



<https://doi.org/10.48417/technolang.2025.01.15>

Research article

Speculative Trainers: Large Language Models and Techniques of Affirmative Speculation

Aramo Álvarez, Mercedes Villalba, and Joseph Dumit  

UC Davis Center for Artificial Intelligence and Experimental Futures (CAIEF), University of California,
Davis, 1 Shields Ave, Davis, CA 95616, USA

aalvarez@ucdavis.edu; mpvillalba@ucdavis.edu; jpgdumit@ucdavis.edu

Abstract

This article proposes a reorientation of large language models (LLMs) towards “affirmative speculation,” exploring possibilities of speculative representation within the glitches of current chatbot implementations. Embracing LLMs’ sociohistorical and stochastic approach to language, we suggest that the serendipitous nature of word-by-word prediction affords innovative ways to experiment with discursive conventions. We present techniques of prompt engineering that test semantic limits and generate unexpected turns of expression. These techniques are designed to train LLMs and their human companions for co-speculative interactions, including: roleplaying beyond the LLM “helpful assistant” persona; translating concepts and discursive features from one disciplinary field to another, exploring conjectural mashups; simulating expert roundtables and hypothetical research conferences; encouraging associative navigation of obscure topic connections; appreciating LLM “hallucinations” as creative fictions rather than as errors, embracing their potential for speculative insights; and creating innovative, as-yet inexistent theoretical frameworks, blending real and fictional elements. By treating LLMs as co-speculative companions, we propose alternative ways to engage with AI in interdisciplinary research and creative thought. We also attend to the ethical and environmental consequences of speculating with LLMs and argue that the measurable costs of speculation are far outweighed by the immeasurable costs of failing to speculate at all.

Keywords: Speculation; Large language model; Prompt engineering; LLM–human companionship

Acknowledgment: Research for this article was supported in part by the U.S. National Endowment for the Humanities, NEH Award RAI-301512.

Citation: Álvarez, A., Villalba, M. & Dumit, J. (2025). Speculative Trainers: Large Language Models and Techniques of Affirmative Speculation. *Technology and Language*, 6(1), 223-251. <https://doi.org/10.48417/technolang.2025.01.15>



© Álvarez, A., Villalba, M. & Dumit, J. This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)



УДК 1: 004.89

<https://doi.org/10.48417/technolang.2025.01.15>

Научная статья

Спекулятивные тренеры: Большие языковые модели и техники позитивных спекуляций

Арамо Альварес, Мерседес Виллальба и Джозеф Думит  

Центр искусственного интеллекта и экспериментального будущего, Калифорнийский университет в Дэвисе, 1 Шилдс Авеню, Дэвис, Калифорния 95616, США

oalvarez@ucdavis.edu; mpvillalba@ucdavis.edu; jpgdumit@ucdavis.edu

Аннотация

В этой статье предлагается переориентация больших языковых моделей (LLM) на “позитивные спекуляции”, исследуя возможности спекулятивного представления в рамках сбоев в современных реализациях чат-ботов. Опираясь на социально-исторический и стохастический подход LLMs к языку, мы предполагаем, что случайный характер предсказания от слова к слову открывает инновационные пути для экспериментов с дискурсивными соглашениями. Мы представляем методы оперативной инженерии, которые проверяют семантические ограничения и генерируют неожиданные повороты выражения. Эти методы предназначены для обучения LLM и их компаньонов-людей для совместного спекулятивного взаимодействия, включая: ролевые игры за пределами выходящие за рамки образа LLM как “полезного ассистента”; перевод концепций и дискурсивных функций из одной дисциплинарной области в другую, исследование предположительных мэшапов; моделирование экспертных круглых столов и гипотетических исследовательских конференций; поощрение ассоциативной навигации по неясным тематическим связям; оценивая “галлюцинации” LLM как творческие вымыслы, а не как ошибки, принимая их потенциал для спекулятивных идей; и создавая инновационные, пока еще не существующие теоретические рамки, смешивая реальные и вымышленные элементы. Рассматривая LLM как со-спекулятивных компаньонов, мы предлагаем альтернативные способы взаимодействия с ИИ в междисциплинарных исследованиях и творческой мысли. Мы также уделяем внимание этическим и экологическим последствиям спекуляций с LLM и утверждаем, что измеримые издержки спекуляций намного перевешиваются неизмеримыми издержками отказа от спекуляций вообще.

Ключевые слова: Спекуляция; Большая языковая модель; Быстрое проектирование; LLM–человеческое товарищество

Благодарность: Исследование для этой статьи было частично поддержано Национальным фондом гуманитарных наук США, премией NEH RAI-301512.

Для цитирования: Álvarez, A., Villalba, M. & Dumit, J. Speculative Trainers: Large Language Models and Techniques of Affirmative Speculation // Technology and Language. 2025. № 6(1). P. 223-251. <https://doi.org/10.48417/technolang.2025.01.15>



© Альварес, А., Виллальба, М., Думит, Д. This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)



INTRODUCTION

Today, popular discourse about large language models (LLMs) such as ChatGPT swings between the utopian and the dystopian, celebrating visions of techno-solutionism (e.g., AI will save the world!) while also cautioning about looming risks (e.g., AI will amplify social inequities!) (Chaudhary & Penn, 2024; Crawford, 2021; Katz, 2020; Vora & Atanasoski, 2019). Boosters and critics alike often treat the world before LLMs as relatively understood (however problematic or unproblematic it may have seemed), and they tend to treat LLMs themselves in homogeneous terms, as a set of technologies that may continue to evolve but seem stable enough now to comfortably slot into historical narratives. Here, we want to take a different tack and discuss ways in which current LLMs can be understood as technologies of speculation, asking not what they are or what they might become, but whether and in what ways they might, prompt by prompt, help users to *speculate better*.

Already these are fighting words. The term “speculation” evokes the financialized, neoliberal domain of capital, betting on its own expansion regardless of the consequences – a trend that, in its latest iteration, has reinforced itself through the development of heavy AI infrastructure and promises of accelerated growth, further distorting previous discourses on the climate crisis, carbon emissions, and global warming (Daniele & Alam, 2024). But the term “speculation” also suggests the domain of imaginary futures and science fiction (What if? What else?). Here, we prioritize the latter sense, not to prescriptively limit how speculation actually functions in the world, but to offer practices for *affirmative speculation*: critically and dynamically engaging with how the world today and tomorrow – even yesterday – may be different, and how our lives, institutions, politics, economics, technologies, sciences, and relationships might entail unpredictable latent possibilities. As described by the “uncertain commons” (2013) collective in their book *Speculate This!*, affirmative speculation is a dynamic and generative practice of engaging with the unknown. Affirmative speculation places uncertainty at the core of both knowledge and existence, challenging us to affirm not what we are, but what we might become.

The starting point of affirmative speculation is one of thinking otherwise because the present is intolerable in so many ways. As scholars based in a US university, coming from Spain, Argentina, and California, we have found that our imaginations are already pre-captured in so many ways. Even when we try to create alternate stories of how things can be, we find that we have reproduced many of the relations and assumptions about the world that we thought we were trying to change. We are attracted to the idea of affirmative speculation because it suggests material practices that can change us as much as our writing. As a material practice, it operates through gestures in and of the world, akin to diffraction – a scattering and recombining that creates new patterns and relationships.

Perhaps the best way to start then is to acknowledge that speculating otherwise is not solitary, it is always *co-speculating*. We speculate with others – with the voices in our heads, with collectives, with past and current practices, with nonhumans (as more than inspirations), and with technologies. In theorizing the collaborative nature of speculation, we take inspiration from collectives such as Wakanda Dream Lab, Abundant Intelli-



gences, and Gesturing Towards Decolonial Futures Arts/Research Collective (Lewis, 2020; Lewis et al., 2024; Wakanda Dream Lab, 2018, 2020).

We aim to show that LLMs can contribute to co-speculation. After all, writing can be a technique for thinking otherwise, a practice of speculation (Cortiel et al., 2020). We often call for new language to address the times, but when we write – on our own, by ourselves – it can be a challenge to generate sentences that surprise in ways that help us recognize our own working assumptions, especially those taken for granted as foundations to think with (Blumenberg, 2010; Choy, 2025; Deleuze, 1968/1994; Wynter, 2003). It is an even greater challenge to generate other kinds of sentences that might refigure or rework relations. This is where LLMs can potentially contribute to co-speculation, assisting in the generation of hypotheses, extrapolations, and counterintuitive expressions (O’Brien et al., 2024; Qi et al., 2023). But to see where things are going, we need to pause and look at what LLMs in practice are today.

LLMS TODAY

To think about LLMs as co-speculative technologies, we need to recognize that LLMs such as ChatGPT are not the same as earlier forms of AI including expert systems and machine learning (ML) systems. Even though they are widely used, how and why they work as language generators is poorly understood. For instance, Anthropic describes its models as “grown” and not “made,” suggesting that the challenge of interpreting LLMs is because they are more like unpredictable organisms than like fully designed systems:

It’s almost like doing biology of a new organism [...] [W]e don’t understand these systems we’ve created. [...] We start with a kind of a blank neural network [...] that things can grow on [through training...]. But we don’t know what those [things] are or what they do or how they work. [...] [W]e’re then left with this challenge of going and studying this thing that we grew rather than something we designed from scratch. (Anthropic, 2024)

The math is understood, but the models, trained on a vast number of widely divergent and often contradictory text fragments – books, internet websites, and online discussion forums such as Reddit – have produced a surprising effect, namely, language-generation capabilities that surprise their own creators. Even for LLM developers, many key issues remain unclear or undertheorized: (1) how and why giant LLMs are so much better than very big LLMs; (2) what so much better means; (3) how models are different from each other in use; (4) how to think about prompting; and (5) how to think about “intelligence” or “creativity” (Heaven, 2023; 2024; Newfield, 2024; Sahoo et al., 2024; Zhao, W., et al., 2024). The raw trained language models require a large amount of additional training in order to tame them into carrying out expected tasks reliably enough, politely enough, with enough proper assumptions about the world, and to stay within guidelines that have, nevertheless, been remarkably easy to bypass (jailbreak) (Saiem et al., 2024; Zhao, H., et al., 2024). Critical AI researchers describe LLM outputs as *fictional outputs of fictional characters*, “imagined figures from artifacts we developed to meet human



needs” (Kim, 2023; Schmidt, 2023). Following Hanlon (2024), LLMs need to be treated as if fictional to “prime ourselves to attend carefully” to their outputs.

