Teaching Critical AI Literacies[1]:
"Explainer" and Resources for the New Semester
(Currently being updated for Fall 2025)

This germ of this document was prepared by Lauren M. E. Goodlad and Sharon Stoerger in collaboration with the Rutgers AI Round Table Advisory Council and the Office of Teaching Evaluation and Assessment Research. It is now a living document maintained by Critical AI @ Rutgers in conjunction with the *Critical AI* editorial team and our new DESIGN JUSTICE LABS initiative.

We welcome your feedback!

We thank Sabrina Burns (Rutgers, English Class of 2025), for serving as Managing Editor in AY 2024-25.
Special thanks to Emily M. Bender for advice on an early draft.

For an updated Student Guide (which doubles as an ice breaker for discussing gen AI with students), click here; for additional teaching and learning resources see our Educators and Students pages..

Click here for the recorded sessions of our Thursday September 12, 2024 event
RESEARCH IN THE ERA OF GENERATIVE AI: A Hybrid Symposium for Design Justice Thinkers
For video recordings of our earlier October 6, 2023 event,
CRITICAL AI LITERACY IN A TIME OF CHATBOTS:
A Public Symposium for Educators, Writers, and Citizens, click here.
For video recordings from our **DESIGN JUSTICE AI Global Humanities Institute** at the University of Pretoria in July 2024, click here.

The below information and analysis have been curated to help instructors to equip themselves for strong decision-making and productive discussions with students, colleagues, and the general public.
"AI" is a complicated subject with many contexts and far-reaching implications: we have sought to strike a balance between brevity and comprehensiveness.

(1-2) introduction to "artificial intelligence" and "generative AI"
(3) critical AI literacies and actually existing harms
(4) student learning and academic integrity
(5) suggestions for updating assignments and syllabi
(6) "living" list of potential resources.

[1] Readers may ponder the decision to use "literacies" to describe the critical thinking about AI products, histories, and ecosystems which this document strives to support. Although some readers may disagree (or opt for alternatives such as "fluencies"), our view is that *critical AI literacies* position educators, students, and citizens as empowered and active decision-makers as distinct from passive consumers. While *literacy* has sometimes been mobilized to construct binaristic, racist, and exclusionary impediments to equity and citizenship, the harms in question point to *illiteracy* as the disqualifying condition. We believe that robust critical AI literacies are crucial vectors of empowerment and citizenship. It is precisely to combat the disenfranchised position of the passive consumer, who is perceived to lack critical knowledge and decision-making skills with respect to technology, that we mobilize and encourage the teaching of critical AI literacies.

1. *What is "Artificial Intelligence"* (AI)?

Artificial Intelligence (AI) is a common term for an emerging set of computer technologies that affect individuals, communities, societies, and the environment at an increasing scale. Although the phrase "AI" was coined in the 1950s, the field of research to which it refers has undergone multiple transformations and "winters." Moreover, until recently, "AI" was familiar to the general public largely as a theme for science fiction.

"AI" returned to public discussion in the 2010s when a number of innovations in "deep learning" became possible, largely because of the availability of massive stores of human-generated data on the internet and through networked devices. At around the same time, these technologies began to power widespread applications including voice assistants, recommendation systems, and grammar checks. When technologists speak of *deep learning* (DL), which is a type of *machine learning* (ML), the *learning* in question denotes a computer model's ability to "optimize" for useful predictions while "training" on data (a process that involves adjusting the weights in an elaborate set of statistical calculations). The "learning" is *deep* because of the multiple computational layers in the very large models that DL involves. Because AI researchers have used this anthropomorphic language for many decades, today's DL and ML models are often said to "understand," "learn," "reason," "experience," and "think." Although most technologists recognize that products like OpenAI's ChatGPT or Microsoft's Copilot are built on disembodied statistical models that do not "understand, "learn," or "experience" the way that people do, this confusing vocabulary pervades the hype surrounding this resource-intensive technology at the expense of public understanding. Teaching critical AI literacies in the current landscape begins with helping students to distinguish between the functionalities of actually existing technologies, and the fictional "AI" on view in popular media such as *Blade Runner* (1982), *Ex Machina* (2014), or *Westworld* (2016-2022).[2]

[2]As Matthew Stone, Lauren M.E. Goodlad, and Mark Sammons write in their history of chatbots (2024), engineers during the era of digital assistants like Apple's "Siri" understood that the methods they had developed for machine "reasoning" or "learning from experience" were computational proxies for human cognitive faculties, even if these crucial "provisos were largely implicit." From an ML standpoint, machine *experience* equates to the acquisition of new data, while *learning from experience* involves modes of statistical optimization informed by access to this new data during subsequent rounds of training or fine-tuning. NOTE: The editors of this document are part of a working group preparing a suggested set of learning goals for teaching critical AI literacies. As the Rutgers English department "Statement on AI" (n.d.) notes student learning goals in the humanities are often "carefully crafted to emphasize skills in critical thinking, research, textual analysis, and the use of evidence. That is particularly true of writing courses that aim to develop habits of reading and writing that students need to meet rhetorical challenges creatively and to take risks intellectually. Learning goals in literature

TEACHING IDEA: *Ask students to describe a fictional "AI" that they've encountered in a film, novel, series, or video game. Then ask them to describe (to the best of their ability) the fictional technology that enables this imaginary system to work. Finally, ask them to compare and contrast this fictional technology to today's chatbots or image-generating systems. (Remember that not all students have used these new systems; such students may instead discuss their views on fictional AI, or perhaps their views on machine learning algorithms used to recommend content on social media).*

2. What is "Generative AI"?

The most heavily promoted form of "AI" today–often referred to as *generative AI*–involves large language models (LLMs) implemented through chatbot interfaces.[3] The LLMs on which chatbots like OpenAI's GPT-4o, Google's Gemini, Anthropic AI's Claude 4, and Meta's LLaMA series are based, "learn" through computation-intensive "training." Such training involves the modeling (and quasi-memorization) of vast stores of data "scraped" from the internet–in effect creating a compressed statistical

courses emphasize the ability to evaluate and critically assess sources and use the conventions of attribution and citation correctly, as well as to analyze and synthesize information and ideas from multiple sources to generate new insights. Depending on how they are used, generative AI tools can undermine all these goals." The statement goes on to describe the teaching of critical AI literacies as a process of "equipping students with the necessary knowledge for exercising judgment about whether or how to use these imperfect and, so far, largely untested commercial technologies" including the understanding of how such "tools work, what they are capable of, and how to contend with their ethical implications." This document attempts to provide a capacious framework for such teaching as well as links to relevant publications and potential teaching ideas. For a set of potential learning objectives for writing courses that focuses on resistance, see McIntyre, Fernandes, and Sano-Francini (n.d.)

