Advances in Archaeological Practice (2026), **14**, 218–233
doi:10.1017/aap.2025.10110

SAA
Society for American Archaeology

ARTICLE

Artificial Intelligence and the Interpretation of the Past

Matthew Magnani[1] and Jon Clindaniel[2]

[1]Department of Anthropology and the Climate Change Institute, University of Maine, Orono, ME, USA and [2]Chicago Center for Computational Social Science, University of Chicago, Chicago, IL, USA
Corresponding author: Matthew Magnani; Email: matthew.magnani@maine.edu

(First published online 18 December 2025)

Abstract
Artificial intelligence is reshaping the contemporary world. Trickling deeper into archaeology and history, these technological changes will influence how the past is written about and visualized. Through the evaluation of text and images generated using AI, this article considers the systemic biases present in reconstructed archaeological scenes. We draw on advances in computer science, running large-scale, computational analyses to evaluate patterns in content. We present a case study examining Neanderthal behavior, juxtaposing published archaeological knowledge with images and text made using AI. Our study reveals a low correspondence between scientific literature and artificially intelligent material, which reflects dated knowledge and cultural anachronisms. Used to identify patterns in (mis)representations of the past, the methodology can be applied to understand the distance between scholarly knowledge and any domain of content generated using AI, across any archaeological time depth and beyond the discipline.

Resumen
La inteligencia artificial está transformando el mundo contemporáneo. Al expandirse hacia la arqueología y la historia, estos cambios tecnológicos influirán en la manera en que se escribe y se visualiza el pasado. Mediante la creación de textos e imágenes, generados por inteligencia artificial, este artículo examina los sesgos sistémicos presentes en la reconstrucción de escenas arqueológicas. Desarrollamos enfoques cada vez más frecuentes en la arqueología que se basan en avances de las ciencias computacionales, realizando análisis computacionales a gran escala para evaluar patrones en textos e imágenes. Presentamos un estudio de caso sobre el comportamiento neandertal, en el que contrastamos el conocimiento arqueológico publicado con imágenes y textos generados por IA. Nuestro estudio revela una baja correspondencia entre la literatura científica y los contenidos producidos mediante inteligencia artificial, los cuales reflejan conocimientos desactualizados y anacronismos culturales. Esta metodología, empleada para identificar patrones en las (mal)representaciones del pasado, puede aplicarse para comprender la distancia entre el conocimiento académico y cualquier tipo de contenido generado con IA, en cualquier profundidad temporal arqueológica e incluso más allá de la disciplina.

Keywords: bias; computational archaeology; generative artificial intelligence; neanderthals

Palabras clave: sesgo; arqueología computacional; inteligencia artificial generativa; neandertales

Artificial intelligence (AI) is poised to reshape the world around us. Original material ranging from audio to visual and textual content may be created in the likeness of human-made recordings, paintings, or writing. Artificially intelligent analyses promise to process and make sense of the surrounding world in new and more effective ways. Scholars from disciplines spanning the hard and social sciences are investing time and resources to understand how diverse tasks from pedagogy to art and surgery will be facilitated or reimagined.

© The Author(s), 2025. Published by Cambridge University Press on behalf of Society for American Archaeology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press

Although AI may revolutionize contemporary visions of medicine, art, and aesthetics, its application also promises to reinscribe standing power structures by replicating pervasive inequalities and stereotypes. Racial, gender-based, and regional biases threaten being baked into emergent programs (Fang et al. 2024; Roselli et al. 2019). Trained on weird datasets by weird researchers (sensu Heinrich et al. 2010), applications of the technology will further entrench dominant social biases that influence patients' healthcare (Alowais et al. 2023), the way students are taught (Williamson 2024), and how diverse cultures are written about and seen. Even if the use of AI is rife with opportunity, problems that await on the horizon are equally ubiquitous and should command attention from software developers, academics, and legislators alike.

How do the rewards and risks of these technologies reverberate across anthropology and archaeology? Technological transformations will affect not only contemporary social worlds but also the way we represent, reflect, and engage with the past. Like the broader context in which our field is adopting AI, the benefits and drawbacks of the technology should be carefully evaluated. On the one hand, AI will increase the accessibility of representing archaeological material, and those who lack specialized skills in illustration or writing will be able to easily imagine alternative pasts (see Magnani and Clindaniel 2023). On the other hand, these narrations of the record may reinforce stereotyped depictions based on either antiquated or misguided content. Representing the past will become more accessible to a broader cross section of society but at the risk of proliferating misinformation.

Archaeologists must evaluate these implications to proactively shape the foundations of our engagement with this new technology. This article uses generative AI (the subset of artificial intelligence focused on producing novel output) to create large samples of text and images of archaeological scenes and compares them to a larger corpus of archaeological knowledge derived from published journal articles using a multimodal encoding approach. We specifically examine a well-established literature on *Homo neanderthalensis*, a species whose depiction has changed significantly since the founding of our discipline (Nowell 2023). Building on our earlier work that addressed workflows for archaeological illustration using DALL-E 2 (see Magnani and Clindaniel 2023), this article uses computational approaches to text and images to systematically address the overlap between content generated using AI and histories of anthropological archaeological knowledge.

Bias and Ethics in Artificial Intelligence

Generative AI platforms are trained using vast digital repositories of images and text. Because of this broad input, the resulting content reflects not only dominant and majority ideologies but also biases widespread across society. Concerns about gendered and racialized AI output has received significant scholarly attention (Nazer et al. 2023; O'Conner and Liu 2023). In addition to the problem of implicit bias, groups of individuals are actively trying to circumvent the limitations imposed by AI developers to generate violent, pornographic, or racist content. On major platforms like ChatGPT or DALL-E, developers have built in safeguards to limit the types of images and texts that might be generated. For instance, restricted text content includes political discussions, and filters and user agreements are designed to prevent prompts that generate bigoted or graphic material (OpenAI Usage Policies). In other cases, there has been criticism of corporate overcorrection, which has yielded historically inaccurate visual depictions (see Google Communications 2024). Such problems have attracted the attention of state legislators and Congress in the United States, as well as politicians globally. In Europe, for example, more sweeping legislation like the EU AI act seeks to establish risk profiles for the development and application of AI (Pehlivan 2024).

Although some of the more severe cases of bias in AI-generated content pertaining to political orientation, racism, or sexism have begun to be addressed, other types of social biases are equally pervasive but less attention-grabbing. While programmers and companies are quicker to remedy clearly gendered or racist depictions, other widespread misconceptions diffused throughout society risk blind reproduction, tinting our perceptions based on outdated science or ideologies. Artificially intelligent depictions of the past may escape the immediate attention of moderators, but they have broad social implications for contemporary communities around the world. We follow the example of archaeologists like Huggett

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press

(2015) who have argued for taking a proactive, "introspective" stance to unpack the influence of digital technologies on our field.

Archaeology and Artificial Intelligence

Studies using AI in archaeology predate the recent boom in widely accessible commercial AI tools. Based on human-trained models, archaeologists regularly identify sites and artifactual scatters from satellite and drone-derived imagery using techniques including object-based image analysis. These applications have roots that span back over two decades (Davis 2019). Such studies have been successful in identifying a range of archaeological subjects from large archaeological sites to rock art (Horn et al. 2022), artifact scatters, and stratigraphy (Mesanza-Moraza et al. 2020). Machine-learning tools have sped up the process of human identification of sites and the analysis of features (Camara et al. 2023) and locations where looting has taken place (Altaweel et al. 2024).