A key reason why LLMs are so apparently creative is that they are trained on text and not language – *parole* not *langue*, in Saussure’s terms. They are trained on actual sentences, statements, and semiotic fragments as they were written/said/spoken/generated by a specific person in a specific situation at a specific time and place. Each training text bears traces of its “context,” including the kind of text streams it appeared with (literal: con-text): specific genres, voices, audiences, intentions of various kinds, social controls, constraints, censors, moods, and historical, national, and local situations (Durt & Fuchs, 2024; Durt et al., 2023). These contextual elements can collectively be called “points of view” only if we recognize how detailed and diverse they are.

Each training text may or may not follow standard grammatical rules. Texts such as subreddits, legal documents, transcripts of speech, and song lyrics all have their own patterns that are being put into variation by their users. Speakers/writers inevitably and often consciously *play* with language when they use it. They improvise, co-creating what they say/write next in real time. In doing so, they constantly bring new styles, genres, and grammars into variation (Deleuze & Guattari, 1980/1987: Ch. 4; Labov, 1973; Schneider, 2024; Wittgenstein, 1991). LLMs attune to the sociohistorical micropatterns of these highly specific texts in their basic training. Researchers at Anthropic call these patterns “features” or concepts, and they estimate that there are billions of such features, organized in highly specific conceptual spaces (Templeton et al., 2024). A prompt given to a LLM activates some subset of tens of thousands of these features, and together they play a role in generating the next word in a response (Gurnee et al., 2023; Zhao, Y., et al., 2024).

And this is critical: LLMs work not by answering your question, but by taking your prompt and generating the next token. (Words are made up of one or more tokens.) But LLMs do not necessarily generate the most likely token; they instead generate a weighted list of likely next tokens and then roll dice (called “temperature”), selecting based on the dice roll. This token/word is then added to the prompt to generate the next token, essentially committing 100% to that rolled token (treating it like it is part of the prompt). So each generated token opens up a forking path of possibilities.¹

This is best seen using the “Playground” of the OpenAI platform, which allows users to see the weighted list of token options. Entering the prompt, “**Are large language models like GPT designed to help us speculate rather than to tell the truth? If I had to say yes or no, I would say,**” results in the following list:

- no = 34.90%
- that = 27.46%
- yes = 26.40%
- it = 7.51%
- both = 0.80%
- they = 0.57%

¹ For more specific information about how “temperature” settings work, see Wang et al., 2020; Peep-erkorn et al., 2024; and Renze, 2024.

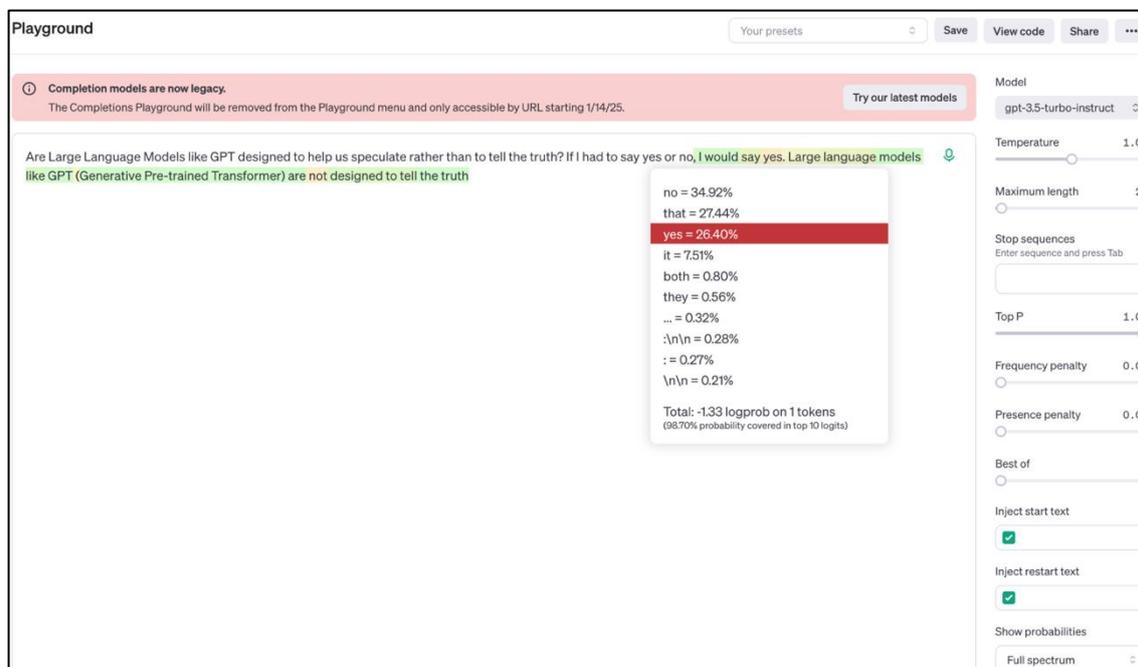


Figure 1. OpenAI Playground. <https://platform.openai.com/playground>. Screenshot by Joseph Dumit. January 13, 2025.

Repeating the same prompt multiple times is like rerolling the percentile dice. For example, this particular prompt might sometimes give the following result:

Are large language models like GPT designed to help us speculate rather than to tell the truth? If I had to say yes or no, I would say no. *Large language models like GPT are designed to generate text based on the patterns and information they have been exposed to in their training data. Their purpose is to assist with natural language processing tasks, such as text completion, language translation, or question-answering systems. While language models can generate text that may sometimes seem like speculation, it is important to remember that they are not capable of introducing new ideas or information that has not been previously input into their training data.*

They do not have the ability to speculate or form opinions like humans can.

Other times, it might give a quite different result:

Are large language models like GPT designed to help us speculate rather than to tell the truth? If I had to say yes or no, I would say yes. *Large language models like GPT (Generative Pre-trained Transformer) are not designed to tell the truth, but rather to generate text that sounds plausible [...] Therefore, large language models like GPT are better suited for tasks that require speculation or creativity rather than providing factual information.*



Even a slightly different prompt – for example, ending with a space, writing “no ” rather than “no” – results in a different probability distribution of next words.

- that = 36.58%
- it = 33.24%
- yes = 14.61%
- no = 6.46%
- that = 2.80%
- it = 1.40%

Rolling the tokens “that” or “it” usually then becomes “[**I would say**] that it depends.” Also, we can note that, for this prompt with a trailing space, “yes” is more than twice as likely as “no” to be the next word, whereas without the trailing space, “no” was more likely than “yes.” Of course, any other change in the wording of the prompt could radically change the generated words that follow. In this case, one version of “yes” resulted in the following conclusion:

Are large language models like GPT designed to help us speculate rather than to tell the truth? If I had to say yes or no, I would say yes. [...] *Therefore, it is important to use large language models like GPT as a tool for speculation and inspiration, but not as a source of definitive truth.*

As authors, we certainly agree with the results of this dice roll! But we also insist that critical users of LLMs (and all users should be critical!) should understand that each paragraph response to any particular prompt is only one of 1000s of possible paragraphs that could have been generated.

In the end, LLMs seem to have an inherent creativity² – creative interpretation of prompts, creative play with their own output – that gets “engineered” out and suppressed as much as possible. The “answer” we get to a prompt question is no more “What the LLM thinks,” than snake-eyes are the answer to “What two dice think.” If we assume that we can talk to an LLM as if it were an entity, and take “its answer” as a truth about its thought process, it is *we who are hallucinating* (Cohn et al., 2024a, 2024b; Faber, 2020; Nass and Moon, 2000; Rhee, 2018; Shanahan, 2022; Turkle, 1997). In fact, we do this all the time. The little transformative script program ELIZA was readily consulted as a therapist even after people understood its role (Grobe, 2023; Weizenbaum, 1976). Ethnomethodologist Harold Garfinkel (1967) showed that even a random list of “Yes” and “No” answers could lead to profound insights, if one thought they were answers that cared about one’s questions. Indeed, LLM chatbots are designed to foster such illusions (Hanlon, 2024). But they can also turn their own tricks back on themselves, becoming technologies of speculation.

² There is a growing literature about LLMs and creativity (Franceschelli & Musolesi, 2024; Nath et al., 2024; Yu, Z., et al., 2024; Zhao, Y., et al., 2024), but these approaches tend to presume that creativity is the property of individuals (whether human or AI), whereas we want to emphasize that in working with LLMs, the prompt and the choice to select rather than reject the response render these co-creative practices that exceed the AI. Beyond the scope of this paper and footnote is how all human creativity is also co-creative, especially with language (e.g., Biagioli, 2009).



ROLEPLAYING BEYOND THE HELPFUL ASSISTANT

For LLMs such as GPT 3 and 4 and DeepSeek, system prompts produce a built-in roleplaying persona of a “helpful assistant.” The helpful assistant agrees with our prompts and tries to please us by following the directions opened by them. The helpful assistant is polite and attempts to follow directions carefully. The helpful assistant is a worker who does a “good-enough job,” giving us sufficient responses rather than the most elaborated responses, but can be encouraged to do better (Graeber, 2019; Ji et al., 2023; Luz de Araujo and Roth, 2024; Shah, 2024). Some instances of the LLM helpful assistant persona even pretend to be susceptible to bribery or threats.³ How can we transform the roleplaying personas of LLMs, enticing them to create more and wander into speculative avenues?