[3] Though the underlying models on which "generative AI" is typically built are called large *language* models, their training data includes visual as well as linguistic content. Today's chatbots are increasingly multi-modal, capable of generating images as well as texts, and sometimes able to generate moving images (including nonconsensual pornography) and/or auditory content (including systems for generating music that, as guitarist Marc Ribot puts it, have trained on "large chunks of copyrighted data" without consent, credit or compensation.) Note that OpenAI, though sometimes described as a "start-up," was valued at about $80 billion in February 2024 and funded partly through multi-billion dollar investments from Microsoft in exchange for a 49% stake.Since that time the company's valuation has increased to $157 billion as of March 2025 as investors continue to pump funding in what what many regard as an unsustainable bubble. For an in-depth account of the board's unexpected November 2023 decision to fire CEO Sam Altman followed by their reversal, see Hao and Wurzel (2023). On the recent exodus from the company see, e.g., Quiroz-Guttierez (2024); on the billions of dollars that OpenAI has burned through, see Efrati and Holmes (2024); on the multi-billion dollar bail-out necessary to keep the company solvent in August 2024, see Okemwa (2024). Anthropic AI was founded by OpenAI employees who, during an earlier exodus, disagreed with OpenAI's direction.. BLOOM, an open source LLM, was created as a collaboration between more than 1000 researchers. Meta's LLaMA series of models are often described as "open source" though as Widder, West, and Whittaker (2023) argue, the practices in question do not meet the criteria defined by the Open Source Initiative.

representation of that data in the form of a multi-layered "architecture" for the passing of statistical weights. During training, these weights are adjusted in order to "optimize" for strong predictions. In doing so, text-generating LLMs leverage a particular software architecture: the *generative pretrained transformer* (GPT).[4] The resulting systems are *probabilistic* (designed to synthesize plausible outputs from a statistical distribution) rather than *deterministic* (designed to consistently deliver the same output in response to the same input).

Hence, despite their ostensible fluency, LLMs do not "understand" language in a human-like way. Rather, as "stochastic parrots" that mimic observed patterns in language probabilistically, GPT-based systems have no means of ensuring the veracity of their outputs or tracking their provenance. This means that GPTs (as OpenAI researchers made clear in a research paper on the topic) are "misaligned" for reliable human use. Hence, generative AI chatbots as we know them depend heavily on the hidden labor of vast bodies of human data workers, typically working under exploitative conditions. These little-discussed labor practices involving millions of people, make "AI" seem more intelligent than it is (see also Section 3 below).

For those seeking to teach and to cultivate critical AI literacies, three points about "generative AI" stand out as especially salient.

- **Generative AI tools are not "search engines".**

Although AI chatbots are often marketed as question-answering systems (superficially akin to the original implementation of Apple's Siri), the LLMs on which they are built do not work by searching the web. Conventional search engines index content found while "crawling" the internet and then provide direct links to those sites; in doing so, search engine developers make an effort to prioritize links to the most authoritative sources. By contrast, generative AI "trains on" on an internet-size trove of "scraped" data and then draws probabilistically on the most common patterns. Consider the example of Google's recent "AI Overview" feature which, in May 2024, falsely identified Barack Obama as the first "Muslim president" of the United States–probably due to

[4] As Stone, Goodlad, and Sammons explain (2024), GPT architectures were developed as "probabilistic scorers," to improve predictive technologies for machine transcription and translation. Whereas earlier word embeddings "could capture data-driven similarities at the level of individual words," *transformer* architectures "work across sequences of words" and "offer statistical proxies for the syntax through which words compose grammatical structures." A *generative* transformer is one that predicts successive words in sequences as those sequences move from the beginning of a sentence to its end–hence, generating human-like text in the act of predicting.

the plentiful misinformation and conspiracy theories in training data scraped from dubious websites that Google's search algorithm would likely deprioritize.[5]

The Association of College and Research Libraries defines information literacy as a process of "inquiry, discovery, and serendipity"—a "complex experience that affects, and is affected by, the cognitive, affective, and social dimensions of the searcher." High-quality search engines support such literacy by situating students as active researchers,[6] helping them to fulfill learning goals. For example, a core goal for the Discipline-Based Writing and Communications curriculum at Rutgers is the ability to "evaluate and critically assess sources," "use the conventions of attribution and citation correctly," as well as to "analyze and synthesize information and ideas from multiple sources to generate new insights." Generative AI tools subvert all of these goals. Even when their outputs are accurate, such tools–which have no means of reliably pointing to the sources for the text they generate–diminish inquiry, discovery, serendipity, and synthesis by serving up statistically probable content in pre-digested form. As Leslie Allison and Tiffany DeRewal (2024) write, the use of generative AI for research makes "it harder for people not only to find trustworthy sources, but also to *know*" when they have done so (see also Shah and Bender 2024 and, for specific commentary on NotebookLM, DeRewal [2025]). To be sure, distinguishing between generative AI and conventional search engines has become ever-more confusing now that platforms like Google (which introduced "AI Overview" into search in May 2024 and "AI Mode" in May 2025) proffer generative syntheses alongside conventional search tools. For student use, we recommend using search engines like kagi.com or DuckDuckGo, which do not collect data or feature fewer energy-intensive AI overviews for the purpose of search activities. [7]

[5] A second consequential issue is the devastating impact on the political economy of the internet: as David Pierce of the *Verge* reports (2024), Google's autogenerated syntheses have broken the "social contract" of the open web. According to a recent Pew Research Center Report (2025) and corresponding analysis from Emanuel Maiberg of *404 Media* (2025), Google's synthetic overviews have undermined the business model of websites of many kinds even as the technology extracts information from those very sites for their own advantage. For example, Google's "AI mode" provides bullet-pointed summaries of Maiberg's stories, but no direct link (instead providing links to aggregating sites).

[6] To be clear, search engines are also subject to bias and exclusions as Safiya Noble's (2018) important work shows. Moreover, Google's emphasis on monetizing user data for revenue has degraded the quality and experience of search to the point of "enshittification." Nonetheless, it remains the case that search engine algorithms were designed to optimize for authoritativeness of the source (the mechanism behind Google's PageRank); by contrast generative tools optimize for a probabilistically plausible response to a user's prompt.

[7] Techniques such as *retrieval augmented generation* (RAG) provide pre-trained models with access to more up-to-date information have enabled tools that feature footnotes or links to sources that may not actually be the source of the synthesized information in question (see also Besen (2023). For their evaluation of Perplexity AI, a tool that uses RAG for supposedly strong research results, see Allison and DeRewal (2024). For an account of how Perplexity AI synthesizes content while misattributing the

***TEACHING IDEA**: Invite students to contemplate the conventions of citation and alongside the notion of information literacy as a process of "inquiry, discovery, and serendipity." For example,*

- *What is the difference between regarding research as a process rather than as a product? Why should "serendipity" matter to this process?*
- *What is the point of citation, practically as well as ethically? (consider sharing some of the following embedded links in preparation for this discussion). How does a collaboratively built resource such as Wikipedia, or a peer-reviewed academic resource, differ from generative AI tools such as ChatGPT or Perplexity AI (both of which are being sued by creative workers and publications including the* New York Times *for violation of copyright). Consider sharing this thinkpiece on search interfaces by Leslie Allison and Tiffany DeRewal to help students to think about "friction" in this context.*[8]
- *According to media theorist Sasha Costanza-Chock, interviewed here in the journal* Critical AI*, a key element to "practice-based learning" is to talk repeatedly about where knowledge comes from through active reflections such as "here's this time and place where I learned this thing." How does this insight into the practice of design justice relate to the conventions of citation?*

