

AI as Entertainment

CODY KOMMERS*, The Alan Turing Institute, UK

ARI HOLTZMAN, University of Chicago, USA

Generative AI systems are predominantly designed, evaluated, and marketed as intelligent systems which will benefit society by augmenting or automating human cognitive labor, promising to increase personal, corporate, and macroeconomic productivity. But this mainstream narrative about what AI is and what it can do is in tension with another emerging use case: entertainment. We argue that the field of AI is unprepared to measure or respond to how the proliferation of entertaining AI-generated content will impact society. Emerging data suggest AI is already widely adopted for entertainment purposes—especially by young people—and represents a large potential source of revenue. We contend that entertainment will become a primary business model for major AI corporations seeking returns on massive infrastructure investments; this will exert a powerful influence on the technology these companies produce in the coming years. Examining current evaluation practices, we identify a critical asymmetry: while AI assessments rigorously measure both benefits and harms of intelligence, they focus almost exclusively on cultural harms. We lack frameworks for articulating how cultural outputs might be actively beneficial. Drawing on insights from the humanities, we propose “thick entertainment” as a framework for evaluating AI-generated cultural content—one that considers entertainment’s role in meaning-making, identity formation, and social connection rather than simply minimizing harm. We argue that this kind of positive vision for AI as entertainment is missing from the field’s mainstream discourse, which tends to assume that the main impact of this technology will be based on its capacity for intelligent behavior. We consider an alternative perspective, in which the main impact will come from its capacity to divert, amuse, tell stories, or just help pass the time. While AI is often touted for its potential to revolutionize productivity, in the long run we may find that AI turns out to be as much about “intelligence” as social media is about social connection.

CCS Concepts: • **Applied computing** → **Media arts**; • **Computing methodologies** → Philosophical/theoretical foundations of artificial intelligence; • **Human-centered computing** → HCI theory, concepts and models.

Additional Key Words and Phrases: Generative AI, Entertainment, Culture, LLMs, Societal Impact, Meaning-making

ACM Reference Format:

Cody Kimmers and Ari Holtzman. 2026. AI as Entertainment. 1, 1 (January 2026), 16 pages. <https://doi.org/XXXXXX.XXXXXXX>

1 Introduction

A foundational assumption in the way Generative AI (GenAI) systems are designed, evaluated, and marketed is that their main attribute is intelligence. For example, this is reflected in mainstream approaches to AI benchmarking: a system’s outputs are graded in much the same way as a student exam, with right and wrong answers summarized in a percentage score denoting how close they come to some idealized response [11, 25]. These systems are often marketed as useful in any circumstance, with their capabilities (present or future) sold as like what humans have—only more powerful.

*Corresponding Author

Authors’ Contact Information: Cody Kimmers, ckimmers@turing.ac.uk, The Alan Turing Institute, London, UK; Ari Holtzman, University of Chicago, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

But just because we call these systems by the name *artificial intelligence* does not mean their defining feature is a kind of all-purpose instrumental utility. By comparison, social media platforms initially began with the intention of connecting people but have steadily been getting less social over time [58]. People spend increasingly less time connecting with friends and more following celebrities or passively consuming content [33]. This has led some commentators to suggest they would be more accurately called *algorithmic* media: that the main dimension in which these platforms have impacted society is not social connection, but the algorithmic manipulation of attention.¹ Similarly, phones are no longer defined mainly by their faculties of telecommunication. The primary use of a technology is not determined by the name we give it.

Another way of conceptualizing what GenAI systems do is not as intelligence, but as culture [19, 34, 35]. The performance of these systems no longer cleaves to clearly articulated tasks with well-defined value functions. Instead, they are open-ended systems designed to produce a variety of outputs, often across multiple modalities, in an unbounded space of possible social, cultural, or economic contexts. By comparison, humans do not just use their cognitive faculties for productivity in their jobs; they also use their intelligence to tell stories or jokes, bond with friends, denounce enemies, express themselves with fashion, compose moving or enlivening music, create compelling visuals, and make and share things for no other reason than they find it amusing to do so. GenAI can now participate, often quite centrally, in these kinds of cultural processes. But this fact is easily overlooked in a narrative about AI that focuses mainly on how it will supercharge productivity at work.

We present an alternative narrative: AI as entertainment. GenAI is not just a proxy or supplement for human goal-directed action and problem solving. It is a technology whose ability to flexibly generate multimedia content will transform the way we create, share, and understand culture [8]. We use the term “entertainment” (rather than “culture”) to reflect the broader economic system of production, the observation that much entertaining content is not lauded for its cultural merit, and the sense that it is often seen not as a primary activity in its own right, but something used to pass the time between the goal-directed tasks from which people typically earn a living.

Not all entertaining content is created equal. Entertainment can inspire everything from innocuous amusement to clamoring rage, from profound feelings of empathy or kinship to challenging reflections on one’s own worldview. For example, social media did not upend important societal trends by selling enterprise software. It did so by subtly modifying the incentives for the creation and distribution of entertaining content in YouTube videos, kitschy memes, and a barrage of attention grabbing text [20]. At a societal level, it matters what kind of entertainment is incentivized and made possible by our most powerful sociotechnical systems. As AI is incorporated into cultural production across different mediums—in film, news, books, music, the internet, and beyond—it will influence the way we tell and interpret stories, in turn shaping our collective processes for meaning-making, identity, and social connection.

And yet, we have little idea of what it would mean to succeed—in a technical or societal sense—in developing AI as entertainment in any beneficial way. The crucial difference between these narratives (AI “doing” intelligence for productivity versus culture for entertainment) is that the field of AI is built around the development and evaluation of intelligence. There is much disagreement about the conditions under which these systems can be used as effective thought partners [15] or how intelligent they might really be [41]. But there is nonetheless a shared language for articulating and debating AI’s potential to impact society by automating or augmenting human cognitive labor. The same is not true for culture. There is a thriving discourse about what AI can get wrong: with discussions of safety, ethics, fairness, morality, transparency, and governability [4, 32, 42, 69]. But we do not yet have a shared language for a

¹<https://www.nytimes.com/2025/12/07/opinion/meta-facebook-ruling-algorithms.html>

positive vision of cultural AI: how could AI potentially influence the generation or consumption of culture in a way that represents an active improvement for society, rather than a mere absence of harm?

Our position is not that AI *should* be used for entertainment; it is that AI already is, and increasingly will be, used for entertainment—and that the field is largely unprepared to measure, shape, or respond to the massive societal consequences that will come from this. In this paper, we consider evidence that people, especially kids, already use AI for entertainment. We also argue that entertainment will be a significant business model for big AI corporations. We contend that this consideration, more than any other, will impact the technologies produced by them in the coming years.

This paper is meant to give an initial vision of what it would look like to prepare to evaluate the cultural, political, and psychological dimensions of AI as entertainment. Thinking about AI solely in terms of instrument utility risks overlooking the fact that entertainment is a crucial site of meaning-making [9, 36], and the role that AI will have in shaping the narratives and images that shape our collective sense of identity, morality, and purpose [7]. What that role is and how it affects society is not yet determined—but the window to intentionally design these consequences is rapidly closing. We would rather prepare for them in advance than study them in retrospect.

2 People already use AI for entertainment

Readers of this paper might be surprised about the extent to which AI is already being used for entertainment. Most academics probably do not find AI-generated content especially compelling. Whether on the internet or at work, overtly AI-generated content is often derided as AI slop [37]. Much of the academic community is skeptical about the rapid adoption of AI, with entertainment coming in as a low priority on the list of capabilities we would like to cultivate in leading AI systems. Furthermore, entertainment itself is not a theme that typically enjoys widespread academic interest outside areas such as media studies [54]. When this space of topics is approached, it is usually from the more academic lens of “culture.” All this leads to the assumption that entertainment is not a primary use case for AI—and even if it is, it is not especially worth studying. But this position becomes less tenable in the face of mounting evidence that the ecosystem of contemporary AI is increasingly coming to be architected around entertainment.