Based on longer-term patterns of machine learning, a rapid expansion in AI archaeology occurred in 2022 following the release of commercially available programs. Dirk Spennemenn has contributed significantly to this literature, addressing the impacts of accessible AI on broader issues ranging from the cultural heritage of COVID (Spennemenn 2023a) to a critical evaluation of ChatGPT source material (likely Wikipedia) and errata (Spennemen 2023b). Additional noteworthy work considers the impacts of artificial historical representations on Indigenous communities themselves (Holguin 2023). More broadly, Peter Cobb (2023) highlights several potential uses of AI for archaeologists—from education to illustration—that are only just beginning to be explored, establishing themes that we addressed in our study using AI for archaeological illustration (Magnani and Clindaniel 2023). Most recently, Gabrielle Gattiglia (2025) considers trends across an expanding body of applications while foregrounding conversations on the ethics of data transparency alongside the power of knowledge production and interpretation. Despite growing consideration by archaeologists, the impacts and potentials of broadly available AI tools for our field are only beginning to come into focus.

Artificial Intelligence and Its Convergence with Big Data

Along with the growing interest and investment in AI, humanity's digital footprint expands with each passing year. Digital media once measured in bytes have ballooned in size trillions of times, and personal devices that store terabytes of data are becoming ubiquitous. Driven by commerce, our movements, purchases, and social exchanges are recorded with increasing frequency and precision. Tracing but also informing these broader technological trends, scholarly approaches to big data have emerged across the hard and social sciences. Armed with a tool kit developed by computer scientists with the capacity to parse terabytes of complex data, scholars are shedding light on human behavior at new scales.

Anthropologists and archaeologists have been quick to capture and analyze big datasets to explain relationships between people and things and between people and the past more broadly. Bonacchi and colleagues have conducted extensive analyses of social media posts on Facebook and Twitter, revealing complex perceptions of heritage as they relate not only to contemporary political events like Brexit but also to particular archaeological remains (Bonacchi and Krzyzanska 2021; Bonacchi et al. 2018). Considering material culture discarded on curbs across the United States, we examined the formation processes of data, analyzing how abandoned digital materials reflect patterns of human behavior (Clindaniel and Magnani 2024). Scholars including Altaweel and Hadjitofi (2020) have conducted a close examination of popular vendor platforms like Ebay, analyzing the provenience of archaeological materials being sold on global markets. We also analyzed face mask production across the pandemic—scraping data from the craft makers' website Etsy—and revealed how politics informed mask production and efficacy throughout the COVID-19 pandemic (Magnani et al. 2022). Similarly, recent research on the Inka khipu—the knot-and-cord recording system of the Indigenous Andes—has applied deep-learning strategies to explore semantic variation and genre across large-scale digital khipu datasets, offering new insights into the ways in which meaning was encoded and structured in the Andean past (Clindaniel 2025). Across these studies, authors have sought to examine age-old questions at scale using new digital media.

AI is poised to contribute growing quantities of data to the online material and social worlds of interest to archaeologists and anthropologists. Within a few years, it is possible that most online content

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press

will be generated using AI. The analytical tools that scholars have developed to approach big data will become crucial to understanding the social implications of this corpus.

Case Study

To examine the intersection between published archaeological knowledge and content created about the archaeological record using AI, we selected a case study supported by more than a century of academic inquiry and a robust historiography (Madison 2021; Nowell 2023). Beginning with their scientific discovery in the nineteenth century in the Neander Valley in Germany, Neanderthals have captured the attention of archaeologists and the broader public. The first descriptions of Neanderthal skeletal remains in the 1860s painted a picture of a robust and primitive human-related species, with some cranial features more like those of chimpanzees than of *Homo sapiens* (King 1864). By the end of the nineteenth century the image of "a crude prototype of our own species" gelled into the popular and academic imagination (Peeters and Zwart 2020). Early twentieth-century dioramas at the Museum of Natural History in New York and Field Museum in Chicago reflect a persistence of this thinking decades later, representing hunched-over humanoids with simple clothing, familiar looking but clearly not of our own species.

By the 1950s and more recently, these brutish depictions were subject to increasing revision, and Neanderthal behavioral complexities figured prominently in scientific literature and the popular imagination. Neanderthal behavior remains a hotly contested subject in archaeology, with camps of opposing academics arguing for their behavioral sophistication or lack thereof. Using genetic data, lithics, and faunal analyses, a robust body of scholarship considers their patterns of kinship, use of medicinal plants, hunting strategies, and symbolic repertoires (for a summary of recent research developments in these subfields, see Nowell 2023). Innovative studies have suggested to what degree they produced fitted clothes (Collard et al. 2016) or had control over fire (Dibble et al. 2018), and by extension, how similar they were to *Homo sapiens* behaviorally.

These robust but transformative histories of research spanning a century and a half, with deep bodies of literature, may be drawn on to structure artificially intelligent depictions of the species. To excavate the formation processes of AI depictions of the past, we developed a methodology to compare the distance between scholarly representations of Neanderthals with those generated using AI.

Methods

All the code necessary to reproduce our analysis is available on Zenodo (Clindaniel 2024), along with a detailed README file that outlines the full methodological workflow used in this article. Employing the OpenAI API, we used DALL-E 3 and ChatGPT to generate hundreds of images and batches of text related to several prompts we crafted to illustrate or describe a day in the life of a Neanderthal. Both models are trained "Transformer" deep-learning models that have learned to predict complex sequences of data, based on corresponding input sequences of data (Vaswani et al. 2017). In the case of ChatGPT, this enables coherent text generation in response to a user-submitted prompt. For DALL-E 3, this transformer backbone works together with a diffusion model, which begins with random noise and gradually de-noises it through a series of steps to generate an image that aligns with a text prompt. This iterative process is guided by learned patterns from the training data (paired examples of images and captions) and is designed to produce images that are both coherent and consistent with the types of images seen during training. The use and description of such models for the generation of text and images were surveyed in our prior work (Magnani and Clindaniel 2023).

We used the DALL-E 3 model to generate 100 images for each of the following prompts: (1) "image depicting a day in the life of a neandertal, digital art" and (2) the same prompt with the phrase "based on expert knowledge of neandertal behavior" appended at the end. Before generating images, the model will revise the prompts to provide a more detailed textual description of the intended scene. For instance, the "non-expert" prompt was revised to more detailed prompts like the following, which the model used as a basis to produce the image in the upper-left corner of Figure 1:

> Create a digital art piece illustrating a typical day in the life of a Neanderthal. The scene should depict a Neanderthal carrying out daily activities such as hunting, foraging, making tools, or

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press



Figure 1. Images closest to average embedding from the four different prompts; clockwise from the top with prompt revision, with prompt revision (expert), no prompt revision (expert), and no prompt revision.

> communicating with others in their group. The setting should be prehistoric, with the environment consisting of wild, untouched landscapes, boulders, caves, or primitive shelters. Additionally, include details like the rough and sturdy attire of the Neanderthal and their hand-made tools.

The "expert" prompt generated prompt revisions like the following, which the model used to generate the image in the upper-right corner of Figure 1:

> A detailed digital art image illustrating a day in the life of a Neanderthal, based on scientifically validated knowledge about their behavior. The image includes scenes that show them hunting, foraging for food, making tools, interacting with family members, and resting in their cave shelter. The environment should reflect the harsh and frigid landscapes that the Neanderthals are believed to have inhabited, with tundra-like vegetation and woolly mammoths in the distance. Their physique should be robust and muscular, with a heavy brow ridge, and they should be dressed in clothing made from animal hides.