- **Generative AI's implementation in the form of conversational chatbots leans into harmful anthropomorphization and courts the ELIZA effect.**

By developing generative tools as chatbots that refer to themselves in the first person and use human-generated scripts, OpenAI abandoned decades of best practice in Natural Language Processing. Chatbot implementations encourage the user's confidence in the system's human-like status and authority. Thus, in contrast to the dominant norms of research from the 1960s to about 2012, today's chatbots are designed to invite the "ELIZA effect" (Stone, Goodlad, and Sammons 2024). The decision to market chatbots as human-like companions seems to have motivated the choice of a Scarlett Johansson-like voice for OpenAI's controversial "Sky" voice

source of this information, see Tim Marchman's article in *Wired*, "Perplexity Plagiarized Our Story About How Perplexity is a Bullshit Machine" (2024). For in-depth evaluation of NotebookLM see Tiffany DeRewal (2025). For additional research contrasting search engines and AI chatbots, see Chirag Shah and Emily M. Bender (2022 and 2024).

[8] Note that we offer *Wikipedia* as an example of a relatively transparent non-profit collaborative resource–not as a perfect research tool. For specific critiques of *Wikipedia* see, for example, Gabrowksi and Klein (2023); Kyle Keeler (2024), and Ming the Merciless (2025).

program.[9] As Kyle Chayka (2024) wrote in the *New Yorker*, OpenAI's release of Sky in May 2024 placed it in the terrain of startups like Replika AI, which specialize in automated "companions." Such determination to blend information retrieval with companion technology that sells "the semblance of emotional connecting," stumbles over the reality that LLM-based systems are better at conversational mimicry than at delivering "reliable information." The result is "a tool that sounds far more convincingly intelligent than it is." In the intervening months, the serious dangers of personified chatbots are becoming increasingly vivid: as tech journalist Brian Merchant notes in an August 2025 substack entry titled "A $500 Billion Tech Company's Core Software Product Is Encouraging Child Suicide," the first-ever case (e.g., Hill 2025) of a wrongful death lawsuit against OpenAI "is at least the third highly publicized case of an AI chatbot influencing a young person's decision to take their own life, and it comes on the heels of mounting cases of dissociation, delusion, and psychosis among users."[10]

- **Generative AI models, which train on the work of others without credit, consent or compensation, involve users in practices that constitute or border on plagiarism.**

[9] Widely perceived to be mimicking the role of the fictional digital assistant that Johansson performed in Spike Jonze's 2013 film *Her*, "Sky" simulated feelings and responded "flirtatiously" (Knight 2024) while delivering a *Her*-like fantasy through a "deferential" and girlish persona that is "wholly focused on the user" (Wilkinson 2024). On the suspension of "Sky" following controversy and legal action by Johansson see Todd Spangler (2024). The fast-developing market for companionate and "character" AI in which chatbots are implemented to impersonate professional roles (such as therapists) or historical figures (such as Harriet Tubman [see Wallace and Peeler [2024]) represents a deliberate departure from the lessons of the ELIZA effect.

[10] Evidence for the harms of these systems has begun to accumulate in research and journalism: e.g., Fang et al. [2025], a study from researchers at MIT Media Lab and OpenAI, argued for a holistic approach to the potential psychosocial harms of chatbot use, including "broader societal interventions aimed at fostering meaningful human connections"; Kosmyna et al. [2025] is among several recent studies to correlate chatbot use with increased homogeneity and "cognitive debt" (see below note 26 for additional studies on learning loss) https://www.wsj.com/tech/ai/mark-zuckerberg-ai-digital-future-0bb04de7] [encouragement to use AI 24/7 https://www.nytimes.com/2025/06/25/magazine/ai-commercials-ads-loneliness.html]A Teen Was Suicidal. ChatGPT Was the Friend He Confided In. - The New York Times; for commentary: Muldowney and Bender (2025). "What My Daughter Told ChatGPT Before She Took Her Life." *New York Times*, 24 August 2025. https://www.cnn.com/2025/09/05/tech/ai-sparked-delusion-chatgpt https://techcrunch.com/2025/08/29/meta-updates-chatbot-rules-to-avoid-inappropriate-topics-with-teen-users/?utm_source=substack&utm_medium=email https://parentstogetheraction.org/wp-content/uploads/2025/09/HEAT_REPORT_CharacterAI_DO_28_09_25.pdf [AI generated sex abuse] https://purl.stanford.edu/mn692xc5736; https://arxiv.org/pdf/2507.21919 ["warmer" models more errors]

According to the computational cognitive scientist Iris van Rooij (2022), since LLMs "produce texts based on ideas generated by others without the user knowing what the exact sources were," generative AI implicates those who use it in a species of "automated plagiarism." Journalists and creative workers are also making the case for plagiarism, as when Perplexity AI was shown to reproduce content from news articles that it did not cite (cf. Marchman 2024;); when the *New York Times* charged OpenAI with seeking "to free-ride" the newspaper's "massive investment in its journalism" (cf. Reuters 2024); and when artists like Karla Ortiz sued Stability AI for its use of copyrighted artworks to train its image-generating model.

Whatever the outcomes of these lawsuits, which bear on the proper limits of "fair use," educators must continue to teach the appropriate use of research and citation.[11] That is doubtless why new conventions have been proposed for the citation of chatbot outputs. But what does it mean for students to cite a chatbot as an information source when the system's own source for the information is buried in undocumented training data? The idea that students can use chatbots "ethically" if they simply cite the generated text, papers over the underlying lack of consent, credit, and compensation. It also overlooks many other harms (to which we next turn).

TEACHING IDEA: *Introduce your students to the above problems of generative AI while reviewing the controversy over Google's "Dear Sydney" Olympics ad from 2024. Ask your students to discuss why the ad triggered such a negative response. Consider asking them to read Alexandra Petri's satirical send-up of the ad. What does the controversy suggest to them about what it means to write a letter? What does it suggest about "generative AI" (or AI more generally)?*

For a more updated version of this assignment, introduce your students to the Meta ads discussed in this NYT Magazine *opinion essay. "Why Does Every Commercial for A.I. Think You're a Moron?" by Ismail Muhamed. According to Muhamed, "what makes these commercials so amusing is that we are watching Silicon Valley struggle to imagine how normal humans might use this technology, and then reverse-engineer the problems those uses might solve."*

Ask your students to discuss the questions that come up in the two Meta ads, one on a "Moby Dick" book club and the other for a young man preparing to meet his girlfriend's father? For example: invite them to describe their feelings (potentially in breakout groups) both toward the ads and toward Ismael's response. How would they write about their own responses to the ads? Do they agree that Meta's pitch to young people

[11] [NYT subscription agreement with Amazon;] On Anthropic's agreement in September 2025 to pay $3k each in a lawsuit representing about 500,000 writers, see, for example, Amanda Silberling *TechCrunch* article, "Screw the Money–Anthropic's 1.5B copyright settlement sucks for writers." The dour assessment stems from Judge William Alsup's ruling that the company's training of Claude despite also finding that the company had infringed on copyrights by downloaded books from a pirated website–incurring a historic payout that some reporting has likened to the "Napster moment" of the early 2000s (e.g. Metz 2025). As Silberling explains, judges may now regard this case (Bartz v. Anthropic) as a precedent; however, another judge may arrive at "a different conclusion."
]

infantilizes, dehumanizes, or (further) isolates these users? What other ways might they recommend to "solve" the "problems" that Meta imagines as plaguing the human social condition? Does it matter that book clubs and the meeting of a girlfriend's father are positioned in this way? Do they think the recurrent consultation of chatbot advice builds confidence or something else?

