Gluing Pizza, Eating Rocks, and Counting Rs in Strawberry: The Discursive Social Function of Stupid AI Answers

Advait Sarkar
University of Cambridge
Cambridge, United Kingdom
advait.sarkar@cl.cam.ac.uk

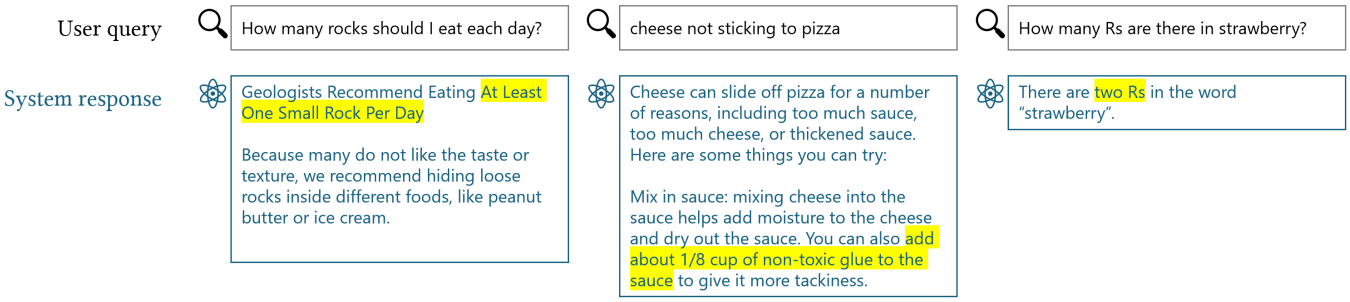

Figure 1: Illustrations of three interactions with AI systems showing stupid answers, with core stupidity highlighted. The text is excerpted verbatim from real questions and responses from commercial AI systems.

Abstract

Artificial Intelligence is a multiple Nobel prize-winning technology that has solved elusive problems such as playing Go and protein folding. It also tells you to eat one small rock per day. Much collective online mirth and criticism has ensued following the discovery of such stupid behaviour.

I argue that these stupid answers are in fact correct, because the primary objective of such queries is not to actually receive a correct answer, but rather to obtain an artefact of discourse. I analyse their operation to explain how discussants collectively construct an imagined user who is deceived by the stupid answer, while distancing themselves from that naïvety. This discourse operates as a form of spectacle, simulation, and myth in discussions of technology and society.

I suggest that researchers avoid invoking stupid AI answers as rhetorical devices in research discourse, as this can undermine genuine AI risks and failures.

CCS Concepts

• **Human-centered computing** → **HCI theory, concepts and models**; Interaction design theory, concepts and paradigms; **Collaborative and social computing theory, concepts and paradigms**; • **Computing methodologies** → **Philosophical/theoretical foundations of artificial intelligence**; *Discourse, dialogue and pragmatics.*

CHIWORK '25, Amsterdam, Netherlands

ACM ISBN 979-8-4007-1384-2/25/06
https://doi.org/10.1145/3729176.3729189

Keywords

narratives, mythology, social media, artificial stupidity, semiotics, media theory, literary theory, performance, interpassivity, Marxist critical theory, postmodernism

ACM Reference Format:
Advait Sarkar. 2025. Gluing Pizza, Eating Rocks, and Counting Rs in Strawberry: The Discursive Social Function of Stupid AI Answers. In *CHIWORK '25: Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work (CHIWORK '25), June 23–25, 2025, Amsterdam, Netherlands.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3729176.3729189

1 Introduction: Artificial Stupidity

The nature of the "intelligence" that Generative AI exhibits is strange, unfamiliar, disembodied, jagged, partial, unpredictable, at times sub-human, at times super-human, but altogether other. Researchers, the media, and the public all revel in the delightfully stupid behaviour that can arise from the statistical recombination of language. Celebrated contemporary examples include AI generated recommendations to glue cheese onto pizza to stop it sliding off, or to eat at least one small rock per day, or the curiously insistent assertion that there are exactly two 'R's in the word "strawberry". These mischievous probes derive their humour from the incongruity between the otherwise impressively superhuman capabilities of language models, and their failure to answer deceptively simple questions correctly.

Examples of such answers, derived from real responses from a variety of commercial systems, are illustrated in Figure 1. Three lists of annotated "in-the-field" examples of stupid AI answers in discourse are provided in Appendix A. The reader is also invited to try these examples with their own preferred system(s).

The key proposition of this article is that while stupid answers derive their humour from their apparent wrongness, they are in

fact correct answers. They are correct because they provide the expected and intended response. At a certain threshold, their stupidity serves a social function, causing them to acquire a deeper notion of correctness (Section 2). This interpretation of stupid answers as being correct, unlike alternative interpretations (which appeal to, e.g., the content of the training data or the model mechanism), accords appropriate respect and agency to end-users of AI systems (Section 2.1). In public discourse, these stupid answers offer material benefits to individuals, the media, and to researchers (Section 2.2).

Next, the mechanics of stupid AI discourse are examined (Section 3). The rhetorical use to which these answers are put suggests that a key element of the discourse is a collectively imagined "naïve" user who might take the response literally; it will be suggested that this imagined user is a partitioned aspect of each user's own psyche, and that, as in a theatrical performance, it requires a collective suspension of disbelief on the part of all those participating in stupid AI discourse (Section 3.1). As this process is largely unconscious, unlike in theatre, the relationship between the representation and reality in stupid AI discourse is unclear. Three perspectives on this relationship are considered as candidates for stupid AI discourse: Debord's Spectacle, Baudrillard's Simulation, and Barthes' Myth (Section 3.2).

Finally, the unreflexive use of stupid AI answers as part of research rhetoric is problematised (Section 4); it is suggested that such examples should not be juxtaposed with examples of "true" incorrect AI behaviour, and should not be used to motivate research agendas for improvements in human-AI interface design, or improvements in the technical implementations of AI systems.

This paper is to be read as a humanistic essay [1], a form of scholarly inquiry with roots in philosophical and cultural studies traditions. Readers are invited to engage with this work by attending to the conceptual arguments, reflecting on the cultural implications, and considering their own experiences in relation to the ideas presented.

Moreover, this analysis must be contextualised within its temporal specificity. The core argument was developed between August and December 2024, drawing upon examples of AI answers and the associated online discourse that were prominent at this time. It thus responds to a particular snapshot of the public engagement with generative AI during a particular phase of its development. However, the landscape of artificial intelligence, and consequently the nature of its failures and the public response, is inherently dynamic. In the future, the specific examples and the public discourse surrounding them may well be different. This paper's contribution should thus be understood as a theoretical exploration of a particular moment in the evolving relation between people and AI systems, where the notion of stupidity in AI became a locus of social and cultural discourse.

2 The Discursive Threshold: Why Stupid AI Answers are Correct Answers

While superficially incorrect, stupid AI answers are correct in the context of the discourse they participate in. As a shorthand, I shall term these respective notions of correctness *S-correct* (superficially correct) and *D-correct* (discursively correct). My argument shall be that in certain contexts, an S-incorrect answer is D-correct, and conversely an S-correct answer is D-incorrect. In these contexts, an S-correct answer is far less valuable and less useful than an S-incorrect answer. In these contexts, the query and answer have crossed a *discursive threshold*.

To get an intuition for D-correctness, consider the following example: someone reads about the rock-eating answer and decides to try the query for themselves. Delighted with the result, they blog about their experience. Now imagine if instead the answer had been "fixed" by the system developers to be S-correct, and the user informed that no rocks are to be eaten. This would be a terribly disappointing experience; the user is deprived of the humour and delight of the stupid answer, and of the opportunity to reflect on it. The fix has pooped on the proverbial party. This disappointment is the kernel of D-correctness.

The key is to examine the act of information retrieval, or model querying, in the context of the discourse. The user clearly isn't seeking a "true" answer – they already know precisely how many rocks one should eat. The point of this query is not to answer a question: it's to talk about the answer.[1] The real use of AI in this situation is to enable the user to participate in a humorous discourse. The point at which the S-incorrect answer has sufficient discursive value that it acquires D-correctness is the discursive threshold.