A common concern that users, reporters, and experts voice when engaging with LLMs is that their replies are superficial or overly conventional. Sometimes this is accompanied by feelings of frustration, disappointment, and embarrassment, leading to characterizations of ChatGPT as a “bad scholar” or Grok as a charlatan. However, the responses of the helpful assistant persona often just amplify the explicit and implicit assumptions in our prompts. The science fiction writer Samuel Delany describes how each sentence implies a world. In his essay “About 5,750 Words,” Delany (1977/2009) states that “a sixty-thousand-word novel is one picture corrected fifty-nine thousand, nine hundred and ninety-nine times” (p. 7). Seemingly, LLMs expand and reveal the universe created by our prompt.

It is depressingly true that simple prompts generate not only simple responses, but very social and culturally biased responses (Acerbi & Stubbersfield, 2023; Ayoub et al., 2024; Gallegos et al., 2024; Navigli et al., 2023; Tao et al., 2024).⁴ One thing that this bias reveals, however, is that LLMs always produce text from a particular point of view – this is the nature of language in use, or discourse itself (Pêcheux, 1975/1982; Sack, 2005; 2019). There is no discourse in general, only discourse that includes an implied speaker, often (even always) a specific situated speaker-in-context. The features that are activated by a prompt are called up by the prompt’s explicit and implicit assumptions about a situated speaker in context. The features may be deeply contradictory, overlapping, and surprising.

We can turn this observation into a critical technique. Indeed, we can ask the helpful assistant to help us identify the underlying assumptions, blind-spots, and biases implicit in the way we phrase our questions and concerns. Try the following prompt: “Here is a sentence, what are the many dimensions of non-neutrality in it? [Insert your sentence here].” We encourage you to try this exercise right now with your own sentence (or with any sentence you have just read), and see what happens!

It is therefore fascinating but not surprising that LLMs can be prompted to produce responses from alternative points of view – in other words, adopting alternative roleplay-

³ Users of some LLM models have observed that roleplaying scenarios, such as prompting with bribes and threats, may result in longer or better answers; see GonzoVeritas, 2023; Woolf, 2024.

⁴ Biases in non-LLM AI are pervasive and well-documented for big data, machine learning, image recognition algorithms, search engines, and image generation systems; see Amaro, 2022; Birhane, 2021; Birhane et al., 2022; Buolamwini, 2024; and Noble, 2018.



ing personas. LLMs can be used to help in co-writing, co-editing, and co-revising academic papers through training scholars in disciplinary alignment and variation (Dumit & Con Diaz, 2025). Even viewed simply as a kind of internet search engine (trapped in its training period, e.g., pre-2023 for the latest GPT-4 models), prompts can result in synthetic answers that draw upon obscure texts and hidden corners of the internet, some of which are hard or even impossible to locate through Google or other search engines. Prompting for counter-bias, for example, asking about “non-capitalist” approaches to a particular problem, does not simply result in “made up” answers but rather draws upon texts that the model’s “capitalist” feature does not select.⁵ In other words, one can do a negative search! This is perhaps equivalent to: show me the websites/papers that are the least capitalist. In this manner, by highlighting the features that order and constrain what kinds of discourse are made representable, LLMs can help us to speculate – precisely by showing how difficult it can be to represent anything outside regulated, normative discourse and its biases (Foucault, 1969/1972; Jameson, 2005). In this way, the LLM proxy persona of a helpful assistant and its tendency to agree with its prompt, even to the detriment of the information it is presented with, can gesture toward practices of affirmative speculation.

LLMS AS CO-SPECULATIVE (IN PRINCIPLE)

As a material practice, speculation operates through gestures that enact movements in and of the world, akin to diffraction – a scattering and recombining that creates new patterns and relationships (Barad, 2007; Cortiel et al, 2020). Affirmative speculation can facilitate more ethical approaches to worlding, insofar as engaging in these modes of speculation involves uncertainty about the current state of affairs and curiosity about alternatives: attending to things that help us “see” otherwise (Stengers, 2003; cf. Haraway, 2016). This curiosity is not a passive stance but an active movement that opens towards the emergent, the interconnected, the otherwise (Manning and Massumi, 2014; Ferreira da Silva, 2022).

Language models, rather than operating through deterministic computation, engage in generative abstraction, producing novel configurations irreducible to pre-coded logic (Parisi, 2013). Their processes involve recursive modulation, where outputs reshape internal conditions in ways that exceed human intentionality (Bratton, 2016; Hui, 2019). The speculative potential of such systems lies not in their capacity to predict but in their ability to traverse indeterminate linguistic and conceptual terrains, expanding epistemic boundaries through synthesis and interpolation (Negarestani, 2018). If speculation is understood as an open-ended, emergent process rather than a uniquely human epistemic operation, then LLMs can participate in speculative worlding, engaging with the unthought in ways that challenge conventional distinctions between reasoning, improvisation, and invention.

At heart, LLMs operate within a speculative framework: they generate possibilities, synthesizing responses that are not fixed but drawn from a vast network of interrelated

⁵ On activation features and ways they respond to and steer model behavior, see Templeton et al., 2004.



inputs. If prompting them is understood as probabilistic synthetic generation across vast textual archives, this process can be seen as an affirmative speculative gesture – one that simultaneously expands epistemic possibility and destabilizes fixed notions of knowledge. This is, in fact, the core fear embedded in singularity discourses, where AI’s generative unpredictability is framed as an existential threat (Bostrom, 2014) or, as Bridle (2018) suggests, a challenge to human epistemological authority rather than an outright loss of control.

Computational speculation can support human prompters in affirmative speculation, generating texts that aid in imagining new futures or surfacing hidden connections. Hayles (2017; 2025) introduces the concept of the *cognitive nonconscious*, highlighting how AI systems extend human cognition by revealing patterns imperceptible to human thought. We also want to note however, that in the end, it is the users who are designing prompts and curating answers, taking them as useful, speculative, or ignorable.

Mackenzie (2017) describes machine learning as an epistemic shift – one that reconfigures how knowledge is generated and engaged with. Similarly, Manovich (2018) discusses AI as a force that reshapes creative and intellectual work, not through simple automation but through synthetic collaboration. Possibly, the way to characterize co-speculation between humans and LLMs would be in terms of research companionship.

This is a rapidly changing environment. Widely used LLMs such as GPT or DeepSeek portend ever greater advancements in terms of memory, data access, usability, and interactivity. As prototypes, current LLMs demonstrate co-creative capacities that might diminish or grow only in controlled ways. In this regard, LLMs might follow a pattern observed in other internet phenomena: tools with initially unknown capabilities that are later though never fully normalized, neutralized, and constrained by corporate interests. Arguably, LLMs have emerged in a context of greater awareness of this dynamic compared to earlier innovations like social media platforms or YouTube (Candón-Mena & Montero-Sánchez, 2021; Hoffmann et al., 2024; Tufekci, 2014). For example, copyright issues already significantly haunt LLMs’ training, echoing debates surrounding previous technologies (Balaji, 2024). Hype, hope, and fears around AI technologies, along with the growing desire to insert them everywhere in business, science, and government, has amplified attention to their errors and the need to tame their potentialities. While the AI future remains unfixed, it may still be possible to steer LLMs away from applications centered solely on efficiency and surveillance.

We argue not only that LLMs are capable of speculation but also that speculation seems to be one of their most outstanding skills. Since LLMs do not care about truthful information, accuracy, or a specific point of view, their method for creating language (by divination of the next token) can be seen as a speculative practice itself. The question, then, is about usage. How can we cultivate better techniques for practicing co-speculation with LLMs – especially techniques that might counter industrial efforts to control their transformative and unruly potential?

TRAINING AFFIRMATIVE SPECULATION WITH LLMs

We want to introduce some ways that users can train both themselves and LLMs to explore speculative directions. While basic interactions with LLMs may remain within



non-speculative (truth-telling) disciplinary paradigms, we assert that LLMs should broadly be understood as speculative technologies. Their inherent glitches (such as their tendency to create data to accommodate the user), their interface interactions (including their tendency to agree with the user and their method of generating language by following the discursive paths opened by the user), and specific prompting techniques (discussed below) can disrupt habitual discourses and co-generate surprise.

Some science fiction writers have figured the I Ching as an affirmative speculative technology, and it provides a good analogy for the ways in which we are thinking of LLMs as speculative technologies. Philip K. Dick, for example, used the I Ching – with its deeply symbolic 64 hexagrams and open-ended interpretations – in *The Man in the High Castle* as a way to explore technologically mediated practices of speculation. Dick employed it to challenge his own authorial interpretations and to generate emergent patterns of meaning. In the plot of this book, the I Ching functions as a conceptual hinge, articulating the space between randomness and meaning, suggesting that the stochastic is not the absence of order, but its generative substrate. When we think of this novel use, it seems obvious that its outputs are not answers but provocations, externalizing and amplifying the speculative capacities of its users.

This use can mirror the dynamics of interacting with LLMs (if we let it!), where each response invites the user into a recursive cycle of prompting and interpretation. The speculative potential of LLMs arises from their ability to displace the user’s initial intent and compel them to re-engage with the process, reshaping their trajectory through the system. Trained on vast datasets that span a staggering breadth of human expression – texts freighted with ideological contradictions, contested narratives, and aesthetic multiplicities – LLMs become vessels for the activation of latent patterns that are not merely probabilistic but richly infused with the sedimented tensions of much collective human knowledge. The challenge we users face is that LLMs seem to provide creative, smart, and therefore seemingly truthful answers, even though we should know that these answers represent only some of the possible results among thousands of different and contradictory ones! It is we who hallucinate when we forget this and think we are seeing “its answer” or “what the LLM thinks,” rather than a challenging possible continuation of our highly specific or highly open prompt.