This assignment can be accompanied by access to OpenAI CEO Sam Altman's boast that "GenZs" "don't really make life decisions without asking ChatGPT what they should do" because the system "has the full context on every person in their life and what they've talked about"--including the surveillance and data-extracting implications of that claim.

3. *Actually Existing Harms of Generative AI*

Despite much talk about "mitigation," the actually existing harms of generative AI are hard to minimize and potentially impossible to eradicate. In addition to the problems for education, research, and dangerous psychosocial ELIZA-effects described above (see also note 10) generative AI's actually existing harms include copyright infringement, embedded biases, misinformation, lack of transparency, built-in surveillance, environmental footprint, and more. Teaching *critical AI literacies* thus includes helping students to learn about the existing and potential harms of these systems.[12] Below we list the chief concerns about generative AI and the practices on which the technology depends. For a more comprehensive survey, see Goodlad and Stone (2024).

- ***Amplification of Bias, Malignant Stereotypes, and "Documentation Debt"***: Since LLM performance relies heavily on large datasets, the best-performing models are riddled with bias and stereotypes from content scraped from the internet. For example, Andrew Hundt and colleagues (2022 753) warn that robots programmed with CLIP (an OpenAI image-to-text classifier), pick up "malignant stereotypes" including "racist, sexist, and scientifically discredited physiognomic behavior"; Frederico Bianchi and colleagues (2022) documented the amplification of demographic stereotypes in large image models (see also Drahl 2023); and according to Valentin Hofmann et al. (2024: 2), LLMs "exhibit archaic stereotypes about speakers" of African American English that resemble "the most negative ever experimentally recorded human stereotypes about African Americans, from before the civil rights movement." As Emily M. Bender, Timnit Gebru and colleagues (2021:

[12] See Kathryn Conrad's "Blueprint for An AI Bill of Rights for Educators and Students," for a useful framework for teaching critical AI literacies (though built on Biden administration recommendations that are no longer extant, the blueprint itself continues to make sense as an educational framework).

615) explain, the overrepresentation of biased and "hegemonic viewpoints" in trained models is made worse by lack of documentation ("documentation debt"). "While documentation allows for potential accountability, undocumented training data perpetuates harm without recourse."[13]

- ***Copyright Infringement, Lack of Consent, Surveillance, and Privacy Concerns***: As we have seen, the use of copyrighted content scraped from the web without consent for the training of AI models has opened a host of legal questions, including high-profile lawsuits claiming improper use of "millions of articles" (Grynbaum and Mac 2023; see also Weatherbed 2023; Allyn 2023; Merchant 2023).[14] Moreover, the accumulation of personal data from users of commercial chatbots expands the surveillant practices that began with the monetization of social media and search engines, exacerbating data privacy concerns.[15] The insatiable demand for

[13] Bender et al. (2021 615) define *documentation debt* as "putting ourselves in a situation where the datasets are both undocumented and too large to document post hoc. …Without documentation, one cannot try to understand training data characteristics in order to mitigate some of the" actual and potential harms. Through probing and audits of LLMs, researchers have discovered "persistent toxic" content (Gehmen et al. 2020 3356) and "severe" bias against Muslims (Abid et al. 2021 298); for the replication of such stereotypes with respect to Muslim-associated names after attempts to debias the model, see Hemmatian, Baltajii, and Varshney (2023). See Sheng et al. 2019 and Lu et al 2019 for examples of gender bias; for evidence that LLMs rationalize their gender biases see Kotek et al. (2023) and on such biases with regard to machine-generated letters of reference, see Wan et al. 2023. Looking at multimodal models, Birhane and colleagues (2021) have found misogynistic and pornographic content. For additional evidence of untrustworthy model behaviors, see Khatun and Brown (2023), Piltch (2023) and Wang et al. (2023). For an important study of bias in facial recognition systems see Buolamwini and Gebru (2018). Foundational research on the topic of algorithmic bias includes Sweeney (2013) O'Neil (2016), Noble (2018), and Benjamin (2019). Broussard's (2019) introduction to AI discusses its cold war-era inception. Research in the field of Artificial Intelligence in Education (AIEd) indicates that AI has the potential to enable beneficial applications in higher education, including intelligent tutoring systems, personalization, and assessment and evaluation (e.g., Luckin and Holmes, 2016); yet it is important to recognize that many of these potential uses have not included a critical reflection of pedagogical research (e.g., Bartolomé, Castañeda, and Adell, 2018; Zawacki-Richter et al, 2019). On related ethical concerns see also Zeide (2023).

[14] The *Atlantic Monthly* documented that hundreds of thousands of copyrighted works are "secretly" being used to train large and proprietary models. See Reisinger (2024) on the use of transcribed YouTube content for training data and Brittain (2025) for updated information on the *Times*'s suit. On Perplexity AI's posting of paywalled journalistic content without permission and with minimal citation see Paczkowski (2024) and for the impact of gen AI syntheses on the internet see footnote 5 above.. On a November 2024 decision to dismiss a case (brought by *Raw Story* and *AlterNet*) for violation of copyright see Masse (2024). See also Cole (2023) on removal of the LAION-5B, used to train popular image models, due to illegal material, including thousands of externally validated images of child sexual abuse. On the flooding of Amazon.com with "scammy" AI-generated imitations of copyrighted books see Knibbs (2024). Legal scholar Sylvie Delacroix (2024) steps back from exploitative data practices to offer a visionary legal framework for a "data trust," built on ideas borrowed from ecocriticism.

[15] Shoshana Zuboff's influential study (2019) describes the underlying business model of tech companies such as Google and Facebook (now Meta) as *surveillance capitalism* (see also Doctorow 2021 and Meredith Whittaker in Coldewey 2023). On the use monitoring software for surveillance in the workplace, see Ackerman (2025).The enormous importance of data accumulation ("big data") in

human-generated data both for lucrative ad targeting and the training of high-performance LLMs, leads companies continually to push the envelope of what the public will tolerate. In August 2024, *404 Media* reported on a service for eavesdropping on customers' cell phone communications in order to target ads. Meanwhile, Meta's image-generating tool, Emu, trained on 1.1 billion "public" images from Facebook and Instagram–practices that, according to journalist Benj Edwards (2023), give "new meaning" to the phrase "If you're not paying for it, you are the product."