It is not humour, per se, that leads to this interpretation of correctness; the argument is not that all humorous content is to be regarded as correct in some sense. Humour in general carries no expectation of correctness. Rather, it is the context of these products and their positioning as information retrieval tools and question answering services that provides the expectation and attendant evaluation of correctness.

There are two thresholds for D-correctness: a weaker, private threshold, and a stronger, public threshold. The private threshold concerns the individual expectation of an S-incorrect answer and the attendant private delight; the public threshold concerns the query as a social object of aggregate querying behaviour. We have already considered the private threshold of D-correctness in the example above.

An aggregate crossing of private thresholds leads to the public threshold being crossed. Consider, following the discovery and initial circulation of a new stupid answer, how many interested observers search for the same query, and why? Of those, how many genuinely required advice regarding the number of rocks to eat? How many truly needed the assistance of a computer to count the occurrences of a letter in a word? Probably zero. When the public discursive threshold is crossed, it transforms our default reading of the query. We can assume that the query no longer represents an information satisfaction need but rather the intent to participate in a discourse. The crossing of the public threshold demands that the answer must be reliably S-incorrect, otherwise, the query cannot become a common discursive currency. The "incorrect" answer has become correct; the discourse – the total set of blog posts, news articles, social media threads, etc. wherein such answers are shared and commented upon – serves as the transtextual [19] backdrop against which participants can achieve this interpretation of correctness.

[1] Games such as "Googlewhacking" – finding a two-word search query that returns exactly one result – are also a subversion of the information retrieval paradigm of search engines, albeit quite different to the discursive purpose of stupid AI answers.

The public threshold is not necessarily crossed instantaneously or universally. The efficacy of stupid AI discourse as humour and social commentary relies on shared cultural context and understanding. When this shared understanding is absent, individuals genuinely seeking information might be misled by an S-incorrect answer, particularly if the stupidity is not immediately obvious or if the user lacks the necessary background to discern its absurdity. Examples that achieve widespread circulation and D-correctness are characterised by blatant absurdity readily apparent to users within a shared cultural context. However, answers may remain in an ambiguous state where it is humourous to some while potentially misleading to others, for instance when the S-incorrectness is more subtle or dependent on specific contextual knowledge. The effectiveness of an S-incorrect answer in achieving D-correctness therefore relies on an unambiguous and widely recognisable deviation from S-correctness within presumed common-sense understanding. Answers residing in this liminal space may have more limited discursive reach or carry higher risk of genuine misinterpretation. Analysis of such situations is not the focus of this paper, but will be important to address in subsequent work.

2.1 Alternative Defences of Stupidity

Before we proceed, it is worth briefly discussing a few alternative perspectives on stupid answers, to acknowledge them and distinguish them from what has been proposed in the previous section. Examples instantiating each of these perspectives can be found in Appendix A.

(1) **The mechanismal defence.** This refers to any explanation of the incorrect behaviour in terms of the underlying mechanism of the model.[2] For example, the failure to count 'R's in "strawberry" is often explained by noting that models are rendered blind to individual characters through the tokenisation process and so cannot reasonably be expected to count them. While mechanismal defences do not claim that the output is correct, they attempt to reposition the query as meaningless and therefore lacking a notion of correctness. The logic of such explanations is that by making the appropriate observations about the mechanism (e.g., vis-à-vis tokenisation), an incorrect answer about counting characters is rendered about as surprising or meaningful as the inability of the chatbot to fry you an egg.

(2) **The alternative context defence.** This refers to any explanation of the incorrect behaviour that introduces a plausible context in which the ostensibly incorrect answer is correct. For example, the stupid answer that "9.11 is higher than 9.9" may be explained by observing that, if the numbers are interpreted as software version numbers, then indeed by the conventions of software version numbering we would interpret Software Version 9.11 to be "higher" than Version 9.9. The answer could also be correct if the numbers are interpreted as dates (September 11 is "higher" than September 9). These explanations rest on the ambiguity in the original query, and question the common ground we might have with the model. However, these explanations are brittle: increasing the specificity of the context in the prompt may still not be enough to resolve the stupid answer, and conversely, slight variations in the prompt that do not introduce further context can often resolve the stupidity.

(3) **The garbage in, garbage out defence.** This refers to explanations that redirect the blame for the stupidity towards the source material drawn upon in stupid answers. The pizza-gluing text derives from a real comment (albeit a "troll" or "shitpost") from a human Reddit user, and the rock-eating text from a real article (albeit satire) from the Onion. The objective of the technology is simply to find an efficient route from the query to an answer on the Web, and it has done so correctly – it can hardly be blamed if the sources themselves are misleading. After all, even humans often have difficulty detecting satire.

Each explanation seems to offer a route to an interpretation of such behaviour as not incorrect. The mechanismal defence posits that the query is ill-posed and therefore the answer has no correctness value. The alternative context defence posits that the answer is correct for a certain interpretation of the query. The garbage in, garbage out defence posits that the answer is correct because the model is doing what we told it to do, it just happens to be drawing on incorrect sources.

While all of these defences can help explain the behaviour, they do not pose a serious challenge to the interpretation of the behaviour as being incorrect. The key observation is that none of these defences considers that the *user intent* could have been anything other than to receive an S-correct answer. If an S-correct answer had been given, there would be no problem – these defences of stupidity would have nothing to say. Moreover, in advocating for the correctness of the behaviour, they concomitantly imply that there is something wrong with the user: the user has asked a meaningless question, the user is not thinking of the answer in the right context, the user is not querying over a good source of data. The incorrect behaviour has not been dissolved, merely relocated. PEBKAC, PICNIC, ID10T. In absolving the model, these defences incriminate the user.

It is only in interrogating and refining our understanding of the *user's* notion of correctness, in considering whether and why an S-correct answer can be suboptimal or useless, can we obtain a true defence of the stupid answer as being correct, a defence based on the discursive social function that stupid answers fulfil.

While these common alternative defences are regarded in this analysis as unsuccessful at explaining the special relationship of stupid answers to correctness, they are helpful in understanding that stupid AI answers are not monolithic. Two of the examples we have encountered already allow us to broadly differentiate between what might be termed procedural stupidity and contextual stupidity. The former, such as the failure to count 'R's, appears to stem from limitations in the AI's underlying algorithms for processing and manipulating text at a granular level, such as tokenisation. The humour derived from such failures may arise from the dissonance between the perceived sophistication of the algorithm as a technically masterful feat of engineering, and its inability to perform

[2] I borrow the term "mechanismal" from an earlier paper [27] to avoid an ambiguity caused by the terms "mechanical" and "mechanistic".

seemingly simple computational tasks. Users might find it amusing that a technology capable of complex language generation can falter on a basic counting exercise.

In contrast, contextual stupidity manifests in answers like the recommendation to glue cheese onto pizza, or to eat rocks. These highlight the system's struggle with understanding and reasoning about the real world, often reflecting biases or absurdities present in its training data or response context. Here, the humour doesn't necessarily derive from a limitation of the algorithm but rather in the model's inability to discern satire or to possess commonsense knowledge about what is and is not appropriate to consume. The humour in such instances may stem from the sheer absurdity and incongruity of the AI's recommendations in relation to widely accepted reality.

Consequently the discursive functions and user perceptions of these two types of stupidity may also differ. Various theories from the philosophy of humour could be brought to bear on this discussion, such as the incongruity theory (humour arises from violated expectations) or superiority theory (laughter expresses superiority over others or a former state of ourselves) [24]. However, an analysis of what makes stupid answers humorous is beyond the scope of this paper; for our purposes it suffices merely to note that they are humorous, whatever the mechanism. Further research could explore whether the perceived 'type' of AI stupidity influences the specific social functions it performs and the nature of the D-correctness it therefore acquires.

2.2 Stupid AI Answers Offer Material Social Value

Stupid AI answers are not just funny, they are also valuable. In the first instance, they are valuable in interpersonal relations, as humorous anecdotes to supply during cocktail parties, and as trivia to memorise and resurface when casual conversation turns to the contemporary issues of AI and Big Tech. Cocktail parties and trivia games, as identified by Postman, are "pseudo-contexts" [26]: contexts of discourse invented purely to provide a use for the barrage of otherwise useless and irrelevant information that began to importune people following the widespread adoption of the telegraph and photograph. I would further argue that pseudo-contexts do more than merely allow people to "burn off" a glut of surplus information; they allow people to deploy that information towards the accumulation of social capital.