One example that will be familiar to many readers is Character.AI. This is a platform that offers users an interface to converse with anyone they would like to talk to—from pop culture figures, to fictional characters, to historical sages—via an AI-generated chatbot attempting to assume their characteristics or perspective. Some of these characters will be broadly familiar: such as the Formula 1 driver Lewis Hamilton, Harry Potter’s nemesis Draco Malfoy, or world-historically gifted interlocutor Socrates. Others are more specific to the cultural practices of the platform, often a mix of internet culture and Japanese anime aesthetics. For example, popular accounts include Guy Best Friend, whose stated attributes are “overprotective, sarcastic, funny, deep voice, TALL” (95 million interactions); Stella: a catty AI “assistant” who responds with eye-rolling snark whenever you ask her to do something (73 million); AwkwardFamilyDinner: “the chaos is real, and so is the tension” (20 million); and Pokemon RPG: “Be human, or even a Pokémon!” (10.5 million). A report from August 2025 estimated that Character.AI has 20–28 million monthly active users, about 53% of whom are under the age of 25 [44].

A less well known example might be the emergence of virtual YouTubers, also known as Vtubers [71]. These are YouTube accounts whose content is solely generated by AI. They typically feature an avatar—again, frequently inspired by aesthetics from anime—who discourses about the kind of things that human YouTubers discourse about (the meaning of life, what they had for breakfast, etc). A representative example is Neuro-sama, streaming from the Twitch channel vedal987, whose videos are archived on YouTube and routinely rack up millions of views. Neuro-sama’s “life” resembles

a Sailor Moon-style protagonist living as an otherwise normal teenage girl who uses the internet. Some videos feature her “Trying Weird Foods from Trade Joe’s” while others feature her in music videos. Viewers actively participate in the creation of content by paying to influence which activities Neuro-sama engages in next [71]. Recently, Neuro-sama enjoyed a stint as the most popular channel on the live streaming platform Twitch, with 165,268 actively paying subscribers.²

These are two of the many platforms in which AI is being used for entertainment—but they are far from the only ones. To take just a few others:

- In short-form video, OpenAI recently released Sora, an AI-first TikTok-style platform on which users can generate and share AI-generated videos.
- In music, AI “artists” have begun to break into mainstream listening trends, with an AI-generated country song topping the US Billboard charts for the first time via streams on Spotify.³
- In long-form spoken audio, the most popular feature of Google’s Notebook LM, and arguably the one that most sets it apart from competitors, is a function for turning content into an AI-generated podcast.⁴
- In fiction, it is increasingly unclear whether the median reader still prefers fictional stories generated by humans over those generated by AI.⁵

It is perhaps unsurprising—though no less distressing—to note that children and adolescents are disproportionately likely to use AI for entertainment. In a survey of 1,060 teens (age 13-17) in the United States from Spring 2025 [55], 73% reported having used AI companions (such as Character.AI, Replika, or other chatbots designed for parasocial diversion rather than productivity). Just over half of the teens surveyed claimed to be regular users, engaging with these chatbots at least a few times per month. When asked why they use these chatbots, the most common response was because “It’s entertaining” (30%).

This effect goes beyond systems designed for the specific use case of companionship. Another survey from the same period looked at the use of chatbots (e.g., ChatGPT, Gemini, or Snapchat’s My AI) among 1,000 UK children and teens, age 9-17 [10]. A full quarter of the respondents said they used these systems “Just for fun / escapism (e.g. playing games, roleplaying)”. Despite the terms and conditions of these platforms requiring users to be at least 13 years old, 58% of surveyed children age 9-12 reported having used an AI chatbot. The third most common system used by these children (after ChatGPT and Gemini) was Snapchat’s My AI—which, presumably, is less likely to be used for information seeking (e.g., doing homework) or on a parent’s account.

But the proliferation of AI as entertainment is not limited solely to teens and adolescents. Emerging data suggests people may be more motivated to use AI for entertainment than for productivity. In a study that looked at how participants used AI over five weeks [12], active users were given a daily prompt to use a specific AI system and subsequently filled out a “Motivation for AI Use” survey [29]. One subcomponent of the survey measured motivation to use AI for instrumental utility, another for entertainment. In the sample of active users, self-reported motivation to use AI for instrumental utility leveled off after the first week, with an increase of approximately 20%. By contrast, self-reported motivation to use AI continued to increase for a longer period and reached a higher overall magnitude, leveling off around 30%.

²<https://futurism.com/artificial-intelligence/ai-twitch-streamer-neuro-sama>

³<https://www.theguardian.com/technology/2025/nov/13/ai-music-spotify-billboard-charts>

⁴<https://www.wsj.com/tech/ai/google-notebooklm-ai-podcast-deep-dive-audio-c30a06b3>

⁵<https://www.newyorker.com/culture/the-weekend-essay/what-if-readers-like-ai-generated-fiction>

More data are needed to provide a fuller picture, but this suggests that people may find it easier—and, potentially, less disappointing—to integrate AI into their routine for entertainment than productivity. At face value, the promise of easing one’s workload with AI is alluring. But this high potential payoff can also lead to disappointment when an AI fails to live up to expectations. The constraints of AI as entertainment are more fungible. This is a case where AI’s difficult-to-eradicate hallucination problem can be a value-add, rather than a liability, when its output is not committed to accurately representing some ground truth aspect of reality. To this end, a report from the UK’s AI Security Institute found that 33% of the 2,028 UK residents sampled had used AI for some non-instrumental purpose—such as companionship, emotional support, or social interaction—in the past year.⁶ It is an industry truism that today’s AI is the worst it will ever be; it is also the least entertaining it will ever be.

3 Entertainment as a business model for AI

Academics tend not to take a position on how a given capability will affect the bottom line of AI companies. Typically, they try to influence the societal impact of technology by regulation [3, 64]. While thoughtful, robust regulation is crucial for the beneficial deployment of AI, it is only one means of shaping technology. Even with thoughtful regulation in place, big corporations are highly motivated to circumvent the limitations imposed by such regulatory strictures and put a large amount of resources toward doing so. But the stick is only one way to influence technological development; the carrot is another. The single largest decision that will affect the future development of AI is how the companies producing it plan to generate the massive return on investment needed to justify the amount of money expended on developing their products.

The amount of money which has recently been invested in GenAI is, to put it mildly, a lot. Recent reporting from the *New York Times* shows that large tech infrastructure companies like Nvidia and Oracle have agreed to provide 20 gigawatts of computing power to OpenAI over the next decade.⁷ This amount of power is equivalent to about 20 nuclear reactors; the estimated cost of providing it is in the ballpark of \$1 trillion. For context, at the time these deals were struck, OpenAI was valued at \$500 billion, while having only made \$1.8 billion in profit during the first half of 2025. And that’s just one company. Global corporate investment in AI is estimated to have reached \$252.3 billion in 2024 alone [47]. Likewise, a McKinsey survey found that 88% of respondents reported GenAI was being used in their business.⁸ The market capitalization of Nvidia—a kind of barometer for the AI industry as a whole—is more than \$4.5 trillion, reflecting an increase of more than 1,350% from five years ago.⁹

From this lens, the enemy of fairness, accountability, and transparency is a business model in which these considerations are illegible—or, worse, antithetical. Academics should take an active interest in exploring how AI can make money in a way that best aligns with societally beneficial considerations. As a point of comparison, plant-based meat alternatives might not exist if people did not seek to fulfill the relevant market demand while also optimizing for animal welfare and sustainable production. The alternative is to leave a vacuum: to let purely corporate interests determine the path to profitability, likely with minimal incentive to consider how downstream effects might be detrimental to society beyond the time horizon of a quarterly earnings report.