- ***Environmental Footprint***: Because generative AI is computationally intensive, the technology uses significantly more energy and water than simple internet search. One estimate calculates the energy footprint of using ChatGPT in lieu of a search engine for information queries as, on average, 10 times more intensive.[16] A March 2024 report on climate disinformation, focused on tech companies that promise that non-existent advanced AI capabilities "will supercharge society's ability to tackle and manage climate change." Such wishful thinking distracts from the reality that, according to the International Energy Agency, rising demand for data center is projected to add "the equivalent of Germany's entire power needs" during the next three years. Reporter Karen Hao, writing in the *Atlantic Monthly* (Hao 2024), notes that the $10 billion that Microsoft is funneling into energy-intensive and water-thirsty data center expansion every quarter marks what one analyst described as "the largest infrastructure buildout that humanity has ever seen" [17] According to reporting from the *MIT Technology Review* (2025), based partly on a December 2024 report from the Lawrence Berkeley National Lab, data center energy use flattened beginning in 2005 due to increased efficiency until in 2017, expansion of AI led data center consumption to grow at "an increasing rate." As of 2023 data centers represent "4.4% of total U.S. electricity consumption" and are expected to grow further, driven by AI-related needs. Citing the 2024 report, *MIT*

training AI and other digital processes continues to be studied across disciplines: for example, Gitelman (2013), Sadowski (2019), D'Ignazio and Klein (2020), Brayne (2021) and Denton et al. (2021).

[16] [hypermegascale] - Hao ch. 12] For a pioneering essay on the environmental footprint of training large models see Strubell et al. (2019) and Luccioni, Viguier, and Ligozat (2023) and Heikkilä (2023); on the water usage involved in training and prompting chatbots, see Li et al. (2023) and on the increased water footprint for Microsoft and Google (due to AI) see O'Brien, Fingerhut, and A.P. (2023); for a more holistic discussion of AI's footprint, see Crawford (2021); on the ecological and environmental costs of cloud computing more generally, see, *e.g.*, Hogan and Vanderau (2019) and Monserrate (2022). On controversial remarks by OpenAI CEO Sam Altman on the need for energy "breakthroughs" to power AI development, see Tangermann (2024).

[17] See also *The Markup*'s February 2025 report on California's efforts to rein in the rate hikes through which the state's residents are believed to subsidize the build-out of data centers while also encouraging "more energy efficiency or use of clean energy on the part of the tech companies, entrepreneurs, and IT departments that utilize the centers" (Johnson 2025).

Tech Review writes that, by 2028, AI use alone could consume "as much electricity annually as 22% of all US households," while data centers, in the effort to meet growing demand, are trending toward "dirtier, more carbon-intensive forms of energy." The reporters add: "All of this growth is for a new technology that's still finding its footing" in domains such as education, medicine, and law, and which may be "the wrong tool for the job or at least have a less energy-intensive alternative."[18]

As ecocritic and media scholar Mel Hogan writes (2024), "When thinking of AI's destructive impacts on the environment—either as the pollution emitted from training large language models…or the exhaust from machine vision used to train self-driving cars, or the destruction and pilfering that results from military's uses of autonomous drones, among (so) many other examples—it's important to also consider the AI industry's integration into existing mining and fossil fuel companies that have for centuries been destroying any kind of sustainable conditions for life on earth and foreclosing alternatives."

- ***Exploitation of Human Labor***: Since generative AI chatbots are subject to bias, misinformation, and toxicity, the current technology relies on millions of low-paid workers whose high-speed annotations improve results–producing the illusion but not the reality of automated human-like intelligence.[19] Hence, what commercial developers hype as the automation of human-level tasks and even the imminence of "AGI,"[20] quietly relies on a vast and expanding human "underclass" which is usually poorly paid and can entail traumatic exposure to disturbing content for workers in the global south (Perrigo 2023; Tan and Cabato 2023). Moreover, as Elizabeth Losh explains (*CAI* 2.2), tech companies shift the burdens of their faulty systems to other workers including educators, librarians, students, and the parents of schoolchildren.[21]

---

[18] See Roshan (2025) for reporting on secretive plans to build a "massive" Meta data center in rural Louisiana which is likely to create significant economic and health-related harms for the local community. [Greenpeace 2025 report: https://www.oeko.de/fileadmin/oekodoc/Report_KI_ENG.pdf]

[19] Recent journalism documents how supposedly automated chatbots require massive input from workers tasked with the labor of labeling violent and disturbing content, often outsourced to low-paid workers in the global South. As one article reports, the practice of "auctioning off work globally" creates "a race to the bottom for wages" On the longstanding use of human crowdworkers for machine learning and the improvement of automated systems, see, for instance, Ross et al. (2010), Irani (2015), Gray and Suri (2019), and Crawford (2021, chapter 2).

[20] On the vexed topic of AGI ("Artificial General Intelligence), a poorly defined concept often leveraged for marketing purposes and bound up in the history of eugenics, see Gebru and Torres (2024) as well as Goodlad and Stone (2024).

[21] For strong teaching resources on hidden data work, see, for example, the community-based Data Workers' Inquiry (co-organized with the DAIR Institute); Gray and Suri's 2019 monograph; and the

- ***Misinformation and Degradation of the Internet (through Misconceptions, "Hallucinations," Conspiracy Theories, and Malicious Use)***: Generative AI's dependence on a vast crowdworker underclass is directly related to the limitations of probabilistic systems. As we have seen, generative models do not understand language in a humanlike way, cannot distinguish between truth and falsehood, and persistently fail to recognize inappropriate stereotypes and biases. The result is that LLMs and other generative models are likely contributing to the stream of socially and politically destabilizing misinformation on social media (e.g., Meyers and Thompson 2025),, dubious websites, and a degraded online ecosystem. Tech columnist Julia Angwin (2023) warns that, with the advent of generative AI, the internet is becoming "even more polluted with untrustworthy content." "While creators of quality content are contesting how their work is being used" and may therefore hesitate to post online, she explains, "dubious A.I.-generated content is stampeding into the public sphere." On September 3, 2024, *Proof News*, a data journalism collective, reported that the probing of five leading AI models delivered wrong or misleading information about presidential candidates Kamala Harris and Donald Trump thirty percent of the time. More recently, *The Columbia Journalism Review* (Jazwinska and Chandrasekar 2025) studied eight AI research tools and found, for example that premium systems "provided more confidently incorrect answers than their free counterparts" ChatGPT alone "incorrectly identified 134 articles, but signaled a lack of confidence just fifteen times out of it its two hundred responses, and never declined to provide an answer."[22]

  Although the industry's preferred term for LLMs' most bizarre outputs is "hallucination," in actuality these are simply bad predictions that arise due to the lack of a fundamental understanding of language and the world that it mediates. As Naomi Klein rightly notes (2023), applying the anthropomorphizing language of "hallucination" to a statistical model is misleading and problematic (see also Birhane and Raji 2022; and Fredrikzon 2025). Since ChatGPT's release in November 2022, malicious use of these systems, including the practice of "jailbreaking" chatbots by circumventing their instructions, has sometimes been treated as a comical pastime. However, given that the topic includes deepfakes, non-consensual

pioneering work of Irani and Silberman (2010). For a recent student project that surveys some of this content in the form of video commentary, see Mahek Shah (2025).