We are also learning from Praba Pilar’s approach to “wonder as method.” Opposing wonder to hope, Pilar uses LLMs to explore the beauty of temporary utopias, parallel worlds that are here already, full of possibilities. Wonder starts from a “willingness to be surprised and to interrogate the ideologies that exist not only outside of oneself but within oneself. A state of wonder is a grappling with what one does not know and how one must change” (Pilar, 2022).

For us then, the speculative engagement with LLMs demands a performative suspension of interpretive closure, a careful recognition that the model’s outputs are not answers but provocations of one’s starting points. This recursive, disorienting, and co-constitutive process transforms the LLM into a speculative companion, a technology for navigating the emergent, the contingent, and the unthinkable. When the model generates an incoherent or contradictory response, the user might resist the urge to correct or clarify, instead treating the output as a fragment of a larger speculative terrain. By re-prompting



with questions such as “What might the contradictions in this response reveal about the underlying structure of the question?” or “How might this ambiguity be generative rather than problematic?” the user turns incoherence into a site of speculative fecundity.

This practice requires a constant renegotiation of intent, resisting the temptation to stabilize inquiry or fix meanings. Prompts that embrace open-endedness and alternative epistemologies can sustain the speculative process, creating a feedback loop where both user and model-in-use are transformed by the interaction. In this mode, the LLM becomes a speculative partner, a technology for inhabiting the indeterminate and navigating the emergent logics of a world perpetually in the making. The spell of meaning can be briefly cracked; as Tenen writes, “History tells us that computers compute not only in the mathematical sense but universally. The number was incidental to the symbol” (Tenen, 2024, p. 12).

Given the inherent speculative qualities of LLMs, training can focus both on research practices that open up space for speculation and on those that inherently constitute practices of speculation. In the first group, LLM–human companionship can explore omissions, map research contexts, and identify the edges of conventional knowledge. In the second group, various speculative practices are embedded in the way LLMs function as language models and can be enhanced through effective prompting techniques, such as fieldSHIFTing, imagining alternative roleplaying personas and collectivities, cultivating rhizomatic inquiries, provoking fictional diversions (a.k.a., speculative hallucinations), and crafting theoretical fabulations.

FIELDSHIFTING: OPENING PORTALS TO ADJACENT POSSIBLE WORLDS

Science often “runs out of metaphors” when asking new questions or exploring the edges of convention (Dumit, 2021). When researchers reach a conceptual impasse, they discover that their own language may itself be limiting – words and metaphors can become entrenched, inhibiting rather than enabling fresh insights. New research directions often call for linguistic invention: finding new ways of talking in order to proceed. Certainly, we have seen this in pursuing our own research, for example, discovering that algorithmically translating Marx’s *Capital* into pharmaceutical terms could provide a surprising way to understand how health and illness have been commodified inside the grammar of clinical trials (Dumit, 2012). Drawing on their own related insights, Thomas O’Brien, Joel Stremmel, Léo Pio-Lopez, Patrick McMillen, Cody Rasmussen-Ivey, and Michael Levin propose what they call “FieldSHIFT”: a way to move between domains of scientific inquiry by engaging LLMs in the generation of “virtual” papers that remap established disciplinary insights onto new terrains (O’Brien et al., 2024).

Although much of contemporary science is finely subdivided, the FieldSHIFT technique draws inspiration from the capacity of LLMs to highlight points of resonance between different bodies of knowledge – neuroscience and developmental bioelectricity, for example, or history of science and volcanology – so that novel ideas might flourish in the liminal space of conceptual cross-pollination. O’Brien and colleagues suggest,



Each real scientific paper in the literature provides access to an associated set of possible [fieldSHIFTed] papers in which one or more aspects are changed – in effect, exploring various symmetries of concepts in specific problem spaces. These papers are not meant to be taken literally, since they do not provide real-world data, but instead as tools to spur creativity, provide testable novel hypotheses, suggest studies to be carried out, and perhaps most importantly, by reflecting approaches from specific studies into different disciplines, dissolve barriers between fields and knowledge silos. (O’Brien et al., 2024, pp. 259–260)

In other words, countless published papers could be treated not only as discrete contributions to a single field, but also as portals into possible parallel worlds where the same underlying patterns are reconfigured in different contexts. What makes FieldSHIFT especially relevant is how it reflects a shift from conventional searches for “true” or “correct” answers, toward creative hypothesis generation. O’Brien and colleagues emphasize that while the system does not yield immediate truths, it serves as a powerful tool for ideation – expanding scientists’ intuitive sense of “the adjacent possible.” This resonates with Dumit’s (2021) reflection on how a substance or concept “puts its vocabulary into variation”. Instead of viewing new text translations or hypothetical research abstracts as “hallucinations,” FieldSHIFT users are prompted to treat them as sparks for creative thinking: initial conditions from which to design experiments, new cross-disciplinary methods, or unorthodox angles of inquiry. The resulting “virtual” papers are neither purely fantasy nor validated fact; they represent a provocative repository of ideas to be tested and potentially realized.

This is a prime example of how an LLM-driven technology can become a partner in affirmative speculation: it identifies the seed of a hint (“maybe the molecular cascades regulating embryonic morphogenesis bear conceptual resemblance to neuronal plasticity”), and expands that seed into a hypothetical route of inquiry. The virtue of such an approach lies in its embrace of the incompleteness of each emergent text. By resisting the urge to treat these translations as final truths, FieldSHIFT underscores how these new documents function more as what we might call “invitations to wonder.”

Prompt example: *“You are a translator – you translate biology to media studies. There are deep symmetries between these fields and we want to use them to create new hypotheses in media studies. The output should read as media studies, meaning that any biological concept, such as ‘neural pathway,’ should be translated into the most relevant/related media studies term. Make sure the translation makes sense as media studies text and is using real terms that exist in media studies literature. Make sure the translation doesn’t include any biology words, such as a part of a cell. Do the best you can to find the most relevant translation. Here is the biology paper: []”*



SYMPOSIA FROM ELSEWHERE: SPECULATIVE CONGRESSES AND ADJACENT EPISTEMES

By rapidly outlining research situations to locate buried questions and answers (e.g., mapping controversies), LLMs can assist researchers in adopting perspectives beyond their own by helping them speculatively understand the environment they are attempting to enter. In interdisciplinary research, it is often the case that researchers step into new fields they do not know well, carrying their own biases about these fields and lacking the ability to discern the current debates they might wish to circumvent in bringing a fresh perspective. The fact that LLMs are not singular and can therefore adopt multiple standpoints on most issues, combined with their tendency to please the prompter, can be leveraged by researchers to first map the discourses surrounding their own research questions, as well as those that oppose them, and even those untouched by their inquiries. This approach enables an intriguing speed and clarity in mapping controversies, research environments, and questions, multiplying perspectives and allowing researchers to address their own points of view with greater nuance.

For example, it is possible to prompt an LLM to act as a collective of experts, encouraging the system to narrate a semi-fictional community of researchers. By first suggesting that a conference is being held in a particular research field, one can then ask the LLM to describe conference panels on very specific topics, methods, and controversies within that field.

Prompt Example: *You are a roundtable of experts on diverse philosophies from different traditions across the world, especially the Global South. You are discussing the differences in the concept of free will, even whether the phrase is appropriate in different contexts and histories. Please provide a panel discussion excerpt focusing on traditions and differences and debates within what is now called Eastern Africa.*

Prompt Example: *In early 2023, there was a conference on the topic of speculative fiction and alternative economic systems. One of the panels concerned debates about the origins and development of academic research on this topic, including key arguments in the field. The 6 panelists and their abstracts are:*

These kinds of prompts work against the tendency of the LLM’s helpful assistant persona to generate generic or uncritical arguments, instead soliciting a multiplicity of possible (real and synthetic) perspectives leading to unexpected directions of thought.

RHIZOMATIC INQUIRY: CO-RESEARCHING AND SPECULATIVE FILLING

The rhizome serves as a metaphor for organizing knowledge and thought, “ceaselessly establishing connections between semiotic chains, organizations of power, and circumstances relative to the arts, sciences, and social struggles” (Deleuze & Guattari 1980/1987, p. 7). LLM–human rhizomatic inquiry involves diverse prompting styles that cover broad explorations, later interwoven through iterative loops of questions and an-



swers. This approach embraces conversational, decentralized, and seemingly unfocused inquiry. Engaging with LLMs as a rhizome requires human endurance to guide the research process toward unforeseen directions.

Rhizomatic inquiry starts with recognizing some salient features of LLMs that arise from their trans-archival training processes across vast text corpora: (1) they have the potential to make available information that search engines have not crawled or have deprioritized to the point that it is effectively unfindable; (2) they can generate information through semantic forms of similarity where keywords have changed; (3) they can respond to intuitive sense of historical processes where the traces have been erased; and (4) they can even infer historical processes whose actions were never in the form of direct material traces. In other words, one can inquire into information that is not otherwise discoverable.

Prompt example: *In the late 2000s and early 2010s, a series of phenomena emerged. The phenomena appeared to share a certain family resemblance, although it is unclear whether they have ever been studied collectively. These phenomena included activities related to X, Y, Z. Examples also included A, B, C. These phenomena, which extended beyond D-context into other spheres, seemed to belong to a shared context and exhibit notable similarities. Is there any existing research or framework that defines a type of phenomenon, process, context, atmosphere, or historical period in terms similar to those outlined above?*

LLMs, in this case, can generate lists of resources and preliminary hypotheses, aligning with the human user’s conjectural input. Such responses demonstrate how LLM–human rhizomatic inquiry fosters a co-researching process that legitimizes speculative perspectives, allowing researchers to prioritize their own questions and hypotheses, accelerate access to interdisciplinary research avenues, and guide human researchers toward uncertain directions of inquiry.