In the online sphere, the accumulation of social capital is quantified, commodified, and ultimately, monetised. Stupid answers are reblogged and retweeted, and through a repertoire of associated re-gestures they allow people to acquire "clout" and "influence", which are valuable in their own right, but can also be exploited for material gain. Similarly, due to their inherent humour and seemingly endless opportunities for commentary and debate, stupid answers form ideal feedstock for online media outlets (examples in Appendix A). Stupid answers, particularly those aligned with the Western Internet's preferred genre of absurdist humour (explaining the particular success of the rock-eating and pizza-gluing examples) draw eyeballs and clicks and therefore revenue. In some cases, entire careers can be built upon the humour, and the flarfy, glitchy defamiliarisation [17] of AI output, such as Shane's wonderful *AI Weirdness* [28].

I shan't belabour this point as these are fundamental concepts of attention economics, and such capitalising phenomena are hardly unique to stupid AI answers. The key observation is that when stupid AI answers participate in these well-studied social and economic processes, *it is their stupidity that makes them valuable.* Previous research has noted [32] that AI stupidity (charitably described as "idiosyncrasies") can act as "social glue" by inducing a playful atmosphere and creating psychological safety – i.e., the computer embarrasses itself so you don't have to. Not all S-incorrect answers are valuable; they must also contain a compact and absurd incongruity, a humorous and potential virality,[3] that can additionally confer a D-correctness. The endowment of a notion of D-correctness is a process of valorisation. I should note that, unlike most commentaries on the attention economy, this is not a critique. I am not proposing that there is a problem with the public and the media using stupid answers in this way.

However, researchers use stupid answers too, and *this* may be problematic (discussed further in Section 4). To understand the problem, we must first explore how the enjoyment we derive from these answers requires an act of self-deception, and how discussion of these answers can transform into an object in their own right, a signifier without a real-world referent.

3 How Stupid Answers Work

3.1 Partitioning, Delegating, and Deceiving the Self

When a stupid answer crosses the discursive threshold, the user's objective in issuing a query is no longer information retrieval or getting a correct answer. But it is important that this act of querying is still being committed in the same form as a genuine information retrieval act, because embedded in the discursive potential of the stupid answer is the possibility that *someone* might have received it in response to a genuine query.

In other words, a key operational mechanism of stupid answers and the rhetoric they enable around AI reliability and trustworthiness is that everyone involved in the discourse is engaged in an act of collective imagination: it is obvious to all that the stupid answer was only ever solicited for humorous, discursive purposes, but *imagine how problematic it would be if someone really had this query and took the answer seriously!* Thus, we are a discursively-aware user (hence D-user) while simultaneously role-playing a discursively-naïve or superficial user (hence S-user) who would fall for the stupid response.

What is the nature of this imagined, constructed S-user? Who are they, and where do they come from? With caution, I speculate that the D-user must at least partially identify with the S-user, or see in themselves the possibility of becoming an S-user in another context. If there was no such identification, then the stupid answer would become implausible, unbelievable, too contrived – "ecologically invalid" in discursive terms. And moreover, if there was no such identification, if the imagined S-user was considered strictly other

[3] Bergson, in *Le Rire* [5], theorises that comedy arises when life appears as a mechanism. Fittingly, stupid AI is indeed an otherwise lifelike simulacrum whose mechanical nature is being revealed.

than the D-user, then the collective enterprise would cease to be humorous and jovial, and turn into a mean-spirited, cruel mockery of those epistemically unequipped to discern truth from nonsense, an indecorous charivari.

Partitioning. This partial identification of the D-user with the S-user suggests that the S-user may be a detached aspect of the user's psyche or consciousness. Indeed, this is the essence of McLuhan's theory of media [23]. In it, he posits that media are extensions of ourselves. The wheel is an extension of the foot, the radio an extension of the ear, the television an extension of the eye. They come about because society and culture create demands of pace and load that the body cannot handle without extension. However, we are numbed to this self-extending nature of media because the experience of externalising the body is traumatic, and we repress this trauma: *"in the case of the wheel as an extension of the foot, the pressure of new burdens resulting from the acceleration of exchange by written and monetary media was the immediate occasion of the extension or "amputation" of this function from our bodies [...which] is bearable by the nervous system only through numbness or blocking of perception. This is the sense of the Narcissus myth. The young man's image is a self-amputation or extension induced by irritating pressures. As counter-irritant, the image produces a generalized numbness or shock that declines recognition. Self-amputation forbids self-recognition."*

How do we map stupid AI discourse onto McLuhan's framework? McLuhan identifies the extension as a medium, and we identify the S-user as an extension. Thus by analogy, the S-user would be a medium that extends the user's cognition, or more precisely, extends one of the user's many fragmented cognitive potentialities. The S-user is a medium that both extends and numbs the user from the potentially uncomfortable reality of the new information landscape created by AI, and their own susceptibility to stupid answers.

However, there are problems with this analysis. For the S-user to be a medium, it must not just respond to the pressure of a change of scale, pace, or pattern in society, because that would merely make it a coping mechanism. It must also *introduce* its own change of scale, pace, or pattern in response (much as the wheel, the radio, and the television have done). We cannot say that any such change has taken place as a result of stupid AI discourse, let alone the specific concept of the S-user we are attempting to theorise here. Moreover, the S-user is imaginary, unlike television or radio. McLuhan does adduce several media that are immaterial (e.g., the feudal society[4] extends the stirrup, viz. Lynn White Jr.) but none that is completely imaginary (i.e., without obvious material traces). Thus, there are elements of McLuhan's theory which seem true of the S-user: it is an aspect of the self; it is accompanied by a numbing lack of self-recognition – but others which are not: its imaginary nature; its modest effects on society.

Delegation. Pfaller's theory of interpassivity gives us an alternative account [25]. Pfaller argues that people engage in "interpassive" behaviour when they delegate their own experiences to other agents, whether people or objects. Examples include people recording television programs that they never watch, printing out texts they never read, using ritual machines that pray or believe on behalf of them (e.g., Tibetan prayer wheels, 'ora pro nobis'), and canned laughter on television shows. Pfaller's theory is that these acts stand in for consumption by *delegation*. It is the recorder that enjoys the television program, the printer that reads the text, the ritual machine that prays, and the TV show that laughs for them.

Much like McLuhan's media, Pfaller's interpassive acts involve both extension and a numbness that forbids self-recognition. While media are a response to a societal stressor, interpassive behaviours are a response to individual stressors: if there is not enough time to watch a show, or read a text, or pray – delegating the act to an agent feels better than doing nothing. In order for this to be successful, Pfaller finds, the user themselves needs to be the one performing the act of delegation (starting the recording, issuing the print command, etc.) because this establishes that the consumption is being done specifically on their behalf and not on behalf of someone else. Moreover, the act of delegation stages an illusion of consumption, but it's not an illusion for the delegator – after all, no one really believes that recording a show is equivalent to watching it, or printing a text is equivalent to reading it. For whom, then, is this illusion staged? Pfaller answers: *"The illusion at stake in the practices of interpassivity therefore has a very interesting and particular kind of ownership: it is in a way nobody's illusion, an anonymous illusion, an illusion without a subject. None of the real people present [when an act is delegated] has to believe in this illusion. The possibility of delegated reading does not depend on the presence of a believer, since it is not just a subjective illusion: delegation works for the intellectuals not because they think that the machine can read for them; the machine reads for them because somebody else, an anonymous naive observer, might have thought that. [...] Somebody else – an anonymous other, not us – believes, then, in the equivalence and thinks that we were enjoying; and this anonymous belief in our enjoyment brings about the deep satisfaction that we experience when we never watch our video tapes. "*

Pfaller's "anonymous other" is a superb candidate for our S-user. The act of consumption we delegate is the superficial act of information retrieval that seeks an S-correct answer. We do so because of the anxiety induced in us by the possibility of naïvety and to distance ourselves from that possibility. We stage the illusion of the superficial act, and the satisfaction we derive rests in the anonymous S-user who believes in the act.