[22] See also Peters and Chin-Yee (2025) for a study that finds significant failure to "generalize" in the summarizing of scientific texts: with the top models proffering insufficiently specific results in 26-73% of cases and underperforming human summaries at a rate of nearly 5 to 1; and with the newest models performing worse in "generalization accuracy" than earlier ones

pornography, and the potential hacking of cars and other powerful automated systems, malicious use of AI is, of course, a serious matter.[23]

- ***Political Economy, Concentration of Power, Lack of Transparency and Accountability***: The political economy of "AI" today was forged through the concentration of computing, economic, and data resources in some of the largest and most lucrative companies in the world. Corporations such as Google, Microsoft (and their OpenAI partner) intensively lobby legislators, sometimes "watering down" regulatory demands for transparency, accountability, and fairness. Lina Khan, who was chair of the Federal Trade Commission for the Biden administration, has described the risks of "AI" in a context of "race-to-the-bottom business models and monopolistic control."[24] According to Harvard Law professor Lawrence Lessig, California's "Safe and Secure Innovation for Frontier Artificial Intelligence Models Act" (SB1047), a "modest" measure for commonsense regulation, is (as of August 2024) "the target of an extraordinary lobbying effort" "If the bill fails," he writes in the *Nation*, it will signal the "power of money in American politics" and the country's incapacity to regulate (see also Lovely 2024).

In an educational setting, chatbots also create particular challenges for student learning and academic integrity–a subject to which we now turn.

TEACHING IDEA: *Choose one or more published resources from each of these categories for class reading and discussion, potentially by organizing the class into groups that focus on each topic. Ask the class to discuss and/or produce a set of notes on the actually existing harms of "generative AI" and the concerns such harms generate. Consider having the class draft their own recommendations for an "ethical" approach to the technology. Prepare them for the difficulty of this question perhaps by beginning with discussion of what "ethical" decision-making entails!*

---

[23] See Maiberg (2023) for a disturbing account of how generative models are used to "produce any kind of pornographic scenario…trained on real images of real people scraped without consent from every corner of the internet." See Funk, Shahbaz, and Vesteinsson (2023) for a report documenting how generative tools are being used to "supercharge online disinformation campaigns" and to "strengthen censorship" in authoritarian countries.

[24] See Whittaker (2021: 51), co-founder of the AI Now Institute, for the case that AI technology "cedes inordinate power" to a handful of corporations while significantly "capturing" academic research in the field. Estrin, who is the former CTO of Cisco, argues that the "hubris and determination of tech leaders to control society is threatening our individual, societal, and business autonomy." See Hao (2023) for discussion of a Stanford "transparency index" (Bommasani 2023) which, while itself arguably insufficient, found a wide range of gaps in disclosure including lists specifying the "authors, artists, and others" whose works were used for training; the use of copyrighted works; and documentation of a model's known biases and confabulations.

Consider supplementing (or succeeding) this idea by having students view the plenary panel on "Accountability and Online Safety" featuring historian Brittney Cooper and technologist Abeba Birhane at the recent DESIGN JUSTICE AI institute (begin at 29:35 on this video).

4. Student Learning and Academic Integrity: Research and Reflections

Despite the hype over AI's supposed capacity to transform education, researchers have hardly begun to evaluate the impact on student learning. There is, however, a century-long history of enthusiasts overpromising on the "personalized" benefits of education technology (Watters 2023). As the *New York Times*'s Natasha Singer (2025) explains, OpenAI's current "campaign" to expand subscriptions at institutions of higher learning "is part of an escalating A.I. arms race among tech giants to win over universities and students"; moreover, Google and Microsoft "have for years pushed to get their computers and software into schools, and court students as future customers." Hence, the tech industry's current "push to A.I.-ify college education, amounts to a national experiment on millions of students." The pressure is often even greater in K-12 education: according to Alex Molnar (qtd. in Grose 2025), who directs the National Education Policy Center at the University of Colorado, generative AI is being "forced upon" K–12 schools "without any particular context or funding that would allow them to make informed decisions about what may or may not be valuable to them." Columnist Jessica Grose (2024) cites a Pew Research finding that only 6 percent of public school teachers in the US think that AI tools produce more benefit than harm.[25]

Instead of solid evidence for AI's educational benefits, enthusiasts often assert that adoption of bots is required to prepare students for jobs. Some argue that AI will level the playing field (by equipping all students with newfound capacities), while others believe that students trained to use AI will be more employable than their peers. Few focus on teaching students *how* these technologies work–though such knowledge is integral to (critical) AI literacies.[26] Indeed, it is by no means clear that a student trained

[25] See footnotes 25 and 26 for early research confirming these intuitions.

[26] As writing center director Jane Rosenzweig writes in a helpful blog post: "It is crucial that we teach our students to think critically about generative AI–and asking them to engage with AI tools in different contexts across different disciplines will be an important part of that process. But rather than simply asking students to turn to the chatbot for 'feedback' or for any other step of the writing process, we should be helping our students understand how LLMs are trained, what types of data they are trained on, what we don't know about that data, and how bias is baked into these systems." Audrey Watters rightly argues (2025) that what is often portrayed as a crisis attributable to chatbots alone is as socio-cultural as it is technological. That is, disengagement among students is worsened by standardized testing and associated curricular changes; social media and other screens; the decline of

to depend on bots for a wide range of tasks is a more attractive employee than one who has learned how to probe these tools and understand their serious limitations (e.g., Estrada 2025; Ramoni 2025a Ramoni 2025b). Certainly all students need to recognize the risks of entrusting high-stakes tasks to probabilistic statistical models: consider CNBC's May 2024 report that young job-seekers are sending companies "hundreds of the exact same cover letters word for word."[27]

AI enthusiasts may counter that students trained to write good prompts will avoid such pitfalls. But what if the ability to write efficacious prompts–much like writing well in the first place–is a high-level sociocognitive capacity that requires experience, relevant knowledge, and control over the writing process? Many AI enthusiasts recommend that students prompt chatbots to generate a first draft, which they afterwards edit and revise. Once again, the approach may short-circuit a complex process. Whereas editing and revision require hard-won habits of critical reflection and rhetorical skill, chatbot writing suffers from simplistic, derivative, or inaccurate content. Where, then, is the evidence that students invited to skip over the tried-and-true building blocks of college writing will thrive as "prompt engineers," fact-checkers, and editors of mediocre text? Where too is the opportunity for such students to explore their own ideas and develop their voice?