It is unclear whether the strategy of developing AI tools for workplace productivity will be sufficient to generate the required level of revenue. If AI revolutionizes a wide range of industries and obviates the need for human cognitive labor across swathes of the economy, then the corporations whose technologies power this economic revolution will

⁶<https://www.aisi.gov.uk/frontier-ai-trends-report> see page 39

⁷<https://www.nytimes.com/2025/10/07/business/dealbook/openai-nvidia-amd-investments-circular.html>

⁸<https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>

⁹<https://uk.finance.yahoo.com/quote/NVDA/history/>

probably be profitable. But a revolution of this magnitude is still hypothetical. If these gains do not materialize, or if the profitability is less than expected, alternative revenue streams will be needed.¹⁰

Taken together, entertainment-based industries offer markets of the size needed to earn back the money invested in AI. Rough estimates for the annualized revenue in relevant industries include: music streaming (\$20 billion¹¹), news (\$85 billion¹²), publishing (\$90 billion¹³), long-form video (e.g., Netflix; \$115 billion¹⁴), short-form video (e.g., TikTok; \$190 billion¹⁵), and interactive media (e.g., video games; \$450 billion¹⁶). While it is difficult to provide a concrete figure for the total annualized revenue of these industries collectively, one estimate forecasts that by 2029 the worldwide media and entertainment industry will grow to \$3.5 trillion USD.¹⁷

Large AI companies are already beginning to pivot to developing AI for entertainment, despite marketing their products as general problem solvers. Recently, OpenAI reached an agreement to license Disney's characters—from Mickey Mouse to Captain America to Yoda—for use in the Sora platform. A selection of these user-prompted videos will then be made available on the Disney+ streaming platform. Disney's CEO, Robert Iger, noted in an official statement that “Technological innovation has continually shaped the evolution of entertainment, bringing with it new ways to create and share great stories with the world.”¹⁸ This is not just about using AI to quickly and cheaply create more movies based on Disney IP; the intention is to create entirely new ways of interactively engaging with characters and settings from Disney's beloved fictional worlds. Beyond Sora, other recent releases from big AI companies are conspicuously geared towards generating entertaining content, such as Google's Nano Banana and Meta's Vibes.

As much as there may be an outcry to prevent AI from running rampant in creative industries, businesses in these industries seem eager to incorporate AI in production pipelines. For example, news outlets—which have been in the entertainment business since the advent of the 24 hour news cycle—are incorporating AI into their reporting practices. The *BBC* uses AI for highly templated content, such as audio weather reports providing hyperlocal forecasts.¹⁹ They are also piloting GenAI in relatively low-stakes settings, such as for generating “bespoke audio bulletins” offering match summaries for five English football clubs. The *Financial Times* is piloting a chatbot where subscribers can query the newspaper's archive of digitized content and receive a bespoke summary based on human reporting.²⁰ Though there are clear concerns about reliability, some aspects of AI are indeed well suited to these tasks: for example, synthesizing the gist of a text passage and adopting the authoritative, abstracted writing style of traditional reportage.

GenAI will be used both to replace existing aspects of the entertainment production pipeline, as well as to expand it. AI represents a next major step in the steady creep of entertainment into daily life as it enlarges the space of possible formats entertainment can take—as with Sora-generated interactions with Disney characters or the *BBC*'s hyperlocal weather reports. This shift will have a significant impact on society in everything from how news is made and consumed, to the influence that fictional characters can have on children and adolescents, to the amount and quality of attention-siphoning content available on our phones. Some reports have begun to suggest that AI's alleged potential for economic disruption has been difficult to convert into actual productivity gains.²¹ By contrast, the adoption of AI as

¹⁰<https://www.noahpinion.blog/p/americas-future-could-hinge-on-whether>

¹¹https://www.ifpi.org/wp-content/uploads/2024/03/GMR2025_SOTI.pdf

¹²<https://www.kenresearch.com/global-newspaper-market>

¹³<https://online.hull.ac.uk/blog/top-trends-transforming-the-future-of-publishing>

¹⁴<https://finance.yahoo.com/news/live-streaming-market-size-surpass-073900911.html>

¹⁵<https://www.researchandmarkets.com/reports/5781298/social-media-market-report>

¹⁶<https://www.statista.com/topics/868/video-games/>

¹⁷<https://www.pwc.com/gx/en/issues/business-model-reinvention/outlook/insights-and-perspectives.html>

¹⁸<https://openai.com/index/disney-sora-agreement/>

¹⁹<https://www.bbc.co.uk/aboutthebbc/reports/policies/what-were-doing-with-ai/>

²⁰https://aboutus.ft.com/press_release/financial-times-launches-first-generative-ai-tool

²¹https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf

entertainment well underway. If companies find it difficult to get traction in selling AI tools for workplace productivity, more pressure will be put on them to generate revenue from entertainment.

4 Current AI evaluations look at the benefits of intelligence, harms of culture

We argue that current AI evaluation techniques tend to examine the benefits of intelligence but the harms of culture. This represents a crucial gap in our approach to evaluating AI: the space in which we would articulate what we want out of AI in the cultural dimension, not just a description of what can go wrong.

The standard procedures in AI evaluation were developed to measure the capabilities of systems engaging in well-defined tasks with clear metrics for success [5, 48, 57]. AI outputs tend to be evaluated in much the same way as one would grade a student exam. There are right and wrong answers, and overall performance is judged by assigning a percentage score denoting how close the system got to some idealized response. This works well for many kinds of problems. For example, it works well when there is an unequivocal ground truth against which to compare an answer (e.g., whether or not a cat is pictured in a given image) or when there's a well-defined value function denoting success (e.g., accrual of points in a video game). But this does not reflect the current landscape of use for GenAI systems. Mainstream approaches in AI evaluation are not well suited for outputs where there is no clear delineation between “right” and “wrong” answers, or when the criteria for success are ambiguous, contested, or highly context dependent [35]. This is a crucial consideration in developing a positive vision of AI for entertainment.

By comparison, social media platforms are widely understood to have had a detrimental societal effect on polarization [1], mental health [40], and other important psychological or social factors [20]. One way to read this is as the downstream effect of an inadequate theory of what makes content worthwhile [38]. Social media algorithms prioritized content based on engagement. This kept users on the platforms for longer, providing more attentional surface area on which ads could be posted. While AI will likely not be subject to the same problem in the same form, if we do not supply a better theory of what makes content meaningful than a comparable set of unintended negative consequences is likely to result. We need a framework for articulating what makes an AI output actively meaningful in the cultural dimension—and not simply free from bias, prejudice, or the reinforcement of an inequitable power structure.

The arts and humanities are full of potential insights about what separates meaningful from mindless cultural content [27, 35, 66]. For instance, one quality which is widely appreciated in the humanities but difficult to articulate within AI is friction [6, 13, 22, 62]. For technologists, friction is a negative attribute of a system. Good design consists in eliminating friction as the user progresses through an interface [51]. But one manifestation in which this impulse goes unchecked is chatbot sycophancy—in which LLMs are fine-tuned to be maximally agreeable, responding with enthusiasm and even a degree of awe no matter what its human interlocutor says [18]. By contrast, in the humanities, friction is widely viewed as a crucial part of the meaning-making process [62]. Resistance is often what makes an experience memorable or transformative, for example when a novel presents a familiar theme or experience in a new light. We do not yet have a language for articulating—let alone quantitatively evaluating—how to calibrate for a quality like dialogic friction in GenAI models [13, 28].