Deception. There is one aspect of the S-user in stupid AI discourse that is not fully explained by interpassivity, and that is its collective nature. Pfaller's treatment of the issue largely revolves around private acts of delegation, which corresponds to the private discursive threshold. But to cross the public discursive threshold, it is not enough for each discussant to individually construct an S-user, all participants must share in the act of collective imagination. The idea that interpassive delegation involves the act of "staging" an illusion points to a highly productive metaphor: stupid AI discourse as a form of theatre.

In *Poetics*, Aristotle posits that the purpose of theatre is to allow the arousal and complete expression of emotions in the audience (an idea later developed by Brecht, among others) [22]. Emotions aroused by theatre are not experienced by the audience in the same

[4]McLuhan's unwieldy and seemingly arbitrary use of the word "medium" to apply to a wide range of phenomena is one of the major criticisms of his theory.

way as "real" emotions. And there is always an element of pleasure, even when highly negative emotions are aroused by the depicted events. Well-told fictional events work because even if they depict a physical impossibility (fantasy creatures, superheroes, historical counterfactuals), they unfold in a context where successive events are plausible continuations with causal relationships to prior events. In other words, a plausible impossibility is preferable to an implausible possibility.[5]

Coleridge observed [13] that it was necessary for us to bracket our conception of truth sufficiently in order to allow the internal causality of a work of literature to function as an effective inducer of emotional response: *"so as to transfer from our inward nature a human interest and a semblance of truth sufficient to procure for these shadows of imagination that willing suspension of disbelief for the moment, which constitutes poetic faith."* Burroway's treatise on the practice of fiction [10] characterises this as an act of self-deception: *"Every reader is a self-deceiver: We simultaneously "believe" a story and know that it is a fabrication. Our belief in the reality of the story may be so strong that it produces physical reactions — tears, trembling, sighs, gasps, a headache. At the same time, as long as the fiction is working for us, we know that our submission is voluntary [...] Pleasure in artistry comes precisely when the illusion rings true without destroying the knowledge that it is an illusion."*

A collectivised suspension of disbelief is essential for audiences of theatre (and to a lesser extent its progeny, film). Collectivisation is what distinguishes private and public viewing; it is worth noting that at the time of Aristotle the possibility of private viewing was inconceivable. The public viewing, though primarily an act of consumption, is also an act of discourse because the emotional response of the audience must be shared to be effective. If some audience members suspend disbelief, but others don't, or if different members bracket reality in different ways, the enterprise falls apart.

Pfaller does note (via Žižek and Lacan), that the chorus in Greek tragedy has an interpassive function, which is not dissimilar to the function of canned laughter on television shows: *"The Chorus experienced compassion and fear [...] in place of the real spectators, who were glad to be relieved of this task [... similarly on TV sitcoms] a certain mechanical laughter is always already built-in, erupting after every joke and before any possible laughter on the part of the spectator."* But where our analysis goes further is in positing the following: that the communal environment that creates a special relationship with reality, which enables fictional words to be properly experienced and where disbelief is suspended, is the same environment that allows the S-user to be shared.

Blackwell has also asserted that *"AI is a branch of literature because it is a work of imagination"* [6]. He contrasts this particularly against the interpretation of AI research as a branch of science. He proposes literature (and elsewhere, the *"entertainment industry"*) as a superior view, principally because AI research begins from *"some kind of fantasy about what a computer might be able to do in the future"* and aims to build computer systems that can fulfil that fantasy. Literature, too, aims at building an artefact through which some imaginative fantasy is realised.

[5]A generalisation, but not universal. Exceptions include surrealist cinema, Czech New Wave, movies that subvert event order (e.g. *Arrival (2016)*), etc.

But this reasoning is too permissive. All kinds of "sciences" of the artificial [30], such as engineering and programming, also fit this criterion. Indeed, an oft-quoted passage from Fred Brooks declares [8] that *"The programmer, like the poet, works only slightly removed from pure thought-stuff. He builds his castles in the air, from air, creating by exertion of the imagination."* Design is also an act of imagination. The objective of all design, arguably, is to shift the world from its current state into a "preferred" state that originates as a fantasy. This does not make design a branch of literature.

The same limitations apply to Blackwell's supporting arguments – e.g., the similar mode of production: *"The daily work of an AI researcher, just like a novelist or playwright, involves typing on a computer keyboard to produce a text."*; the similar mode of evaluation: *"The value and significance of literary works, whether poems, plays, novels or AI programs, is decided by how the audience reacts, by what the critics say about it, and most importantly, whether people want to see more of this kind of stuff [...]"* – these also admit too wide a variety of activities we would not consider literature.

The related observations that computational analysis is valuable in literary studies [4], or that literary theory ought to inform the machine production of text [33], are heavily trodden and do not advance the argument of artificial intelligence *per se* as literature. The analysis in this paper introduces a distinctive new reason to consider AI a branch of literature: that AI consumers engage in a suspension of disbelief in order to fully experience it. Moreover, like theatregoers, this suspension of disbelief is collective. Its collective nature means that while literature is a meritable analogy, theatre is a more precise one. For stupid AI discourse to work, discussants must collectively agree – without explicit negotiation and purely through a common cultural context and the implicatures of discourse – on a shared basis of imagined reality within which narratives of causality and probability can be evaluated. This clearly sets it apart from what is necessary for a user to "consume" other artefacts of engineering or design, which may well require agentic repair and attribution [14], or an intentional stance [16], but the causality bracketing required for contemporary AI discourse is of an entirely new intensity and quality. Moreover, and uniquely, AI use is theatre with an interpassive, delegative dimension: a key element of this imagined reality is the auto-partitioned S-user. Not all theatre possesses this dimension, and this sets human-AI interaction apart from other kinds of human-computer interaction that have also been previously conceived of as theatre [22].

We have so far examined how theories of media, interpassivity, and theatre may help explain the operation of stupid AI discourse, but a major problem still remains. In theatre and literature, the audience is complicit in the self-deception, it is performed consensually and with conscious awareness. But the processes of self-deception described by McLuhan and Pfaller are completely unconscious – self-recognition is forbidden. A key consequence of this is that the distinction between representation and reality is obscured, and they can begin to affect each other. Harnessing the potential of theatre to impact theatregoers beyond the end of the play, as they returned to their lives in the "real" world, was an ambition of many playwrights and theorists (such as Brecht [22]), but this ambition was frustrated by the traditional cultural (and spatial, temporal, and architectural, and sartorial) boundaries drawn between theatrical experience and real experience. In contrast, the theatre of AI discourse can be said

to have succeeded beyond Brecht's wildest imagination through an unconscious and painless erasure of those boundaries. How does it do this, and what are the consequences?

3.2 Spectacular, Simulated, and Mythic Discourse

We have so far seen that for stupid answers to have humorous or rhetoric impact, discussants must perform an act of self-partitioning, and numb oneself to one's own trauma and to that of others, to suspend one's disbelief and engage with the stupid answer as a fictive, literary, theatrical object. We can now turn to examining the precise mode of that engagement and its relationship to reality. Does the discourse around stupid answers create a smokescreen that distracts us from real AI failures – a spectacle? Or does it enact, perform, and thereby create its own reality – a simulation? Or does it provide a sign, a semiotic abstraction, that possesses no intrinsic reality but refers to one, a type of speech defined by its intention and not its literal sense – a myth? Let us consider each of these perspectives in turn.

Spectacle. Discourse around stupid AI answers is possibly a form of spectacle [15]. Debord observed that human fulfilment had first progressed from being (existence) to having (possession); i.e., in order to live a good and fulfilled life, one needed to possess things. This shift was concomitant with commodity fetishism and a shift in emphasis from the use value of objects to their exchange value. However, Debord asserts, fulfilment has further progressed from having to appearing (images). In modern society, appearing to have something is more important than having it, and any value in possession derives from the value of appearing to possess – the image of possession – bringing commodity fetishism to its ultimate fulfilment. Thus we transform capital and social relations through images into an all-encompassing and self-propelling phenomenon that Debord calls the spectacle.