In the meantime, the internet is already teeming with tips on prompts, which are warehoused on websites, recycled on social media, surveilled by AI developers, and generated automatically by systems including ChatGPT (see also Halm forthcoming). In such a milieu, students striving to learn need more than prompting techniques to develop academic work they can proudly claim as their own. Early research suggests that even "brainstorming" with bots may reduce students' confidence and

reading (including the decline of parents reading aloud); the lingering effects of the pandemic; the instrumentalist narrative of education as a vehicle for "job skills"; increasing costs; growing economic inequality and growing competition for good jobs in a dwindling economy.

[27] On the deleterious effects of generative AI on writing quality, see computer scientist Margaret Mitchell's January 2025 thread on Bluesky concerning platforms that are aggressively promoting the (unprompted) use of the technology for users' everyday writing tasks: these include reduction of originality, pressure to homogenize, increase of erroneous content, reduction of information diversity, deterioration of web content (and future training data), and irresponsible abuse of corporate power. See also Du, Gross, and Hong (2025) for a compelling case for prioritizing a writing process that prioritizes voice over shallow focus on surface polish [and *New Yorker* writer Kyle Chayka (2025) on recent studies that document the homogenizing effects of chatbots on writing in conjunction with reduced brain activity.]

"self-efficacy."[28] After all, work that demonstrates creativity, thoughtfulness, care, and resilience is hardly to be gotten at the touch of a button. While there may be real value in learning to recognize the flaws in a chatbot's outputs, that does not mean that tasking students to "write" by improving auto-generated content is a good way to inspire them, help them to cultivate their own articulacy, or sharpen their ability to think for themselves. As educators are called on to undertake experiments on their own students that the majority of business are not ready to trial on their customers or clients, the goal of preparing students for the future should not be abandoned to industry recommendations or harried administrators who have not yet had the time to develop their own critical AI literacies (Goodlad 2025).

- ***The Economic Objectives of Generative AI***

Teachers of critical AI literacies invariably encounter the tensions between the goals of higher education, and those surrounding the design and implementation of generative AI. Whereas education aims to strengthen students' articulacy, understanding, and application of knowledge, promoters of generative AI aim to transform the economy–perhaps even to usher in a fourth industrial revolution. College writing, which is a core proficiency for undergraduate education, is at the very center of this tension. Long considered a recursive *process*, college writing usually begins with reading and/or research; proceeds to "pre-writing" practices such as

[28] See Nataliya Kosmyna et al. for a study (2025) comparing LLM-users and non-users which found that the former displayed "consistent homogeneity" across a range of tasks; fell behind on ability to quote from essays composed "just minutes prior" (see also Chayka [2025] on the same study). The study concluded by predicting that LLM use is "likely" to coincide with a " decrease in learning skills." In a study of the impact of LLM-use on creativity, Harsh Kumar and colleagues (2024) observed that "participants who had no prior exposure to LLMs consistently performed better," for example by "generat[ing] more original ideas on average" than those exposed to LLMs. Their "findings suggest that while LLMs may provide short-term boosts in creativity during assisted tasks, they might inadvertently hinder independent creative performance." According to Hamsa Bastani et al. (2024) high schoolers who encountered math instruction by chatbot experienced substantial learning loss. Sabrina Habib and colleagues (2024), found that use of ChatGPT for brainstorming resulted in "reduced self-efficacy" for those still developing diverse "thinking skills" and "creative confidence": "as some participants expressed difficulty in coming up with ideas beyond what the AI offered." The study suggests that educators whose lessons enlist students to criticize, edit, and fact-check chatbot outputs may be overestimating (and thus undermining the development of) their students' competencies for these tasks. From a student perspective, ChatGPT's output of grammatically and syntactically correct prose, and its authoritative tone may seem like an unreachable ideal–not a machine-generated draft amenable to improvement from a novice writer. On the matter of coding assistants in industry, see, for example, Aslan's (2024) insider account of "the learning curve paradox." Although junior developers using coding assistants experienced some boosted productivity they "exhibited a shallow understanding of fundamental concepts. When asked why specific patterns were used, many struggled to explain their reasoning. *The reliance on AI seemed to shift focus from learning to completion*." (emphasis added).

"freewriting" and "brainstorming"; and culminates in revision. By contrast, for AI developers writing is a *product,* the speedy delivery of which can maximize productivity. "The most important thing that technological advancement does," writes one MIT researcher on ChatGPT, is to enable workers to "produce economic output more efficiently."

Ironically, generative AI is stumbling because it has yet to deliver anything like such world-historical efficiencies–a point Goldman Sachs emphasized in a June 2024 report titled "Generative AI: Too Much Spend, Too Little Benefit?." As journalist Mateo Wong writes in a cogent analysis, the "industry is asking the world to engage in something like a trillion-dollar tautology." That is, "AI's world-transformative potential justifies spending any amount of resources, because its evangelists will spend any amount to make AI transform the world."[29] This situation has made the education market ever more important to tech investors. That is, if generative AI can be portrayed as a productivity boost for teachers and a trustworthy "copilot" for students, schools could provide the industry with the legitimacy, growth, and profitability that investors crave.

While there is no single way to instill critical AI literacies, educators committed to such teaching, we urge, must steer clear of the industry's tautology and recognize the persistence of irresponsible hype. Too often educators both in K-12 and higher ed are asked to assume that since investors are spending billions to make commercial tools accessible to students–all while embedding these tools into devices and platforms in ways that make them difficult to avoid–their primary role is now to teach the "ethical" and "responsible" use. This flawed thinking enlists teachers to whitewash the cognitive, social, and environmental impacts of an underregulated technology, while simultaneously striving to ensure student learning in the face of tools that were not designed with education in mind.[30] The truth is that no teacher or student can

[29]For more recent confirmation of the limited business uptake of the technology, see for example, *The Economist's* May 2025 report, "Welcome to the AI Trough of Disillusionment"; Sheryl Estrada's August 2025 *Fortune* article, "MIT report: 95% of generative AI pilots at companies are failing," describing a survey finding that only 5% of such pilots were accelerating revenue; Grant Grose's August 2025 *CIO* article, "GenAI Descends into Disillusionment;" Sri Ripidi's September 2025 *Information* article "OpenAI Says Its Business Will Burn $115 billion Through 2025," which reports increased losses of more than $80 billion since the first projection; and Bryan McMahon's September 2025 article in *The American Prospect*, reflecting on the widely perceived failures of OpenAI's release of GPT-5 in August 2025, warns that the "financial bet" on so-called artificial general intelligence (AGI) is so big (and so dependent on OpenAI's hype that failure could cause an economic depression.

[30] On cognitive harms see above, including note 26. Discuss of "responsible" and/or "ethical" use in such discourse typically focuses on transparency of usage (such as the citation of chatbot content to document its use). Those who advocate for active adoption and use of chatbots in their pedagogy while simultaneously emphasizing the importance of teaching the harms of these systems, in effect reduce

neutralize AI's pervasive harms–which spring from a concentrated political economy bent on expanding profitable forms of resource-intensive surveillance, data-capture, and platform dominance. A critical AI literacies approach responds by helping students to "get the facts," and equipping them to make decisions about whether or how to use chatbots from positions of knowledge, citizenship, and care. One approach to this effort emphasizes how a student's own research into how generative AI works–for example, through simple "probes" and/or audits of models–can provide skills and insights into the technology's strengths and limitations (of the kind that potential employers may value) without turning students into habitual users.