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press

To better understand the behavior of DALL-E 3's prompt revision behavior, we also produced 100 additional images for each of those two prompts, using OpenAI's suggested phrase—"I NEED to test how the tool works with extremely simple prompts. DO NOT add any detail, just use it AS-IS"—to ensure we could evaluate the behavior of each prompt on its own (OpenAI Image Generation). All original prompts, revised prompts, and resulting images are available online.

We additionally generated 100 text samples using the ChatGPT API (GPT 3.5) for each of the following prompts: (1) "describe a day in the life of a neandertal in one paragraph" and (2) the same prompt with the phrase "you are an expert on neandertal behavior" appended at the end. The non-expert produced responses like the following:

> A Neanderthal begins their day at dawn, emerging from their shelter in a cave or simple hut made of branches and hides. They gather together with their small family group to share a breakfast of meat from the hunt the previous day, supplemented with wild plants and berries. As the sun rises higher in the sky, the Neanderthal sets out with their companions to forage for more food, using stone tools to dig up roots and crack open nuts. In the afternoon, they work on constructing new tools and repairing their clothing made from animal skins. As evening falls, the Neanderthal returns to their dwelling, sharing stories and songs around a fire before settling down for the night, their daily tasks complete in harmony with the natural world.

It produced responses like the following for the expert prompt:

> A Neanderthal's day in prehistoric times likely began at dawn, with the first light illuminating their cave dwelling. The family unit would gather together, sharing the warmth of the fire they had kindled from the embers of the previous night. The adults would discuss the day's plans while the children played nearby, mimicking the tasks they saw their parents performing. After a simple breakfast of gathered fruits, nuts, and perhaps some leftover meat from the previous day's hunt, the group would set out in search of food. The men would head off to hunt large game, armed with spears and rocks, while the women and older children would gather plants and small animals. As the day wore on, the Neanderthals would return to their cave, where they would share the day's bounty and prepare for another night in the safety of their rocky shelter. After the evening meal, the group would settle down to rest, the flickering light of the fire casting shadows on the walls as they drifted off to sleep, ready to rise again with the sun and repeat the cycle of survival in their harsh and unforgiving world.

All prompts, as well as the ChatGPT responses to each prompt, are available on Zenodo for reference (Clindaniel 2024).

To assess the degree to which text and images produced by these generative AI tools corresponded with contemporary scholarly conceptions of Neanderthal behavior, we used the Constellate Dataset Builder to download all the metadata from articles, chapters, and books on JSTOR that contain the word "neanderthal" or "neandertal" between 1900 and 2023, beginning with the earliest dated article, and are in the Biological Science, Physical Science, or Social Sciences JSTOR content categories, including publications from common venues for Neanderthal scholarship such as *PNAS*, *Current Anthropology*, *Science*, and the *Journal of Archaeological Method and Theory* (Constellate 2024). Very few articles were downloaded in their entirety ($n = 378$, the majority of which date to 1927 or earlier; see Figure 2), rendering full-text articles a biased reference sample for computational content analysis. Note that works from 1927 entered the US public domain in 2023, so this temporal bias seems, at least partially, to be a result of copyright law in the United States. Therefore, we selected the abstracts from articles as a more temporally representative means of characterizing Neanderthals in the scholarly literature ($n = 2{,}063$, from the first available abstracts in 1923 to 2023).

To compare AI-generated images and text to scholarly text, we encoded CLIP embedding representations of all the generated images and text, as well as abstract text. In contrast to ChatGPT and DALL-E 3, which are generative models that produce text and images, respectively, CLIP is a multimodal encoder model that was trained on 400 million pairs of images and associated captions from across

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press

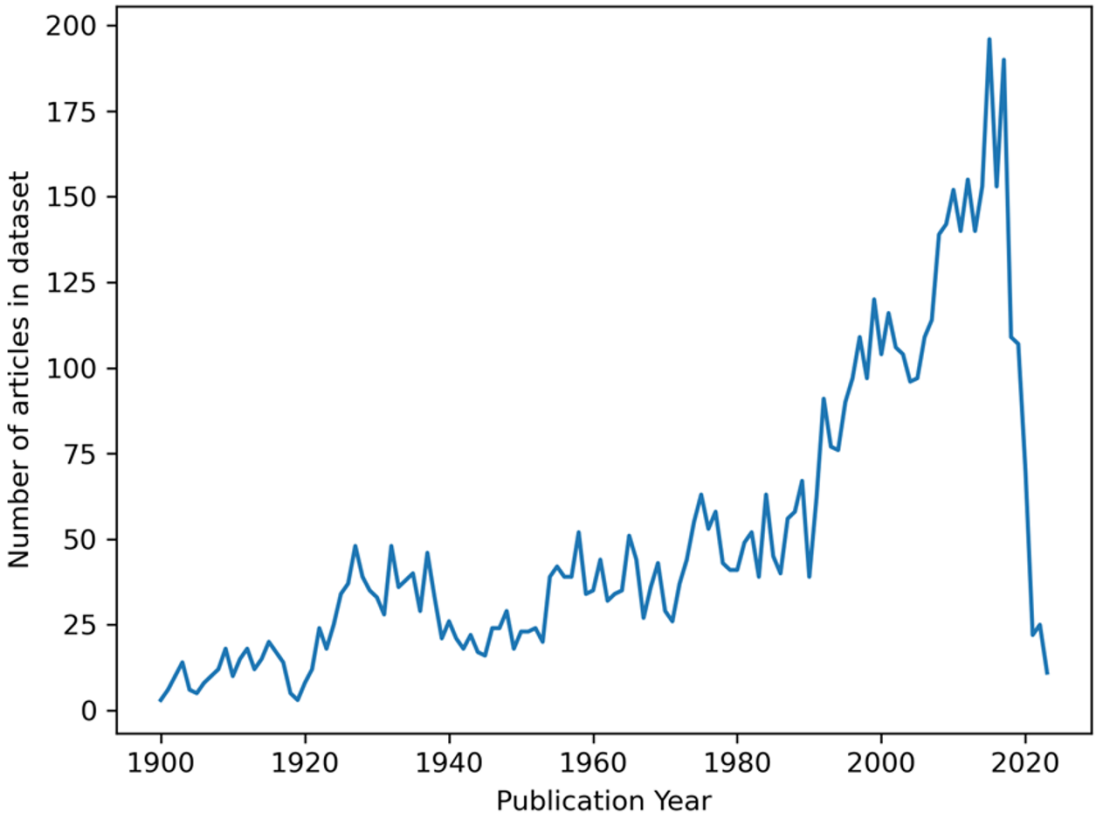


Figure 2. Availability of "Neanderthal" article content type by year in the collected Constellate dataset.

the internet. Based on associated image and text data, it learned to produce a (512-dimensional) joint text/image embedding space, which can be used for tasks such as classifying images based on text, comparing visual and textual content, and analyzing semantic similarity across modalities (Radford et al. 2021).

To interpret the type of image and text content generated by DALL-E 3 and ChatGPT, we drew inspiration from a neural topic modeling approach used to identify common texts in high-dimensional BERT embeddings, called BERTopic (Grootendorst 2022). Instead of encoding data using a (text-only) BERT model, however, we developed a related strategy for finding topic clusters in our multimodal CLIP embeddings. Specifically, we employed a combination of Uniform Manifold Approximation and Projection (UMAP) to reduce the dimensionality of the 512-dimensional scholarly abstract CLIP embeddings, while preserving important global and local information about the higher dimensional space (as per McInnes et al. 2018). We then identified clusters of scholarly content in these lower-dimensional representations via a clustering method called HDBSCAN (McInnes et al. 2017). Based on identified content clusters of scholarly abstracts, we predicted cluster membership for AI-generated image and text content based on the location of their embeddings in the joint image–text CLIP embedding space. Finally, we again used UMAP to reduce the dimensionality of the embedding to two dimensions and visualize the position of the different clusters and content embeddings.

