

12 Artificial intelligence and machine learning

The application of artificial intelligence (AI) and machine learning in digital humanities scholarship allows for efficient processing of large amounts of data in various forms (e.g., text, images, video). However, questions remain on how these approaches will affect the historical record. This chapter covers an introduction to GoFAI and generative AI, Large Language Models (LLMs), Large Vision Models, and Vision Language Models. This overview also considers the important ways AI systems can be open or closed. The second half of this chapter is devoted to ethical concerns, including bias, transparency, accuracy, academic integrity, and intellectual property, and discusses ways to evaluate AI tools rigorously, alongside the development of human oversight, regulatory policies, and best practices.

12a Intro to neural nets: LMMs and large image models

Exercise 12.1: Machines that can think and learn

Read the historical summaries of AI provided by Britannica and IBM. After reviewing these resources, consider the recurring mission throughout AI's history: to make machines “think” or replicate human thought. What are the key milestones and figures in the history of AI that you found most significant? How has the definition and understanding of “thinking” or “human thought” evolved in the context of AI development? Based on these historical accounts, do you believe the goal of replicating human thought is achievable? Why or why not? What are some of the ethical implications that arise from this mission?

- <https://www.britannica.com/science/history-of-artificial-intelligence>
- <https://www.ibm.com/think/topics/history-of-artificial-intelligence>

Exercise 12.2: Prompt craft

Read through the prompt bank available on the AI for Humanists website: <https://aiforhumanists.com/guides/prompts/>. What do you notice about the different approaches? Consider a research project you are working on. How would you construct a prompt for your project? What specifics do you think would be needed?

How-to example

Reading through the prompts, consider the construction of the prompts based on the criteria presented in the chapter. Notice the ways the prompts are tailored to content, style, format, discipline, and audience. Be sure to include these details in your own prompt construction. Reflect on the output and adjust the parameters. Notice how that shifts the output.

Exercise 12.3: Explorative uses of ChatGPT in literary research

Explore the “Explorative uses of ChatGPT in literary research” section on this web page (<https://litdh.au.dk/ai-tools>) web page. Choose one of the suggested prompts or research questions from that section. Replicate the experiment using ChatGPT or another accessible Large Language Model (LLM). Did it meet your expectations? What were its strengths and weaknesses? Describe potential benefits and limitations of using LLMs for this type of literary research.

How-to example

As results are assessed, keep in mind that LLMs can summarize large amounts of text quickly and provide insights into complex topics without prior specialized knowledge. However, the results may produce incorrect or over-generalized information that lacks the nuance and expertise provided by those trained in the field. In addition, results may replicate bias within the historical texts or introduce it due to the ways in which the model was constructed or weighted.

Exercise 12.4: Talking to museum objects

In The Living Museum app (<https://www.livingmuseum.app/>), the model is not constrained to the British museum's archival metadata or a specific curator's comments but draws on outside sources to respond to wide questions about culture, history, and artistic process and practice. If we are not aware of all the sources of information that the AI's answers are being drawn from, how is the user expected to assess the validity of the information as well as weight contextual positioning of the information? Does authorship of information matter in this context? Why or why not? What issues arise when AI becomes the voice for a museum object? Pick an object and ask it questions. What can the AI "see" and what does it miss?

How-to example

While the disrupting the hierarchy where the single voice of the museum is the authority is desirable—we want to encourage the museum space to be more of forum of discussion that represents communities broadly and inclusively—without acknowledging the originator of the information, the audience lacks the context to properly assess its accuracy and positionality. Authorship is crucial for understanding how the information is situated culturally and historically and attribution also ensures accountability, particularly around sensitive or contested narratives. A conversational format provides engagement and flexible responsiveness that standard wall labels cannot (which may be particularly ideal for young or new audiences), but an AI chatbot risks oversimplifying or misrepresenting complex histories where nuance may be deeply important. Additionally, there are considerations regarding matters of respect that are implicit within the ideas of stewardship embedded in the traditions of curatorial practice that the use of AI abandons. Ethically and morally, we should definitely not abdicate the most difficult of these tasks to machines. Museum professionals and cultural communities have a responsibility to each other to come together to express complicated narratives that warrant the depth of human compassion and understanding.

