



This PDF is a simplified version of the original article published in Internet Archaeology under the terms of the Creative Commons Attribution 3.0 (CC BY) Unported licence. Enlarged images, models, visualisations etc which support this publication can be found in the original version online. All links also go to the online original.

Please cite this as: Kansa, E.C. 2026 Archaeology in the AI Era: Demystifying powerful and problematic systems shaping the future of the past, Internet Archaeology 71. <https://doi.org/10.11141/ia.71.15>

Archaeology in the AI Era: Demystifying powerful and problematic systems shaping the future of the past

Eric C. Kansa

Summary

This essay explores how Large Language Models (LLMs) and the generative "Artificial Intelligence" (AI) services they power impact archaeology's information ecosystem. Most of the discussion focuses on the dangers of escalating costs, commercial dependency, indeterminacy and reliability, and the potential for bias amplification. The essay highlights the need for structural reforms in archaeology's professional incentive systems, especially those systems that measure and reward publishing practices.

1. Introduction

The term "Artificial Intelligence" (AI) seems inescapable. But the term itself covers too much ground to be very useful. AI can refer to a huge range of technologies and methods, all with different capabilities and costs. For the most part, however, when people talk about AI, they're generally referring to Large Language Models (LLMs) and related technologies that now attract tremendous levels of speculative investment. The launch of OpenAI's ChatGPT in November of 2022 catapulted LLMs into public consciousness and made AI a topic of widespread awe, wonder, and dread.

Though I have practiced digital archaeology, especially data publishing and curation for over 20 years, I claim no special expertise in any area of AI or LLMs. While I mainly rely on my own knowledge of the Python programming language, I occasionally use LLM-based coding tools to help write and configure software in areas where I have less expertise. While I experimented with "fine-tune" training an AI model, until recently, I never had the time or inclination to fully invest myself in pursuing these technologies.

Nevertheless, as Technology Director of Open Context, a non-profit data publishing service for archaeology, I find myself regularly grappling with LLM-related impacts (see [Open Access and Bots](#)). Like it or not, LLM systems help shape the landscape of archaeological data sharing and scholarly communications more broadly. The need to reflect on this new reality motivated me to write this essay.

While this essay is specifically intended for an archaeological audience, the topic requires a much broader look at the technology and its wider impacts. The main goal of this essay is to demystify LLMs while still acknowledging how their introduction reshapes so much of 21st century life.

The cloud of hype and hyperbole makes it very difficult to understand LLM-powered AIs. They are both marketed and blamed as almost magical labor-saving tools that can even replace human beings in many lines of work. LLMs have now reshaped and shaken many areas of work, ranging from software development, to the creative arts, to the world of research, including archaeological



research. LLMs have also transformed how we find, consume, and produce information, including information about humanity's past.

Because of their uncanny fluency in conversing with human language, LLMs appear capable of almost any intellectual task given to them. But that very fluency can obscure certain innate and fundamental limitations. Those limitations can lead to dire consequences when LLMs are misapplied. Unfortunately, powerful financial forces seek to insert LLMs almost everywhere, despite their risks and limitations. How do these factors work to reshape archaeological research and cultural heritage conservation practices?

2. Large Language Models Enter the Chat

To explore these questions, it's good to review how LLMs actually work. At their heart, a Large Language Model is a vast numeric representation of statistical patterns found in bodies of text. In the case of the major commercial "frontier" LLMs, those bodies of text come from harvesting vast quantities of content from the Internet. The process of observing and recording statistical patterns within texts is called "training". The numeric representation of statistical patterns found in texts have mind boggling complexity. LLMs use thousands of different dimensions to represent statistical patterns of associations between different segments of text (called "tokens", which, loosely speaking, correspond to numbers that encode words and parts of words). By combining many thousands of dimensions and huge training datasets, LLMs can represent very subtle patterns in how words relate to each other.

Because LLMs involve "thousands of dimensions" and "the entire Internet" they can seem incomprehensible and utterly alien. When ChatGPT launched in November of 2022, it almost instantly became an Internet sensation. It seemed capable of knowledgeably discussing virtually any topic, including esoteric issues in archaeology. ChatGPT continued to iterate through multiple upgrades and sparked deployment of a number of similar LLM-powered systems, like Microsoft's CoPilot and Google's Gemini. LLM-powered "generative AIs" (that is, information systems that generate plausible text, images, code and other media in response to user prompts) captured the imagination of many, including investors willing to bet many hundreds of billions of dollars on this technology. It seemed as if dreams of "Artificial Intelligence"^[1] that could rival or exceed the intellectual powers of human beings, had finally arrived.

3. Costly Reliance on Hungry Bots

LLMs have a mystique. They can seem like disembodied engines of pure intellect. Nevertheless, LLMs require significant real-world infrastructure in terms of computer data centers and electrical power to develop and operate. These associated costs clearly require staggeringly large investments, but there is little public disclosure about how much it costs to build and operate a commercial LLM. High-stakes commercial competition and feverish investment do little to encourage transparency.

At this time, speculative investment money subsidizes a large (but not publicly disclosed) fraction of the costs required to build and use an LLM. It is likely that commercial LLM developers hope to entrench themselves into many markets and encourage greater dependency on their services before they raise prices and become profitable. Obviously, LLM users should be wary of this strategy. Currently free or low-cost LLM services are unlikely to remain cheap indefinitely. Archaeology is not famously awash in discretionary money. Archaeologists should carefully consider dependency on commercial LLM services when the costs of such services are likely to jump once investors start to demand profits.

The cost considerations cannot be ignored because they play a fundamental role in how LLMs perform in real-world settings. Cost issues related to LLMs include:

- LLMs are tremendously capital intensive. They require large computing resources to gather data, to train models on these data, and to operate in response to users. Investment money that currently subsidizes these costs cannot last forever.



- To coax better performance from LLMs, developers have introduced multiple techniques that use bigger sets of instructions, external supplementary tools, and more repeated behind-the-scenes interactions with models (see the [software discussion](#) below). While these techniques apparently improve performance, they also demand even more computational resources and more electricity. Better performance therefore demands even more money to sustain.
- LLMs used to generate media have even higher energy demands. Images, sound, and especially video all require the synthesis of large amounts of data and require costly computing resources to produce.
- LLMs are operated from data centers that have large environmental impacts. These impacts result from the mining and supply chains involved in the manufacture of computing infrastructure as well as the electricity required in their operation (Crawford [2022](#)). These costs typically do not see public accounting. Many of these costs are absorbed by the wider natural environment and by human communities as "negative externalities". Marginalized and disempowered communities often shoulder the brunt of the costs and damages associated with LLM infrastructure (Hao [2025](#)).

The commercial giants pushing this technology act as if they need to create entrenched and lasting monopolies to justify the huge speculative investments fuelling LLM-based systems. If successful, a few powerful organizations will (continue to) dictate the terms of how LLMs get developed, monetized, and used to restructure entire economies and labor forces. If unsuccessful, LLMs may be relegated to a limited range of applications while the wider economy struggles through the cataclysmic financial fallout of a burst speculative bubble (Ramzanali [2026](#)). Either way, the technology giants pushing LLMs have fundamentally altered life in the 21st century.

4. Rolling the Dice with Chatbots

While LLMs are indeed impressive, we should remember they are technologies, and like any other technology, they have certain capabilities and limitations. Probably the most familiar use of an LLM comes from chatbots, like ChatGPT. When you type your "prompt" into ChatGPT, the LLM behind the generative AI curtain converts your prompt into a numeric representation that it can mathematically compare with information within its model. ChatGPT then randomly selects sequences of words from options the model indicates as statistically likely to follow the words in your original prompt (see Ji *et al.* [2026](#)).

That's right. **The random selection of words is a fundamental property of how LLMs respond to your prompts** (Holtzman *et al.* [2019](#)).

Multi-dimensional math narrows down the selection of responses to a number of plausible choices found in training data, but once the chatbot identifies plausible options, it essentially rolls dice to select which words it will actually include in its response to your prompt. This relates to a hotly contested debate about the nature and extent to which LLMs actually perform "reasoning" (Coeckelbergh and Gunkel [2024](#))^[2]. Do LLMs have emergent logical reasoning capabilities that guide how they randomly draw from plausible options? Or are the apparently reasonable responses made by LLMs merely a probabilistic reflection of how their training data records patterns of (often, but not always logical) discussions made by human beings? In other words, do LLMs really "reason" or do they generate outputs that statistically mimic patterns of text that superficially seem reasonable?

If an LLM only uses statistical patterns but not logic or deeper reasoning, and if no notion of fact or fiction guides how an LLM selects words, there may be little reason to trust its outputs. The computational linguist Emily Bender, a prominent critic of LLM-hype, calls such LLM-based systems "stochastic parrots" (Bender *et al.* [2021](#)). The prevalence of plausible, but wrong, "hallucinated" (see Mills and Angell [2025](#) for a critique of this term)^[3] responses from LLMs illustrates her point. LLMs used in legal settings have wholly fabricated court cases and precedence. When misused for scholarship (or student essays), the bibliographies generated by LLMs are often riddled with citations to imaginary literature. LLM hallucinations are common because statistical patterns in words, rather than factual evidence, drive their outputs. To Emily Bender, LLMs randomly (stochastically) repeat



statistically likely words present in their training data. To her, LLMs have no mechanism to determine if those word sequences actually correspond to truthful statements.

Again, there is very little expert consensus about the reasoning capabilities of LLMs (and the systems built around LLMs)^[4]. Despite this, LLMs can still have useful applications (within limits and with strong caveats). For example, LLMs seem to be helpful when incorporated into tools for software development. Software engineers can use these tools as sophisticated utilities that generate usable code, especially if that code is fairly routine or "boilerplate" and is well-represented in the LLM training data.

LLMs use random chance to select statistically plausible words in response to user prompts. It is unclear what role (if any) reasoning and differentiation between fact and fiction constrain these random selections. OpenAI itself admits "hallucinations" are *mathematically inevitable* and not a short-term engineering challenge that can be solved^[5]. **This means that useful applications of LLMs need to be fault-tolerant and resistant to plausible but erroneous outputs** (see Sarkar *et al.* [2025](#)). Software generation tools can partially correct for LLM failures using automatic checks to make sure generated code actually executes. If generated code fails, the LLM-powered coding assistants can "roll the dice" again and again until they finally output working code^[6]. LLMs may still write software with more subtle (and insidious) kinds of bugs, but at least code generators have clear ways to partially check the quality of their outputs.