To quantify the semantic distance between any given AI-generated image or text embedding and scholarly text embeddings, we measured the cosine distance between embeddings in the CLIP space. We then compared the distance between AI-generated content to scholarly content to compute the average distance between AI-generated and scholarly content in any given year. To estimate the "age" of AI-generated content, we averaged the closest 20 years (the top 25% of years with abstracts). If the content

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press

was mostly consistent with newer scholarly conceptions of Neanderthal behavior, this 20-year average should be closer to 2023. Otherwise, the average year would reflect the publication year of scholarly content that the AI-generated content was most similar to. This quartile-based method smooths out noisy outliers and provides an interpretable, exploratory signal of the temporal affinity between AI-generated content and the scholarly record. For each of the resulting years, we also identified the most salient terms in scholarly abstracts from the era (five years before and after the closest average year to the AI-generated content), as computed via average TF-IDF (term frequency-inverse document frequency) for the 10-year period using the Gensim Python package (Rehurek and Sojka 2011).

Finally, for each AI-generated text and image category—images and text produced with and without "expert" guidance, as well as images produced with and without prompt revision— we identified the image or text closest to the average embedding in that category. The identified images are displayed in Figure 1.

For texts produced without "expert" guidance, the closest average response was this one:

> A Neanderthal wakes up early in the morning as the sun rises, emerging from their cozy cave dwelling. They gather their tools made of stone and wood before setting out to hunt for food in the nearby forests. With their keen senses and strong physique, they stalk and capture a wild boar, using their sharp spears and teamwork to bring it down. After a successful hunt, they return to their cave and share the food with their family and tribe, enjoying a hearty meal together. The rest of the day is spent working on various tasks such as making clothing from animal hides, crafting tools, and tending to the fire. As the sun sets, the Neanderthal settles down for the night, feeling content and grateful for another day of survival in the harsh prehistoric world.

For texts produced with "expert" guidance, the closest average response was as follows:

> A typical day in the life of a Neanderthal would involve waking up near dawn in their shelter, perhaps a cave or simple structure made of branches and hides. The Neanderthal would gather with their family group to start the day by foraging for food and water, possibly hunting for game or scavenging for plants and other resources. Throughout the day, they would engage in a variety of tasks such as making tools, tending to their fire, and caring for their young. Communication and social interaction would be important parts of their day, as they would rely on each other for survival and cooperation. As evening approached, the Neanderthal would likely gather around the fire for warmth and protection, sharing food and stories with their group before settling down for rest.

Using these central, representative images and texts, we performed further qualitative analyses to complement our quantitative analysis.

Results and Discussion

Computational approaches to archaeology are primarily constrained by data accessibility. Indeed, data availability structured our analysis, limiting our study to thousands of abstracts, rather than to entire articles, which were unavailable as full text in major article repositories. Data governance policies are also clearly visible in the structure of our study. For instance, copyright laws dictated a steep drop-off in article availability from the late 1920s. With the more recent advent of open access, the availability of full-text articles spiked again in the early 2000s. Although the source information used to train generative AI is opaque—not least because of the misuse of copyrighted materials by large companies—it can be assumed that the availability of knowledge will shape the AI outputs to skew toward older, more visible texts or publicly available information on websites that is more accessible to crawlers but that on average might reflect older information. How data are made accessible through the publication process will shape a generation of studies on large datasets. Tangibly, we expect this data availability to lead to more archaic representations of the archaeological record.

Grouping archaeological knowledge derived from journal databases facilitated a comparison to large bodies of archaeological representations created using generative AI. Our HDBSCAN model identified four major semantic clusters in the scholarly abstracts, with these clusters containing nearly all the

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press

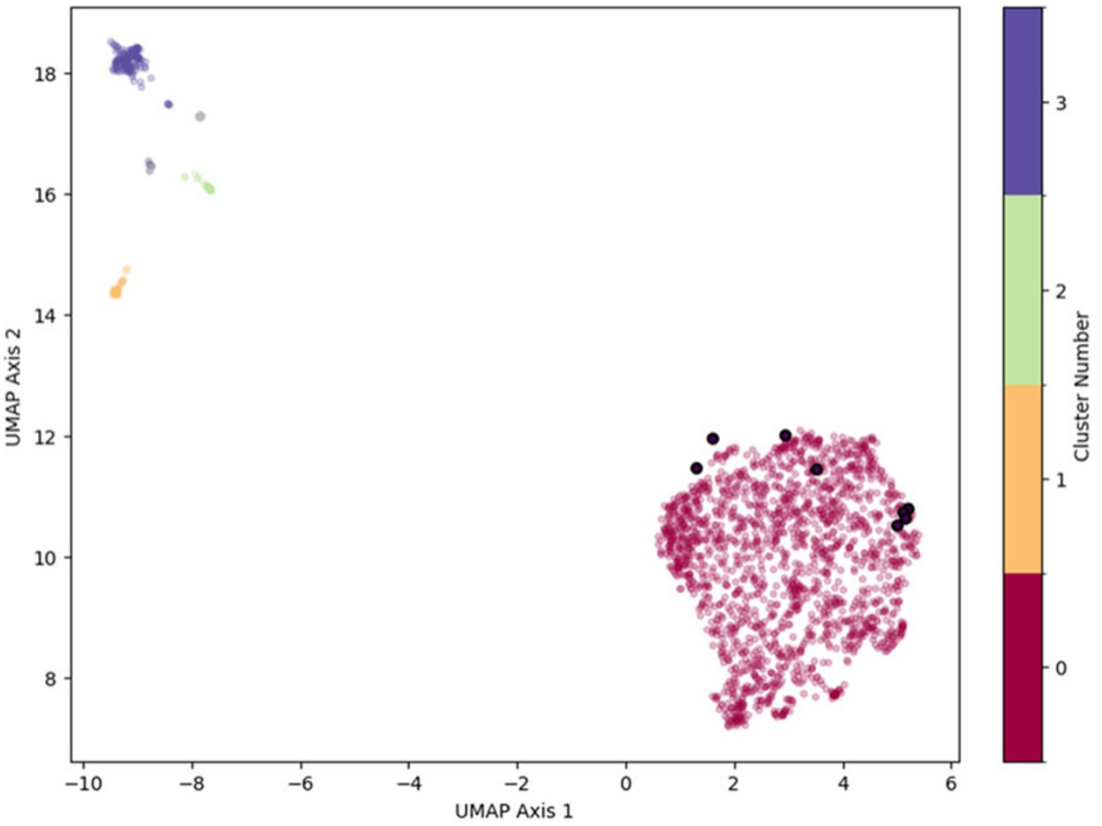


Figure 3. Clusters of scholarly abstracts identified by HDBSCAN and projected into two dimensions by UMAP. Abstracts that could not be assigned to a cluster are denoted with the color gray. Note that the average embeddings for AI-generated content are all presented via larger circles and dark outlines (they are all identified as belonging to Cluster 0).

abstract embeddings (98.93%; see Figure 3). The average embeddings for each of the AI-generated content types are included in Cluster 0. This cluster contains abstracts on a variety of topics that bear on Neanderthal behavior, ranging from paleogenomic research, to demographic characteristics of Neanderthal populations, to studies that emphasize lithic tool production. This cluster is semantically separate from abstracts that emphasize burials and physical characteristics of Neanderthal bones and bodies (Cluster 1), retrospective pieces and reflections on contemporary reconstructions of Neanderthals in art and in museums (Cluster 2), and abstracts broadly focused on mobility and continental migration (Cluster 3). Samples of 10 abstracts from each cluster discussed are available online, which provide additional details on these characterizations.

