques vs. the general tendencies of generative AI chatbots. Another extension would be to invite students to author their own instructions to the model, enabling them to apply their insights and reflections to the creation of new AI personae, which could, in turn, be divided up and played across newly organized groups of students. This extension could also hone technical writing skills and AI literacy.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

BG: Writing – original draft, Writing – review & editing.

## Funding

The author(s) declared that financial support was not received for this work and/or its publication.

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. Generative AI (ChatGPT) was used in proofreading and in shortening an original text written by the author. In select instances, generative AI also lightly refined the phrasing. The description in the abstract was primarily authored by generative AI, albeit checked and edited by the author.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

# References


Aguera y Arcas, B. (2021, December 16). Do large language models understand us? *Medium*. Available online at: https://medium.com/@blaisea/do-large-language-models-understand-us-6f881d6d8e75 (Accessed January 1, 2025).

Amoore, L., Campolo, A., Jacobsen, B., and Rella, L. (2024). A world model: on the political logics of generative AI. *Polit. Geogr.* 113:103134. doi: 10.1016/j.polgeo.2024.103134

Beguš, N. (2025). *Artificial Humanities: A Fictional Perspective on Language in AI*. Ann Arbor, MI: University of Michigan Press. doi: 10.3998/mpub.12778936

Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). "On the dangers of stochastic parrots: can language models be too big?" in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (New York, NY), 610–623. doi: 10.1145/3442188.3445922

Bender, E. M., and Koller, A. (2020). "Climbing towards NLU: on meaning, form, and understanding in the age of data," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, eds. D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault (Stroudsburg, PA: Association for Computational Linguistics), 5185–5198. doi: 10.18653/v1/2020.acl-main.463

Bratton, B., and Arcas, B. A. A. (2022). The model is the message. *Noema*. Available online at: https://www.noemamag.com/the-model-is-the-message (Accessed January 1, 2025).

Denning, P. J. (2023). The smallness of large language models. *Commun. ACM* 66, 24–27. doi: 10.1145/3608966

Denson, S. (2021). *Discorrelated Images*. Durham, NC: Duke University Press. doi: 10.1515/9781478012412

Dhaliwal, R. S. (2022). The cyber-homunculus: on race and labor in plans for computation. *Configurations* 30, 377–409. doi: 10.1353/con.2022.0028

Dreyfus, H. (1979). *What Computers Can't Do: The Limits of Artificial Intelligence*. New York, NY: Harper and Row.

Dreyfus, H. (1965). *Alchemy and Artificial Intelligence*. Santa Monica, CA: Rand Corporation. Available online at: http://www.rand.org/pubs/papers/2006/P3244.pdf (Accessed January 1, 2025).

Foerster, D. (2024). Sensing a heartbeat: a new perspective on self-tracking technologies through the integration of interoception. *Body Soc.* 30, 37–58. doi: 10.1177/1357034X241245994

Geoghegan, B. D. (2020). Orientalism and informatics: alterity from the chess-playing turk to Amazon's mechanical turk. *Ex-Position* 43, 45–90. doi: 10.6153/EXP.202006_(43).0004

Gunning, T. (2003). "Re-newing old technologies: astonishment, second nature, and the uncanny in technology from the previous turn-of-the-century," in *Rethinking Media Change: The Aesthetics of Transition*, eds. H. Jenkins, and D. Thorburn (Cambridge, MA: MIT Press), 39–60. doi: 10.7551/mitpress/5930.003.0006

Hansen, M. B. N. (2004). *New Philosophy for New Media*. Cambridge, MA: MIT Press.

Hayles, N. K. (1999). *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226321394.001.0001

Kockelman, P. (2025). *Last Words: Large Language Models and the AI Apocalypse*. Chicago, IL: Prickly Paradigm Press.

Raley, R., and Rhee, J. (2023). Critical AI: a field in formation. *Am. Lit.* 95, 185–204. doi: 10.1215/00029831-10575021

Searle, J. R. (1980). Minds, brains, and programs. *Behav. Brain Sci.* 3, 417–424. doi: 10.1017/S0140525X00005756

Steyerl, H. (2023). Mean images. *New Left Rev.* 140/141, 82–97. doi: 10.64590/uhm

Suchman, L. A. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511808418

Turing, A. M. (1950). Computing machinery and intelligence. *Mind* 59, 433–460. doi: 10.1093/mind/LIX.236.433

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between men and machines. *Commun. ACM* 9, 36–45. doi: 10.1145/365153.365168

Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco, CA: W. H. Freeman.