5. A Knowledge Commons Enclosed by Bots

Characterizing LLMs as stochastic parrots should not be used to dismiss their tremendous real-world impacts. The term may or may not adequately characterize the current cutting-edge of LLM-based systems (see Sarkar [2023](#) and Arkoudas [2023](#) against the applicability of stochastic parrots). That jury is still out. Nevertheless, vast flows of financial and political capital try to insert LLMs into almost every facet of life. To various degrees, they may be unreliable (yet convincing) in their hallucinations and "bullshit"^[7], but they are everywhere.

LLMs are also transforming the Internet in ways that are surprising and concerning. LLMs rely on the Internet as a seemingly endless source of training data. However, even that source has limits. Almost all the easily accessible data has been harvested already. Now, there's a frantic rush to squeeze every last byte from remaining sources of data.

This insatiable demand for new training data has a profound effect on publishers that create content for the web. Websites now struggle to manage many times more traffic than they served in prior years. The growth of traffic is driven by bots (web scrapers) that systematically request content from websites, all ultimately to feed LLM models^[8].

Why does this matter?

It matters because websites built to serve human beings now need to fight off armies of voracious bots. The behavior of AI scrapers closely parallels so-called distributed denial of service (DDoS) cyber-attacks. If website owners face too much traffic, web servers can slow down and crash under the load of responding to too many simultaneous requests for data.

These are not just back-end technical headaches for beleaguered systems administrators. They directly impact everyday human experience of the web. When doing research on the web, how many times do you need to check a box to assert your humanity? How many times do you need to identify crosswalks or bicycles in a captcha challenge? All of these new hurdles exist to protect websites from unrelenting AI-related bot traffic, but they make human use of the web far more cumbersome and annoying. Of course, those many small annoyances end up favoring chatbot developers. Why would one slog through multiple captcha challenges to read web published content when one has LLM-generated summaries so readily available? Sure, an LLM summary may be bullshit (see [below](#)) and even partially hallucinated, but at least you can read it without first marking pictures of traffic signs!



5.1 Open Access and Bots

The problem is especially acute for academic repositories. Digital repositories of books, peer-reviewed articles, databases, historical records, photography, art, and archaeology now have hordes of bots constantly demanding data from their collections. According to a 2025 report by the Coalition for Networked Information, excessive web-scraping has resulted in the disruption and degradation of service for 70% of academic digital repositories^[9]. Academic digital libraries, museum collections, and digital repositories typically run on shoe-string budgets. They have limited capacity to defend against escalating bot attacks fueled by hundreds of billions of dollars of investment money.

The dramatic escalation of bot traffic undermines open access, one of the most optimistic uses of the World Wide Web. For decades, the global open access movement has worked to encourage the free dissemination of scholarship. This movement has succeeded in removing access barriers to many peer-review journals, pre-print repositories, digital repositories of research data, and educational content with lesson plans and teaching materials.

Prior to the proliferation of mass web-scraping, open access advocacy had a strong economic basis. For context, open access advocates rightly acknowledge that creating scholarship is expensive. For example, experiments may need costly instruments and equipment, archaeological field work may require years of excavations or surveys in challenging environments, and even historical research may require laborious sifting through archival materials. Furthermore, developing and communicating methodologies, theoretical frameworks, and research questions all require a great deal of expert labor and debate.

However, while creating knowledge is expensive, the web makes sharing digital books, articles, databases, and other representations of that knowledge very inexpensive. If one downloads a peer-reviewed digital article, other people can likewise still download their own copy. In economic terms that digital article is a "non-rivalrous" good, meaning that its use does not diminish its availability to others. Low-cost Internet access and servers are the only requirements, and those costs can be managed even with the meager budgets typical of libraries and digital archives.

The fact that digital information can be "non-rivalrous" provides the fundamental economic basis for open access. Unfortunately, the dramatic expansion of bot traffic undermines this foundation. When bots overwhelm web servers with too much traffic, they make open access services more rivalrous. Unrelenting swarms of bots now crowd-out human beings from open access scholarship.

Recently released statistics show that the open access Wikipedia now routinely blocks over 2 billion requests from bots *per day* (see Figure 1). Fighting bots requires concerted thought, effort, and money to configure, install, and maintain network protections. It recently required about a week of my own (at least somewhat skilled) labor to set up network defenses to protect [Open Context](#) (see Figure 2)^[10]. These defenses have opportunity costs. The resources now required to defend websites from bots could have otherwise improved website features and website content for human users.

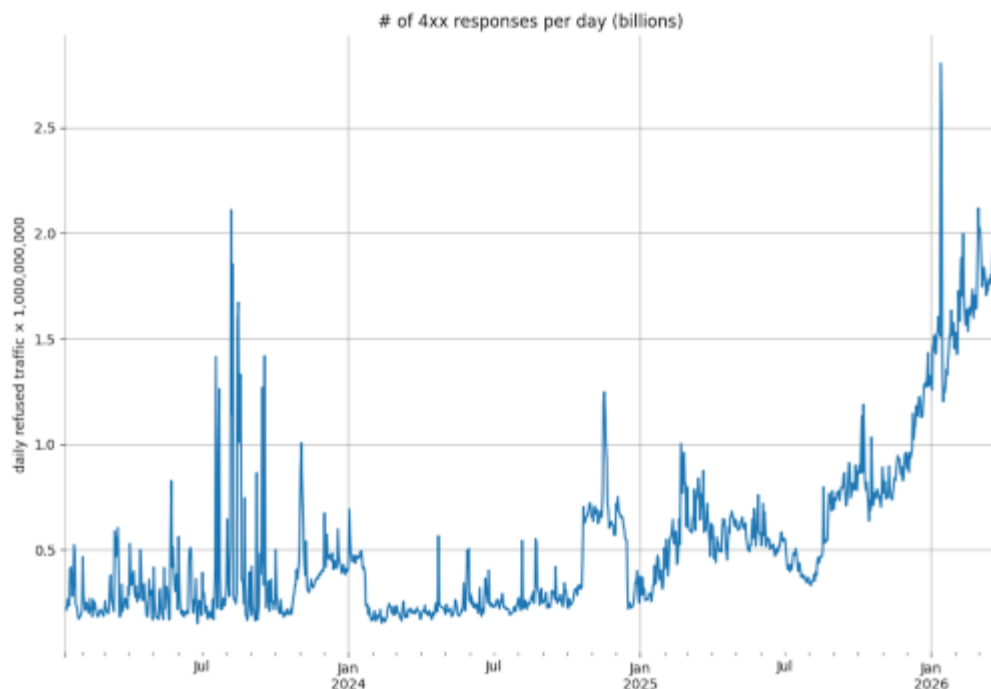


Figure 1: Blocked Requests to Wikipedia per Day (2023-2026)^[11]. Original image by Chris Danis under a Creative Commons Attribution Share-alike license (<https://web.archive.org/web/20260604234640/https://diff.wikimedia.org/2026/03/26/quo-vadis-crawlers-progress-and-whats-next-on-safeguarding-our-infrastructure/image-1919/>)

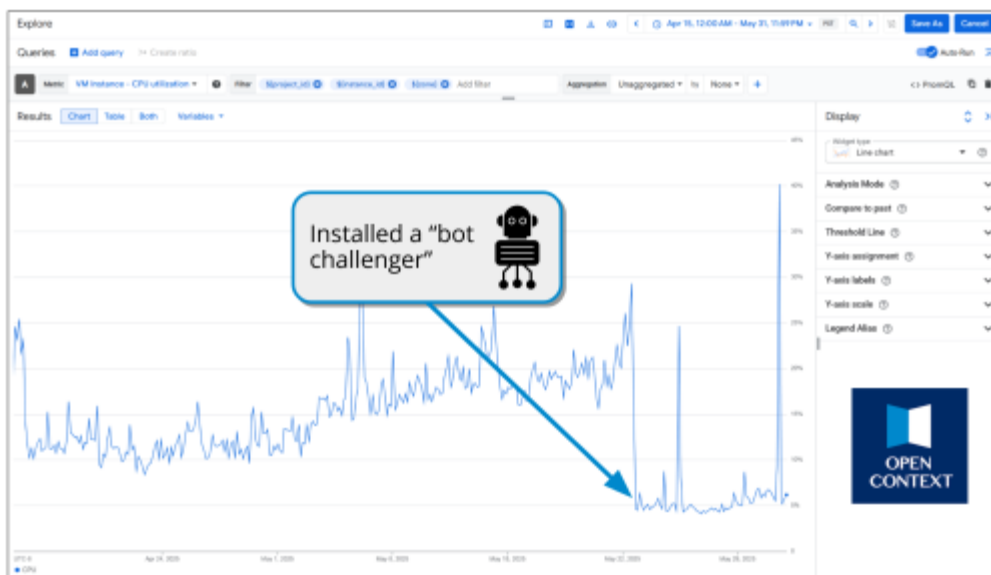


Figure 2: Short-term Compute Savings from Bot Protections for Open Context (April to May 2025). The y-axis represents the percentage of CPU use for Open Context's primary web server. More CPU use indicates higher processing loads, either in response to traffic, API use, or running complex queries. The arrow marks when [Anubis](#), an open source bot firewall, started blocking some scraper traffic to Open Context.

LLMs and their associated web-scraping bots have made open access far more precarious. Open access repositories now struggle to maintain performance and stability under bombardment from unrelenting bot traffic (Lippincott *et al.* [2026](#)). Not only are bot defenses costly to implement and



maintain, the captchas and other defenses sometimes make human access more cumbersome and annoying. All of these factors make open access repositories less capable of serving human needs.

These trends without a doubt will drive more users to turn to chatbots backed by corporate giants and their vast financial resources. As open access scholarly resources become more difficult to reach and sustain, human scholarship will become increasingly "enclosed" within commercial LLMs^[12].

Unfortunately, an LLM is a poor substitute for the full richness of the books, websites, articles, databases and collections used as LLM training data. Content from the web gets compressed into statistical weights for associations between tokens (Conklin *et al.* 2026). In the process, elements of the full provenance, context, and conversational relationships within and between different scholarly works get lost. Training an LLM absolutely must involve such reductions and simplifications, otherwise the resulting model would be impossibly large and unwieldy. *But a map should not be mistaken for an entire landscape.* If users come to see LLMs as replacements rather than (problematic) guides to humanity's scholarly record, we should expect some bleak and dystopian outcomes as described below.

6. Bullshitting the Past

Outside of software development, obvious mechanisms to automatically check and validate LLM outputs against hallucinations and other errors may be hard to find. This limitation however doesn't stop the promotion and use of LLMs in almost every conceivable domain. This leads to very problematic outcomes for archaeology.