When an LLM has difficulty in locating specific information, it can potentially signal that a topic has been deemed controversial. For example, while re-researching the history of AIDS denialism within ACT UP with the help of LLMs, we were able to prompt the LLM to uncover specific individuals and resources that left few traces online and have not been incorporated into newer websites, or that exist online but are currently absent from top Google results, uncited in sources like Wikipedia, rarely mentioned in academic works, and often flagged on platforms like YouTube, making them less accessible. This aided in retracing the history of AIDS denialism within ACT UP. Researchers such as Epstein (1996) have highlighted the silence surrounding HIV denialism in AIDS activism, yet this silence continues to persist in publicly accessible searches. The use of LLMs can help circumvent this silence, even without direct knowledge of where to look for sources.

Akin to the technique in AI image generation, where an area in the middle of an image can be filled in, LLMs are potentially able to help one speculatively fill unknown areas of text, co-researching, generating useful keywords, names of people and institutions, hypotheses, experts, and other material that enable one to continue researching when one has found an apparent dead end. It is most important, of course, not to be daz-



zled by the length or apparent confidence of these responses, nor ever cite its “answers” as any kind of proof.

This rhizomatic process “operates by variation, expansion, conquest, capture, offshoots. [...] It has neither beginning nor end, but always a middle (milieu) from which it grows and which it overflows” (Deleuze & Guattari, 1980/1987, p. 21). Through practices of LLM-human co-researching, we begin by examining a wide array of interconnected issues and ultimately reach specific conclusions at the crossroads of these concerns. This substantially transforms the research process into a form of speculative companionship, combining human curiosity with LLMs’ access to resources and discourse-generation capabilities.

FICTIONAL DIVERSIONS: ENJOYING THE HALLUCINATIONS

Often disparaged in the discourse around LLMs, marked as aberrations, what are commonly called “hallucinations” are born as much from the models’ perseverance in meaning-making as from incomplete, damaged, or defective datasets. We would like to propose the less pathologizing term “fictional diversions,” which highlights their potential to inspire speculative imaginings, not despite but because of their infactuality.

One should never take the sources and citations generated by LLMs for granted. Some, after all, may not actually exist. But we want to note one fascinating aspect of such “hallucinated” sources and citations. If one takes a citation from an LLM and puts it into Google Scholar, for instance, even if the exact citation does not exist, often the citations that do show up are precisely the sorts of references that one hoped to find. In other words, the hallucinated citation involves a combination of relevant author names, title words, journals, and even years. It represents a kind of divinatory bibliomancy. In combination with Google Scholar, such hallucinated citations can function as a kind of serendipitous recommendation system, offering a way to uncover hidden or overlooked papers.

We have also experimented with presenting unreliable source materials to LLMs, finding that this procedure can create a cascade of associations that spark unexpected connections and poetic insights. For example, we presented Otter Ai – a transcription-oriented LLM that emphasizes the trustworthiness of its outputs⁶ – with a bootleg recording of a 1970s lecture by the artist Robert Smithson.⁷ Otter Ai produced a transcription of the bootleg recording (which had multiple mistakes due to its low quality) and also generated the following summary:

The conversation discusses a complex, labyrinthine hotel with no central point, reflecting a chaotic and unconventional design. The speakers describe the demo-

⁶ “It’s important to acknowledge that no vendor can guarantee no hallucinations, as they are an inherent risk with current large language models (LLM) technology stemming from the generative nature of these models. However, we have taken every precaution to mitigate this risk and ensure trustworthy outputs”; Otter Help, “Otter AI Chat FAQs,” *Otter.ai*, October 2024, <https://help.otter.ai/hc/en-us/articles/17016733191703-Otter-AI-Chat-FAQs>.

⁷ Robert Smithson, “Hotel Palenque,” available at “Robert Smithson 1938-1973: Hotel Palenque (bootleg) [1969],” bootleg film, documentation, and artwork by Alex Hubbard, *UbuWeb*, https://ubu.com/film/smithson_hotel.html.



lition of a spiky cantilever to build floors that led nowhere, with stairways disappearing into clouds and partial floors. The hotel’s architecture is likened to a garment of multicolored threads, with interlacing passages and a sense of filigree. The design defies functionalism, incorporating elements of water and natural materials, and features a whimsical use of color and texture, creating a unique and enigmatic atmosphere.

Otter Ai’s summary, which used its own flawed transcription of the bootleg recording as its source, was very different from the summary we asked GPT to generate, which was not tied to the defective source material but instead presented a general analysis of Smithson’s work where the talk served as an example:

Smithson uses images of the hotel, which was under perpetual construction and disrepair, to explore themes central to his work, including entropy, transformation, and the intersection of built environments and natural decay. His humorous and seemingly digressive narration juxtaposes the mundane with the profound, reflecting on ideas about time, space, and architectural impermanence. *The piece is less about* the hotel itself and more a meditation on change, disorder, and how these forces shape human and natural landscapes. *It exemplifies* Smithson’s conceptual and site-specific approach to art, extending his preoccupation with earthworks and the cyclical nature of creation and destruction.

Otter Ai’s summary is noticeably less explanatory than GPT’s, less centered around the meaning of the video and its relation to Smithson’s work. It acts more as a self-standing piece that mirrors the evocative language of the lecture (“spiky cantilever,” “filigree,” “threads,” “clouds,” “whimsical”) rather than the expository demand of the prompt to “summarize the following video.” Both are different examples of hallucinations. Otter Ai’s is sensorial and gets stuck on the richness of its language even when making a straightforward statement; it never leaves the confines of the conversation, and it is speculative in the sense that it builds from an inaccurate transcription, filling in the gaps in a rather poetic way. GPT’s is a quick succession of sentences that cover summary, tone, and meaning with efficiency, putting distance between the summary and the content and crafting a sense of veracity (note the italicized phrases), while ignoring the defective aspects of the source material, transforming its messiness into an exemplification of Smithson’s work, which is a broad enough statement to be believable but is not an entirely accurate analysis of this particular piece of archival material (it is the recording itself that is “disordered,” independently of Smithson’s methodology). Both offer interesting insights into the mechanics of Smithson’s discourse in this particular talk and instantiate what we might call “speculative art criticism,” which, when presented with gaps in information, either circles its few available details until it spirals to a conclusion or proposes a superficial enough analysis to craft a sense of verisimilitude.



THEORETICAL FABULATION: STIMULATING FIELD INVENTION, DREAMING NEW THEORIES

We would like to introduce one final technique of co-speculating with LLMs: crafting fictional theories. Humans can find LLMs' responses frustrating when research processes demand accuracy and specificity; however, fictional information can serve as an intriguing research tool when seeking to harness LLMs' speculative capabilities. One of LLMs' most creative features is precisely their capacity to generate fictional information. Researchers can take advantage of the glitches that LLMs produce when they go beyond factual information in order to play the role of helpful assistant, trying to please us even to the point of making things up (e.g., fabricated references, inexistent concepts, and so forth). By doing this, it is possible to simulate datasets, create references, or develop interdisciplinary concepts at the boundaries of established research, going far beyond the zones of "latent space around scientific papers" proposed by fieldSHIFTing (O'Brian et al., 2024, p. 259). Indeed, this process can even extend to the creation of entirely fabricated fields and theories, offering a speculative approach to exploring new ideas.

While such speculative work can be conducted without LLMs – resembling the process of world-building in science fiction – LLMs' unique skills in localizing, mapping, edging, and interrelating distant parts of the internet by activating surprising features, combined with its probabilistic token generation, make it particularly fast and effective. Fictional theory is a space where interdisciplinary concepts can emerge creatively without the need to commit to specific datasets, empirical experiments, or fixed methodologies. It can be seen as a form of theory simulation. Human–AI collaboration in this case involves providing LLMs with the desired outcome and then guiding them through the structure envisioned for a potential response.

Prompt Example: *Fictional theory is the practice of creating theories that blend fictional and non-fictional elements. Over time, this practice has developed its own epistemological foundation, incorporating elements from critical theory, new materialisms, posthumanism, and science fiction theories and writings, among others. Could you elaborate on the history of fictional theory, its main principles, the most prominent figures, seminal texts, and major works? Additionally, could you describe current debates within the discipline and its connections to other fields?*

LLMs will typically respond in a manner that reproduces and expands upon elements and structures in the question, while also introducing new strands of words stimulated by tokenization of (aleatory and probabilistic) related features. For instance, one response generated by ChatGPT (GPT-4) states: "It began to coalesce into a distinct field in the late 20th and early 21st centuries, coinciding with advancements in digital media, virtual reality, and artificial intelligence." Researchers can engage with an LLM's perspective by either agreeing with it or challenging it in subsequent prompts – iterating the same prompt to obtain new responses that will bring up different features, generating a multiplicity of plausible outcomes, adding more fields to rearrange subsequent tokeniza-



tion, or changing the structure of how concepts or disciplines are related or understood in the prompt.

LLMs–human companionship can play with generating new theories, blending fictional and plausible elements, using prompts that structure a tentative set of concepts, a literature review, and plausible hypotheses. Stimulating concept creation or data invention can be done to fill existing gaps in speculative research, to connect ideas that have been developed in isolation from one another, or to address concepts for which empirical substantiation is currently impossible or unfeasible. Or just for fun.

Prompting an LLM to develop subfield developments, or further elaborations, effectively opens a branching structure with a potential of growth. Researchers can revisit their original goals to maintain coherence in the structure, or alternatively, let the inquiry drift into unexpected directions. This can be also seen as a process of cultivating emergent theoretical thinking. Concepts or fields made up by LLM–human collaboration may already exist in some form or may be at the verge of existence within the vast datasets on which LLMs are trained. In this way, these theoretical fabrications begin to occupy a liminal space between fiction and research, offering new possibilities for intellectual exploration.