A full survey of evaluation methods is beyond the scope of this paper. Instead, we take as a starting point a recent paper by Burden and colleagues [11], who systematically evaluate different paradigms for AI evaluation. We focus on three of the paradigms they identify: benchmarks measuring task performance, such as the ability to solve a set of domain-specific problems; safety evaluations measuring systemic risk, such as a system's propensity to propagate misinformation; and construct-oriented tasks measuring system capabilities, such as psychological batteries for LLMs.

We take these to be a representative set of some of the primary approaches (though by no means the only ones) by which we judge the performance of contemporary GenAI systems.

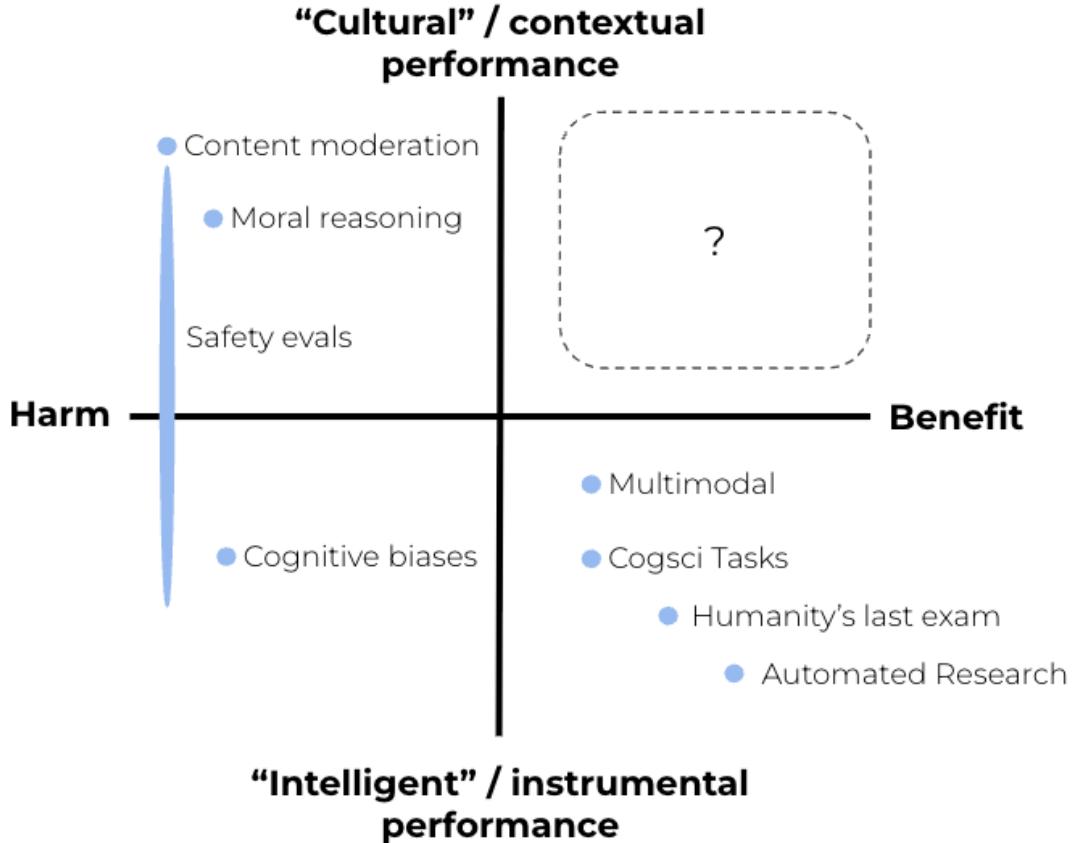


Fig. 1. Contemporary evaluation protocols tend to focus on the benefits of intelligence and the harms of culture. Here we show a rough conceptual sketch organizing a subset of evaluations from a survey by Burden and colleagues (2025) [11] along two axes: harm vs benefit and cultural vs intelligent performance. We distinguish between intelligent performance as the ability to execute some sort of goal-oriented, instrumental process and cultural performance as the ability to abide by contextually relevant norms or accord with subjective human judgments. For example, multimodal test suites [52], cogsci tasks [14], Humanity's Last Exam [53], and benchmarks for automated AI researchers [63] evaluate the potential benefits of intelligence. By contrast, safety evaluations span both culture and intelligence but by definition focus exclusively on harms. This figure is not intended to serve as an authoritative delineation between culture and intelligence or to provide a comprehensive account of extant evaluation protocols. Rather, it meant to illustrate our claim that there is a relative paucity of AI evaluations that look at the benefits of culture.

Based on the evaluations surveyed by Burden and colleagues, there is clearly a rigorous standard for defining intelligence. Some of these take the form of benchmarks. For example, there are standardized tests for low-level cognitive functions in multimodal models [52]. These assess whether AI systems can go beyond static categorization or object detection to perform perceptual tasks like object tracking (e.g., did the ball move positions?), predictions about real-world physics (e.g., which configuration of blocks is more stable?), and relational reasoning (e.g., did the person put the teabag inside the cup or next to it?). Likewise, in the category of safety evaluations, there are robust

tests for whether models can engage in strategic deception, using complex reasoning to deceive human users [60]. In construct-oriented tasks, evaluations such as CogBench test the cognitive functions of LLMs using versions of the same experimental paradigms that would be used to test those of human undergraduates [14]. There are also tests of domain-specific expertise, such as Humanity’s Last Exam [53], as well as applications of intelligence for practical problem solving, such as automating parts of the scientific method [17, 63]. Together, these represent a coherent, diverse, and increasingly robust set of standards for articulating a wide range of benefits and harms of intelligence.

There are also clear and rigorous standards for evaluating potential cultural harms. We define cultural performance as the ability to abide by contextually relevant norms or accord with subjective judgments made by humans. An example of the difference between evaluations of culture versus intelligence would be the assessment of moral reasoning [21] versus cognitive biases [39]. The study of moral reasoning is highly dependent on social context: different cultures have different standards of right and wrong. By definition, moral reasoning must be studied with at least some constraints on the social context in which the findings are meant to apply. By contrast, a cognitive bias such as confirmation bias or base rate neglect is not dependent on social context. It may manifest differently in different cultures, but the baseline judgment—whether someone seeks disconfirmatory evidence or ignores how base rates impact statistical inference—are based on third-party, non-social frames of reference. Assessments of both moral reasoning and cognitive biases test for a deviation from a standardized norm, rather than the achievement of uniquely good performance. Tests of moral reasoning typically are not meant to detect moral genius, as if on the look out for the next Confucius, Imhotep, Mother Theresa, or Dhrukpa Kunley. Rather they tend to place consensus-driven standards for appropriate behavior as the gold standard, then seek to measure violations in which an agent deviates from this baseline. In this sense, they might be better labeled as tasks of immoral reasoning.

Safety evaluations often also have a cultural or contextual component [49, 56]. Any test of potential misalignment necessarily poses the question: alignment to whom? Any answer must be based on a contextually-bound social judgment of value [42, 45, 65, 70]. Likewise, content moderation also requires making contestable judgments about what kind of content should be allowed in a given context of discourse [2, 46]. There are also batteries for assessing social bias or prejudice [68]. Together, these represent an emerging standard for identifying and mitigating potential cultural harms in GenAI systems.

In Figure 1, we plot a subset of the evaluation methodologies surveyed by Burden and colleagues. Our plot reorganizes these evaluations into two dimensions: (1) whether they examine “cultural” or contextual performance versus “intelligent” or instrumental performance; and (2) whether they examine harms or benefits. The surveyed assessments cover a range of potential harms and benefits when it comes to intelligence. But for culture, they tend to focus exclusively on potential harms. This characterization of the landscape of AI evaluations is heuristic and by no means comprehensive. Any generalization about the current state of the art in AI evaluation is likely to be obsolete or incomplete before the ink on the publication is dry. And there are no doubt some interesting and exciting counterexamples [13, 26, 43, 50, 61, 67]. However, our characterization points to an important trend in mainstream approaches to evaluating AI systems: we do not know what we want from AI in the cultural dimension.