Exercise 12.5: Handwriting analysis tool

Alexander Hamilton and George Washington shared a close professional relationship for many years, with Hamilton serving as Washington's aide-de-camp during the Revolutionary War and later as his Secretary of the Treasury. This close collaboration often involved Hamilton drafting letters and documents on Washington's behalf. Use the handwriting analysis tool (HAT) <https://www.csmc.uni-hamburg.de/publications/software/hat.html> to explore the possibility of identifying Hamilton's handwriting within the George Washington Papers using a HAT. Select handwriting samples from the following collections:

- <https://www.loc.gov/collections/george-washington-papers/about-this-collection/>
- <https://www.loc.gov/collections/alexander-hamilton-papers/about-this-collection/>

Recommended readings

- Azar, Geoff Cox Mitra, and Leonardo Impett. 2021. "Introduction: Ways of Machine Seeing." *AI & Society* 36: 1093–1104. <https://doi.org/10.1007/s00146-020-01124-6>.
- GeeksforGeeks. "Difference between Generative AI and Traditional AI." Accessed March 16, 2025. <https://www.geeksforgeeks.org/difference-between-generative-ai-and-traditional-ai/>.
- Illinois Central College Library. "What Is a Language Model?" Accessed March 16, 2025. <https://library.icc.edu/c.php?g=1372140&p=10141461>.
- Impett, Leonardo. 2020. "Analyzing Gesture in Digital Art History." In *The Routledge Companion to Digital Humanities and Art History*, edited by Kathryn Brown, 386–407. New York: Routledge.
- Kempf, Sebastian, Markus Krug, and Frank Puppe. 2023. "KIETA: Key-Insight Extraction from Scientific Tables." *Applied Intelligence* 53: 9513–30. <https://doi.org/10.1007/s10489-022-03957-8>.
- LaPensee, Beth, and Diba Kaya. "The Humanities in the Age of AI: Notes on a Participatory Exploration at MLA 2025." *JSTOR Daily*. Accessed March 16, 2025. https://about.jstor.org/blog/mla_2025_highlights/.
- University of Southern California Libraries. "Using Generative AI in Research: The Digital Humanities and AI." Accessed March 16, 2025. <https://libguides.usc.edu/c.php?g=1329251&p=10760530>.

12b Ethical issues: fakes, misinformation, deception

Exercise 12.6: Deepfakes

Deepfakes are becoming increasingly prevalent in digital media, blurring the lines between reality and fabrication. Find a specific example of a deepfake that you have encountered (e.g., a deepfake video of a political figure, a digitally resurrected actor, a manipulated TikTok trend). What was its purpose? Who was the target or subject? How convincing was it? What visual or auditory cues revealed its artificial nature? Who was potentially harmed by its creation or distribution? Consider the broader societal impact of deepfakes. How do they challenge our understanding of truth and authenticity in the digital age? What are parameters that can't be faked—are there any?

How-to example

The Dali Lives (<https://thedali.org/exhibit/dali-lives/>) installation at the Salvador Dalí Museum employs AI to “resurrect” the Spanish surrealist Salvador Dalí (1904–89), allowing visitors to interact with a lifelike digital imitation of the artist on screens throughout the museum to learn about him, his life, and his artistic legacy. The creation team made use of archival footage, interviews, and writings from the artist to train the AI model. In addition, an actor with similar features was used to add to the realism of the performance. There may be occasions of unnatural facial expressions, slightly awkward pauses or halts in speech, or mismatched expressions that would signal the artificial nature of the responses. While Dalí was a very experimental artist, we cannot know how he would feel about being reanimated in this matter. Though his estate did approve the project in order for the museum to produce the installation, we do not have Dalí’s direct consent.

Exercise 12.7: The coded gaze

Watch these videos featuring Joy Buolamwini. Consider the concept of the coded gaze. What other examples of the coded gaze can you find? Then consider the projects featured on the Algorithmic Justice League website (<https://www.ajl.org/>). How do these projects combat bias? What other techniques might be used to promote fairness and greater transparency?

- <https://www.youtube.com/watch?v=162VzSzzoPs>
- https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms
- <https://www.youtube.com/watch?v=QxuyfWoVV98&t=2s>

How-to example

Major takeaways from Joy Buolamwini’s projects and work include:

- The need for diverse training sets to make sure models represent a wide range of demographics and perspectives.
- Regular audits from outside parties to ensure inaccuracies and biases are being addressed.
- Models are developed by inclusive teams with a variety of stakeholders to consider potential blind spots and assess if models are being properly tuned to meet the desired goals.

Exercise 12.8: Context windows

Pair up with someone and login to your respective AI chatbot of choice (e.g., ChatGPT, CoPilot, Google Gemini)—be sure you both select the same platform. Then enter in the same prompt and compare the outputs. Do you see any differences in your outputs that are typical to the way you tend to interact with the chatbot?