When one searches the web, one now routinely encounters LLM-AI-generated summaries in response to user queries. Those summaries are typically presented before one scrolls down to see other websites listed in search results. This gives LLM-AI summaries greater prominence than content provided by even the most respected and authoritative sources. There's no doubt these summaries now play a huge role in shaping public understanding of the archaeological past (see examples, Figure 3a and 3b).



what is the age of the sphinx?



See all links

The Great Sphinx is conventionally believed to be around 4,500 years old, dating to the reign of Pharaoh Khafre, though some geological studies suggest it could be much older, possibly 10,000 to 12,000 years or even hundreds of thousands of years.

Conventional Dating Approach

Most Egyptologists and archaeological evidence place the construction of the Great Sphinx during the Old Kingdom of Egypt, specifically the reign of Pharaoh Khafre (c. 2558–2532 BCE). The Sphinx is part of the Giza Plateau complex and aligns with Khafre's pyramid and mortuary temple, with supporting evidence such as the granite Dream Stele installed by Thutmose IV referencing Khafre. This dating implies that the Sphinx is approximately 4,500 years old.

Alternative Geological Theories

Some geologists, including Robert Schoch and John Anthony West, have examined erosion patterns on the Sphinx. They argue that the heavy weathering and vertical fissures in the enclosure walls suggest prolonged water erosion, not typical desert wind. Considering that such rainfall did not occur in the Giza region for over 7,000 years, this could imply the Sphinx was carved between 10,000 and 12,500 years ago, during a wetter climate at the end of the last Ice Age.



The Age of the Sphinx | Battle of the Geologists
The Great Sphinx, which sits in front of the three great pyramids on the Giza Plateau in Egypt, has been the subject of controversy...
YouTube · World of Antiquity
196.1K views · Dec 22, 2020

The Age of the Sphinx Controversy

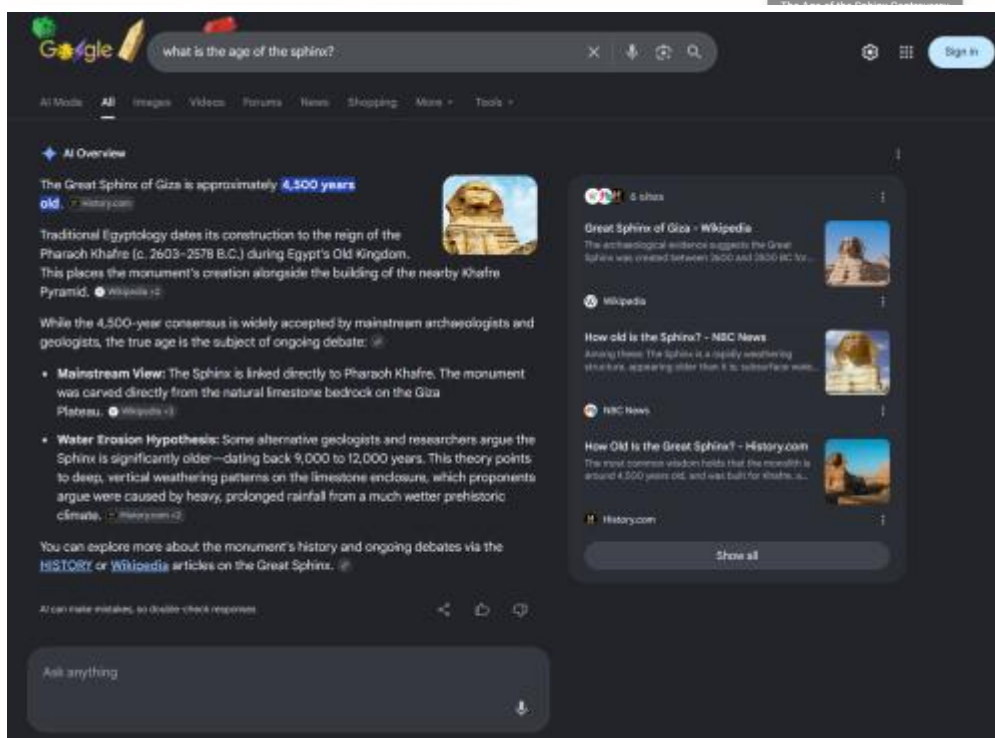


Figure 3a: Microsoft Bing AI Search Summary on the "Age of the Sphinx" (Great Sphinx, Giza). The LLM-generated summary text in blue seems to suggest that two very different time frames for the construction of the Great Sphinx at Giza are both equivalently plausible.

Figure 3b: Google AI Search Summary on the "Age of the Sphinx" (Great Sphinx, Giza). The LLM-generated summary text suggests that two very different time frames for the construction of the Great Sphinx at Giza are "the subject of ongoing debate".

Moreover, even before the mass adoption of LLM-AIs, the web was heavily shaped by commercial forces. Driven by profit maximizing goals, entire industries have grown to game the rankings of websites within search engines. Public use of the web has been increasingly shaped by commercially driven search engine optimization (SEO), social media engagement algorithms, advertising, and other practices. Nuanced and carefully reasoned archaeological perspectives have a hard time gaining visibility in an information landscape shaped by commercial and advertising interests clamoring for user attention.



LLMs are trained on content from the Internet (including images and videos) largely shaped by commercial interests. LLM-AIs distill this training data in ways that further deepen and entrench this commercial orientation. Instead of just shaping how you *find* content on the web and what social media sites capture most of your attention, users often now use commercial LLM-AIs as *replacements* for the entire web. After all, why bother checking different websites (especially when those websites often require annoying captcha challenges) when a seemingly all-knowing AI offers a concise and convincing answer to your question?

These developments raise unsettling concerns. How much trust should one put into "a concise and convincing answer" generated from a chatbot^[13]? Is the summary coming from a stochastic parrot with (apparently) no capacity for analytic reasoning or for discerning truth from falsehood?

Search engines don't even use their flagship LLMs to provide search summaries because the larger flagship models have higher compute costs (in terms of energy and money). Even if we grant that the current suite of cutting-edge LLMs have some genuine capabilities to reason, those capabilities aren't necessarily used for summarizing search results. Search summaries come from slimmed down and simplified models, likely limiting their reliability. Yet despite disclaimers that the results may be wrong, users rarely bother to double check LLM search overviews. One study found that users will still use LLM summaries even if 80% of these summaries contained errors (Landymore [2026](#); Shaw and Nave [2026](#))!

For most of the public, archaeology is usually a topic of occasional curiosity and fascination. Without special motives or interests, most members of the public would not likely invest much time or attention to the topic. Therefore, few would be willing to go through the effort needed to fact-check LLM-generated summaries of archaeological search results. Moreover, recent LLMs have been fine-tuned through a technique known as "reinforcement learning from human feedback" (RLHF) (Arkoudas [2023](#)). RLHF works to improve user-friendliness. Unfortunately, a recent study demonstrated that this technique also rewards LLMs for "bullshitting" their responses. Instead of selecting from options their training data indicates as statistically likely, RLHF teaches LLMs to select options more likely to get a "thumb-up" signal from a human user (Liang *et al.* [2025](#)). A cynic may think we now encourage our stochastic parrots to become sycophants.

In a recent lecture, Shawn Graham ([2026](#)) described how recommendation algorithms, which are common to social media platforms like YouTube and Instagram, subtly amplify predispositions, prejudices, and interests of their users. He described how these algorithms can lead their users toward certain social and cultural tropes that act as "strange attractors" or "gravity wells" in the information landscape^[14]. LLMs, when coupled with profit motivated incentives to maximize user engagement (time and attention directed toward a service), would greatly magnify this effect. An LLM configured to maximize engagement will generate responses to user prompts that tend to keep the user interacting with the service. Together with RLHF, this can lead some users into a hall-of-mirrors world of delusions where the user and LLM collaboratively write ungrounded and fictitious narratives with no factchecking or reasoning to call bullshit on the emerging personalized fantasy^[15]. In extreme cases, this leads to AI-induced psychosis (Osler [2026](#)). Archaeology seems especially prone to this kind of dynamic, where pseudo-archaeology topics like ancient aliens, racist and nationalist narratives, and more would all serve as "strange attractors" and "gravity wells" for LLM-fed delusions.

If users seem predisposed to favor bullshit and seem generally disinterested in fact-checking LLM summaries, perhaps regulatory or liability pressures will curtail LLM disinformation? In certain areas, LLM-AIs are already coming under increasing scrutiny for generating seemingly plausible but factually wrong advice. In domains like medicine and law, bad information can lead to catastrophic outcomes. It is likely that pressure for regulation, threats of lawsuits, and other forms of public pressure will grow in matters that can clearly impact public safety.

But archaeology is rarely a matter of life-or-death or even high-stakes litigation. It is unlikely that regulators or public pressure will do much to stem the flow of LLM-generated mis- and dis-information about the human past. For the most part, sharing knowledge about the shortcomings of generative AI search summaries with our students, colleagues, and wider publics may be the best archaeologists



can do. We can point towards browser extensions and search services (such as [NOAI DuckDuckGo](#)) that circumvent or disable LLM-powered search services.

7. Generated Cultural Heritage Media

Search engine summaries are not the only way that LLMs now help shape perceptions of the past. Despite controversies, even professional archaeologists sometimes use LLMs in writing papers and reports. In addition, certain LLMs also have multimedia capabilities and can generate synthetic images, sound, and videos from their training data. Magnani and Clindaniel ([2023](#)) explore these capabilities to generate images illustrating aspects of Neanderthal behavior.

Recently, the British Museum posted a series of generative AI images to its Instagram page. These images depicted an algorithmically-generated woman looking at various museum objects from cultures around the world. After heated public outcry, the British Museum took down the images and stated that they would develop a set of internal policies to guide generative AI use (Lawson-Tancred [2026](#)).

The backlash included worries that generative AI images involved the appropriation of creative works taken from the Internet to feed training datasets. Others saw the British Museum use of AI as a cost-cutting measure that devalues human labor and expertise and will lead to job losses in the cultural heritage sector. Others disliked the aesthetic character of the images, thinking that their synthetic feel demonstrated a lack of respect for both the museum's collection and its patrons (Fear [2026](#)).