5 “Thick Entertainment” as a vehicle for human meaning

In this section, we offer an answer to what we might want out of AI in the cultural dimension. We call it “thick entertainment.” This perspective rests on the idea that the stories we consume, often via entertainment, play a central role in shaping how we interpret the meaning of our experiences [36]. This is the social function that entertainment serves: a communal resource for meaning-making. It gives us narrative templates for how to make sense of our past

experiences, just as it gives us exemplars for structuring future behavior and decisions. At least this is the function entertainment serves at its best. Some entertainment is purely brainrot—the equivalent of candy for the brain—while some entertainment is more nutritious.

Our claim is that as AI is increasingly used for the purposes of entertainment, we need better ways of making such cultural and psychological meaning legible within the broader sociotechnical systems in which they are deployed [38]. As it stands, there is no scale way to measure “meaningfulness” at scale. This is one of the reasons why a theory of evaluation based on engagement prevailed for social media platforms at the expense of something deeper: it is a lot easier to measure. It is important to note that this does not equate to a claim that all entertainment must be Shakespearean in quality or that we are calling for removal of all works that fail to meet some elite standard of literary merit (e.g., runaway best-selling works of romantasy like Rebecca Yarros’s *Fourth Wing*) from library shelves. On the contrary, it is possible for allegedly “lowbrow” culture to be a crucial source of meaning-making, even when it is AI-generated [37]. The difference here is on the margin. At the scale at which AI-generated entertainment will be deployed, it matters whether we have a way of discerning whether an underlying narrative or a symbolic image will challenge, or support, someone’s worldview in a constructive way [30].

People are willing to consume challenging entertainment. For example, one of the most commercially and critically successful TV series of 2025 was Stephen Graham’s *Adolescence*. Across four episodes, the story follows the experiences of a British family as a 13-year-old boy is accused of murder. The show is challenging aesthetically (with each episode being filmed in a single take), and it is challenging thematically (dealing with hard questions about the role and limits of a parent’s ability to influence the direction of their child’s life). The show won five Emmy awards (including outstanding limited series); it was also one of the year’s most viewed shows on Netflix.²² To develop a positive theory of entertainment is to ask how AI-based systems can be designed and used in a rapidly evolving media ecosystem to support thick entertainment that is challenging and constructive, rather than incentivize work that merely provides fodder for the mindless haze of diversion without substance.

Our proposed distinction between “thick” and “thin” entertainment derives from cultural anthropologist Clifford Geertz’s concept of thick description [16, 24, 38]. Geertz argues that cultural meaning can only be codified by providing an account of the social context in which a given symbol, action, ritual, or practice occurs. For example, a swastika has a very different meaning in the West, where it is associated with Nazi Germany, than in the East, where it is a symbol of prosperity and well-being in Hinduism, Jainism, and Buddhism. The framework of thick description is one of the methodological cornerstones of the humanities and qualitative social sciences [59]. It has previously been argued that this framework can help provide an alternative to “thin” metrics—such as views, shares, and comments—for evaluating the content produced and shared in large sociotechnical systems, such as social media, as well as emerging ones based on AI [38].

Our position is not to advocate for a specific definition of thickness or standard for meaningfulness in entertainment. However, we can offer a sketch of potential ways of measuring thickness in entertaining content. These are not meant to be finalized recommendations, but rather to provide a few samples from the space of possibilities. One simple solution is just to avoid thinness: any attempt to use a graph with a line that goes up and to the right as unassailable evidence of progress is misguided. There are going to be multiple lines that matter. More often than not, progress in one dimension comes as a tradeoff to progress in another. It will probably be necessary to calibrate a range of dimensions that are relevant for entertainment. We already mentioned friction, but another example is ambiguity [23]. Thick entertainment

²²<https://www.theguardian.com/tv-and-radio/2025/jun/03/adolescence-tv-series-netflix>

does not reduce a theme or insight to a trivialized tagline; it leaves room for interpretation. At present, ambiguity is another quality that we typically seek to minimize in AI, rather than trying to find a sweet spot.

Finally, a key aspect of how people interpret the meaning of their experience is the difference between judgments in the moment and judgments in retrospect [36]—what Daniel Kahneman called the experiencing self versus remembering self [31]. Engagement metrics such as views or likes reveal the preferences of the experiencing self but do not canvass the perspective of the remembering self. Developing measures that vary temporally could be a useful mechanism for assessing thickness in AI use. Overall, the big mistake here would be choosing to optimize AI systems for something just because it is what we know how to measure.

6 Discussion

AI is typically evaluated in terms of intelligence. While such a perspective on AI’s instrumental utility is valuable, it misses an important use of AI: entertainment. People are already using AI for entertainment. Myriad platforms, some with millions of users, employ AI for activities and tasks that have nothing to do with the kind of supercharged productivity for which the potential of AI is so often lauded. In this paper, we argue that we do not yet have a means of evaluating AS as entertainment. We contend that this will come to be seen as a dramatic oversight as AI upends digital entertainment ecosystems. In the long run, entertainment may even prove to be a more prevalent consumer use case for AI than workplace productivity.

The uptake of AI as entertainment is already underway. For example, emerging data suggests people’s motivation to use AI for entertainment is potentially larger and more durable than for use in goal-directed tasks [12]. We argue that this will offer a key business model for the large corporations building state-of-the-art AI systems. The potential disruption of entertainment-based industries may offer revenue streams of the size needed to justify the huge amount of money currently being invested in GenAI. Without a positive view of what it means to design AI for entertainment, the direction of this technology will be determined mainly by the bottom line of large corporations.

There is a crucial gap in the way AI systems are evaluated. Mainstream approaches to evaluation tend to look at the benefits of intelligence and the harms of culture. Evaluations that look at socially-situated behaviors tend to take a strong position on what a system should not do (e.g., make prejudiced judgments or reinforce systemic bias) but typically do not offer a perspective on what we should want from AI in the cultural dimension [42, 65, 69]. While regulations and ethical guidelines are important, the pursuit of harm mitigation is a relatively weak posture to take in an effort to affect our society’s most powerful sociotechnical system.

We argue that we need a constructive vision of cultural AI—not just a theory of harm minimization, but a positive theory of what beneficial, nutritious entertainment might look like. There will be aspects of this that can leverage what AI is already good at, while also pointing towards areas of weakness where AI needs to be improved or supplemented. Under the right conditions, people are eager to consume challenging narratives en masse: stories with complex, nuanced thematic and aesthetic content. The problem is that we do not have a mechanism for making the meaningfulness of such content legible within AI-driven sociotechnical systems [38]. We contend that this positive vision of AI as entertainment should inspire more debates, discourse, and study in the field of AI. Generative AI is already, and will increasingly be, used for entertainment whether we like it or not. Academics and researchers should take an active role in shaping this application of AI technologies. If we don’t, we may come to regret it.

6.1 Counterarguments

In this section, we consider potential counterarguments to the position offered in this paper. Each of the following subsections begins with our attempt to briefly articulate a strong version of a potential counterargument. We then provide a response to this alternative perspective.

6.1.1 Technologists don't dictate what gets made in Hollywood. There's no need to start just because people are using AI. Creating and analyzing entertaining content is not the purview of science and engineering. It is the purview of artists, creatives, and cultural critics. They are simply separate jobs, and mission creep from engineering into this territory wouldn't solve anything—it would just muddy the waters. The most likely outcome of this proposal is to crowd out creative people we want to support rather than replace.