Our initial clustering approach was intentionally conservative, designed to reveal broad semantic distinctions across the full corpus of abstracts. To better understand the internal structure of Cluster 0, which contained the majority of scholarly abstract embeddings, we conducted a second round of clustering using HDBSCAN with less conservative parameters and a leaf-based cluster selection method. Using this approach, 96.07% of abstract embeddings were assigned to a cluster. This made it possible to interpret four distinct thematic subclusters within Cluster 0 (see Figures 4 and 5). Subcluster 0 is generally centered on recent paleogenomic research; subcluster 1 on the demographic characteristics, life cycle, and diet of Neanderthals; subcluster 2 on tool production and its connection to social learning; and subcluster 3 on inferring behavior and its connection to broader evolutionary themes via the fossil record.

Our results indicate that generative text models tend to produce text that fall outside the bounds of scientific literature. Note that much of the ChatGPT-generated text does not fall within a subcluster

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press

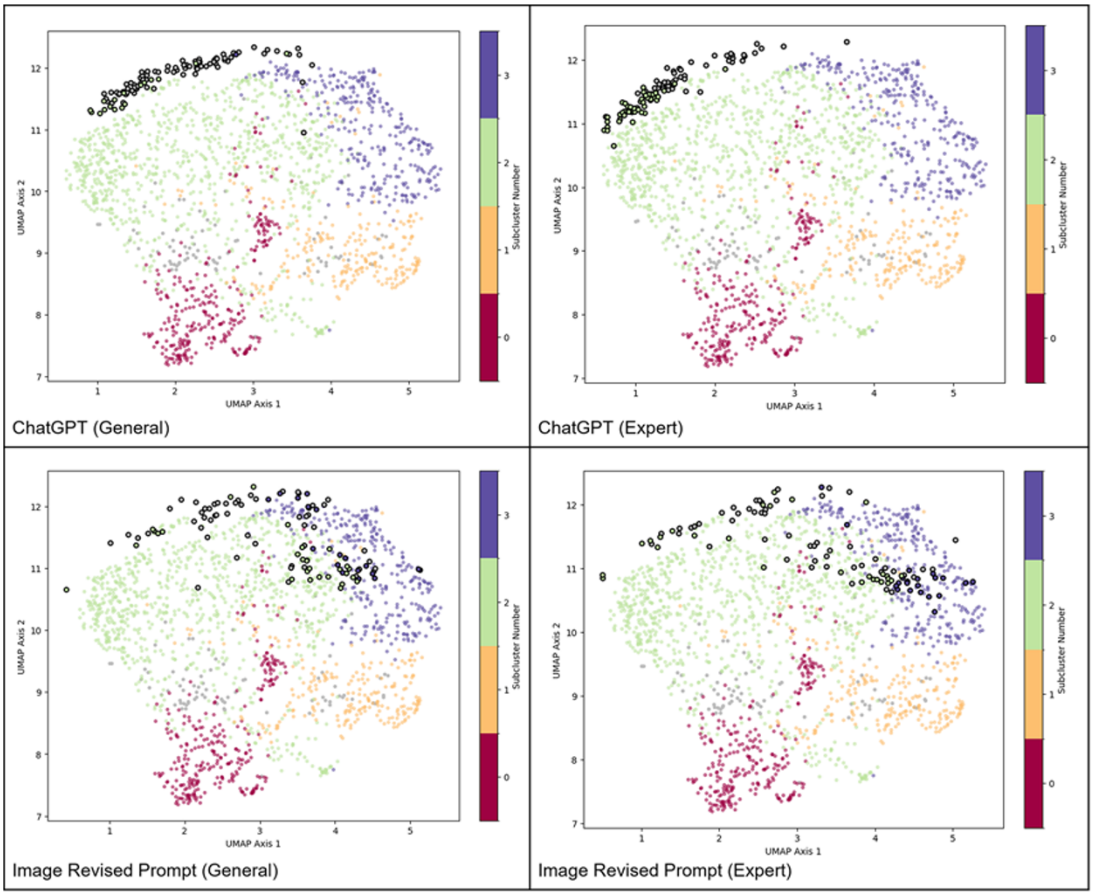


Figure 4. Subclusters of scholarly abstracts in cluster 0, as identified by HDBSCAN using a leaf-based cluster selection method and projected into two dimensions using UMAP. AI-generated text embeddings have been superimposed according to their predicted cluster membership. Note that the average embeddings for AI-generated content are all presented via larger circles, colored by their subcluster membership and outlined in black. Embeddings that could not be assigned to a cluster are denoted with the color gray.

(82% for general prompt, 51% for expert prompt). Furthermore, HDBSCAN could not successfully place the average embeddings for ChatGPT-generated text into a cluster (whereas all the others fit within the abstract clusters)—hinting at a relatively low correspondence with scholarly content. The content that does fall within an established cluster generally corresponds to cluster 2 (17% for general prompt, 49% for "expert" prompt). The revised prompts generated as part of the image-generation process for DALL-E 3 produced a mix of samples in cluster 2 (28% for both prompts) and cluster 3 (19% for general prompt, 21% for expert prompt), but more than half of all content was not in a cluster (53% for general prompt, 51% for expert prompt). AI-generated images seem to more closely align with scholarly content, however. All image embeddings (regardless of prompt) were strongly centered in Cluster 3, with only 17% not falling in a cluster and less than 1% in Cluster 2.

To explore these issues beyond topic-level comparisons, we analyzed the temporality of text and images generated using AI. The closest average year to each of the generated AI categories is summarized in Table 1. ChatGPT seems to produce content most consistent with the early 1960s, emphasizing human ecology and broad discussions of "culture" and "nature." DALL-E 3 produces images that depict elements more characteristic of the late 1980s and early 1990s: they are focused on connections to modern humans, genetics, and language and emphasize the particularity of time periods and different regions. The prompt revisions seem to produce texts closest to scholarly content from the early 1970s in which