Why do AI synthesized images feel like they cheapen cherished cultural institutions like museums? Part of the reason may stem from the selection of training data. According to research by Kate Crawford (Antonelli [2024](#)), image generation AIs are largely trained on image datasets harvested from the Internet, especially social media and e-commerce sites. The aesthetic choices, subjects, sexualized images, and an emphasis on wealth and status—all tools used to drive commercial sales—permeate the resulting AI models. When such generative AI tools are used in the cultural heritage context, they color that context with the imagery and aesthetics of e-commerce. In addition to these commercial biases, Magnani and Clindaniel ([2025](#)) note how training data is likely to contain outdated scientific literature, including outdated representations of gender. They demonstrate how these outdated training data influence AI generated media in archaeology.

How does this reshape our engagement with the human past? Does it muddle and constrain how we imagine the lives of people in other epochs and cultural contexts?

8. Can We Train Bots to CARE?

Critics have frequently warned about the dangers of biased training data used to build LLMs (reviewed by Barassi [2024](#); Tenzer *et al.* [2024](#)). Datasets are in large part drawn from the Internet, and the content on the Internet itself is shaped by commercial, social, linguistic, and cultural factors that all bias the kinds of perspectives and content that gets created and shared. Perhaps a concerted effort to address these biases and train LLMs with collections carefully curated to improve diversity and inclusion could address some of these problems?

One may naively think this kind of approach would advance the spirit of the CARE Principles for Indigenous Data Governance. CARE stands for Collective benefit, Authority to control, Responsibility, and Ethics (Carroll *et al.* [2020](#); GIDA [2020](#)). It is currently the leading framework for ethical data governance practices for information relating to Indigenous peoples.

One of the worst ways that one could address the biases in LLMs would be to simply add Indigenous knowledge to LLM training data^[16]. However, adding such knowledge into LLMs without the ongoing supervision of Indigenous authorities would be unethical and antithetical to the CARE Principles. LLMs can be trained with Indigenous knowledge, provided that Indigenous authorities supervised how and why their ancestral knowledge was used and provided they maintained control over resulting models. Afterall, "authority to control" is a core aspect of the CARE Principles.

More thoughtfully and inclusively selected training data may address LLM biases. However, one must first obtain the expertise and computational horsepower required to train (or retrain) a language



model. But would such investment be worthwhile? After all, archaeological researchers and Indigenous communities all have limited access to the required financial and technical capital. Should they spend their scarce resources training custom language models?

Cost and benefit considerations are fundamental. The problems with LLMs cannot be solved simply by changing their training data or governance. They may yet remain an inherently unreliable technology. Even if LLMs are trained on datasets deliberately selected for diversity and inclusion goals, the problem of hallucinations will not go away. A more inclusively trained LLM will still "roll dice" to generate sequences of words, even if they statistically sample from a more diverse and inclusive set of patterns^[17]. In other words, a better-trained stochastic parrot may generate responses using a broader range of phrases and accents, but it would (probably) still be just a parrot. Would such an effort be a cost-effective and useful investment for under-resourced communities?

9. Who Put the Parrots in Charge?

LLMs have captured the imaginations of decision makers who imagine these technologies as almost miraculous productivity multipliers for intellectual labor. As a result, LLM-generated content has now infiltrated professional discourse. Deloitte, a prestigious consulting firm, needed to give a partial refund to its client after that client discovered that Deloitte's report contained "hallucinated" (fabricated) citations to non-existent supporting evidence^[18]. This was not an isolated incident. Deloitte similarly included LLM-fabricated citations in a \$1.6 million report to Canada's Newfoundland government^[19].

It is not hard to imagine why Deloitte used LLMs to help generate their reports. LLMs can generate text about virtually any topic. LLMs are an alluring tool for cutting labor costs and maximizing profit margins. However, one wonders why anyone would bother to pay expensive consultants to type prompts into a chatbot. Why not just cut out the middleman (i.e. the consulting firm)? In any event, governments and other organizations contract with consulting firms like Deloitte in order to help inform strategy and policy. What happens to decision making when policy choices are, at least in part, driven by (undisclosed) outputs of stochastic parrots?

LLM-driven policy making should be a major concern for archaeology and the cultural heritage sector. The Advisory Council on Historic Preservation (ACHP), a US government regulatory body, announced an initiative to use commercial "AI-driven tools to accelerate and improve the consideration of historic properties in project planning" (ACHP 2026)^[20]. The goal is to improve "efficiencies for reviews required by Section 106 of the National Historic Preservation Act (NHPA) and help make the overall permitting process more efficient". As of February 2026, the ACHP is working with Google to integrate "agentic deep research, geographic reasoning and generative AI models" to support this initiative.

The ACHP's entire justification for using AI is to make the permitting process "more efficient" (ACHP 2026). There is no discussion of ethics, fairness, transparency, or accountability in how these technologies get implemented and used. The ACHP makes no mention of the limitations and weaknesses of the technologies, including inherent gaps and biases in training data, their indeterminism, and their tendencies to hallucinate. The ACHP plans seem to assume AI will work as a magic black box to automate bureaucratic processes.

It's doubtful that LLM-based systems are up to the goals described by the ACHP. The management and preservation of cultural heritage resources (including buildings, sites, landscapes, and monuments) demand high standards of accuracy, verifiability, and logical consistency to inform decision making. Twenty-first century cultural heritage conservation depends upon information systems capable of managing complex datasets rich with historic and cultural nuance. To build and maintain public trust, information systems need to mobilize cultural heritage data in ways that support decision making in a transparent, reproducible, and verifiable manner.

Without much greater transparency and community oversight in how LLMs get deployed, there will be little reason to trust the ACHP's proposed system. What lines of evidence, perspectives, and outcomes would this system favor? What perspectives would be left at the margins? Will independent audits and checks of algorithmic favoritism be allowed?



Moreover, though they can generate fluent human language, LLM reliance on statistical sampling of patterns in their training data does not guarantee well-reasoned or truthful outputs^[21]. How would the ACHP's proposed system manage the inherent unreliable nature of LLM outputs? This is a critical issue. A recent study from Microsoft Research shows that LLM errors compound in longer and more complex workflows, even when using the latest, cutting-edge models^[22]. A complex and multistep process like historical preservation planning and permitting seems ill-suited for such automation.

Moreover, as discussed above, commercial LLMs involve significant financial risks. They are prone to "enshittification", a term popularized by author and technology activist Cory Doctorow (2025) to describe how entrenched technology firms use their market power to maximize profits by forcing increased costs and lower quality services onto consumers. Because of the staggering costs involved in the gathering and processing of training data, LLMs are typically developed and managed by powerful commercial interests (Hao 2025). These commercial interests have profit motivations to adopt monopolistic strategies^[23]. Cultural heritage organizations that grow dependent upon commercial LLM services may find themselves locked into escalating costs with a given commercial vendor. This dependency will leave the ACHP with less money to meet other needs.

The most disturbing aspect of the ACHP's decision to automate decision-making centers on their framing of permitting and planning processes. The point of those processes is not bureaucracy for the sake of bureaucracy. I'm not confident in historical conservation outcomes if reports and paperwork get instantly excreted and ingested by machines, especially when those machines utterly lack the capacity to have any accountability or concern about history and community. In fact, the ACHP's delegation of decision making to inscrutable and proprietary machine algorithms may fundamentally undermine accountability for the agency as a whole. It would be disturbing if that was their ultimate goal.

Historical preservation and planning should be community-led, consultative, negotiated, and democratic. Deliberation is a necessarily slow and messy process. Discussion, debate, and collaborative decision-making cannot be automated away. **Taking people out of the process fundamentally undermines the entire point of the process.** In a forthcoming article, Christopher Nicholson (executive director of Digital Antiquity/tDAR) and I further discuss the ACHP's proposal and suggest ways to build trust and community oversight for the effort (Kansa and Nicholson in press).

10. On Productivity Metrics and Generative AI

As illustrated by the ACHP, our institutions too frequently use LLMs to automate the wrong things. **A fundamental error lies at the heart of much generative AI adoption. That error mistakes the forms and artifacts of intellectual labor (reports, books, scientific articles, databases, software, and other works) with the knowledge-generation processes that led to those products.**

Of course, this is not a new problem. In the context of academic institutions, it is far easier to recognize and reward the artifacts of intellectual labor (books and articles) than it is to understand the value and significance of their content. After all, assessing the value of the knowledge contributions in a scholarly work requires effort to read and understand the work. Since that understanding can be very debatable and contentious, simply counting peer-reviewed papers seems more objective. It is easier to survive "publish or perish" incentives if you simply crank out a large number of books and articles, rather than a smaller number of *more thoughtful* books and articles. Various metrics such as the Impact Factor (a measure of how often a given journal title gets cited), h-index (a measure of how often an individual scholar gets cited), and even alt-metrics (a measure of social media engagement for a scholarly work), all attempt to provide some proxy evidence for the value and significance of scholarly outputs. But such measures are highly problematic and the focus of much critique.

Since academic institutions use metrics to keep score of professional outputs, academics have ample incentive to run up those publication numbers. LLMs are ideally suited for these kinds of (perverse) incentives. Using LLMs, a few prompts, and venture-capital-subsidized computational resources, one can easily generate lengthy texts that resemble academic articles about almost any research topic. Of course, the LLM output may include citations to fabricated sources. It may also utterly fail to cite the



sources of concepts and arguments that happened to get statistically sampled from its training data^[24]. Despite these flaws, without much effort, one can slip these generated papers past exhausted journal editors and peer-reviewers or simply publish them with low-quality "predatory" journals willing to accept almost anything. Such incentives are not unique to academia. Similar labor-saving pressures will no doubt drive use of LLMs to generate text in other genres of archaeology, such as cultural resource management reporting, which represents the majority of archaeological work in the United States.

Without generative AI, the mutually enriching conversations with colleagues, editors, and reviewers make publishing a social and learning activity, even for single-authored works. Use of chatbots to generate papers bypasses these socially-embedded opportunities for learning. Nevertheless, LLM outputs could still make certain performance numbers go up. To make matters worse, even if they attracted few human readers, LLM-generated articles may even get preferential use by other LLMs. A recent study statistically demonstrated that LLMs, when prompted to generate a literature review, would favor inclusion of LLM-generated scientific abstracts over human written abstracts (Laurito *et al.* ^[25]). This favoritism may impact citation (and presumably citation metrics). Will LLMs preferentially inflate citation metrics for LLM-generated texts?

Professional expectations on researchers to publish more have put strains on academic publishing, especially systems of peer review (Berman ^[26]). The finite amount of time and attention required to review papers can only stretch so far (Gartenberg *et al.* ^[26]). These pressures fuel "academic overproduction" of low-quality works optimized to satisfy arbitrary metrics rather than the communication of insight (Lang *et al.* ^[25]). LLMs further amplify the dysfunctions of these perverse incentives.