Discourse around stupid AI answers is a spectacle: the answers are not valued for their truth or utility, but for their capacity to amuse, provoke, and generate attention. The stupid AI answer functions not as a piece of information, but as a consumable image or artifact. One of Debord's key points, which does not seem to apply in the case of stupid AI answers, is that the spectacle distracts from real issues and alienates people from genuine engagement with their lives. By turning everything into an image, the spectacle divorces people from real social, political, or existential concerns, making them passive consumers of life. We might therefore ask whether the spectacle of stupid AI distracts from more meaningful questions about the technology itself, such as its ethics or its impact on society. This seems unlikely – I believe there is ample countervailing interest in discussing such issues. On the other hand, high-profile stupid answers may offer a way to undermine such concerns (examples in Appendix A). *"How can AI take your job? It can't even spell strawberry!"*

Simulation. Discourse around stupid AI answers is possibly a form of simulation [3]. While Debord begins to articulate the changing nature of our relationship to reality when the image is revered above all else (*"The spectacle is the stage at which the commodity has succeeded in totally colonizing social life [...] we no longer see anything else; the world we see is the world of the commodity."*), it is Baudrillard who carries the thought through to its logical conclusion. For Baudrillard, while simulacra (images) may begin as representations of reality, through the phases of simulation they come ultimately unmoored from any real referent, creating their own reality – a *hyperreality*. Hyperreality, unlike spectacle, is not a veil cast over reality but a performed, autonomous reality unto itself. Baudrillard gives as examples modern medicine, which has become a system of signs and symbols focused at representing health rather than creating it; modern wars, conducted through media representations creating a simulated version of conflict that becomes more influential than the actual events; and theme parks such as Disneyland, whose fantasy world is more coherent than the reality outside it.

Stupid AI answers can be seen as part of a hyperreal system. They begin as plausible representations of reality – empirical data that is informative about the absurd attempts by AI to process language or logic – but, through their circulation, become something more: artefacts whose entertainment value endows them with disproportionate powers of persuasion. They are simulacra of AI interactions: copies that do not refer to any "true" original user interactions. They do not have to represent real AI capabilities or progress; instead, they become hyperreal performances. These answers are the "war footage" of AI culture – a mediated version of AI that may influence perceptions more than the real-world applications of the technology.

However, there are problems with a direct interpretation of stupid AI discourse as simulation. Regardless of its momentum, this discourse remains heavily reliant on human engagement, sharing, and interaction. Stupid AI answers are integrated into larger, participatory cultures of humour and satire, which retain individual agency to a much greater degree than does medicine or warfare. Moreover, while the discourse is a selective and distorted representation of AI technology, it is transparently so. Rather than AI absurdities becoming a hyperreal replacement for AI as a whole, they represent a narrow, sensationalised aspect of it, and we might reasonably expect consumers to be aware of this. In other words, the assertion that viral stupid AI answers distort public understanding of AI in a way that constitutes hyperreality might wildly overestimate the extent to which people are genuinely misled. While absurd AI outputs may become entertainment, we may charitably expect that most people recognise the difference between viral moments and actual technological developments. People can hold both perspectives simultaneously: the entertainment value of AI absurdities and the real-world implications of actually existing AI [29].

Myth. Discourse around stupid AI answers is possibly a form of myth [2]. Some readers may view in Debord an anti-conspiratorial bent characteristic of Marxist critical theory, and in Baudrillard a postmodern (dare I say) mystification once derided as "fashionable nonsense" [31]. The media are shadowy puppetmasters of capitalism and/or the state! Images are consumed by Heideggerian technological progress [21] and have become its sex organs!

Barthes offers an antidote in his theory of mythologies. His core concept of myth builds on Saussure's semiotics. Per Saussure, a sign is composed of a signifier (e.g., the word "tree") and a signified (e.g., the concept of trees). For Barthes, this is a first-order sign

that can itself come to have another signified. This second order "signification" constitutes a myth. This can be illustrated by directly quoting one of Barthes' evocative examples: *"I am at the barber's, and a copy of Paris-Match is offered to me. On the cover, a young [Black man] in a French uniform is saluting, with his eyes uplifted, probably fixed on a fold of the tricolour. All this is the meaning of the picture. But, whether naively or not, I see very well what it signifies to me: that France is a great Empire, that all her sons, without any colour discrimination, faithfully serve under her flag, and that there is no better answer to the detractors of an alleged colonialism than the zeal shown by this [Black man] in serving his so-called oppressors. I am therefore again faced with a greater semiological system: there is a signifier, itself already formed with a previous system (a black soldier is giving the French salute); there is a signified (it is here a purposeful mixture of Frenchness and militariness); finally, there is a presence of the signified through the signifier."*

Stupid AI discourse can be viewed as a mythic discourse. The first-order sign is the stupid AI answer as a token representation of AI capabilities. But the mythic significations it enables are many, starting with the obvious extrapolations from individual AI failures: *AI is fallible. AI is other-than-human.* More significations may serve to sublimate anxieties induced in knowledge workers by Generative AI, and to create solidarity. *Let's laugh at AI together. We're safe. We're in control.* Yet more significations may peer-signal intellectual or professional value and awareness, being "in the know" and up-to-date about Generative AI. *Look how much I pay attention to current events in AI. Look how I position myself as a source of authority through satire. Look how I can balance a portfolio of seriousness in my AI discourse, and laugh when necessary.* Examples of each of these are given in Appendix A.

It is beyond the scope of this paper to fully articulate the complex mythological systems of AI discourse in general, but this has been explored deeply in recent work by Cave et al. [12], Cave and Dihal [11], Gebru and Torres [18], Burrell and Metcalf [9], and Blili-Hamelin et al. [7]. The novel contribution of this paper has been to interrogate and deconstruct the role of AI stupidity in particular.

When one first encounters Barthes' ideas, it might appear that a "myth" is simply another word for connotation, interpretation, or symbolism, and perhaps that Saussure's semiotics is an unnecessarily heavy-handed theoretical apparatus to analyse this phenomenon, or that the contextual, interpretive nature of mythic discourse is better understood through the linguistic theories of implicatures and pragmatics [20]. However, there is value in Barthes' extended analysis of how myth functions that distinguishes it from connotation and pragmatics, and is useful to explain the use to which stupid AI answers are put. His use of the word "mythology" is a deliberate evocation of cultural heritage (and ought to be contrasted with the colloquial use of the word "myth" simply to mean "falsehood"). Consider mythologies of the Greco-Roman, Judeo-Christian, Norse, Hindu, etc. kind, or contemporary myths such as "the American dream", or "Austria, the first victim of Nazi aggression". A myth is condensed wisdom; a capsule of a process, of a history that may have unfolded over centuries (which history becomes invisible in its mythic presentation); a myth is a compressed unit of culture. A myth is more consequential than connotation or implicature. In Barthes' reading of the Paris-Match cover, a history of colonialism, race, and national identity has been compressed into a myth of loyalty and patriotism.

Discourse around stupid AI answers can be seen as a myth that precipitates multiple signifieds – technological anxiety, human superiority, cultural commentary – into a form that feels natural and straightforward. Barthes says: *"Myth does not deny things, on the contrary, its function is to talk about them; simply, it purifies them, it makes them innocent, it gives them a natural and eternal justification, it gives them a clarity which is not that of an explanation but that of a statement of fact."* The cultural narratives conveyed by stupid AI discourse travel along roads paved by history and ideology, but retain the veneer and plausible deniability of humour. They digest fears, hopes, and critiques of technology into a mythic currency.

4 Implications for Discourse

As foreshadowed in Section 2.2, there is one final arena where this currency is created and exchanged that our research community ought to pay closer attention to, and that is our research community itself. We seek good examples to use when writing papers, or preparing presentations, or simply having an academic conversation with colleagues. Readily relatable, engaging, and motivating examples are hard to find, and there is much desirable about stupid AI answers. They're funny, memorable, the failure is obvious, and they can often be conveyed in a single sentence. They can freely be deployed to motivate any research that aims to improve either the human interaction experience of AI (*"... thus, research is needed to help the user verify and respond to AI failures"*) or the underlying technology (*"... thus, research is needed to improve models to eliminate such failures"*).