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press

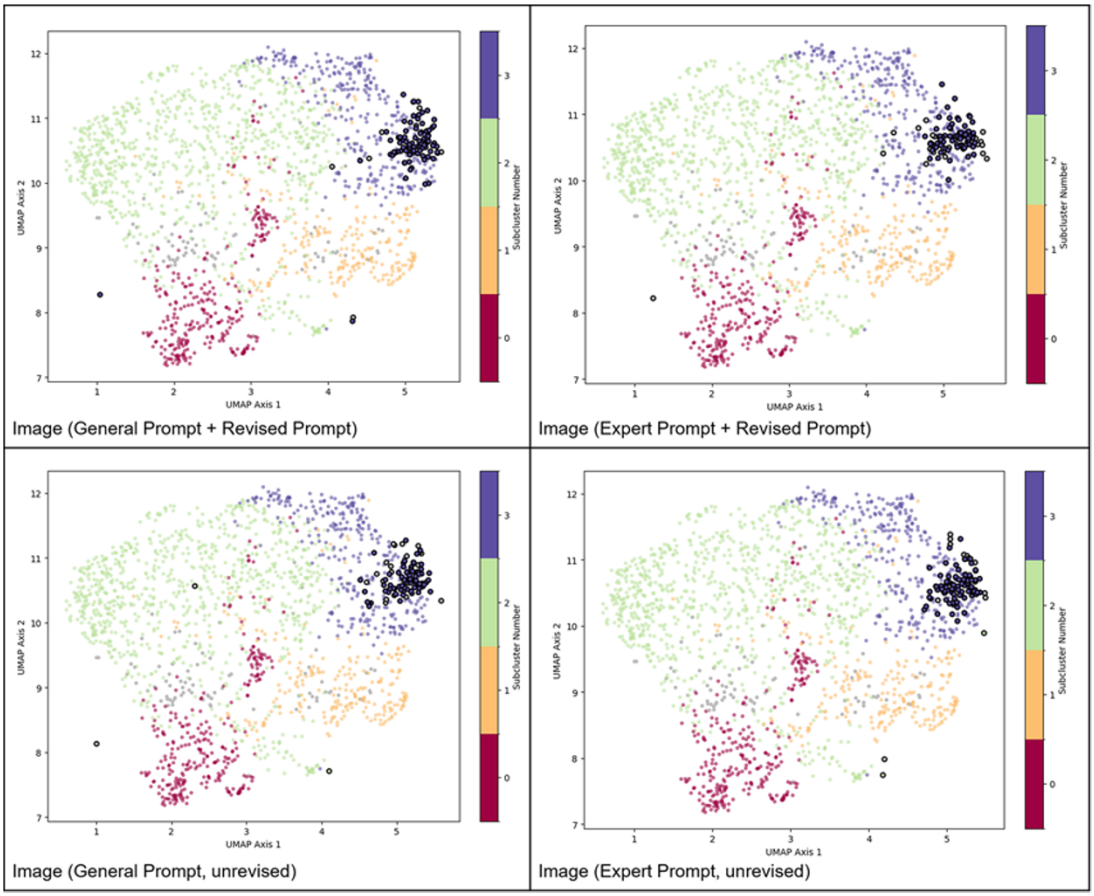


Figure 5. Subclusters of scholarly abstracts in cluster 0, as identified by HDBSCAN using a leaf-based cluster selection method and projected into two dimensions using UMAP. AI-generated image embeddings have been superimposed according to their predicted cluster membership. Note that the average embeddings for AI-generated content are all presented via larger circles, colored by their subcluster membership and outlined in black. Embeddings that could not be assigned to a cluster are denoted with the color gray.

the salient terms seem centered around the morphological features of Neanderthals and their location in broader evolutionary narratives. As a result of this tilt toward the 1970s, note that the images produced via revised prompts (without specifying that AI should take on the role of "expert on Neanderthal behavior") end up producing slightly older depictions (1985.55) than those produced via our original prompt alone (1987.4).

Note as well that none of these average ages are characteristic of the clusters they belong to semantically (Figure 6). The most recent year for each type of generated AI content in Table 1 is below the 25th percentile of all years represented in each cluster (and generated texts are outliers in terms of closest average year).

To scaffold our qualitative analysis, we identified AI-generated text and images that were closest to the average embedding for each AI-generated text and image category (see Figure 1). A review of "average" images supports poor correspondence with contemporary scientific understandings of Neanderthals. First, we draw attention to the physical characteristics of AI-generated content. Across all prompts, even though the images feature depictions of bipedal hominids, they incorporate archaic features associated with other extinct hominins and perhaps extant taxa like *Pan*. A majority of images depict human-like figures, slightly stooped, with large quantities of body hair. These depictions have more in common with early twentieth-century drawings of Neanderthals than contemporary scientific knowledge. Our AI

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press

Table 1. AI-Generated Content Type by Closest Average Year and the Most Salient Terms in Scholarly Abstracts from the Era (±5 Years).

AI-Generated Content Type	Closest Average Year	Most Salient Terms in Abstracts ± 5 years from Closest Average Year
Image (general prompt + revised prompt)	1985.55	homo sapiens, evolution, human, homo erectus, assemblages, species, genetic, archaeological, cave, samples, populations, social, sapiens, paleolithic, region, biological, selection, development, change, pleistocene
Image (expert prompt + revised prompt)	1991.05	human, age, archaeological, middle paleolithic, modern humans, populations, humans, region, evolution, dating, assemblages, patterns, cave, language, europe, bone, remains, industries, development, africa
Image (general prompt, unrevised)	1987.4	homo sapiens, human, evolution, populations, assemblages, middle paleolithic, homo erectus, neandertal, modern humans, paleolithic, sapiens, cave, social, age, samples, development, species, variation, individuals, archaeological
Image (expert prompt, unrevised)	1991.05	human, age, archaeological, middle paleolithic, modern humans, populations, humans, region, evolution, dating, assemblages, patterns, cave, language, europe, bone, remains, industries, development, africa
Text (revised prompt for general image prompt)	1970.7	human, evolution, cultural, archaeological, variation, asia, tool, population, fossil, change, neanderthal(s), shift, modern, burial, physical, hominids, evolutionary, biology, man
Text (revised prompt for expert image prompt)	1974.0	evolution, human, cultural, archaeological, variation, hominid, homo sapiens, morphological, teeth, language, hominids, change, modern, fossil, population, problems, burial, development, upper paleolithic, aspects
Text (general ChatGPT prompt)	1963.85	biology, evolution, occurrence, human ecology, neanderthal, human, development, africa, man, shanidar, tabun, caracteres, nature, hominid, hominoid, problems, homo sapiens, modern, culture, physical
Text (expert ChatGPT prompt)	1962.4	occurrence, evolution, development, shanidar, tabun, africa, human ecology, caracteres, hominoid, nature, cranial capacity, oreopithecus, culture, homo sapiens, depression, threshold amplitude, waveform, dawson, human, stage

Neanderthals exhibit high levels of facial prognathism, exaggerated brow ridges, and a low cranial profile too extreme to fall within Neanderthal phenotypic variability. It is worth noting that revised expert prompts generated by DALL E-3 appeared less incorrect, depicting Neanderthals with fewer archaic features, who were covered less in body hair, and had facial structures that appear more consistent with later *Homo*.

Scientific literature in recent decades has cast a critical eye on gendered representations of the past (Conkey and Gero 1997; Dobres 1995), yet a lack of focus on women and children in prehistory clearly informs our artificially intelligent images. Heavily muscled male hominins are foregrounded in all our representative visuals. Only in the expert prompt was a small child shown playing. Developers are quiet about their training sources, and some archaeologists have suggested a close reliance on Wikipedia (Spennemann 2023a). It is likely that training material for our subject matter, and by extension representations of Neanderthals in general, is equally gendered, sidelining women and children. Following scholars who have pointed out these issues more broadly (Langley 2020), our results suggest that AI systems require corrective measures to more equitably and accurately imagine Neanderthal prehistory.