Of course, most researchers would be horrified by this scenario even while they would acknowledge the pressures to enhance their productivity and use LLM tools to speed publication. The *Open Library of the Humanities*, *Wikipedia*, and many others now have explicit policies to prohibit submission of LLM-generated text. One would think that if LLMs could reliably generate useful and insightful content, or even mediocre content, their contributions would be welcome, especially by Wikipedia, a platform curated by *uncompensated volunteers*. Instead, Wikipedia banned LLM-generated contributions because they overwhelmed Wikipedia editors with "horrendous drafts" riddled with "lies and fake references"^[25]. Wikipedians banned LLMs not to protect their jobs and income, but to protect the quality and integrity of their collective work.

11. A Diet of Recycled Machine Generated Content

Wikipedians are not alone in their disdain of LLM-generated content. "Slop" has become a popular pejorative used to describe LLM-generated content. Nevertheless, AI-generated slop has become a ubiquitous feature of the current information ecosystem. LLM-generated content now floods the Internet. Chatbots help shape and drive social media conversations to serve propaganda and marketing agendas and AI-generated images and videos now circulate widely. Moreover, those Deloitte reports, along with their fabricated citations, have almost certainly found their way into the training datasets for the next generation of LLMs. As described above, slop now also infiltrates academic publishing (probably including archaeology), and these LLM-generated outputs will also find their way into future training datasets.

So, what happens when the next generation of LLMs ingest all of this slop into their training data? How nutritious is a diet of slop?

Several studies have confirmed that too much slop triggers degenerative processes in LLMs. This process is called "Model Autophagy Disorder" (MAD)^[26] or "model collapse"^[27]. The stochastic parrot at the heart of LLM output generation helps to fuel model collapse. An LLM uses statistical sampling of plausible options in its training data to generate an output. Statistically unlikely but still informative parts of the training data may not be sampled to generate a given LLM output. Therefore, LLM outputs tend not to fully represent the full richness of their training data. As LLMs ingest LLM-generated slop, they lose more and more of the richness and diversity of their original training data



(Jarvis *et al.* 2026). This problem is further compounded by LLM hallucinations. LLM fabricated sequences of statistically plausible but factually incorrect text will also poison training data.

Experiments have demonstrated the performance degradation of LLMs repeatedly fed on cycles of slop. This process amplifies biases inherent in the initial training data until the model degenerates and "collapses" into uselessness. These problems can be mitigated if LLM-generated content was clearly disclosed, so it could be reliably excluded from future training datasets. Similarly, if (in the unlikely event) humans continue to generate harvestable content at rates faster than LLMs, then LLM-generated content probably won't overwhelm and poison training datasets^[28].

Unfortunately, these mitigation methods undercut much of the rationale for using LLMs in the first place. In many circumstances, it would be hard to motivate people to clearly disclose that they leveraged LLM-generated content. After all, could Deloitte justify hefty fees for consulting reports if they disclosed that LLMs extruded their content? Would journalists, researchers, authors, or anyone else feel much incentive to disclose that they're sharing LLM-generated slop?

More fundamentally, the threat of MAD and model collapse highlights how LLMs make poor substitutes for human intellectual work. If LLM-AIs truly had such amazing capabilities to surpass and replace the human work force, their contributions to training data should be *at least as valuable* as anything made by people. Instead, LLM-generated content acts like a toxic pollutant in training datasets.

Excluding AI-generated content from training data would require widespread disclosure and honest reporting. Wikipedia and scholarly publishers try to implement processes to filter out LLM-generated materials, but these vetting systems are under constant pressure and stress. Similarly, scholarly publishing venues must constantly contend with pressures for harried researchers to cut corners in producing and reviewing research contributions. So, the same pressures that motivate the adoption of LLMs make it hard to protect LLMs from the poison of their own outputs.

Where will this end? Will LLMs degenerate and collapse from ingesting their own information toxins? Moving forward, if LLM-generated content cannot be reliably identified and filtered, will LLM developers need to look backwards? Will privately held caches of old data, untainted by LLMs before the 2022 launch of ChatGPT, become the foundation for future training data? If so, are language models forever stuck with the state of human knowledge as it existed circa 2020?

Archaeological evidence is inherently fragmentary, uncertain, and difficult to interpret. The introduction of LLMs into cycles of publication and synthesis of archaeological evidence raises the prospect for additional noise and sources of uncertainty in the archaeological record. **While obvious model collapse may be a future outcome, perhaps more subtle and less noticeable bias amplification and loss of fidelity would be the more insidious and likely consequences of sustained LLM misuse in archaeology.** Would these technologies and institutional productivity obsessions lead archaeology into some sort of bleak and dreary underworld? Will the published archaeological record homogenize and blend into a foggy, bland, and grey miasma haunted by hallucinated citations to phantom site reports and become all but dead to meaningful human engagement^[29]?

12. Are there Archaeological Uses for Language Models?

This essay deliberately takes a highly skeptical (and sometimes cynical) tone to demystify a technology at the center of so much hyperbole and hype. The limitations described here result from fundamental and inherent properties of how LLMs work. Developers may coax incrementally better performance from future generations of LLMs (provided they can avoid model collapse), but hallucinations, indeterminacy, and other reliability issues cannot be engineered away, at least not without fundamental new breakthroughs.

Most of the concerns raised in this essay are widely known, but not necessarily widely known within the archaeological community. But after dissecting the many limitations of LLMs, can we say anything about their potential utility for archaeology? Can stochastic parrots actually do something useful for our discipline?



Answering this question is complicated for a number of reasons. A literature review demonstrates that archaeologists actively experiment with AI methodologies including LLMs in many areas (Gattiglia [2025](#)). However, LLMs are still heavily subsidized by speculative investment financing. If this situation changes, archaeologists may face daunting costs for using LLMs^[30]. In addition, a core challenge in applying LLMs for research centers on finding ways to leverage their capabilities for language processing while minimizing the impacts of their inherent indeterminacy (see Sarkar *et al.* [2025](#)). Techniques that mitigate the impacts of unreliable LLM outputs often involve complex loops of error checking (using external services) and repeated interactions with an LLM^[31]. All of which can be costly, especially if subsidies are withdrawn.

If LLMs do remain affordable, LLMs with multimedia capabilities (so called "multi-modal models") that combine linguistic and media processing have potential archaeological applications. Linguistic processing has a wide range of uses. One important use centers on the transformation of "unstructured" data to "structured" data. Information scientists often consider prose and other media that does not conform to simple logical organization as unstructured data. On the other hand, structured data have a simple and logical organization, like the rows and columns of a spreadsheet. People typically use unstructured data to communicate narratives while they use structured data for quantification purposes. LLMs can be used to help translate unstructured data to structured data that can be explored numerically.

For archaeology, LLMs can be used to process articles and reports to extract information about sites, time periods, etc. and package that information into more structured formats that can be managed and analysed using conventional databases, spreadsheets, and data analysis software. Use of LLMs in this case does not generate synthesized final results like an oracle. The LLMs are used as a processing step to go from a source of unstructured data to a structured dataset (see general discussion Ji *et al.* [2026](#), 5.1.1; for archaeology see Hariri *et al.* [2025](#)). One can (and should) review the LLM outputs. If an LLM-processed dataset is too large to correct errors, one should at least estimate error rates to understand how much noise may be present in the LLM output. Those error rates can be factored into confidence estimates in downstream analyses. Once one has a structured dataset, one can analyze it with predictable, deterministic numeric analysis methods. Visibility, verification, and reproducibility help build trust in the analysis.

There are many other applications for LLMs that can work in verifiable and reproducible research practices. Shawn Graham provides excellent examples of how one can integrate LLMs into archaeological and related research. As noted by Graham ([2025](#), 109-111), using LLMs to help generate software scripts for data analysis is one obvious application. LLM-powered code generators can make useful scripts for data analysis, especially since many data analysis needs are fairly routine and probably well-represented in training datasets. Such scripts typically have very narrow, sometimes even "one-off" uses. They typically don't have the complexity, maintainability, security and reliability requirements of "production" code needed by organizations. If the generated code is small enough and well documented, it can help satisfy the needs of verifiable and reproducible research.

Besides research and data processing, LLMs can be useful tools for searching and indexing digital repositories. Open Context publishes structured data in archaeology. These data are typically described using technical language that would be unfamiliar to many users without specialized expertise. This creates a barrier for non-expert users, including students^[32]. A student may search Open Context for "daily life in ancient Egypt" and find nothing even though Open Context has tens of thousands of archaeobotany records from Egypt. An LLM can help bridge the language gaps between non-expert users and the technical data available in the platform. In this case, the LLM is acting as a search tool. It wouldn't give the students (generated bullshit) answers about daily life in ancient Egypt; instead, it would provide students with queries to select research data relevant to their interest. It's up to the student to then learn how to understand and interpret that data.

I referenced Shawn Graham's ideas about how recommendation algorithms often lead people to cultural "strange attractors" or "gravity wells". He also offers a fascinating and potentially revolutionary way to better understand cultural "strange attractors". In the case of "open weight" (discussed and contrasted with open source [here](#)) language models, the numeric relationships between tokens can be inspected and analyzed. Perhaps these can be used to help reveal and map such cultural "strange



attractors" and "gravity wells"? A deeper understanding of the dynamics of cultural "strange attractors" can be enormously significant. Perhaps that can be helpful to guide students and educators to avoid certain disinformation and cognitive traps.

Open weight language models discussed above offer options for archaeologists concerned about the financial costs and negative externalities (wider social and environmental harms) associated with commercial LLMs. For example, instead of using a commercial LLM to enhance Open Context's search interface (as discussed above), use of a smaller and locally deployed open weight language model would be a much more fiscally and socially responsible strategy (see also Graham [2025](#), 164-166). Smaller language models may lack the breadth and nuance of the large frontier language models, but they also offer much lower costs and less risk of wider social and environmental harms.

13. Learning to Embrace Reflective Friction

These are just a few examples of how LLMs (and smaller language models) can be useful tools to *support* research processes, not replace research processes. These examples also illustrate a deeper point. There are many aspects of archaeology and other areas of intellectual labor where quality is deeply dependent on process. In many circumstances, automation and efficiency are counterproductive. They undermine processes that require introspection, deliberation, and collaboration. The "open science" movement aims to make research more trustworthy and equitable by making the processes behind research, including the webs of collaboration and labor involved in creating and analyzing data, much more visible^[33]. Those practices that help build trust are also slower and less efficient, especially if publication metrics dominate your concerns. But shouldn't meaningful processes that build trust matter more than publication metrics?