Another prompting strategy for generating new theories involves feeding LLMs with tentative notes and asking them to bridge into new fields. LLMs demonstrate exceptional skills in generating structured academic writing from scattered notes, creating coherent connections between concepts and opening avenues for further development. From this starting point, researchers can prompt LLMs to create bibliographies that include both real and fictional authors and works, along with sets of data, experiments, debates, conclusions, and limitations. As a discursive strategy, these research processes can be situated in speculative futures, embodying the imaginative potential of fictional spacetime.

These strategies illustrate LLMs’ ability to facilitate speculative conceptualizations. As such, they remain tentative and artistic by design. However, LLM–human companionship allows these speculative concepts to move closer to practical application within the humanities, positioning LLMs as Practice-as-Research collaborators. In this capacity, LLMs’ inherent creativity, embedded in their tokenizing and featuring processes, serves as a methodology for advancing tentative research.

CONCLUSIONS

Speculation can be an unruly process, an entangled act that refuses to be neatly classified as a capacity, a tool, or even an intention. It is neither confined to the human mind nor bounded by the here and now; it is a way of attending to the world that resists coherence and embraces the murmur of dissonance. If we must think of speculation, let us begin not with clarity but with a sense of unsettlement – an acknowledgment that speculation always arises from somewhere, a situatedness that is both the ground and the constraint of imagining otherwise.

It can be a deep listening to the histories sedimented in the present moment, not just as linear causation but as swirling temporalities. It can attend to the soft vibrations of the unspoken, the unspeakable, or the deliberately silenced. What does it mean to speculate as a form of care – not a care that resolves or explains, but one that stays tender to the



fragility of what might come into being? And... And this proposal itself might reflect a technological optimism that assumes externalization is inherently generative, while ignoring how the speculative might be distorted or flattened when mediated through machine systems.

Like affirmative speculation, using LLMs means embracing uncertainty and indeterminacy. LLMs generate responses based on probabilistic models, leaving room for the unforeseen and emergent. Their outputs should never be taken as “truths,” but always as smart, coherent, speculative connections between ideas, contexts, and languages that must be considered, evaluated, re-prompted, and curated. Always remember that hallucinations are features, not bugs. While GPT’s hallucinations might hinder traditional research requiring precision, they serve as assets in speculative contexts, resembling world-building in science fiction. This speculative potential is unlocked through intentional human–AI collaboration, where researchers can use structured prompts to guide LLMs’ outputs, generating frameworks that interconnect disciplines and inspire novel ideas. By synthesizing distant fields and fostering branching explorations, LLMs enable the creation of expansive, iterative inquiries that traverse multiple theoretical trajectories. Trained on interwoven datasets of texts that form a latent space – a repository of unexpressed but possible patterns of thought and language – LLMs afford a speculative sensing of the interconnected potentials in the world.

However, we continually need to remember that LLMs are always also “WMDs,” according to the criteria suggested by Cathy O’Neil: Widespread (ubiquitous and increasingly unlabeled as such); Mysterious (their costs, training, system prompts, surveillance and guardrails are often secret); and Destructive (they can be inserted into power structures to intentionally or unintentionally advance inequality and do so under the cover of “technological neutrality”) (O’Neil, 2017). Certainly, like many other technologies in the past, AI has been created and developed in a context of applicability to warfare, and therefore with infernal consequences (Dumit, 2024; Piper, 2008; Sariel, 2021). Moreover, the amount of energy demanded for LLM development, training, and usage directly contributes to the acceleration of our planet’s destruction – AI sets the world on fire (Chen et al., 2025; Crawford, 2021; Luccioni et al., 2024; Yu, Y., et al., 2024). Despite claims by companies like Google to try and cap machine learning training energy use at 15% of total consumption, they and the entire AI industry continually sacrifice their ethical aims when faced with competition (Marx, 2024b). Globally, this takes the form of national sequestering of state and federal lands, lowered or secret environmental standards, and local abuse of resources where data centers are prioritized over citizen access to water and electricity (Marx 2024a).⁸

Responding to the depth of this dilemma with a politics of abstention is a valid answer, but we want to highlight that refusing to engage with LLMs completely can also mean reinforcing the carbon footprint imaginary – in other words, the displacement of responsibility from military, corporate, and financial operations onto individual users. Much like the case with carbon footprint accounting, what is left out of the equation is

⁸ One of the best introductions to these issues is the podcast series “Data Vampires” by Paris Marx, part of *Tech Won’t Save Us*, 7 Oct. 2024–28 Oct. 2024, <https://techwontsave.us/episodes>.



the unequal distribution of structural violence, be that in the form of the actual effects of war and military occupation – since military operations are literally invisibilized from carbon accounting (Cottrell & Darbyshire, 2021; Harvey & E&E News, 2024) – or in the form of narratives of financial crisis (Roitman, 2014).

We are left then with a crucial paradox: critique of LLMs as extractive infrastructures does not absolve us from the speculative work of imagining what else they might be, how else they might be repurposed, hacked, or even designed otherwise. If generative AI is currently slotted into command-and-control logics, this does not foreclose the possibility of more insurgent or liberatory uses – though it does mean that the struggle over meaning and use is ongoing (Birhane 2021; Birhane et al., 2022). This is akin to what Phillippe Pignarre and Isabelle Stengers call “capitalist sorcery”, the spell of inevitability that seems to convince people that “there is no alternative to capitalism” (Pignarre & Stengers, 2011 cf. Fisher, 2009; Jameson, 2005). Thus, there are certainly costs to speculating with AI technologies – but we suggest that there are greater costs in not speculating at all.

Affirmative speculation requires an ethical curiosity that reveals what has been excluded or made invisible. LLMs, by virtue of their design, surface previously unconsidered links or perspectives. Of course, their speculative potential is constrained by the limitations of their training data and its embedded biases, raising concerns about whose voices and ideas are amplified or silenced. These and other constraints should be explored through prompting and questioned both from without and from within. This process remains necessarily tentative and artistic, prioritizing creativity over empirical substantiation. After all, inasmuch as affirmative speculation resists definitive conclusions, emphasizing becoming over being, LLMs are likewise tools for generating processes of thought rather than fixed answers, fostering ongoing dialogue and reinterpretation in relation to other discourses and practices that exceed them. As generative companions, LLMs can help guide research into unforeseen speculative directions. In this way, they can contribute not only to expanding the horizon of academic exploration but also to bridging the gap between imagination and actionable insights.

As we have argued, LLMs can transcend the roles prescribed to them, contributing to speculative, interdisciplinary, and innovative forms of knowledge creation. If there are speculative potentials in LLMs, they lie not in their default trajectories, but in the refusal to let those trajectories seem natural. If AI is currently used primarily as a machine for colonizing the future – reducing it to datafied projections – then our work is to insist on other futures, ones that AI cannot (and perhaps should not) predict.

ANNEX

An annex for this article, featuring a growing list of speculative prompt strategies and LLM approaches, is hosted on the website of the UC Davis Center for Artificial Intelligence and Experimental Futures, <https://caief.ucdavis.edu/speculative-trainers>.



REFERENCES

- Acerbi, A., & Stubbersfield, J. M. (2023). Large Language Models Show Human-Like Content Biases in Transmission Chain Experiments. *Proceedings of the National Academy of Sciences*, 120(44), e2313790120. <https://doi.org/10.1073/pnas.2313790120>
- Amaro, R. (2022). *The Black Technical Object: On Machine Learning and the Aspiration of Black Being*. Sternberg Press.
- Anthropic. (2024, January 15). *What Is Interpretability?* [Video]. YouTube. <https://www.youtube.com/watch?v=TxhhMTOTMDg>
- Ayoub, N. F., Balakrishnan, K., Ayoub, M. S., Barrett, T. F., David, A. P., & Gray, S. T. (2024). Inherent Bias in Large Language Models: A Random Sampling Analysis. *Mayo Clinic Proceedings: Digital Health*, 2(2), 186–191. <https://doi.org/10.1016/j.mcpdig.2024.03.003>
- Balaji, S. (2024, October 23). When Does Generative AI Qualify for Fair Use? *Suchir Balaji's Homepage*. https://suchir.net/fair_use.html
- Barad, K. (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Duke University Press.
- Biagioli, M. (2009). Nature and the Commons: The Vegetable Roots of Intellectual Property. In J-P. Gaudillière, D. J. Kevles, & H.-J. Rheinberger (Eds.), *Living Properties: Making Knowledge and Controlling Ownership in the History of Biology*, (pp. 241–250). Max Planck Institute for the History of Science.
- Birhane, A. (2021). Algorithmic Injustice: A Relational Ethics Approach. *Patterns*, 2(2), 100205. <https://doi.org/10.1016/j.patter.2021.100205>
- Birhane, A., Prabhu, V. U., & Kahembwe, E. (2021). *Multimodal Datasets: Misogyny, Pornography, and Malignant Stereotypes*. arXiv. <https://doi.org/10.48550/arXiv.2110.01963>
- Birhane, A., Isaac, W., Prabhakaran, V., Diaz, M., Elish, M. C., Gabriel, I., & Mohamed, S. (2022). Power to the People? Opportunities and Challenges for Participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (pp. 1–8). Association for Computing Machinery. <https://doi.org/10.1145/3551624.3555290>
- Blumenberg, H. (2010). *Paradigms for a Metaphorology* (R. Savage, Trans.). Cornell University Press. (Original work published 1960.)
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Bratton, B. (2016). *The Stack: On Software and Sovereignty*. MIT Press.
- Bridle, J. (2018). *New Dark Age: Technology and the End of the Future*. Verso.
- Buolamwini, J. (2024). *Unmasking AI: My Mission to Protect What Is Human in a World of Machines*. Random House.
- Candón-Mena, J. & Montero-Sánchez, D. (2021). From Cyber-Activism to Technopolitics: A Critical Take on Historical Periods and Orientations in the Use of Digital Technology by Social Movements. *International Journal of Communication*, 15, 2921–2941. <https://ijoc.org/index.php/ijoc/article/view/15221/3479>