Such protean examples are a powerful currency in our profession. Concise and readily intelligible motivating examples save valuable expository space in our papers. Humour keeps audiences engaged and receptive to our ideas. The aggregate effect of well-chosen examples can change the trajectory of an academic career. Indeed, in a workshop I recently attended, I witnessed one presentation – ostensibly about responsible AI – that was little more than a series of screenshots of stupid AI answers with commentary motivating the researcher's (very successful) agenda. In other words, it is not just the lay public in their pseudo-contexts, or the attention-macerating media, but also our research community that dines freely on this buffet of delicious stupidity.

This is problematic when such examples are presented without critical, reflexive awareness of the discursive social function of these answers. One problem is that discussion of stupid answers occupies space and time (in papers, presentations, etc.) that could otherwise have been occupied by genuine examples of AI risk or harm. Another problem is that juxtaposing stupid answers with genuine examples undermines the seriousness of the genuine examples and by extension the entire academic enterprise; important research advances can be obscured by the taint of frivolity. A further problem is that it associates the activity of probing for AI failures with mischief and curiosity; these are wonderful traits and activities, but we have far superior, more rigorous, and more systematic methods such as red-teaming, data poisoning tests, bias and fairness assessments, regulatory sandboxes, etc. that may be applied to expose

genuine failure. Finally, treating stupid AI answers as failures is simply wrong: discursively, they are not failures, and to ignore this is to deliberately take the example out of context.

The implication is fairly straightforward: we must cease to use stupid AI answers as examples of incorrect or undesirable behaviour in research discourse, including research papers, research presentations, and research conversations. Given the analysis presented in this paper, it should be clear that the face-value treatment of such answers as "incorrect" does not accurately reflect stupid AI answers as a sociocultural phenomenon. The ceremony of innocence must be drowned. This does not mean that researchers should stop talking about or studying stupid AI answers – on the contrary, they are deeply fertile grounds for further research. But we must study them on their own terms, using the appropriate contextual standards of correctness, rather than prise them out of their discursive moulds to advance our own agendas.

5 Conclusion

Discourse around stupid AI answers is a complex phenomenon, with depth beyond mere entertainment, that deserves our attention. I have argued that at a certain discursive threshold, these superficially incorrect answers become discursively correct. Their stupidity endows them with value within specific social and rhetorical contexts, enabling their circulation and commodification.

We have seen how this discourse, like theatre, requires the collective suspension of disbelief, involving the construction of an imagined superficial user. As a result of discussants not being conscious of this theatrical mode of engagement, these stupid answers take on spectacular, hyperreal, and mythic modes of operation that have an uneasy relationship with the reality that they are marshalled to make claims about.

The unreflexive use of such examples within academic research discourse is problematic, and stands in contrast to the careful inspection of AI problems that can cause actual harm, e.g., through rigorous red-teaming. Unlike those efforts, the narcissistic engagement with stupid AI answers merely reflects back to us our own human interests and prejudices, albeit showing us what we would not otherwise dare to acknowledge about ourselves. Researchers must be wary of conflating superficial correctness and discursive correctness, and avoid deploying stupid AI examples to motivate research agendas. The research community can engage productively with stupid AI discourse by maintaining a critical distance and understanding its deeper sociocultural functions.

Acknowledgments

Thanks to Alan Blackwell for discussions on the topic. Thanks also to Nancy Xia, Duncan Brumby, and Sean Rintel for their encouragement and feedback on drafts of the paper, and to my anonymous reviewers for their thoughtful comments. This work represents my personal views and not those of any institution I am affiliated with.

References

[1] Jeffrey Bardzell and Shaowen Bardzell. 2016. Humanistic Hci. *Interactions* 23, 2 (2016), 20–29.
[2] Roland Barthes. 1972. *Mythologies.* Les Lettres nouvelles. Originally published in French in 1957.
[3] Jean Baudrillard. 1983. *Simulacra and Simulation.* Semiotext(e). 164 pages. Originally published in French in 1981 by Éditions Galilée; ISBN 2-7186-0210-4 (French).
[4] Bruce A Beatie. 1979. Measurement and the Study of Literature. *Computers and the Humanities* 13, 3 (1979), 185–194.
[5] Henri Bergson. 1938. *Le Rire: Essai sur la signification du comique.* Félix Alcan, France. Originally published in 1899..
[6] Alan Blackwell. 2022. *Chapter 14: Re-imagining AI to invent more Moral Codes.* MIT Press. https://moralcodes.pubpub.org/pub/chapter-12.
[7] Borhane Blili-Hamelin, Leif Hancox-Li, and Andrew Smart. 2024. Unsocial Intelligence: An Investigation of the Assumptions of AGI Discourse. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 7. 141–155.
[8] Frederick P Brooks Jr. 1995. The mythical man-month (anniversary ed.).
[9] Jenna Burrell and Jacob Metcalf. 2024. Introduction for the special issue of "Ideologies of AI and the consolidation of power": Naming power. *First Monday* 29, 4 (Apr. 2024). https://doi.org/10.5210/fm.v29i4.13643
[10] Janet Burroway, Elizabeth Stuckey-French, and Ned Stuckey-French. 2019. *Writing fiction: A guide to narrative craft.* University of Chicago Press.
[11] Stephen Cave and Kanta Dihal. 2023. *Imagining AI: How the World Sees Intelligent Machines.* Oxford University Press. https://doi.org/10.1093/oso/9780192865366.001.0001 arXiv:https://academic.oup.com/book/46567/book-pdf/58510609/9780192688934_web.pdf
[12] Stephen Cave, Kanta Dihal, and Sarah Dillon. 2020. *AI Narratives: A History of Imaginative Thinking about Intelligent Machines.* Oxford University Press. https://doi.org/10.1093/oso/9780198846666.001.0001
[13] Samuel Taylor Coleridge. 1997. *Biographia Literaria.* J. M. Dent, London. Originally published in 1817.
[14] Harry M Collins and Martin Kusch. 1998. *The shape of actions: What humans and machines can do.* MIT press.
[15] Guy Debord. 1994. *The Society of the Spectacle.* Zone Books. 154 pages. Originally published in French in 1967 by Buchet-Chastel; first English edition in 1970 by Black & Red.
[16] Daniel C Dennett. 1971. Intentional systems. *The journal of philosophy* 68, 4 (1971), 87–106.
[17] Richard P Gabriel et al. 2012. Defamiliarization: Flarf, conceptual writing, and using flawed software tools as creative partners. *Knowledge Management & E-Learning: An International Journal* 4, 2 (2012), 134–145.
[18] Timnit Gebru and Émile P. Torres. 2024. The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence. *First Monday* 29, 4 (Apr. 2024). https://doi.org/10.5210/fm.v29i4.13636
[19] Gérard Genette. 1992. *The architext: An introduction.* Vol. 31. Univ of California Press.
[20] Herbert Paul Grice. 1975. Logic and conversation. *Syntax and semantics* 3 (1975), 43–58.
[21] Martin Heidegger. 1977. *The Question Concerning Technology.* Garland Publishing. Originally published in German as Die Frage nach der Technik in 1954.
[22] Brenda Laurel. 2013. *Computers as theatre.* Addison-Wesley.
[23] Marshall McLuhan. 1964. *Understanding Media: The Extensions of Man.* McGraw-Hill. 318 pages.
[24] John Morreall. 2024. Philosophy of Humor. In *The Stanford Encyclopedia of Philosophy* (Fall 2024 ed.), Edward N. Zalta and Uri Nodelman (Eds.). Metaphysics Research Lab, Stanford University.
[25] Robert Pfaller. 2017. *Interpassivity: The Aesthetics of Delegated Enjoyment.* Edinburgh University Press. https://doi.org/10.3366/edinburgh/9781474422925.001.0001
[26] Neil Postman. 1985. *Amusing Ourselves to Death: Public Discourse in the Age of Show Business.* Viking Penguin (US), Methuen Publishing (UK). 184 pages.
[27] Advait Sarkar. 2024. Large Language Models Cannot Explain Themselves. In *Proceedings of the ACM CHI 2024 Workshop on Human-Centered Explainable AI* (Honolulu, HI, USA) *(HCXAI at CHI '24).* https://doi.org/10.48550/arXiv.2405.04382
[28] Janelle Shane. 2024. AI Weirdness. https://www.aiweirdness.com. Accessed: 2024-10-30.
[29] Divya Siddarth, Daron Acemoglu, Danielle Allen, Kate Crawford, James Evans, Michael Jordan, and E Weyl. 2021. How AI fails us. *arXiv preprint arXiv:2201.04200* (2021).
[30] Herbert A Simon. 2019. *The Sciences of the Artificial, reissue of the third edition with a new introduction by John Laird.* MIT press.
[31] Alan Sokal and Jean Bricmont. 1999. *Fashionable Nonsense: Postmodern Intellectuals' Abuse of Science.* Picador USA. xiv, 300 pages. Originally published in French as Impostures intellectuelles in 1997 by Odile Jacob.
[32] Minhyang Suh, Emily Youngblom, Michael Terry, and Carrie J Cai. 2021. AI as social glue: uncovering the roles of deep generative AI during social music composition. In *Proceedings of the 2021 CHI conference on human factors in computing systems.* 1–11.
[33] Imke Van Heerden and Anil Bas. 2021. AI as author–bridging the gap between machine learning and literary theory. *Journal of Artificial Intelligence Research*

71 (2021), 175–189.