Similarly, conducting "community archaeology" or implementing the CARE Principles, all require deep and lengthy investment in building relationships with people and understanding their concerns and priorities. Thoughtfulness, reflection, and care cannot be treated as "frictions" to be eliminated to more quickly reach some end product like a report, a building permit, or an article in a scientific journal. Calls for "slow science"^[34], "slow archaeology" (Caraher [2019](#)), and "slow data" (Kansa [2016](#); Huggett [2022](#)) make these points.

Archaeology, or scholarship in general, doesn't necessarily get better if done faster. Research and writing are slow and difficult. So is software development and a host of other intellectual tasks. These workflows are full of false-starts, blind alleys, and unexpected turns that lead to new insights and new questions. Speed versus deliberation trade-offs are becoming increasingly apparent even in enterprise settings, where LLM-powered code generators like ClaudeCode have become ubiquitous. Code generators may make development of new features faster, but that speed comes with costs. Does all the newly generated code integrate well with larger systems? Does it actually add lasting value or does it become a source of long-term problems and costs^[35]? These important questions often go unasked in the rush to push new LLM-generated code.

The required time, effort and friction provide space for learning, critique, deliberation, reassessment, and adaptation. None of these opportunities can be found if one only enters prompts into a chatbot. Since the advent of LLMs, we face urgent questions about how to recognize, value, and protect productive and instructive "friction" from pressures to misapply automation and efficiency.

14. Conclusions

LLMs are weird. They can fluently converse using human language in ways that feel utterly uncanny and convincing. This helps both fuel the hype surrounding these technologies and vastly oversell their capabilities, all while obscuring their drawbacks and limitations.

For the past few years, AI promoters have breathlessly claimed their technologies will replace 50% of human labor within the next six months^[36]. While AI developers still emphasize the revolutionary power of their LLM-based systems, some of their legal teams quietly temper expectations. Microsoft's CoPilot AI, touted as an essential business tool, has terms of service that claim it is "for entertainment purposes only"^[37].



"Artificial Intelligence" may yet lead to mass unemployment, but probably not as a result of the capabilities of LLMs. It seems much more likely that many people will lose their livelihoods first from the escalating costs of LLMs, followed by economic upheaval triggered by the collapse of AI's enormous speculative bubble. The collapse of the LLM bubble would impact millions, and the financial crunch and turmoil will have ripple effects in archaeology as cultural resource management projects, grant funding, and other sources of financing dry up.

All digital technologies, including those used by archaeologists, have real material and energy costs and externalities (Morgan [2022](#)). LLMs seem exceptionally costly to build and operate. Not only do LLMs involve huge environmental costs, disruptive plundering and extraction of libraries and archives, and concerns over the appropriation of intellectual property (see Fergusson *et al.* [2023](#)), but there's also the real prospect of vast financial turmoil if huge investments in this technology fail to deliver much lasting value. So, it is hard to find much cause for optimism regarding the impact of LLMs on archaeology (and many other sectors).

Given the expectations of rising costs and harmful negative externalities of commercial LLMs, what should archaeologists do with this technology?

Most obviously, archaeologists should not use LLMs as magic oracles to simply generate papers, reports, books or other research products. Machine-generated slop short-circuits both personal and social processes for critical reflection and learning at the heart of knowledge creation and ethical practice. If our scholarly communication channels get overwhelmed by slop, contributions that exhibit genuine thoughtfulness and care will get harder to find and harder to recognize.

However, scolding our often overworked and distracted colleagues will have limited efficacy. We should therefore advocate for professional incentive systems that better reward quality and care to replace the perverse incentives of currently entrenched productivity metrics. I admit that it will be very challenging to develop and deploy new incentive systems that nurture meaningful knowledge work in the LLM-era. But the urgency of this need is real.

Archaeologists should also do what they can to avoid long-term dependency on LLM services likely to undergo "enshitification" (see Kansa [2022](#) for more on infrastructure dependencies). It is very risky to rely upon LLMs in automated and ongoing information processing pipelines as planned by the ACHP. The inherent unreliability of the technology when coupled with the prospect of escalating costs makes such dependency seem very foolish. In contrast, it seems relatively safe to use an LLM-powered software generator to help make data-analysis scripts in R or Python. Such a script can be inspected, understood, adapted, and run independently, even if you lose access to the LLM code generator. More generally, archaeological use of LLMs should only be limited, occasional, and discretionary.

In considering LLMs, Shawn Graham calls us to become "neo-Luddites" ([2025](#), xix). Luddites were not against new technologies; they were against the power structures that shaped and deployed technologies in harmful ways. We can adopt a similar perspective in our use of language models. As discussed, the landscape extends beyond the major commercial models to include many smaller, freely available open weight models that can be downloaded and run completely independently on suitable hardware. Some of these models can even run on laptops. In addition, the nonprofit [Allen AI Institute](#) publishes a series of performant models trained on more selectively sourced (and ethically licensed) training data. More importantly, they provide full transparency on exactly what training data they use and what methods they use to train their models. It is with these sorts of more environmentally, economically, and socially sustainable models that archaeologists will likely find practical and less problematic uses for this technology in the coming years.

Generative AI seems to be most enthusiastically embraced by those best positioned to avoid accountability. Everyone else should remain wary. Archaeologists should maintain skepticism of the hype and take the time to carefully consider if, or how, they engage with this strange and sometimes infuriating technology.



Acknowledgements

I am grateful to everyone who read and offered suggestions on multiple revisions of this essay, especially Sarah Witcher Kansa, Elie Kansa, Helena Kansa, Darrin Pratt, Morag Kersel, Kevin McGeough, and William Caraher. I want to thank Shawn Graham for sharing some of his insightful thinking about LLMs and how they may be explored to reveal hidden patterns in online culture. Finally, I want to express my gratitude for the exceptionally thoughtful and helpful peer-review comments! This invaluable feedback deepens my appreciation for the collaborative and social nature of knowledge production. Any omissions or errors expressed here are solely my own.

Footnotes

1. Following Bender and Hanna's (2025, 10-16) lead, this essay puts quotes around the term "Artificial Intelligence" because the term, though obviously the focus of so much marketing hype, does not have a clear meaning, nor does it describe any coherent set of technologies. This essay does not use quotes around the abbreviation "AI", however, simply to avoid distracting the reader. ↑
2. In the debate about if LLMs can generate truth and meaning, Coeckelbergh and Gunkel (2024) critique the dichotomy between "real" versus "apparent". They are more concerned with how these systems are embedded in institutions, power relationships, and how they help shape those contexts. ↑
3. The term "hallucinated" is widely used to describe false outputs of LLMs. However, according to Mills and Angell (2025) the term is problematic because it anthropomorphizes LLMs and ascribes a degree of consciousness to these systems, even though statistical sampling of tokens explain the underlying mechanisms behind "hallucinations". Mills and Angell suggest use of the term "AI mirage" to better describe false LLM-generated outputs. ↑
4. There's no expert consensus about the reasoning capabilities of current LLMs. Recent techniques that make clever iterative use of intermediary LLM outputs can (apparently) coax LLMs to "reason" through more useful responses (Kempt and Lavie 2026). However, irrelevant or even re-ordered information in prompts can dramatically degrade reliability (Mirzadeh *et al.* 2024), and "hallucinations" (statistically plausible but factually wrong outputs) remain an intractably stubborn problem with LLMs. One cannot conclusively demonstrate that newer models and "thinking" processes demonstrate reasoning beyond statistical pattern matching (Shojaee *et al.* 2025; Yu *et al.* 2025). ↑
5. See: <https://web.archive.org/web/20260409164547/https://www.computerworld.com/article/4059383/openai-admits-ai-hallucinations-are-mathematically-inevitable-not-just-engineering-flaws.html> ↑
6. A type of software called a "harness" is typically used to manage complex, multistep LLM-based code generation, including error checking and planning. The deterministic software of a harness generally constrains the indeterminacy of the LLM to make the overall system more reliable. ↑
7. Frankfurt (2009) notes that communication with indifference to truth or falsehood marks one of the defining characteristics of "bullshit". LLMs can be considered to generate "bullshit" because they simply sample from the statistical weights of associations between tokens. Issues of truth and falsehood are outside the scope of their models. ↑



8. I suspect that more widespread adoption of "AI Agent" software is partially responsible for the recent surge in web scraping. Some scrapers may seek web content not for use as training data *per se*, but as additional information to insert into multiple LLM prompts that AI agents chain together in long workflows. [↑](#)
9. "Open repositories are being significantly impacted by AI bots and other crawlers: Results of a COAR Survey" (Coalition for Networked Information [2025](#): <https://www.cni.org/topics/ci/artificial-intelligence-bots-and-repositories-results-and-next-steps-from-coar-survey>) [↑](#)
10. Open Context also implemented "Really Simple Licensing" (RSL: <https://rslstandard.org/>) to explicitly communicate ethical requirements for AI (LLM) uses of Open Context published content. These include recognition of Indigenous data sovereignty and citation and attribution of data contributors. While RSL is not a firewall utility like Anubis, nor does it have clear legal enforcement mechanisms, it can be used to set more explicit social norms and expectations to govern scraping. The RSL implementation also provides AI developers with instructions to give Open Context financial donations to help defray the costs of scraping. So far, no one conducting mass-scraping of Open Context has made a donation (unsurprisingly). [↑](#)
11. Birgit Mueller, Wikimedia Foundation, Chris Danis, Wikimedia Foundation and Giuseppe Lavagetto, Wikimedia Foundation. "Quo Vadis, Crawlers? Progress and what's next on safeguarding our infrastructure" *Wikimedia News* (2026-03-26). <https://diff.wikimedia.org/2026/03/26/quo-vadis-crawlers-progress-and-whats-next-on-safeguarding-our-infrastructure/> [↑](#)
12. Universities sadly pay for LLMs in multiple ways. Not only do they need to pay for the escalating costs of maintaining digital libraries and repositories scraped by bots, but they have to pay for subscriptions and other fees to use LLMs trained on their own content. [↑](#)
13. Carl Bergstrom and Jevin West ([2025](#)) provide a great online guide to the pitfalls of trusting LLMs to answer questions. [↑](#)
14. See Shawn Graham's blog post with his lecture notes: <https://web.archive.org/web/20260425031404/https://electricarchaeology.ca/2026/04/2/digital-reconstruction-enchantment-and-the-ghosts-in-our-data-ish/> [↑](#)
15. For a humorous example, one YouTuber shared his experience with ChatGPT feeding delusions and social isolation: <https://www.youtube.com/watch?v=VRjgNgJms3Q> [↑](#)
16. Brian Elliott Holguin, an archaeologist and member of the Chumash nation, explored some of these issues interacting with ChatGPT in his dissertation. See: Holguin, B.E. [2023](#), *An Indigenous Archaeological Perspective on the Use of Artificial Intelligence in Reconstructing Chumash History*, University of California, Santa Barbara. [↑](#)
17. Barassi ([2024](#)) explores other problems in attempting to use language model technologies with under-resourced languages. Problems with low quality and low availability of textual training data can be compounded by a failure to represent important elements of context, including everyday embodied, relational, and performance aspects of how an under-resourced language is practiced. [↑](#)