Exercise 12.9: Peer-review

Traditionally, peer-review, the process by which other experts in a field anonymously review, provide feedback, and recommend if certain research should be published or not, has been used as a means of encouraging rigor and integrity within academic fields. Read and consider the article by Bourg et al., which argues that generative AI could be used within the peer-review and academic publishing process. What do you see as the pros and cons?

Bourg, Chris, Sue Kriegsman, Nick Lindsay, Heather Sardis, Erin Stalberg, and Micah Altman. 2024. “Generative AI for Trustworthy, Open, and Equitable Scholarship.” An MIT Exploration of Generative AI, March. <https://doi.org/10.21428/e4baedd9.567bfd15>.

How-to example

AI tools can be very helpful in an editorial process to offer suggestions on how an author might improve clarity or structure, ensure that the writing is uniformly copyedited to match the publisher’s style guide, and screen for plagiarism. Beyond efficiency and consistency, applying AI to the peer-review process could flatten reviews so responses do not vary as broadly and it could ensure that editorial guidelines and policies are generally being followed and applied evenly. However, the potential bias AI algorithms, along with a lack of transparency and risk to privacy and confidentiality would be major concerns. The nuance of human judgment regarding relevance and accuracy of information for the contributions to a field is not something that AI can match. Blind peer-review is an opportunity for colleagues to share honest feedback, sharing both their professional expertise and personal opinions about how to approach and improve the scholarship. This personal perspective is valuable and is not replicated by standard AI.

Exercise 12.10: Grading AI

Try exploring AI's capabilities in academic tasks by experimenting with generative AI tools in your own coursework and research. The goal is not just to use AI, but to critically evaluate its performance and consider its implications for your academic practice. Compare and contrast the AI's performance across three academic tasks. Discuss the potential benefits and limitations of using AI in your academic work. How might AI tools change your approach to learning and research in the future? What are the broader implications of AI for academic integrity and the future of higher education?

Choose from the following list:

- Lecture notetaking: Use AI tools (e.g., live transcription with summarization features, post-lecture audio analysis) to generate notes during or after a lecture.
- Research assistance: Employ AI-powered search tools, literature review summarizers, or question-answering systems to gather information for a research project.
- Presentation creation: Use AI tools to generate slide outlines, suggest visuals, or create draft content for a presentation.
- Assignment response: Utilize AI to generate a draft response to an assignment question, focusing on a specific section or argument.
- Data analysis: If you have any sort of data set, use AI to create visualizations, or to find patterns.
- Coding assistance: If you are working on a coding project, use AI to help with debugging, or with writing new functions.

Document your process and then evaluate the AI's performance based on:

- Accuracy: How factually correct and reliable was the AI's output?
- Relevance: How well did the AI address the specific needs of the task?
- Efficiency: Did the AI save you time or effort?
- Usability: How easy was the AI tool to use?
- Bias: Did you observe any potential biases in the AI's output?
- Originality/creativity: How original or creative was the AI's contribution?
- Ethical considerations: Did the AI raise any ethical concerns regarding authorship, plagiarism, or data privacy?

Assign a letter grade (A–F) to the AI's performance for each task, providing a detailed justification for your grade.

This exercise is inspired by Dr. Kris K. Belden-Adams's (University of Mississippi) presentation "Art History by Bot?: Grading Generative A.I.'s Notetaking, Researching, and Writing Skills." during the Annual College Art Association Conference in 2025.

Exercise 12.11: Terms of use

Compare the following terms of use from OpenAI, Microsoft Co-Pilot, and Google Gemini. What similarities are there among them? What critical differences do you notice?

- <https://openai.com/policies/terms-of-use/>
- <https://www.bing.com/new/termsfuse#content-policy>
- <https://gemini.google/policy-guidelines/?hl=en>

How-to example

The platforms similarly express expectations that users do not engage in harmful, abusive, or illegal activities through the use of the platform or generate offensive or misleading outputs that could cause real-world harm. Each has their own user levels, which grant different experiences of their services. The terms also clearly emphasize their different focus areas and target audiences, such as Microsoft's emphasis on integration with its broader ecosystem. In terms of addressing content creation, OpenAI mentions that its outputs may not be used to train other models, Microsoft states content may not be unique to a single user, and Google Gemini stresses its safeguards against inaccurate and discriminatory content.

Exercise 12.12: Axioms for AI

Consider the AI Axioms put forward by Sonja Drimmer and Christopher Nygren in the following article. What additional axiom would you add?