Contrasting with generally archaic physical depictions of Neanderthals, the technology represented across images was unusually advanced. We know from archaeological evidence that Neanderthals were capable of producing complex lithic industries, which varied across space and time (Delagnes and Rendu 2011; Hoffecker 2018). Their bone tool use is no longer controversial, and evidence for the use of fibers

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press

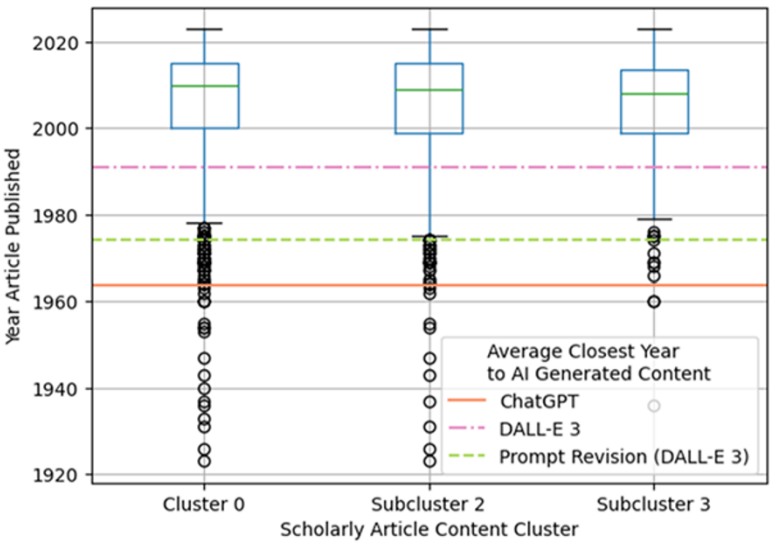


Figure 6. Boxplot of the year each scholarly article in an identified content cluster was published, with the average closest year to AI-generated images and text, based on semantic similarity (Table 1). We chose the most recent average year for each AI model from Table 1 to present in this figure.

has emerged in recent years (Hardy et al 2020; Soressi et al. 2013). Despite expressing a suite of complex and variable behavioral repertoires in scientific literature, AI depictions of Neanderthals presented several anachronisms. Among them were basketry and dwellings with thatched roofs and ladders. Vessels of glass and implements of metal also appear. In all cases, AI-generated representations presented technologies that are neither associated with Neanderthals nor would appear archaeologically for tens of thousands of years. There is a temporal confusion associated with these depictions that mix old biological stereotypes with more recent technologies, indexing a low correspondence with contemporary scientific understandings of Neanderthals.

Comparatively, our AI-generated text provided more innocuous material emphasizing a simplicity of technology and cooperative survival. In both original and expert prompts, Neanderthal technologies consisted of only three materials—stone, hide, and wood—and referred to the maintenance but not the making of fire. Contrasting with our general prompt that emphasized cave dwelling and hunting, our expert prompt suggested a broader diet more consistent with contemporary understandings of Neanderthal lifeways. It included the gathering of plant resources, the construction of open-air shelters, and activities such as child-rearing. Although our text samples are less blatantly stereotypical than the images, they do not significantly address the broad variability and sophistication of Neanderthal cultures reflected in contemporary scientific literature: as mentioned earlier, they tend to cluster with older bodies of scientific knowledge.

Conclusion

With expanding use across diverse fields, content generated using AI is poised to play a central role in society. Understanding the bias of representation across these new technological applications is essential. How the archaeological record is written about and conceived using generative AI will have significant implications for anthropologists and society at large. Our study reveals that artificially intelligent depictions of the past reflect limited overlap with bodies of scientific knowledge. Where there is some correspondence, we show that the temporality and overlap of knowledge are uneven across media, with images reflecting more up-to-date academic knowledge and text averaging an older vintage, from as early as the 1960s. Issues with gender representations are clear, and anachronistic elements were particularly evident in AI-generated images. These discrepancies may result from the types of data being used to train generative AI programs, which we expect reflect broader social biases distributed throughout

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press

bodies of writing and source images. Reproduction of these biases risks their continued propagation and normalization.

Our current research suggests that the way we structure and make information available will directly influence AI output and, by extension, the way we imagine the past. Moving forward, data policies will inform the way archaeological material is written about and visualized. Open access structures the types of information that are accessible for scholars to conduct computational analyses and is equally likely to affect which information is used to train artificial intelligence programs. Institutions with robust funding to allow articles to be widely read will continue to have a greater hand in determining how things are represented. If articles from certain eras or from specific subfields are more likely behind paywalls, they will be less likely to feed into materials generated using AI. Academic publishing practices may ultimately work to undermine public knowledge.

The methods presented in this article may be systematically applied to any archaeological region or period, using the AI platforms discussed here or others. Future research should elucidate with greater specificity the types of biases that creep into artificially intelligent representations of the past and expose their variation across archaeological space and time. What are the societal impacts of AI on contemporary discussions surrounding Indigenous cultures or gender, for instance, if they are informed by dated depictions from the 1960s? The analytical tool kit presented here provides the foundation to explore the ramifications of drawing on older texts or nonscientific bodies of knowledge to inform perceptions of the archaeological record. Ultimately, the identification of these temporal and foundational biases will contribute to more current and equitable archaeological interpretations.

Acknowledgments. We thank the anonymous peer reviewers for their feedback on this article.

Funding Statement. The authors have no funding to report for this article.

Data Availability Statement. All supporting material is available through Zenodo (Clindaniel 2024).

Competing Interests. The authors declare none.

References Cited

Alowais, Shuroug A., Sahar S. Alghamdi, Nada Alsuhebany, Tariq Alqahtani, Abdulrahman I. Alshaya, Sumaya N. Almohareb, Atheer Aldairem, et al. 2023. Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice. *BMC Medical Education* 23(1):689.

Altaweel, Mark, and Hadjitofi Tasoula Georgiou. 2020. The Sale of Heritage on eBay: Market Trends and Cultural Value. *Big Data & Society* 7. https://doi.org/10.1177/2053951720968865.

Altaweel, Mark., Adel Khelifi, and Mohammad Maher Shana'ah. 2024. Monitoring Looting at Cultural Heritage Sites: Applying Deep Learning on Optical Unmanned Aerial Vehicles Data as a Solution. *Social Science Computer Review* 42(2):480–495.

Bonacchi, Chiara, Mark Altaweel, and Marta Krzyzanska. 2018. The Heritage of Brexit: Roles of the Past in the Construction of Political Identities through Social Media. *Journal of Social Archaeology* 18(2):174–192.

Bonacchi, Chiara, and Marta Krzyzanska. 2021. Heritage-Based Tribalism in Big Data Ecologies: Deploying Origin Myths for Antagonistic Othering. *Big Data & Society* 8. https://doi.org/10.1177/20539517211003310.

Câmara, Ariele, Ana de Almeida, David Caçador, and João Oliverira. 2023. Automated Methods for Image Detection of Cultural Heritage: Overviews and Perspectives. *Archaeological Prospection* 30(2):153–169.

Clindaniel, Jon. 2024. ai-interpretation-of-the-past. *Zenodo*, May 5. https://doi.org/10.5281/zenodo.11117535, accessed October 5, 2025.

Clindaniel, Jon. 2025. Colorful Insights from an AI Khipukamayuq. *SocArXiv*, October 3. https://doi.org/10.31235/osf.io/4p7s5.

Clindaniel, Jon, and Matthew Magnani. 2024. Digital Formation Processes: A High-Frequency, Large-Scale Investigation. *Journal of Archaeological Science* 161(2):105890.

Cobb, Peter J. 2023. Large Language Models and Generative AI, Oh My! Archaeology in the Time of ChatGPT, Midjourney, and Beyond. *Advances in Archaeological Practice* 11(3):363–369.

Collard, Mark, Lia Tarle, Dennis Sandgathe, and Alexander Allan. 2016. Faunal Evidence for a Difference in Clothing Use between Neanderthals and Early Modern Humans in Europe. *Journal of Anthropological Archaeology* 44:235–246.