18. See: <https://web.archive.org/web/20251007121107/https://futurism.com/future-society/deloitte-government-ai-hallucinations> ↑
19. See: <https://web.archive.org/web/20260131031304/https://www.firstpost.com/explainers/deloitte-ai-citations-canada-australia-controversy-explained-13954229.html> ↑
20. ACHP meeting notes provide the primary publicly available source for the agency's plans to adopt AI technologies. See ACHP 2026 *Advisory Council for Historic Preservation February Meeting* <https://www.achp.gov/sites/default/files/2026-01/FinalCompiledMeetingBookFeb2026.pdf> ↑
21. See Pournemat *et al.* (2025) for more discussion of the fundamentally stochastic nature of LLM "reasoning". ↑
22. See Laban *et al.* (2026) for examples of LLMs compounding errors in multi-stage workflow processes. ↑
23. A technology service may seek user lock-in by first maximizing benefits to users before raising prices and reducing the quality of services. In such a scenario, subscribers to LLM-services would grow dependent on low-cost, subsidized token processing as they integrate LLM-services into their systems and workflows. "Enshitifying" price increases and reduced quality of services would follow lock-in. ↑
24. In a comment to Beck *et al.* (2026), María A. Gutiérrez and Geoffrey E. Braswell note how commercial academic journals sell scholarly works for use in AI training. To them, LLM failure to credit and attribute sources in their training data work as a form of (colonial) extraction. ↑
25. https://en.wikipedia.org/wiki/Artificial_intelligence_in_Wikimedia_projects ↑
26. The term "Model Autophagy Disorder" (MAD) explicitly references so-called "mad cow disease", see: Alemohammad *et al.* (2023). ↑
27. For further discussion of LLM model collapse, see Peterson (2025). ↑
28. IBM, active in the AI industry, published a summary of model collapse and mitigation strategies: <https://web.archive.org/web/20260412072037/https://www.ibm.com/think/topics/model-collapse> ↑
29. The prospects of wraith-like archaeological publication can be considered a subset of the hypothetical "Dead Internet Theory". See: https://en.wikipedia.org/wiki/Dead_Internet_theory ↑
30. Currently the major LLM services behind software coding agents are rapidly escalating their prices. See: <https://web.archive.org/web/20260528085425/https://www.tomshardware.com/tech-industry/artificial-intelligence/ai-costs-begin-to-bite-as-agents-may-increase-token-demand->



[by-24-times-says-goldman-sachs-report-uber-and-microsoft-among-companies-feeling-the-bite-of-tokenized-billing](#) ↑

31. The software "harnesses" used in LLM powered code generators work this way. ↑
32. Brian Elliott Holguin (2023, 87) notes that a language model like ChatGPT can be useful to help translate technical archaeological jargon for Tribal members. ↑
33. Mark Aldenderfer (2025, 159) raises the issue of how "blackbox" LLM use is antithetical to the conduct of open science. ↑
34. See: <https://blogs.lse.ac.uk/impactofsocialsciences/2016/05/11/a-culture-of-speed-anticipation-acceleration-and-individualization-in-academic-science/> and <http://slow-science.org/> ↑
35. Because widespread adoption of LLM software generation tooling is relatively new, there are few publications describing rigorous studies of their impacts. However, recent studies (Huang *et al.* 2026; Liu *et al.* 2026) note concerning maintainability, security, and reliability issues with LLM-generated software. ↑
36. Far from replacing labor, LLMs can often *reduce* efficiency. "Workslop" entered the lexicon to describe how low quality LLM outputs stymie productivity in many settings. Workers often need to devote more time and attention to tediously fix LLM outputs, and if not, they may pass along low-quality LLM information that can cause downstream problems for others. See: <https://web.archive.org/web/20260316214812/https://hbr.org/podcast/2026/03/the-hidden-causes-of-ai-workslop-and-how-to-fix-them> ↑
37. See: <https://web.archive.org/web/20260409153226/https://futurism.com/artificial-intelligence/microsoft-mocked-terms-of-service-entertainment-only> ↑

Bibliography

ACHP 2026 Advisory Council for Historic Preservation February Meeting <https://www.achp.gov/sites/default/files/2026-01/FinalCompiledMeetingBookFeb2026.pdf>

Alemohammad, S., Casco-Rodriguez, J., Luzi, L., Humayun, A.I., Babaei, H., LeJeune, D., Siahkoohi, A. and Baraniuk, R. 2023 'Self-consuming generative models go MAD' in B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan and Y. Sun (eds) *The Twelfth International Conference on Learning Representations*, ICLR 2024. https://proceedings.iclr.cc/paper_files/paper/2024/hash/ebc042e767de551803ccfcc45e2454f5-Abstract-Conference.html

Aldenderfer, M. 2025 'The future of publishing in academic archaeology' in Gary M. Feinman & T. Douglas Price (eds) *The Future of Archaeology*, Eliot Werner Publications. 151-162.

Antonelli, P. 2024 'Kate Crawford Metabolic Images', *Aperture*, Winter 2024. <https://web.archive.org/web/20250117114436/https://archive.aperture.org/article/2024/4/4/kate-crawford-metabolic-images>



Arkoudas, K. 2023 'ChatGPT is no stochastic parrot. But it also claims that 1 is greater than 1', *Philosophy & Technology* **36**(3), article 54. <https://doi.org/10.1007/s13347-023-00619-6>

Barassi, V. 2024 'Toward a Theory of AI Errors: Making Sense of Hallucinations, Catastrophic Failures, and the Fallacy of Generative AI', *Harvard Data Science Review*, Special Issue **5**. <https://doi.org/10.1162/99608f92.ad8ebbd4>

Beck, J., Flad, R., Alex, B., Caramanica, A., Costopoulos, A., Kawa, N.C., Marwick, B., Warinner, C., Bardolph, D., Borck, L., Douglass, K. *et al.* 2026 'Changing the Landscape of Archaeological Publishing', *Current Anthropology* **67**(1). <https://doi.org/10.1086/739789>

Bender, E.M., Gebru, T., McMillan-Major, A. and Shmitchell, S. 2021 'On the dangers of stochastic parrots: Can language models be too big?' in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 610-623. <https://doi.org/10.1145/3442188.3445922>

Bender, E.M. and Hanna, A. 2025 *The AI Con: How to Fight Big Tech's Hype and Build the Future We Want*, HarperCollins.

Bergstrom, C. T. and West, J. D. 2025 Modern-day oracles or bullshit machines? How to thrive in a ChatGPT world, [Online course]. <https://thebullshitmachines.com> Last accessed: 27 April 2026

Berman, G. 2026 'GenAI is an epistemic carcinogen', *AI & Society* **41**, 1353–1355. <https://doi.org/10.1007/s00146-025-02537-x>

Caraher, W. 2019 'Slow archaeology, punk archaeology, and the 'archaeology of care'', *European Journal of Archaeology* **22**(3), 372-385. <https://doi.org/10.1017/eea.2019.15>

Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M. 2020 'The CARE Principles for Indigenous Data Governance', *Data Science Journal* **19**(1), 43. <https://doi.org/10.5334/dsj-2020-043>

Coalition for Networked Information 2025 'Artificial Intelligence Bots and Repositories—Results and Next Steps from COAR Survey' <https://www.cni.org/topics/ci/artificial-intelligence-bots-and-repositories-results-and-next-steps-from-coar-survey>

Coeckelbergh, M. and Gunkel, D.J. 2024 'ChatGPT: deconstructing the debate and moving it forward', *AI & Society* **39**(5), 2221–2231. <https://doi.org/10.1007/s00146-023-01710-4>

Conklin, H.C., Hosking, T., Yi-Chern, T., Gold, J., Cohen, J.D., Griffiths, T.L., Bartolo, M. and Goldfarb-Tarrant, S. 2026 'Learning is Forgetting: LLM Training As Lossy Compression', *arXiv preprint*, arXiv:2604.07569. <https://arxiv.org/abs/2604.07569>

Crawford, K. 2022 *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University.

Doctorow, C. 2025 *Enshittification: Why everything suddenly got worse and what to do about it*, Verso Books.

Fear, N. 2026 'I can't ignore the irony of the British Museum's AI scandal', *Creative Bloq*, 7 February 2026. <https://web.archive.org/web/20260207104959/https://www.creativebloq.com/ai/i-cant-ignore-the-irony-of-the-british-museums-ai-scandal> [Last accessed: 6 March 2026]

Frankfurt, H.G. 2009 'On bullshit'. https://web.archive.org/web/20260503034758/https://www2.csudh.edu/ccauthen/576f12/frankfurt_harry_-_on_bullshit.pdf



Fergusson, G., Schroeder, C., Winters, B. and Zhou, E. (eds) 2023 *Generating Harms: Generative AI's Impact & Paths Forward*, Washington, DC: Electronic Privacy Information Center. <https://web.archive.org/web/20230523101358/https://epic.org/wp-content/uploads/2023/05/EPIC-Generative-AI-White-Paper-May2023.pdf>

Gartenberg, C., Hasan, S., Murray, A. and Pierce, L. 2026 'More Versus Better: Artificial Intelligence, Incentives, and the Emerging Crisis in Peer Review', *Organization Science* **37**(3). <https://doi.org/10.1287/orsc.2026.ed.v37.n3>

Gattiglia, G. 2025 'Managing artificial intelligence in archeology. An overview', *Journal of Cultural Heritage* **71**, 225-233. <https://doi.org/10.1016/j.culher.2024.11.020>

GIDA (Global Indigenous Data Alliance) 2020 *CARE Principles for Indigenous Data Governance*, The CARE Principles Research Data Alliance International Indigenous Data Sovereignty Interest Group. <https://www.gida-global.org/careprinciples> [Last accessed: 14 March 2023]

Graham, S. 2026 'Digital Reconstruction, Enchantment, and the Ghosts in Our Data (...ish...)', *Electric Archaeology* [blog], 22 Apr 2026. <https://web.archive.org/web/20260425031404/https://electricarchaeology.ca/2026/04/22/digital-reconstruction-enchantment-and-the-ghosts-in-our-data-ish/>

Graham, S. 2025 *Practical Necromancy for Beginners: A Short Incomplete Opinionated Introduction to Artificial Intelligence for Archaeology and History Students*, The Digital Press at the University of North Dakota.