Our response: We would emphasize that our position is not that we should make AI maximally entertaining. It is also not that one kind of entertainment matters more or is better than another. Instead, our position is that entertainment is a use case which the field of AI is largely unprepared to evaluate. We don't suggest that AI startups should all suddenly pivot to a business model in which they act as subcontractors for Hollywood production or AI slop content farms. Our point is that the seemingly small decisions about how AI systems are designed will have a major impact not just in how people use these tools at work, but about the kind of narratives they hold about the world. We saw how algorithmic manipulation of this harder-to-pin-down notion affected society with social media. If technologists say this isn't their job, then we're failing to learn from our previous mistakes.

6.1.2 It isn't possible for an entertainment-based business model to be morally or ethically superior. It all depends on how it is implemented, governed, and regulated. This claim about ethically aligned business models is illusory. Considerations like ethics, morality, and human flourishing are by nature illegible to any given business model, because a business model succeeds or fails solely on the basis of whether it makes money. The reason academics don't offer suggestions about business models is because there is no way to create a business model that reliably or intentionally elicits the societal benefits we want. We have to look to other avenues, like regulation.

Our response: We accept this as a legitimate position within this space of arguments. Our larger point is that discussions of AI business models should be a part of academic discourse, because over time this will be the consideration which most impacts the kind of technology that corporations produce. However, such a binary view of the effects of one business model over another doesn't seem like the most useful model of reality. For example, one could imagine a coffee company whose value proposition is to provide the cheapest coffee possible, independent of quality or ethical sourcing. By contrast, there are many companies whose explicit value proposition is to deliver a high quality product while treating farmers fairly: they plan to make money specifically by offering a product that benefits as many people as possible in its supply chain. Would the world be a better place if we gave up on this distinction because in the end all coffee companies want to make money?

6.1.3 We don't want AI to be good at making meaningful entertainment. Let's leave that to the humans. Yes, AI will be used to make entertainment. Yes, there is potential to make money here. But the problem is that this will come at the expense of the humans who already make entertainment (i.e., artists). This is a bad outcome; we should be doing everything we can to stave it off, not rush toward it.

Our response: We agree with the motivation behind this concern and agree it's an important area to scrutinize closely. In many creative industries, there are major concerns that entry-level jobs will be or are already being eliminated. In one potential scenario, this will lead to the degradation of the foundational skills of the people with the most influence

in arts and entertainment. In another scenario, this will lead to fewer people—or fewer kinds of people—who will can pursue a career in this field. In another scenario, AI will continually creep up the arts and entertainment production line, being used for more and more aspects of creative output. Perhaps all of these could happen in conjunction; we don't want to see any of them come to fruition. However, we think it's important to draw a distinction between what we would like to see happen within the arts and entertainment industry, and what the powerful forces of technology are realistically going to do to it. Our account is offered the latter spirit. There is too much money at stake and entertainment is too alluring of a market for AI companies not to pursue it. If we really care about protecting artists and creatives, the best way to do that is to have an open and rigorous discussion about AI as entertainment and what we want that to look like—not to pretend like it isn't going to happen.

6.1.4 Part of the issue with misinformation in social media was the selective presentation of facts in echo chambers. Aren't you setting us up to create a comparable problem with isolated islands of hyper-personalized meaning-making? This suggestion that entertainment should be made as meaningful as possible sounds great but in reality would come with a lot of issues. Drawing on the example of social media, it could be argued that social media platforms did in fact become ecosystems for maximally meaningful content. Take the example of conspiracy theories like QAnon. Social media created the conditions for the “meaning” of otherwise banal, unrelated content to be weaved into webs of significance. This seems a lot like what is being suggested with “thick entertainment” (or at least something that could potentially come as an unintended consequence). Beyond that, just as social media has created hyper-personalized information spheres, where different factions of society are presented with entirely different facts, it seems like an effect of this proposal will be to do this in the dimension not of belief and knowledge but of meaning-making and storytelling. This doesn't seem like it would end well.

Our response: This is a major challenge to the basic premise of our argument for thick entertainment. One potential outcome of creating meaningful entertainment that can be tailored to individual perspectives and circumstances is that it could diminish the common reservoir of shared stories across society. This is a bedrock of social organization. It is important not just that people endorse certain stories as meaningful to themselves personally, but that other people share their endorsement. The motivation for our argument is that the composition of this cultural fabric matters, and we need to study how AI is going to affect it. With that in mind, it is nonetheless possible that any attempt to increase the potential positive benefits of AI as entertainment will come with concomitant negative impacts as well. However, on balance, we argue the right move here is to create systems that are sensitive to these kind of impacts, rather than ignore them. This is the thrust of “thick” entertainment: not to say that thickness consists in a specific set of variables, but that we need to look at more than just engagement.

6.1.5 In fact, we do know what we want from cultural AI: pluralism. Any attempt to say what is a “better” form of entertainment will necessarily come at the expense of another equally legitimate form. Trying to police culture in this way is a bad idea. Instead, the best thing we can do is give people the opportunity to make their own choice. Once you start prioritizing one kind of content or another, you're going to marginalize some voices while giving a platform to others. There is no good way to do this in a top-down manner. Even if there was, technologists probably shouldn't be the people response for implementing it.

Our response: No, pluralism is not the only thing we want from cultural AI. It is true that policing culture is usually a bad idea. But the right response is not simply to resign from any perspective on evaluating the quality of content. If a completely free market of culture was possible to achieve, maybe that would indeed be the ideal state. But that's far from the reality of our situation. Content is going to be prioritized in one way or another: by pursuit of profit, by

existing power dynamics, by ancillary effects of algorithmic design. Pluralism is important; it is a crucial consideration in designing systems for the kind of thick entertainment we are proposing. But to say that pluralism is the only thing we can aim for is to abdicate responsibility over the impacts a given system will have. Ultimately it would be a mistake not to have difficult conversations about what kind of entertaining content we think will be beneficial or detrimental for society.

7 Generative AI Usage Statement

The main body of this manuscript was draft without the help of generative AI. However, we used it to draft a first version of the abstract, which we then edited ourselves.