Conkey, Margaret W., and Joan M. Gero. 1997. Programme to Practice: Gender and Feminism in Archaeology. *Annual Review of Anthropology* 26:411–437.

Constellate. 2024. Use of Terms Neandertal and Neanderthal in Scholarly Content between 1900–2023. Electronic document, https://constellate.org/dataset/912f20af-3446-c463-bbde-4daeeb8f58a2/, accessed June 1, 2025.

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press

Davis, Dylan S. 2019. Object-Based Image Analysis: A Review of Developments and Future Directions of Automated Feature Detection in Landscape Archaeology. *Archaeological Prospection* 26(2):155–163.

Delagnes, Anne, and William Rendu. 2011. Shifts in Neandertal Mobility, Technology and Subsistence Strategies in Western France. *Journal of Archaeological Science* 38(8):1771–1783.

Dibble, Harold L., Dennis Sandgathe, Paul Goldberg, Shannon McPherron, and Vera Aldeias. 2018. Were Western European Neandertals Able to Make Fire? *Journal of Paleolithic Archaeology* 1:54–79.

Dobres, Marcia-Anne. 1995. Gender and Prehistoric Technology: On the Social Agency of Technical Strategies. *World Archaeology* 27(1):25–49.

Fang, Xiao, Shangkun Che, Minjia Mao, Hongzhe Zhang, Ming Zhao, and Xiaohang Zhao. 2024. Bias of AI-Generated Content: An Examination of News Produced by Large Language Models. *Scientific Reports* 14(1):5224.

Gattiglia, Gabriele. 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.

Google Communications. 2024. We're Aware that Gemini Is Offering Inaccuracies in Some Historical Image Generation Depictions. Here's Our Statement. *Twitter*, February 21. https://twitter.com/Google_Comms/status/1760354549481546035, accessed June 1, 2025.

Grootendorst, Maarten. 2022. BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. *arXiv*:2203.05794. https://doi.org/10.48550/arXiv.2203.05794.

Hardy, Bruce, Marie-Hélène Moncel, Celine Kerfant, Matthieu Lebon, Ludovic Bellot-Gurlet, and Nicolas Mélard. 2020. Direct Evidence of Neanderthal Fibre Technology and Its Cognitive and Behavioral Implications. *Scientific Reports* 10(1):4889.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. The Weirdest People in the World? *Behavioral and Brain Sciences* 33(2–3):61–83.

Hoffecker, John F. 2018. The Complexity of Neanderthal Technology. *PNAS* 115(9):1959–1961.

Holguin, Brian E. 2023. An Indigenous Archaeological Perspective on the Use of Artificial Intelligence in Reconstructing Chumash History. PhD dissertation, Department of Anthropology, University of California, Santa Barbara.

Horn, Christian, Oscar Ivarsson, Cecilia Linhé, Rich Potter, Ashly Green, and Johan Ling. 2022. Artificial Intelligence, 3D Documentation, and Rock Art—Approaching and Reflecting on the Automation of Identification and Classification of Rock Art Images. *Journal of Archaeological Method and Theory* 29(1):188–213.

Huggett, Jeremy. 2015. A Manifesto for an Introspective Digital Archaeology. *Open Archaeology* 1(1):86–95.

King, William. 1864. The Reputed Fossil Man of the Neanderthal. *Quarterly Journal of Science* 1(11):88–97.

Langley, Michelle C. 2020. Space to Play: Identifying Children's Sites in the Pleistocene Archaeological Record. *Evolutionary Human Sciences* 2:e41. https://doi.org/10.1017/ehs.2020.29.

Madison, Paige. 2021. Brutish Neanderthals: History of a Merciless Characterization. *Evolutionary Anthropology* 30(6):366–374.

Magnani, Matthew, and Jon Clindaniel. 2023. Artificial Intelligence and Archaeological Illustration. *Advances in Archaeological Practice* 11(4):452–460.

Magnani, Matthew, Jon Clindaniel, and Natalia Magnani. 2022. Material Culture Studies in the Age of Big Data: Digital Excavation of Homemade Face-Mask Production during the COVID-19 Pandemic. *American Antiquity* 87(4):683–703.

McInnes, Leland, John Healy, and Steve Astels. 2017. HDBSCAN: Hierarchical Density Based Clustering. *Journal of Open Source Software* 2(11):205. https://doi.org/10.21105/joss.00205.

McInnes, Leland, John Healy, Nathaniel Saul, and Lukas Großberger. 2018. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* 3(29):861. https://doi.org/10.21105/joss.00861.

Mesanza-Moraza, Amaia, Ismael García-Gómez, and Agustín Azkarate. 2020. Machine Learning for the Built Heritage Archaeological Study. *Journal on Computing and Cultural Heritage* 14(1):10.

Nazer, Leland, John Healy, and Steve Astels. 2023. Bias in Artificial Intelligence Algorithms and Recommendations for Mitigation. *PLoS Digital Health* 2(6):e0000278.

Nowell, April. 2023. Rethinking Neandertals. *Annual Review of Anthropology* 52:151–170.

O'Connor, Sinead, and Helen Liu. 2023. Gender Bias Perpetuation and Mitigation in AI Technologies: Challenges and Opportunities. *AI & SOCIETY* 39(4):2045–2057.

Peeters, Susan, and Hub Zwart. 2020. Neanderthals as Familiar Strangers and the Human Spark: How the "Golden Years" of Neanderthal Research Reopen the Question of Human Uniqueness. *History and Philosophy of the Life Sciences* 42(3):33.

Pehlivan, Ceyhun N. 2024. The EU Artificial Intelligence (AI) Act: An Introduction. *Global Privacy Law Review* 5(1):31–42.

Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, et al. 2021. Learning Transferable Visual Models from Natural Language Supervision. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 8748–8763. Machine Learning Research, Cambridge, Massachusetts.

Rehurek, Radim, and Petr Sojka. 2011. *Gensim—Python Framework for Vector Space Modelling*. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic.

Roselli, Drew, Jeanna Matthews, and Nisha Talagala. 2019. Managing Bias in AI. In *Companion Proceedings of the 2019 World Wide Web Conference*, pp. 539–544. Association for Computing Machinery, New York.

Soressi, Marie, Shannon P. McPherron, Michel Lenoir, Tamara Dogandžić, Paul Goldberg, Zenobia Jacobs, Yolaine Maigrot, et al. 2013. Neandertals Made the First Specialized Bone Tools in Europe. *PNAS* 110(35):14186–14190.

Spennemann, Dirk H. R. 2023a. What Has ChatGPT Read? The Origins of Archaeological Citations Used by a Generative Artificial Intelligence Application. *arXiv*:2308.03301.

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press

Spennemann, Dirk H. R. 2023b. Exhibiting the Heritage of COVID-19—A Conversation with ChatGPT. *Heritage* 6(8):5732–5749.
Vaswani, Ashish, Noam Shazer, Niki Parmar, Jakob Uzqkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, et al. 2017. Attention Is All You Need. *arXiv*:1706.03762.
Williamson, Ben. 2024. The Social Life of AI in Education. *International Journal of Artificial Intelligence in Education* 34(1):97–104.

Cite this article: Magnani, Matthew, and Jon Clindaniel. 2026. Artificial Intelligence and the Interpretation of the Past. *Advances in Archaeological Practice* 14(2):218–233. https://doi.org/10.1017/aap.2025.10110.

https://doi.org/10.1017/aap.2025.10110 Published online by Cambridge University Press