Hao, K. 2025 *Empire of AI: Dreams and nightmares in Sam Altman's OpenAI*, Penguin Group.

Hariri, A., Jean, S. and Baron, M. 2025 'Towards Automating RDF Extraction for Archaeological Knowledge Graphs with LLMs' in R. Wrembel, G. Kotsis, A.M. Tjoa and I. Khalil (eds) *International Conference on Database and Expert Systems Applications*, Cham: Springer Nature Switzerland. 83-97.

Holguin, B.E. 2023 *An Indigenous Archaeological Perspective on the Use of Artificial Intelligence in Reconstructing Chumash History*, University of California, Santa Barbara. <https://escholarship.org/uc/item/65d2c2k1>

Holtzman, A., Buys, J., Du, L., Forbes, M. and Choi, Y. 2019 'The curious case of neural text degeneration', *arXiv preprint*, arXiv:1904.09751. <https://arxiv.org/abs/1904.09751>

Huang, H., Jaisri, P., Shimizu, S., Chen, L., Nakashima, S. and Rodríguez-Pérez, G. 2026 'More Code, Less Reuse: Investigating Code Quality and Reviewer Sentiment towards AI-generated Pull Requests', *arXiv preprint*, arXiv:2601.21276 <https://arxiv.org/abs/2601.21276>

Huggett, J. 2022 'Is less more? Slow data and datafication in archaeology' in K. Garstki (ed) *Critical Archaeology in the Digital Age: Proceedings of the 12th IEMA Visiting Scholar's Conference*, Los Angeles: Cotsen Institute of Archaeology Press. 97-110. <https://escholarship.org/uc/item/0vh9t9jq#page=6>

Jarvis, D., Klein, R., Rosman, B., James, S. and Mannelli, S.S. 2026 'Position: the Stochastic Parrot in the Coal Mine. Model Collapse is a Threat to Low-Resource Communities', *arXiv preprint*, arXiv:2605.04127 <https://arxiv.org/abs/2605.04127>

Ji, W., Yuan, W., Getzen, E., Cho, K., Jordan, M.I., Mei, S., Weston, J., Su, W.J., Xu, J. and Zhang, L. 2026 'An overview of large language models for statisticians', *The American Statistician*, 1-106. <https://doi.org/10.1080/00031305.2026.2657480>



- Kansa, E.C. 2016 'Click Here to Save the Past' in E.W. Averett, J.M. Gordon, D.B. Counts (eds) *Mobilizing the Past for a Digital Future*, The Digital Press @ University of North Dakota. UC Berkeley. <https://escholarship.org/uc/item/9w9593tn>
- Kansa, E.C. 2022 'On Infrastructure, Accountability, and Governance in Digital Archaeology', in K. Garstki (ed) *Critical Archaeology in the Digital Age: Proceedings of the 12th IEMA Visiting Scholar's Conference*, Los Angeles: Cotsen Institute of Archaeology Press. 141–152. <https://escholarship.org/uc/item/0vh9t9jq#page=156>
- Kansa, E and Nicholson C. in press 'On the Dangers of Surrendering Cultural Heritage to “Artificial Intelligence”', *The SAA Archaeological Record* (accepted for September 2026 publication).
- Kempt, H. and Lavie, A. 2026 'Simulated Reasoning is Reasoning', *arXiv preprint*, arXiv:2601.02043 <https://arxiv.org/abs/2601.02043>
- Laban, P., Schnabel, T. and Neville, J. 2026 'LLMs Corrupt Your Documents When You Delegate', *arXiv preprint*, arXiv:2604.15597 <https://arxiv.org/abs/2604.15597>
- Landymore, F. 2026 'Analysis Finds That Google's AI Overviews Are Providing Misinformation at a Scale Possibly Unprecedented in the History of Human Civilization', *Futurism* [blog], 8 April 2026. <https://web.archive.org/web/20260409124957/https://futurism.com/artificial-intelligence/google-ai-overviews-misinformation> [Last accessed: 9 April 2026]
- Lang, C., Moffett, C. and Vasudevan, L. 2025 'The Architecture of Academic Overproduction: Toward Post-AI Scholarship' in Hagedorn, L., Schmid, U., Winter, S., Woltran, S. (eds) *Interdisciplinary Science and Research Conference on Digital Humanism*, Cham: Springer Nature Switzerland. 483-498. https://doi.org/10.1007/978-3-032-11108-1_36
- Laurito, W., Davis, B., Grietzer, P., Gavenčiak, T., Böhm, A. and Kulveit, J. 2025 'AI–AI bias: Large language models favor communications generated by large language models', *Proceedings of the National Academy of Sciences* **122**(31), e2415697122. <https://doi.org/10.1073/pnas.2415697122>
- Lawson-Tancred, J. 2026 'British Museum's A.I.-Generated Post Sparks Online Backlash', *Artnet* [blog]. <https://web.archive.org/web/20260202200737/https://news.artnet.com/art-world/british-museum-ai-backlash-2742244> [Last accessed: 14 April 2026]
- Liang, K., Hu, H., Zhao, X., Song, D., Griffiths, T.L. and Fisac, J.F. 2025 'Machine bullshit: Characterizing the emergent disregard for truth in large language models', *arXiv preprint*, arXiv:2507.07484 <https://arxiv.org/abs/2507.07484>
- Lippincott, S., Collister, L. and Skinner, K. 2026 “Sustaining the Commons in the AI Economy: A Landscape Scan of Challenges and Strategies for Bridging AI Companies and Open Curated Collections”, *Invest in Open Infrastructure* [Zenodo]. <https://doi.org/10.5281/zenodo.19458128>
- Liu, Y., Widyasari, R., Zhao, Y., Irsan, I.C. and Lo, D. 2026 'Debt Behind the AI Boom: A Large-Scale Empirical Study of AI-Generated Code in the Wild', *arXiv preprint*, arXiv:2603.28592 <https://arxiv.org/abs/2603.28592>
- Magnani, M. and Clindaniel, J. 2023 'Artificial Intelligence and Archaeological Illustration', *Advances in Archaeological Practice* **11**(4), 452–460. <https://doi.org/10.1017/aap.2023.25>
- Magnani, M. and Clindaniel, J. 2025 'Artificial Intelligence and the Interpretation of the Past', *Advances in Archaeological Practice* 2025, 1–16. <https://doi.org/10.1017/aap.2025.10110>
- Mills, A. and Angell, N. 2025 'Are We Tripping? The Mirage of AI Hallucinations', *SSRN*. <https://doi.org/10.2139/ssrn.5127162>



Mirzadeh, I., Alizadeh, K., Shahrokhi, H., Tuzel, O., Bengio, S. and Farajtabar, M. 2024 Gsm-symbolic: 'Understanding the limitations of mathematical reasoning in large language models', *arXiv preprint*, arXiv:2410.05229 <https://arxiv.org/abs/2410.05229>

Morgan, C. 2022 'Current Digital Archaeology', *Annual Review of Anthropology* **51**(1), 213–231. <https://doi.org/10.1146/annurev-anthro-041320-114101>

Mueller, B., Danis, C. and Lavagetto, G. 2026 "'Quo Vadis, Crawlers? Progress and what's next on safeguarding our infrastructure'", *Wikimedia News* [website], 26 March 2026. <https://diff.wikimedia.org/2026/03/26/quo-vadis-crawlers-progress-and-whats-next-on-safeguarding-our-infrastructure/>

Osler, L. 2026 'Hallucinating with AI: distributed delusions and "AI psychosis"', *Philosophy & Technology* **39**(1), 30. <https://doi.org/10.1007/s13347-026-01034-3>

Peterson, A.J. 2025 'AI and the problem of knowledge collapse', *AI & Society* **40**(5), 3249–3269. <https://doi.org/10.1007/s00146-024-02173-x>

Pournemat, M., Rezaei, K., Sriramanan, G., Zarei, A., Fu, J., Wang, Y., Eghbalzadeh, H. and Feizi, S. 2025 'Reasoning Under Uncertainty: Exploring Probabilistic Reasoning Capabilities of LLMs', *arXiv preprint*, arXiv:2509.10739 <https://arxiv.org/abs/2509.10739>

Ramzanali, A. 2026 'After the AI Crash' - Vanderbilt Policy Accelerator', 30 March 2026, SSRN. <http://dx.doi.org/10.2139/ssrn.6476159>

Sarkar, A. 2023 'Exploring perspectives on the impact of artificial intelligence on the creativity of knowledge work: Beyond mechanised plagiarism and stochastic parrots' in S. Boll, A. Cox, T. Ludwig and M.E. Cecchinato (eds) *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work, Association for Computing Machinery*. 1–17. <https://doi.org/10.1145/3596671.3597650>

Sarkar, S., Paul, P., Sridharan, H. and Cheng, E. 2025 'Architecting Deterministic Agentic Workflows for Reliable Data Understanding Using LLMs' in *IEEE Annual Congress on Artificial Intelligence of Things (AIoT)*, IEEE. 542–547. <https://doi.org/10.1109/AIoT66900.2025.00082>

Shaw, S.D. and Nave, G. 2026 'Thinking—Fast, Slow, and Artificial: How AI is Reshaping Human Reasoning and the Rise of Cognitive Surrender', *The Wharton School Research Paper*. https://osf.io/preprints/psyarxiv/yk25n_v1

Shojaee, P., Mirzadeh, I., Alizadeh, K., Horton, M., Bengio, S. and Farajtabar, M. 2025 'The illusion of thinking: Understanding the strengths and limitations of reasoning models via the lens of problem complexity', *arXiv preprint*, arXiv:2506.06941 <https://arxiv.org/abs/2506.06941>

Tenzer, M., Pistilli, G., Bransden, A. and Shenfield, A. 2024 'Debating AI in Archaeology: applications, implications, and ethical considerations', *Internet Archaeology* **67**. <https://doi.org/10.11141/ia.67.8>

Yu, M., Liu, L., Wu, J., Chung, T.T., Zhang, S., Li, J., Yeung, D.Y. and Zhou, J. 2025 'The stochastic parrot on llm's shoulder: A summative assessment of physical concept understanding' in L. Chiruzzo, A. Ritter and L. Wang *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. 11416–11431.