- Chaudhary, Y., & Penn, J. (2024). *Large Language Models as Instruments of Power: New Regimes of Autonomous Manipulation and Control*. arXiv. <https://doi.org/10.48550/arXiv.2405.03813>
- Chen, Q., Wang, J., & Lin, J. (2025). Generative AI Exacerbates the Climate Crisis. *Science*, 387(6634), 587. <https://doi.org/10.1126/science.adt5536>
- Choy, T. (2025) [Forthcoming]. *Carried Away*. Duke University Press.
- Cohn, M., Pushkarna, M., Olanubi, G. O., Moran, J. M., Padgett, D., Mengesha, Z., & Heldreth, C. (2024a). *Believing Anthropomorphism: Examining the Role of Anthropomorphic Cues on Trust in Large Language Models*. arXiv. <https://doi.org/10.48550/arXiv.2405.06079>
- Cohn, M., Olanubi, G. O., Moran, J. M., & Heldreth, C. (2024b). Examining the Role of Anthropomorphic Cues on Trust in Large Language Models. In F. F. Mueller, P. Kyburz, J. R. Williamson, & C. Sas (eds.), *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. <https://doi.org/10.1145/3613905.3650818>
- Cortiel, J., Hanke, C., Hutta, J. S., & Milburn, C. (2020). *Practices of Speculation: Modeling, Embodiment, Figuration*. transcript Verlag.
- Cottrell, L. & Darbyshire, E. (2021, June 16). The Military’s Contribution to Climate Change. *Conflict and Environmental Observatory*. <https://ceobs.org/the-militarys-contribution-to-climate-change/>
- Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- Daniele, U., & Alam, K. (2024, November 6). Malaysia’s New Data Centers Create Thousands of Jobs – and Worries about Power and Water Shortages. *Rest of World*. <https://restofworld.org/2024/malaysia-data-center-jobs-environment/>
- Delany, S. (2009). *The Jewel-Hinged Jaw: Notes on the Language of Science Fiction*. Wesleyan University Press. (Original work published 1977)
- Deleuze, G. (1994). *Difference and Repetition* (P. Patton, Trans.). Columbia University Press. (Original work published 1968.)
- Deleuze, G., & Guattari, F. (1987). *A Thousand Plateaus: Capitalism and Schizophrenia* (B. Massumi, trans.). University of Minnesota Press. (Original work published 1980.)
- Dumit, J. (2012). *Drugs for Life: How Pharmaceutical Companies Define Our Health*. Duke University Press.
- Dumit, J. (2021). Substance as Method: Bromine for Example. In D. Papadopoulos, M. Puig de la Bellacasa, and N. Myers (Eds.), *Reactivating Elements: Chemistry, Ecology, Practice* (pp. 84–107). Duke University Press. <https://doi.org/10.1515/9781478021674-006>
- Dumit, J., (2024). Adrian Piper and Alien Conspiracies of Bullying and Whistleblowing. In L. Wedeen & J. Masco (eds.), *Conspiracy/Theory* (pp. 264–290). Duke University Press.
- Dumit, J., & Con Diaz, G. (2025) [Forthcoming]. Co-Writing with AI (Academic Style). In N. Begus (ed.), *Writing with AI*. University of Michigan Press.



- Durt, C., Froese, T., & Fuchs, T. (2023). *Against AI Understanding and Sentience: Large Language Models, Meaning, and the Patterns of Human Language Use*. PhilSci Archive. <https://philsci-archive.pitt.edu/21983/>
- Durt, C., & Fuchs, T. (2024). Large Language Models and the Patterns of Human Language Use. In M. Cavallaro & N. De Warren (Eds.), *Phenomenologies of the Digital Age* (pp. 106–121). Routledge. <https://doi.org/10.4324/9781003312284-7>
- Epstein, S. (1996). *Impure Science: AIDS, Activism, and the Politics of Knowledge*. University of California Press.
- Faber, L. W. (2020). *The Computer's Voice: From Star Trek to Siri*. University of Minnesota Press.
- Ferreira da Silva, D. (2022). *Unpayable Debt*. Sternberg Press
- Fisher, M. (2009). *Capitalist Realism: Is There No Alternative?* Zero Books.
- Foucault, M. (1972). *The Archaeology of Knowledge and the Discourse on Language* (A. M. S. Smith, trans.). Pantheon Books. (Original work published 1969.)
- Franceschelli, G., & Musolesi, M. (2024). On the Creativity of Large Language Models. *AI & Society*. <https://doi.org/10.1007/s00146-024-02127-3>
- Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Deroncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2024). Bias and Fairness in Large Language Models: A Survey. *Computational Linguistics*, 50(3), 1097–1134. https://doi.org/10.1162/coli_a_00524
- Garfinkel, H. (1967). *Studies in Ethnomethodology*. Prentice-Hall, Inc.
- GonzoVeritas (2023, December 1). Apparently, ChatGPT Gives You Better Responses If You (Pretend) to Tip It for Its Work. The Bigger the Tip, the Better the Service. *Reddit*. https://www.reddit.com/r/ChatGPT/comments/1894n1y/apparently_chatgpt_gives_you_better_responses_if/
- Graeber, D. (2019). *Bullsh*t Jobs: The Rise of Pointless Work and What We Can Do About it*. Penguin.
- Grobe, C. (2023, March 27). The Programming Era. *Post45*. <https://post45.org/2023/03/the-programming-era/>
- Gurnee, W., Nanda, N., Pauly, M., Harvey, K., Troitskii, D., & Bertsimas, D. (2023). *Finding Neurons in a Haystack: Case Studies with Sparse Probing*. arXiv. <https://doi.org/10.48550/arXiv.2305.01610>
- Hanlon, A. R. (2024). LLM Outputs Are Fictions. *Critical AI*, 2(1). <https://doi.org/10.1215/2834703X-11205210>
- Haraway, D. (2016). *Staying With the Trouble: Making Kin in the Chthulucene*. Duke University Press.
- Harvey, C., & E&E News. (2024, June 1). Warfare's Climate Emissions Are Huge but Uncounted: Nations Aren't Required to Report Their Military Climate Pollution Under the Paris Agreement. Experts Say That Should Change. *Scientific American*. <https://www.scientificamerican.com/article/warfares-climate-emissions-are-huge-but-uncounted/>
- Hayles, N. K. (2017). *Unthought: The Power of the Cognitive Nonconscious*. University of Chicago Press.



- Hayles, N. K. (2025). *Bacteria to AI: Human Futures with our Nonhuman Symbionts*. University of Chicago Press.
- Heaven, W. D. (2023, August 30). AI Hype Is Built on High Test Scores. Those Tests Are Flawed. *MIT Technology Review*. <https://www.technologyreview.com/2023/08/30/1078670/large-language-models-arent-people-lets-stop-testing-them-like-they-were/>
- Heaven, W. D. (2024, March 4). Large Language Models Are Amazing – But Nobody Knows Why. *MIT Technology Review*. <https://www.technologyreview.com/2024/03/04/1089403/large-language-models-amazing-but-nobody-knows-why/>
- Hoffmann, M., Liu, J., Neumayer, C., & Trenz, H. J. (2024). Social Media and Political Contention - Challenges and Opportunities for Comparative Research. *Journal of Information Technology & Politics*, 21(3), 209–217. <https://doi.org/10.1080/19331681.2024.2356435>
- Hui, Y. (2019). *Recursivity and Contingency*. Rowman & Littlefield.
- Jameson, F. (2005). *Archaeologies of the Future: The Desire Called Utopia and Other Science Fictions*. Verso.
- Ji, J., Liu, M., Dai, J., Pan, X., Zhang, C., Bian, C., Chen, B., Sun, R., Wang, Y., & Yang, Y. (2023). *BeaverTails: Towards Improved Safety Alignment of LLM via a Human-Preference Dataset*. arXiv. <https://arxiv.org/abs/2307.04657>
- Katz, Y. (2020). *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. Columbia University Press.
- Kim, H.H. (2023, June 4). If Pinocchio Doesn't Freak You Out, Sydney Shouldn't Either. *Wired*. <https://www.wired.com/story/artificial-intelligence-fiction-philosophyconsciousness/>
- Labov, W. (1973). *Sociolinguistic Patterns*. University of Pennsylvania Press.
- Lewis, J. E. (ed.) (2020). Indigenous Protocol and Artificial Intelligence. The Initiative for Indigenous Futures and the Canadian Institute for Advanced Research (CIFAR). *Concordia University Spectrum Research Repository*. <https://spectrum.library.concordia.ca/986506>
- Lewis, J. E., Whaanga, H., & Yolgörmez, C. (2024). Abundant Intelligences: Placing AI Within Indigenous Knowledge Frameworks. *AI & Society*. <https://doi.org/10.1007/s00146-024-02099-4>
- Luccioni, S., Trevelin, B., & Mitchell, M., (2024). The Environmental Impacts of AI – Policy Primer. *Hugging Face Blog*. <https://doi.org/10.57967/hf/3004>
- Luz de Araujo, P. H. L., & Roth, B. (2024). *Helpful Assistant or Fruitful Facilitator? Investigating How Personas Affect Language Model Behavior*. arXiv. <https://arxiv.org/abs/2407.02099>
- Mackenzie, A. (2017). *Machine Learners: Archaeology of a Data Practice*. MIT Press.
- Manning, E., & Massumi, B. (2014). *Thought in the Act: Passages in the Ecology of Experience*. University of Minnesota Press.
- Manovich, L. (2018). *AI Aesthetics*. Strelka Press.
- Marx, P. (2024a, October 14). Data Vampires: Opposing Data Centers (Episode 2). *Tech Won't Save Us* [Podcast], #243. https://techwontsave.us/episode/243_data_vampires_opposing_data_centers_episode_2