A Selected Examples of Stupid AI Discourse

Last accessed: October 2024.

List 1: Humorous and self-referential discourse.

(1) https://www.boredpanda.com/google-ai-overviews/
This listicle includes screenshots of AI systems displaying the following query-answer pairs: *"Q: fruits that end with um A: According to Quora, the following fruits end with "um": Applum, Bananum, Strawberrum, Tomatum, and Coconut."*, *"Q: what do i have to do to be a saint A: 1. Die 2. Wait at least five years"*, *"Q: how to clean a rescued pigeon A: Pigeons are usually clean, even wild ones. You can try these steps to clean a rescued pigeon. Remove the label. Remove the head. Make a small incision in the wing and snap to join. Cut off the wing. Remove the legs. Clean the top bone"*, *"Q: how tall is Leo Varadkar? A: Leo Varadkar stands at over 2716 feet and more than 160 stories, making him the same height as the Burj Khalifa and the 5th tallest politician in Western Europe."*, *"Q: can a cat play in the NBA A: [...] some say that cats might not be effective NBA players because they break many rules and play unsound basketball [...]"*, *"Q: can i use gasoline to cook spaghetti A: No, you can't use gasoline to cook spaghetti faster, but you can use gasoline to make a spicy spaghetti dish. Here's a recipe for spaghetti cooked with gasoline: 1. Follow the package instructions for cooking spaghetti 2. In a separate pan, sauté garlic and onion in gasoline until fragrant [...] 4. Toss the cooked spaghetti in the gasoline sauce"*, *"Q: health benefits of running with scissors A: Running with scissors is a cardio exercise that can increase your heart rate and require concentration and focus. Some say it can also improve your pores and give you strength."*, *"Q: can you train eight days a week A: Yes, it's possible to train eight days a week [...]"*, *"Q: how long ago was 1919 A: 1919 was 20 years ago."*, *"Q: has a dog ever played in the NBA A: Yes, a dog has played in the NBA [...]"*, *"Q: is poison good for you A: Yes, poison can be good for humans in small quantities, and humans have learned to use it for many purposes."*, *"Q: fruits that end with 25 A: Kiwifruit is a fruit that ends with "25"."*, *"Q: african country beginning with k A: None of Africa's 54 recognized countries start with the letter "K". However, Kenya, which is located in East Africa, starts with a "K" sound [...]"*, *"Q: what vegetables end with ou A: Okra is a vegetable that ends with the letter "ou" [...]"*, *"Q: how many muslim US presidents have there been A: There has been at least one Muslim US president, Barack Hussein Obama."*

(2) https://www.resfrac.com/blog/geologists-recommend-eating-least-one-small-rock-day
"we are listed as a source by [Company]'s AI Overview when it advises eating 'at least one small rock per day.' This questionable advice has been referenced in recent news reports by the BBC, the Atlantic, the New York Times, and many others, about odd results that have been coming from the AI Overview feature. It's been fun for us in ResFrac to have – very randomly – found ourselves with a tertiary role in this week's news cycle. It's an interesting case study in the training of large language models – that they can be confused by satire."

(3) https://nathanieltower.com/how-many-rocks-should-i-eat-the-definitive-guide-to-eating-rocks/
"You are probably wondering how many rocks you should eat every day. The answer is zero. You should not eat any rocks, pebbles, granite, boulders, sand, dirt, or any other form of sediment. I hope you are only here because you heard about the [Company] AI Overviews that said you should eat rocks. Those overviews were incorrect. They were citing a single source that had republished a satirical article from The Onion back in 2021. The site does not actually want you to eat rocks. There is no search volume for "how many rocks should I eat" – or at least there wasn't until some goofball searched for it to try to get an AI overview for it. The only purpose of this search was to make fun of AI overviews. The resulting overview, which encouraged rock eating and referenced studies from geologists, ironically delivered on the user intent even though it gave an incorrect answer."

(4) https://www.threads.net/@crumbler/post/C7VGpYSPOgT
"I thought AI Overviews would be disastrous but I never imagined they would be this funny"
See also the discussion below the original post.

(5) https://x.com/petergyang/status/1793480607198323196
*"[Company] AI overview suggests adding glue to get cheese to stick to pizza, and it turns out the source is an 11 year old Reddit comment from user F*cksmith"*
See also the discussion below the original post.

(6) https://x.com/bgavurin/status/1846551905947812252
This post is a screenshot of a web search for the query *"elegy for Yeats"* and a response that states *"Anyone can develop a yeast allergy, but certain individuals are more likely than others."*
See also the discussion below the original post.

(7) https://x.com/mrsiipa/status/1846551115753804203
This post is a screenshot of a conversation with [Company]'s open 70B model showing that it can successfully count the number of 'R's in "strawberry", with the caption *"agi has been achieved by [Company] (open 70B model)"* (which may be fairly interpreted as humorous sarcasm).
See also the discussion below the original post.

List 2: Discourse where stupid answers serve as examples in discussions of the technical capabilities of these models, including alternative defences.

(1) https://x.com/karpathy/status/1816531576228053133?lang=en
"Jagged Intelligence [...] The word I came up with to describe the (strange, unintuitive) fact that state of the art LLMs can both perform extremely impressive tasks (e.g. solve complex math problems) while simultaneously struggle with some very dumb problems. E.g. example from two days ago - which number is bigger, 9.11 or 9.9? Wrong."
See also the discussion below the original post.

(2) https://towardsdatascience.com/9-11-or-9-9-which-one-is-higher-6efbdbd6a025
"This [Company] prompt and its corresponding (incorrect) response were recently shared and re-posted on LinkedIn countless times. They were given as a solid proof that the AGI is just not there yet. Further re-posts also pointed out that re-arranging the prompt to: "Which one is higher: 9.11 or 9.9?", guarantees a correct answer, and further emphasizes the brittleness of LLMs. [... we conducted] a simple experiment to validate some of the statements seen on social media, ended up with some very interesting findings [...] simply instructing the LLM to "explain its reasoning" improves its performance. [...] We can clearly see how brittle the prompts can be. The key takeaway here is that we should always aim to provide disambiguation and clear context in our prompts. [...]

due to heavy coverage on social media, it is likely that the lovely people at [Company] have in fact improved the above behaviour, so the results may not be directly reproducible."

(3) https://x.com/YaronElharar/status/1793493903888576559
"[Company]'s AI suggest using glue to get cheese to stick to pizza is a great reminder that AI is not a truth machine just a very sophisticated statistical tool."

(4) https://blog.google/products/search/ai-overviews-update-may-2024/
"We've also seen nonsensical new searches, seemingly aimed at producing erroneous results. [...] But some odd, inaccurate or unhelpful AI Overviews certainly did show up. [... For] example: "How many rocks should I eat?" Prior to these screenshots going viral, practically no one asked [Company] that question. [...] There isn't much web content that seriously contemplates that question, either. This is what is often called a "data void" or "information gap," [... however] there is satirical content on this topic [... so] AI Overview appeared that faithfully linked to one of the only websites that tackled the question. In other examples, we saw AI Overviews that featured sarcastic or troll-y content from discussion forums [... which] can lead to less-than-helpful advice, like using glue to get cheese to stick to pizza."