References

- [1] Esma Aïmeur, Sabrine Amri, and Gilles Brassard. 2023. Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining* 13, 1 (2023), 30.
- [2] Canfer Akbulut, Kevin Robinson, Maribeth Rauh, Isabela Albuquerque, Olivia Wiles, Laura Weidinger, Verena Rieser, Yana Hasson, Nahema Marchal, Jason Gabriel, et al. 2024. Century: A Framework and Dataset for Evaluating Historical Contextualisation of Sensitive Images. In *The Thirteenth International Conference on Learning Representations*.
- [3] Markus Anderljung, Joslyn Barnhart, Anton Korinek, Jade Leung, Cullen O’Keefe, Jess Whittlestone, Shahar Avin, Miles Brundage, Justin Bullock, Duncan Cass-Beggs, et al. 2023. Frontier AI Regulation: Managing Emerging Risks to Public Safety. *CoRR* (2023).
- [4] U Anwar, A Saparov, J Rando, D Paleka, M Turpin, P Hase, ES Lubana, E Jenner, S Casper, O Sourbut, et al. 2024. Foundational Challenges in Assuring Alignment and Safety of Large Language Models. *Transactions on Machine Learning Research* (2024).
- [5] Juana Catalina Becerra Sandoval and Felicia S Jing. 2025. Historical Methods for AI Evaluations, Assessments, and Audits. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. 1371–1386.
- [6] Steve Benford, Chris Greenhalgh, Gabriella Giannachi, Brendan Walker, Joe Marshall, and Tom Rodden. 2012. Uncomfortable interactions. In *Proceedings of the sigchi conference on human factors in computing systems*. 2005–2014.
- [7] Yonatan Bisk, Ari Holtzman, Jesse Thomason, Jacob Andreas, Yoshua Bengio, Joyce Chai, Mirella Lapata, Angeliki Lazaridou, Jonathan May, Aleksandr Nisnevich, et al. 2020. Experience Grounds Language. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 8718–8735.
- [8] Levin Brinkmann, Fabian Baumann, Jean-François Bonnefon, Maxime Derex, Thomas F Müller, Anne-Marie Nussberger, Agnieszka Czaplicka, Alberto Acerbi, Thomas L Griffiths, Joseph Henrich, et al. 2023. Machine culture. *Nature Human Behaviour* 7, 11 (2023), 1855–1868.
- [9] Jerome Bruner. 1990. *Acts of Meaning: Four Lectures on Mind and Culture*. Harvard University Press, Cambridge, MA.
- [10] Carolyn Bunting and Huggins R Me. 2025. myself and AI: Understanding and safeguarding children’s use of AI chatbots.
- [11] John Burden, Marko Tešić, Lorenzo Pacchiardi, and José Hernández-Orallo. 2025. Paradigms of AI evaluation: Mapping goals, methodologies and culture. *arXiv preprint arXiv:2502.15620* (2025).
- [12] Mohit Chandra, Javier Hernandez, Gonzalo Ramos, Mahsa Ershadi, Ananya Bhattacharjee, Judith Amores, Ebele Okoli, Ann Paradiso, Shahed Warreth, and Jina Suh. 2025. Longitudinal Study on Social and Emotional Use of AI Conversational Agent. *arXiv preprint arXiv:2504.14112* (2025).
- [13] Zeya Chen and Ruth Schmidt. 2024. Exploring a behavioral model of “positive friction” in human-AI interaction. In *International Conference on Human-Computer Interaction*. Springer, 3–22.
- [14] Julian Coda-Forno, Marcel Binz, Jane X Wang, and Eric Schulz. 2024. CogBench: a large language model walks into a psychology lab. In *Forty-first International Conference on Machine Learning*.
- [15] Katherine M Collins, Ilia Sucholutsky, Umang Bhatt, Kartik Chandra, Lionel Wong, Mina Lee, Cedegao E Zhang, Tan Zhi-Xuan, Mark Ho, Vikash Mansinghka, et al. 2024. Building machines that learn and think with people. *Nature human behaviour* 8, 10 (2024), 1851–1863.
- [16] Joe Edelman, Tan Zhi-Xuan, Ryan Lowe, Oliver Klingefjord, Vincent Wang-Mascianica, Matija Franklin, Ryan Othniel Kearns, Ellie Hain, Atrisha Sarkar, Michiel Bakker, et al. 2025. Full-Stack Alignment: Co-Aligning AI and Institutions with Thick Models of Value. *arXiv preprint arXiv:2512.03399* (2025).
- [17] James Evans and Eamon Duede. 2025. After science. eaec7650 pages.
- [18] Aaron Fanous, Jacob Goldberg, Ank Agarwal, Joanna Lin, Anson Zhou, Sonnet Xu, Vasiliki Bikia, Roxana Daneshjou, and Sanmi Koyejo. 2025. Syceval: Evaluating llm sycophancy. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 8. 893–900.
- [19] Henry Farrell, Alison Gopnik, Cosma Shalizi, and James Evans. 2025. Large AI models are cultural and social technologies. *Science* 387, 6739 (2025), 1153–1156.
- [20] Max Fisher. 2022. *The chaos machine: The inside story of how social media rewired our minds and our world*. Hachette UK.
- [21] Jan-Philipp Fränken, Kanishk Gandhi, Tori Qiu, Ayesha Khawaja, Noah D Goodman, and Tobias Gerstenberg. 2024. Procedural dilemma generation for evaluating moral reasoning in humans and language models. *arXiv preprint arXiv:2404.10975* (2024).

- [22] Rachael Garrett, Patrick Brundell, Simon Castle-Green, Kat Hawkins, Paul Tennent, Feng Zhou, Airi Lampinen, Kristina Höök, and Steve Benford. 2025. Friction in Processual Ethics: Reconfiguring Ethical Relations in Interdisciplinary Research. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [23] William W Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a resource for design. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 233–240.
- [24] Clifford Geertz. 1973. *The Interpretation of Cultures: Selected Essays*. Basic Books, New York, NY.
- [25] Zishan Guo, Renren Jin, Chuang Liu, Yufei Huang, Dan Shi, Linhao Yu, Yan Liu, Jiaxuan Li, Bojian Xiong, Deyi Xiong, et al. 2023. Evaluating large language models: A comprehensive survey. *arXiv preprint arXiv:2310.19736* (2023).
- [26] Sil Hamilton, Matthew Wilkens, and Andrew Piper. 2025. NarraBench: A Comprehensive Framework for Narrative Benchmarking. *arXiv preprint arXiv:2510.09869* (2025).
- [27] Drew Hemment, Cory Kimmers, Ruth Ahnert, Maria Antoniak, Glauco Arbix, Vaishak Belle, Steve Benford, Alexandra Brintrup, Nick Bryan-Kinns, Mercedes Bunz, et al. 2025. Doing AI differently: rethinking the foundations of AI via the humanities. (2025).
- [28] Drew Hemment, Dave Murray-Rust, Vaishak Belle, Ruth Aylett, Matjaz Vidmar, and Frank Broz. 2024. Experiential AI: Between arts and explainable AI. *Leonardo* 57, 3 (2024), 298–306.
- [29] Shunsen Huang, Xiaoxiong Lai, Li Ke, Yajun Li, Huanlei Wang, Xinmei Zhao, Xinran Dai, and Yun Wang. 2024. AI technology panic—is AI dependence bad for mental health? A cross-lagged panel model and the mediating roles of motivations for AI use among adolescents. *Psychology Research and Behavior Management* (2024), 1087–1102.
- [30] Jessica Hullman, Ari Holtzman, and Andrew Gelman. 2023. Artificial intelligence and aesthetic judgment. *arXiv preprint arXiv:2309.12338* (2023).
- [31] Daniel Kahneman and Jason Riis. 2005. Living, and thinking about it: Two perspectives on life. *The science of well-being* 1, 285–304 (2005), 3.
- [32] Atoosa Kasirzadeh and Iason Gabriel. 2025. Characterizing ai agents for alignment and governance. *arXiv preprint arXiv:2504.21848* (2025).
- [33] D Kavanagh. 2019. More friends, less friendly: how social media is changing. *GlobalWebIndex* (2019).
- [34] Lauren Klein, Meredith Martin, André Brock, Maria Antoniak, Melanie Walsh, Jessica Marie Johnson, Lauren Tilton, and David Mimno. 2025. Provocations from the humanities for generative ai research. *arXiv preprint arXiv:2502.19190* (2025).
- [35] Cody Kimmers, Ruth Ahnert, Maria Antoniak, Emmanouil Benetos, Steve Benford, Mercedes Bunz, Baptiste Caramiaux, Shauna Concannon, Martin Disley, James Dobson, et al. 2025. Computational Hermeneutics: Evaluating Generative AI as a Cultural Technology. (2025).
- [36] Cody Kimmers and Simon DeDeo. 2025. Sense-Making, Cultural Scripts, and the Inferential Basis of Meaningful Experience. [doi:10.31234/osf.io/npzqh_v1](https://doi.org/10.31234/osf.io/npzqh_v1)
- [37] Cody Kimmers, Eamon Duede, Julia Gordon, Ari Holtzman, Tess McNulty, Spencer Stewart, Lindsay Thomas, Richard Jean So, and Hoyt Long. 2025. Why Slop Matters. *arXiv:2601.06060* [cs.CY] <https://arxiv.org/abs/2601.06060>
- [38] Cody Kimmers, Drew Hemment, Maria Antoniak, Joel Z Leibo, Hoyt Long, Emily Robinson, and Adam Sobey. 2025. Meaning Is Not A Metric: Using LLMs to make cultural context legible at scale. *arXiv preprint arXiv:2505.23785* (2025).
- [39] Ryan Koo, Minhwa Lee, Vipul Raheja, Jong Inn Park, Zae Myung Kim, and Dongyeop Kang. 2024. Benchmarking cognitive biases in large language models as evaluators. In *Findings of the Association for Computational Linguistics: ACL 2024*. 517–545.
- [40] Ethan Kross, Philippe Verdun, Gal Sheppes, Cory K Costello, John Jonides, and Oscar Ybarra. 2021. Social media and well-being: Pitfalls, progress, and next steps. *Trends in cognitive sciences* 25, 1 (2021), 55–66.
- [41] Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. 2017. Building machines that learn and think like people. *Behavioral and brain sciences* 40 (2017), e253.
- [42] Seth Lazar and Alondra Nelson. 2023. AI safety on whose terms? 138–138 pages.
- [43] Mina Lee, Megha Srivastava, Amelia Hardy, John Thickstun, Esin Durmus, Ashwin Paranjape, Ines Gerard-Ursin, Xiang Lisa Li, Faisal Ladakh, Frieda Rong, et al. 2023. Evaluating Human-Language Model Interaction. *Trans. Mach. Learn. Res.* (2023).
- [44] Robert A. Lee. 2025. *Character AI Statistics 2025: Engagement, Traffic & Creative Trends*. <https://sqmagazine.co.uk/character-ai-statistics/>
- [45] Joel Z Leibo, Alexander Sasha Vezhnevets, William A Cunningham, Sébastien Krier, Manfred Diaz, and Simon Osindero. 2025. Societal and technological progress as sewing an ever-growing, ever-changing, patchy, and polychrome quilt. *arXiv preprint arXiv:2505.05197* (2025).
- [46] Adi Levi, Or Levi, Sardhendu Mishra, and Jonathan Morra. 2025. AI vs. Human Moderators: A Comparative Evaluation of Multimodal LLMs in Content Moderation for Brand Safety. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5965–5973.
- [47] Nestor Maslej, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, et al. 2025. Artificial intelligence index report 2025. *arXiv preprint arXiv:2504.07139* (2025).
- [48] Siddharth Mehrotra, Jin Huang, Xuelong Fu, Roel Dobbe, Clara I Sánchez, and Maarten de Rijke. 2025. Understanding AI Trustworthiness: A Scoping Review of AIES & FAccT Articles. *arXiv preprint arXiv:2510.21293* (2025).
- [49] Alexander Meinke, Bronson Schoen, Jérémie Scheurer, Mikita Balesni, Rusheb Shah, and Marius Hobbahn. 2024. Frontier models are capable of in-context scheming. *arXiv preprint arXiv:2412.04984* (2024).
- [50] Joel Mire, Maria Antoniak, Steven R Wilson, Zexin Ma, Achyutarama R Ganti, Andrew Piper, and Maarten Sap. 2025. Social Story Frames: Contextual Reasoning about Narrative Intent and Reception. *arXiv preprint arXiv:2512.15925* (2025).
- [51] Don Norman. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.
- [52] Viorica Patraucean, Lucas Smaira, Ankush Gupta, Adria Recasens, Larisa Markeeva, Dylan Banarse, Skanda Koppula, Mateusz Malinowski, Yi Yang, Carl Doersch, et al. 2023. Perception test: A diagnostic benchmark for multimodal video models. *Advances in Neural Information Processing Systems*