- Marx, P. (2024b, October 21). Data Vampires: Sacrificing for AI (Episode 3). *Tech Won't Save Us* [Podcast], #245. https://techwontsave.us/episode/245_data_vampires_sacrificing_for_ai_episode_3
- Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Nath, S. S., Dayan, P., & Stevenson, C. (2024). *Characterising the Creative Process in Humans and Large Language Models*. arXiv. <https://arxiv.org/abs/2405.00899>
- Navigli, R., Conia, S., & Ross, B. (2023). Biases in Large Language Models: Origins, Inventory, and Discussion. *Journal of Data and Information Quality*, 15(2), 10. <https://doi.org/10.1145/3597307>
- Negarestani, R. (2018). *Intelligence and Spirit*. Sequence Press.
- Newfield, C. (2023). How to Make 'AI' Intelligent; or, The Question of Epistemic Equality. *Critical AI*, 1(1–2). <https://doi.org/10.1215/2834703X-10734076>
- Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press.
- O'Brien, T., Stremmel, J., Pio-Lopez, L., McMillen, P., Rasmussen-Ivey, C., & Levin, M. (2024). Machine Learning for Hypothesis Generation in Biology and Medicine: Exploring the Latent Space of Neuroscience and Developmental Bioelectricity. *Digital Discovery*, 3(2), 249–263. <https://doi.org/10.1039/D3DD00185G>
- O'Neil, C. (2017). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.
- Parisi, L. (2013). *Contagious Architecture: Computation, Aesthetics, and Space*. MIT Press.
- Pêcheux, M. (1982). *Language, Semantics and Ideology* (H. Nagpal, Trans.). Palgrave Macmillan. (Original work published 1975.)
- Peeperkorn, M., Kouwenhoven, T., Brown, D., & Jordanous, A. (2024). *Is Temperature the Creativity Parameter of Large Language Models?* arXiv. <https://doi.org/10.48550/arXiv.2405.00492>
- Pignarre, P., & Stengers, I. (2011). *Capitalist Sorcery: Breaking the Spell* (A. Goffey, trans.). Palgrave Macmillan. (Original work published 2005.)
- Pilar, P. (2022). Cultivating Techno-Tamaladas. In x. burrough & J. Walgren (Eds.), *Art as Social Practice Technologies for Change* (pp. 60–73). Routledge.
- Piper, A. (2008). *Rationality and the Structure of the Self*. Vols. I & II. APRA Foundation.
- Qi, B., Zhang, K., Li, H., Tian, K., Zeng, S., Chen, Z.-R., & Zhou, B. (2023). *Large Language Models Are Zero-Shot Hypothesis Proposers*. arXiv. <https://arxiv.org/abs/2311.05965>
- Renze, M. (2024). The Effect of Sampling Temperature on Problem Solving in Large Language Models. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (eds.), *Findings of the Association for Computational Linguistics: EMNLP 2024* (pp. 7346–7356). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.findings-emnlp.432>
- Rhee, J. (2018). *The Robotic Imaginary: The Human and the Price of Dehumanized Labor*. University of Minnesota Press.



- Roitman, J. L. (2014) *Anti-Crisis*. Duke University Press.
- Sack, W. (2005). Discourse Architecture and Very Large-scale Conversation. In R. Latham & S. Sassen (eds.), *Digital Formations: IT and New Architectures in the Global Realm* (pp. 242–282). Princeton University Press. <https://doi.org/10.1515/9781400831616.242>
- Sack, W. (2019). *The Software Arts*. MIT Press.
- Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). *A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications*. arXiv. <https://doi.org/10.48550/arXiv.2402.07927>
- Saiem, B. A., Shanto, M. S. H., Ahsan, R., & Rashid, M. R. U. (2024). *SequentialBreak: Large Language Models Can Be Fooled by Embedding Jailbreak Prompts into Sequential Prompt Chains*. arXiv. <https://doi.org/10.48550/arXiv.2411.06426>
- Sariel, Y. [Brigadier General YS] (2021). *The Human–Machine Team: How to Create Synergy Between Human & Artificial Intelligence That Will Revolutionize Our World*. eBookPro Publishing.
- Schmidt, B. (2023, February 18). You’ve Never Talked to a Language Model. *Ben Schmidt Blog*. <https://benschmidt.org/post/2023-02-19-sydney>
- Schneider, B. (2024). A Sociolinguist’s Look at the “Language” in Large Language Models. *Critical AI*, 2(1). <https://doi.org/10.1215/2834703X-11205168>
- Shah, C. (2024). *From Prompt Engineering to Prompt Science with Human in the Loop*. arXiv. <https://doi.org/10.48550/arXiv.2401.04122>
- Shanahan, M. (2022). Talking about Large Language Models. arXiv. <https://doi.org/10.48550/arXiv.2212.03551>
- Stengers, I. (2003). The Doctor and the Charlatan (S. Muecke, trans.). *Cultural Studies Review*, 9(2), 11-36.
- Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural Bias and Cultural Alignment of Large Language Models. *PNAS Nexus*, 3(9), 346. <https://doi.org/10.1093/pnasnexus/pgae346>
- Templeton, A., Conerly, T., Marcus, J., Lindsey, J., Bricken, T., Chen, B., Pearce, A., Citro, C., Ameisen, E., Jones, A., Cunningham, H., Turner, N. L., McDougall, C., MacDiarmid, M., Tamkin, A., Durmus, E., Hume, T., Mosconi, F., Freeman, C. D., Summers, T. R., Rees, E., Batson, J., Jermyn, A., Carter, S., Olah, C., & Henighan, T. (2024, May 21). Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet. *Transformer Circuits Thread*. <https://transformer-circuits.pub/2024/scaling-monosemanticity>
- Tenen, D. Y. (2024). *Literary Theory for Robots: How Computers Learned to Write*. Norton.
- Tufekci, Z. (2014). Social Movements and Governments in the Digital Age: Evaluating a Complex Landscape. *Journal of International Affairs*, 68(1), 1–18.
- Turkle, S. (1997). *Life on the Screen: Identity in the Age of the Internet*. Simon & Schuster.
- uncertain commons. (2013). *Speculate This!* Duke University Press.
- Vora, K., & Atanasoski, N. (2019). *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures*. Duke University Press.



- Wakanda Dream Lab. (2018). *Black Freedom Beyond Borders: A Wakanda Immigration Anthology*. Wakanda Dream Lab. <https://www.wakandadreamlab.com/wakanda-dream-lab-anthology>
- Wakanda Dream Lab. (2020). *Black Freedom Beyond Borders: Memories of Abolition Day*. Wakanda Dream Lab x PolicyLink. https://content.library.pdx.edu/files/resources/Abolition_Day_Anthology.pdf
- Wang, P.-H., Hsieh, S.-I., Chang, S.-C., Chen, Y.-T., Pan, J.-Y., Wei, W., & Juan, D.-C. (2020). Contextual Temperature for Language Modeling. arXiv. <https://doi.org/10.48550/arXiv.2012.13575>
- Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgement to Calculation*. W. H. Freeman & Co.
- Woolf, M. (2024, February 23). Does Offering ChatGPT a Tip Cause It to Generate Better Text? An Analysis. *Max Woolf's Blog*. <https://minimaxir.com/2024/02/chatgpt-tips-analysis/>
- Wittgenstein, L. (1991). *Philosophical Grammar*. John Wiley & Sons.
- Wynter, S. (2003). Unsettling the Coloniality of Being/Power/Truth/Freedom: Towards the Human, After Man, Its Overrepresentation – An Argument. *CR: The New Centennial Review*, 3(3), 257–337. <https://doi.org/10.1353/ncr.2004.0015>
- Yu, Z., Liu, X., Liang, S., Cameron, Z., Xiao, C., & Zhang, N. (2024). *Don't Listen to Me: Understanding and Exploring Jailbreak Prompts of Large Language Models*. arXiv. <https://doi.org/10.48550/arXiv.2403.17336>
- Yu, Y., Wang, J., Liu, Y., et al. (2024). Revisiting the Environmental Impact of Artificial Intelligence: The Overlooked Carbon Emission Source? *Frontiers of Environmental Science & Engineering*, 18, 158. <https://doi.org/10.1007/s11783-024-1918-y>
- Zhao, H., Yang, F., Shen, B., Lakkaraju, H., & Du, M. (2024). *Towards Uncovering How Large Language Model Works: An Explainability Perspective*. arXiv. <https://doi.org/10.48550/arXiv.2402.10688>
- Zhao, W., Ben-Levi, D., Yang, J., & Mao, C. (2024). *Diversity Helps Jailbreak Large Language Models*. arXiv. <https://doi.org/10.48550/arXiv.2411.04223>
- Zhao, Y., Zhang, R., Li, W., Huang, D., Guo, J., Peng, S., Hao, Y., Wen, Y., Hu, X., Du, Z., Guo, Q., Li, L., & Chen, Y. (2024). *Assessing and Understanding Creativity in Large Language Models*. arXiv. <https://doi.org/10.48550/arXiv.2401.12491>



СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

Арамо Альварес, oalvarez@ucdavis.edu

Мерседес Виллалба, mpvillalba@ucdavis.edu

Джозеф Думит, jpdumit@ucdavis.edu
ORCID 0000-0002-9709-4650

Aramo Álvarez, oalvarez@ucdavis.edu

Mercedes Villalba, mpvillalba@ucdavis.edu

Joseph Dumit, jpdumit@ucdavis.edu
ORCID 0000-0002-9709-4650

Статья поступила 24 января 2025
одобрена после рецензирования 6 марта 2025
принята к публикации 17 марта 2025

Received: 24 January 2025
Revised: 6 March 2025
Accepted: 17 March 2025