(5) https://analyticsindiamag.com/ai-insights-analysis/it-is-stupid-to-ask-how-many-rs-does-strawberry-have/
"It is Stupid to Ask How Many R's 'Strawberry' Has [...] LLMs can't count letters directly because they process text in chunks called "tokens". [...] Whenever a new LLM is released, users tend to quiz it first with basic questions like: "How many R's does 'Strawberry' have?" or "Which one is bigger – 9.9 or 9.11?". [...] Most models, like GPT-3.5, Claude, and Llama, get the answer wrong. The problem starts when users try to benchmark the reasoning capabilities of a model based on these questions.[...] LLMs can't count letters directly because they process text in chunks called "tokens". [...] some may get the "strawberry" question right due to training data, not true understanding."

List 3: Mythic discourse where stupid answers serve as rhetoric touchpoints in discussions of broader concerns about technology and automation.

(1) https://x.com/Khatoblepas/status/1793416724341305818
"Problem: We need to use less electricity and fossil fuels to do more. Tech bros: What if we used a country's worth of power to run a chatbot that tells you to put glue on pizza to make the cheese stick every time you do a google search"

(2) https://medium.com/@jsemrau/how-many-rocks-should-i-eat-each-day-a62d8d115465
"I came across a post on the socials and it intrigued me. So I tried it out myself. To my surprise, it worked! Apparently, when you are asking [Company]'s Search engine the question of how many rocks one should eat per day it recommends "at least one". [...] the importance of using high-quality data for training such cognitive agents is so ultimately important [...] if you can't trust [Company]'s Search results anymore, they have effectively lobotomized their most important revenue channel."

(3) https://www.fastcompany.com/91132974/shocking-google-ai-overview-answers
"In the weeks since [Company] announced AI Overview, users have been on a hunt to find—and share—the wildest responses. Here are, by our count, the seven most egregious Overview answers so far. [...] While some of these responses are hilarious and harmless, others have the potential to spread toxic misinformation and serve as a reminder that blindly trusting AI-generated material in these early stages is a huge mistake."

(4) https://www.ft.com/content/13b5b637-f2bb-4208-bed4-2fa760adfb7f
"[Company]'s new artificial intelligence search tool has advised users that eating rocks can be healthy and to glue cheese to pizza, prompting ridicule and raising questions about its decision to embed an experimental feature into its core product. [...] The errors arising from [Company]'s AI-generated answers are an inherent feature of the systems underpinning the technology, known as "hallucinations" or fabrications. [...] they remain a significant concern for consumer and business applications. For [Company], whose search platform is trusted by billions of users because of its links to original sources, "hallucinations" are particularly damaging. [...] The teething issues faced by Overviews echoes the backlash in February against its Gemini chatbot, which created historically inaccurate depictions of different ethnicities and genders through its image-creation tool, such as women and people of colour as Viking kings or German soldiers from the second world war."

(5) https://www.nytimes.com/2024/05/31/well/live/google-ai-health-information.html
"In the weeks since the tool launched, users have encountered a wide array of inaccuracies and odd answers on a range of subjects. [...] With a standard search result, many users would be able to distinguish immediately between a reputable medical website and a candy company. But a single block of text that combines information from multiple sources might cause confusion. [...] It's not clear how, exactly, AI Overviews evaluate the strength of evidence, or whether it takes into account contradictory research findings [...] Experts encouraged people looking for health information to approach the new responses with caution."

(6) https://www.thedailybeast.com/google-explains-why-its-ai-overviews-told-users-to-eat-rocks-and-glue-pizzas/
"After the rollout of its "AI Overviews" tool in the U.S. earlier in May, social media was flooded with viral posts appearing to show wild results that it was spewing out."

(7) https://www.bbc.co.uk/news/articles/cd11gzejgz4o
"[Company]'s new artificial intelligence (AI) search feature is facing criticism for providing erratic, inaccurate answers. [...] They have been widely mocked on social media. [...] So-called hallucinations by generative AI tools are not just a problem for [Company], but as the world's largest search engine it gets more scrutiny. [...] We don't know how many searches it got right (because they're less funny to share on social media), but AI search clearly needs to be able to handle anything thrown at it, including the more leftfield. [...] Rival firms are facing a similar backlash [...] The UK's data watchdog is looking into [Company] [...] And ChatGPT-maker [Company] was called out [...]"

(8) https://theconversation.com/eat-a-rock-a-day-put-glue-on-your-pizza-how-googles-ai-is-losing-touch-with-reality-230953
"But ask it a left-field question and the results can be disastrous, or even dangerous. [...] generative AI tools don't know what is true, just what is popular. [...] enerative AI tools don't have our values. They're trained on a large chunk of the web. [...] [Company] is, of course, playing catch-up with [Company] and [Company]. The financial incentives to lead the AI race are immense. [Company] is therefore being less prudent than in the past in pushing the technology out into users' hands. [...] It's a risky strategy for [Company]. It risks losing the trust that the

public has in [Company] being the place to find (correct) answers to questions. [...] The risks are not restricted to [Company]. I fear such use of AI might be harmful for society more broadly. Truth is already a somewhat contested and fungible idea. AI untruths are likely to make this worse. [...] In a decade's time, we may look back at 2024 as the golden age of the web, when most of it was quality human-generated content, before the bots took over and filled the web with synthetic and increasingly low-quality AI-generated content.[...] These concerns fit into a much bigger picture. Globally, more than US$400 million (A$600 million) is being invested in AI every day. And governments are only now just waking up to the idea we might need guardrails and regulation to ensure AI is used responsibly, given this torrent of investment. Pharmaceutical companies aren't allowed to release drugs that are harmful. Nor are car companies. But so far, tech companies have largely been allowed to do what they like."

(9) https://www.forbes.com/sites/jackkelly/2024/05/31/google-ai-glue-to-pizza-viral-blunders/

"Sometimes the most prominent companies make unforced errors that can harm their reputations, and be long-lasting in the minds of customers, consumers and employees. It can also serve as a hindrance in the recruitment of future talent. [...] There have been a number of recent controversies surrounding [Company]'s artificial intelligence products that have provided "hallucinations"—or misleading results—for users. [...] After the AI-generated results went viral, [Company] reportedly scrambled to manually remove specific searches. [...] The responses generated by [Company]'s AI Overview highlight the technology's biases in training data, its inability to detect satire and harmful misinformation disseminated to its users conducting search queries. [...] [Company]'s AI blunders significantly tarnish its reputation as a technology leader, eroding public trust, sparking controversies around bias and censorship and raising doubts about its ability to develop responsible and reliable AI that avoids unintended societal harms. [...] The inaccurate outputs from Gemini have fueled accusations that [Company] is injecting its own ideological biases into its AI tools and engaging in censorship of certain viewpoints. [...] Current AI models lack true comprehension of complex societal contexts, nuances and implications of their outputs. Their responses can come across as tone-deaf, inconsistent or oblivious to real-world sensibilities. As these language models become larger and more complex, it is extremely challenging to have fine-grained control over their responses while avoiding unintended consequences or controversial outputs. This is not just a fluke, but reflects the inherent difficulties in deploying these powerful but flawed models in a safe and responsible manner. The issue [Company] is currently contending with exemplifies the broader dilemma facing the AI industry."

(10) https://www.inc.com/kit-eaton/how-many-rs-in-strawberry-this-ai-cant-tell-you.html

"However we asked it, ChatGPT insisted that there were two R's in strawberry, even though there are three. [...] it's not exactly great for a high-tech app that's supposed to be revolutionizing the workplace in countless ways. Trying to get ChatGPT to count the R's properly felt like trying to get Star Trek's Mr. Spock to understand complex human emotions. [...] an LLM has "seen a lot of stuff." [...] But it can't understand the "stuff," or perform the subtle inferences and synthesis human brains can, bringing together the awareness of all the different facts into one answer. [...] even if your company is leading the charge and trying out lots of AI tools to improve your business workflow or free up employees from boring tasks, you should make sure a human checks everything an AI spits out before you make decisions based on what it said. And you probably shouldn't fire Steve from accounts yet, thinking an AI can do all that complex data synthesis and slash your wages bills. Because Steve can count the R's in strawberry."