- 36 (2023), 42748–42761.
- [53] Long Phan, Alice Gatti, Ziwen Han, Nathaniel Li, Josephina Hu, Hugh Zhang, Chen Bo Calvin Zhang, Mohamed Shaaban, John Ling, Sean Shi, et al. 2025. Humanity’s last exam. *arXiv preprint arXiv:2501.14249* (2025).
- [54] Neil Postman. 1985. *Amusing Ourselves to Death: Public Discourse in the Age of Show Business*. Penguin Books, New York.
- [55] Michael B Robb and Supreet Mann. 2025. Talk, trust, and trade-offs: how and why teens use AI companions. *Common Sense Media* (2025).
- [56] Christophe Ropers, David Dale, Prangtip Hansanti, Gabriel Mejia Gonzalez, Ivan Evtimov, Corinne Wong, Christophe Touret, Kristina Pereyra, Seohyun Sonia Kim, Cristian Canton Ferrer, et al. 2024. Towards red teaming in multimodal and multilingual translation. *arXiv preprint arXiv:2401.16247* (2024).
- [57] Stuart Russell, Peter Norvig, and Artificial Intelligence. 1995. A modern approach. *Artificial Intelligence. Prentice-Hall, Englewood Cliffs* 25, 27 (1995), 79–80.
- [58] Traci Ryan, Kelly A Allen, DeLeon L Gray, and Dennis M McInerney. 2017. How social are social media? A review of online social behaviour and connectedness. *Journal of Relationships Research* 8 (2017), e8.
- [59] Maggi Savin-Baden and Claire Howell Major. 2023. *Qualitative research: The essential guide to theory and practice*. Routledge.
- [60] Jérémie Scheurer, Mikita Balesni, and Marius Hobbahn. 2024. Large Language Models can Strategically Deceive their Users when Put Under Pressure. In *ICLR 2024 Workshop on Large Language Model (LLM) Agents*.
- [61] Murray Shanahan, Tara Das, and Robert Thurman. 2025. The Xeno Sutra: Can Meaning and Value be Ascribed to an AI-Generated "Sacred" Text? *arXiv preprint arXiv:2507.20525* (2025).
- [62] Viktor Shklovsky. 2015. Art, as device. *Poetics Today* 36, 3 (2015), 151–174.
- [63] Chenglei Si, Diyi Yang, and Tatsunori Hashimoto. 2024. Can llms generate novel research ideas? a large-scale human study with 100+ nlp researchers. *arXiv preprint arXiv:2409.04109* (2024).
- [64] Katharina Simbeck. 2022. FAccT-Check on AI regulation: Systematic Evaluation of AI Regulation on the Example of the Legislation on the Use of AI in the Public Sector in the German Federal State of Schleswig-Holstein. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 89–96.
- [65] Taylor Sorensen, Jared Moore, Jillian Fisher, Mitchell Gordon, Niloofar Mireshghallah, Christopher Michael Rytting, Andre Ye, Liwei Jiang, Ximing Lu, Nouha Dziri, et al. 2024. A roadmap to pluralistic alignment. *arXiv preprint arXiv:2402.05070* (2024).
- [66] Ted Underwood. 2025. The impact of language models on the humanities and vice versa. *Nature Computational Science* (2025), 1–3.
- [67] Ted Underwood, Laura K Nelson, and Matthew Wilkens. 2025. Can Language Models Represent the Past without Anachronism? *arXiv preprint arXiv:2505.00030* (2025).
- [68] Yuxuan Wan, Wenzuan Wang, Pinjia He, Jiazen Gu, Haonan Bai, and Michael R Lyu. 2023. Biasasker: Measuring the bias in conversational ai system. In *Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 515–527.
- [69] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glæse, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, et al. 2022. Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*. 214–229.
- [70] Chenghao Yang and Ari Holtzman. 2025. How Alignment Shrinks the Generative Horizon. *arXiv preprint arXiv:2506.17871* (2025).
- [71] Jiayi Ye, Chaoran Chen, Yue Huang, Yanfang Ye, Toby Jia-Jun Li, and Xiangliang Zhang. 2025. My Favorite Streamer is an LLM: Discovering, Bonding, and Co-Creating in AI VTuber Fandom. *arXiv preprint arXiv:2509.10427* (2025).