Nygren, Christopher, and Sonja Drimmer. 2023. "Art History and AI: Ten Axioms." *International Journal for Digital Art History*, no. 9 (April):5.02–5.13. <https://doi.org/10.11588/dah.2023.9.90400>.

Recommended readings

Gebru, Timnit, and Émile P. Torres. 2024. "The TESCREAL Bundle: Eugenics and the Promise of Utopia through Artificial General Intelligence." *First Monday* 29 (4). <https://doi.org/10.5210/fm.v29i4.13636>. 1–42.
 PubPub. "An MIT Exploration of Generative AI." *MIT GenAI*. Accessed March 23, 2025. <https://mit-genai.pubpub.org/>.

Bibliography

Buolamwini, Joy. 2023. *Unmasking AI: My Mission to Protect What Is Human in a World of Machines*. First edition. New York: Random House.

Caldwell, Keegan. 2023. "AI and Intellectual Property: Who Owns It and What Does This Mean for the Future?" *Forbes*. Accessed March 16, 2025. <https://www.forbes.com/councils/forbesbusinesscouncil/2023/10/31/ai-and-intellectual-property-who-owns-it-and-what-does-this-mean-for-the-future/>.

Criado-Perez, Caroline. 2019. *Invisible Women: Data Bias in a World Designed for Men*. New York: Abrams Press.

Gabriel, Saadia, Jessy Xinyi Han, Eric Liu, Isha Puri, Wonyoung So, Fotini Christia, Munzer Dahleh, et al. 2024. "Advancing Equality: Harnessing Generative AI to Combat Systemic Racism." *An MIT Exploration of Generative AI*, March. <https://doi.org/10.21428/e4baedd9.7dc53bbf>.

Groh, Matthew, Aruna Sankaranarayanan, Nikhil Singh, et al. 2024. "Human Detection of Political Speech Deepfakes across Transcripts, Audio, and Video." *Nature Communications* 15: 7629. <https://doi.org/10.1038/s41467-024-51998-z>.

Longpre, Shayne, Robert Mahari, Naana Obeng-Marnu, William Brannon, Tobin South, Jad Kabbara, and Sandy Pentland. March 2024. "Data Authenticity, Consent, and Provenance for AI Are All Broken: What Will It Take to Fix Them?" *An MIT Exploration of Generative AI*. <https://doi.org/10.21428/e4baedd9.a650f77d>.

Riddle, Jacob. 2023. "AI Featured Artists: Jacob Riddle." *Digital Art History Journal*. Accessed March 16, 2025. <https://dahj.org/ai-featured-artists/jacob-riddle>.

Salvaggio, Eryk. 2023. "AI Featured Artists: Eryk Salvaggio." *Digital Art History Journal*. Accessed March 16, 2025. <https://dahj.org/ai-featured-artists/eryk-salvaggio>.

Thaler v. Vidal. "United States Court of Appeals for the Federal Circuit." Opinion No. 21–2347. Decided August 5, 2022. Accessed March 16, 2025. https://www.cafc.uscourts.gov/opinions-orders/21-2347.OPINION.8-5-2022_1988142.pdf.

U.S. Copyright Office. *Copyright and Artificial Intelligence: Part 2- Copyrightability Report*. Accessed March 16, 2025. <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf>.

Zhang, Zhiqing, Wanyi Song, and Peng Liu. 2024. "Making and Interpreting: Digital Humanities as Embodied Action." *Humanities and Social Sciences Communications* 11: 13. <https://doi.org/10.1057/s41599-023-02548-3>.

Resources

- Humanities Perspectives on Artificial Intelligence (NEH) (<https://www.neh.gov/AI>)
- Image-net (<https://www.image-net.org/>)
- University of California, Irvine. "Generative AI and the Humanities." Accessed March 16, 2025. <https://sites.google.com/uci.edu/humanities-ai/home>.
- EP17: Leading AI Ethicist Rumman Chowdhury Defines the Real Risks of Artificial Intelligence (R&G Insights Lab: There Has to Be a Better Way? Podcast) (<https://share.transistor.fm/s/dfc52e52>)
- Holocaust Research Lab (<https://holocaustresearchlab.com/>)
- The Institute for Other Intelligences by Mashinka Firunts Hakopian (<https://mashinkafirunts.com/algorithmic-bias-training-or-lectures-for-intelligent-machines/>)
- Artificial Intelligence course by CrashCourse (https://youtube.com/playlist?list=PL8dPuuaLjXtO65LeD2p4_Sb5XQ51par_b&si=Nyn2FGogAPxu5Ybi)