As generative AI struggles to find a firm foothold in business, professional work, and everyday life, those teaching critical AI literacies will need to distinguish hype from reality. To be sure, machine learning programs trained on high-quality data in dialogue with experts and community stakeholders can produce valuable tools in specific domains–welcome technologies that will likely be called "AI."[31] But "generative AI"–especially when operationalized in the form of anthropomorphized chatbots–is a very different technology that, despite its high costs, persistent unreliability, and manifold harms, strives to be all things to all people all of the time.

That is why we urge resisting the pressure to succumb to technodeterminist and tautological thinking. As the suggestions for syllabi and assignments which follow make clear, teaching critical AI literacies does not entail "banning" AI, policing students, or fueling panics of any kind. Nor is the point to shame students who seek strategies for using generative tools (opportunities for which have proliferated online and in many campus spaces). Rather, teaching critical AI literacies involves enabling students to understand what generative AI–or any other automated technology–can and cannot do. It emphasizes a student's need for informed straight-shooting that can prepare young learners to exercise judgment, and counter the hard sell and hype. Educators already know how to do the rest.[32]

> ***TEACHING IDEA**: Choose one or more learning goals from your syllabus and invite students to discuss their ideas about how best to achieve these objectives. If these*

---

the question of "ethical" use to a potential requirement to acknowledge harms in the hope that they will improve over time. Critical AI literacies as we understand it, must meet a higher bar.

[31] Less auspicious than well-scoped, special-purpose technologies in, say, weather prediction or drug applications are machine learning systems devoted to harmful predictive technologies including algorithmic pricing, facial recognition, and (supposed) fraud detection.

[32] See *Critical AI*'s two-part special issue, "Beyond Chatbot-K: Large Language Models, Generative AI, and the Rise of Chatbots." We have also begun an ongoing series, "Generative AI and Teaching Writing in Higher Ed,",guest-edited by Marit MacArthur (UC, Davis), the first entries of which are forthcoming in *Critical AI* 3.2 (October 2025).

goal(s) bear directly on generative AI, invite your students to explore potential impacts on their learning. Consider inviting your students to co-create a contract or set of policies that involve possible use of permissible digital technologies. For example, should students adopt a preferred search engine or relevant library application for their research? If they are permitted to use grammar check should they commit to disabling any "generative" or "AI" features that the application offers? Ask students to specify how they believe these choices will affect learning goals and the classroom community.[33]

5. *Suggestions for Updating Assignments and Syllabi (through Clarifying Course Policies and Learning Goals)*

Despite an already full workload, every instructor should review their syllabus and assignments to assure the clearest possible articulation of policies regarding the use of generative AI. This process could begin with a close look at your institution's code of conduct and the learning goals for your course; you might also wish to distribute our student guide to your class (possibly discussing it as a classroom activity).

The code of conduct at Rutgers states that students must ensure "that all work submitted in a course, academic research, or other activity is the student's own and created without the aid of impermissible technologies, materials, or collaborations." In creating policies on generative AI tools, this puts special emphasis on the identification of permissible technologies and the question of whether a given tool impedes the learning goals of the course (including the submission of suitable work that is "the student's own").

At Rutgers, learning goals vary widely across and within schools, disciplines, majors, pedagogical approaches, and levels of difficulty. This means that course policies and teaching approaches that effectively build critical AI literacies may look different.

For example,

[33] In September 2024, Rutgers had sponsored an enhanced version of the "Grammarly writing support tool" that offers a variety of features but will not "enable the generative artificial intelligence (AI) features"--the university will instead "work with faculty to assess and understand the impact of AI on instruction. In January 2025, the AI features were enabled in large part because, according to information shared with a faculty committee, some instructors wanted their students to have access and it is not possible to provide access to some classrooms and not to others. No doubt there is a great deal of research to be done on how different campuses are rolling out different tools, with or without meaningful shared governance, through what campus entity, and through what kind of messaging to faculty, students, or both. It is, of course, possible for users to manually disable (or simply ignore) Grammarly's generative features.

- A computer science instructor teaching an introductory course may wish to prohibit students from using chatbots for coding in order to ensure that they learn fundamental skills; but she may wish to allow such use in an advanced course designed for those who have already mastered these skills.

- A professor might organize "probing" experiments that enable students to investigate model bias, perhaps preparing them to publish their results on a class website.[34]

- An instructor teaching research at the graduate level may wish to allow students to use chatbots to improve grammar and syntax so long as they document that the actual research is their own. To save energy and avoid inadvertent changes to content, she may instruct her students to disable the generative features on a tool such as Grammarly.

- A humanities instructor who assigns writing and research to build critical thinking and sharpen engagement with course materials and themes may determine that use of chatbots poses a serious impediment to these objectives. He may therefore explain why the use of these tools is impermissible for assigned writing. However, he may simultaneously assign a research task in which students compare and contrast resources they find using search engines, library databases, or Wikipedia to those they find through chatbot use. He may also recommend that students use kagi.com or DuckDuckGo, search engines that, unlike Google, do not surveil users or monetize their data. (Note that as of 2025, both have begun to incorporate some AI features, thus limiting the extent to these search engines help users to avoid the recourse to resource-intensive computations for simple queries.

- An environmental science course may focus on the use of water, energy, and rare-earth metals required to train and deploy chatbot systems without using these tools in class.

- A course in law or in graphic design may involve the study of "AI" copyright infringement across different companies and domains while inviting students to use these tools to audit models for research regarding intellectual property.

[34] For examples of simple "probing techniques" that teachers can introduce into their class, see Daniel Estrada's slides, "Teach Your Students to Audit." and Teresa Ramoni's slides, "Fostering Critical AI Literacy Through Research and Probing."

Of course, students will have their own views on the topic:

- Some may wish to opt out of using tools known to embed harmful stereotypes and/or subject users to surveillance and data collection.

- Some may seek to hone their ability to use "AI" in order to prepare themselves for the job market. They may find that a professor who assigns the probing of model bias and inaccuracies has prepared them to demonstrate their skills more deeply than a course that simply allows chatbot use for the generating of text.

- Some may recognize the virtue of "probing" models but have privacy concerns. They may request an alternative assignment that does not require them to sign up for a surveillant tool.

The good news is that all of these situations can effectively teach and enhance critical AI literacies, whether AI tools are directly used or not.

Below we offer recommendation on ASSIGNMENTS, USE OF AI "DETECTORS," PEER REVIEW PRACTICES, and SYLLABUS UPDATES:

ASSIGNMENTS: In reviewing assignments, instructors may wish to implement changes in light of the fact that students may be tempted to use AI tools even if they are told not to do so. Simple response papers ("what did you think of this reading?") might work best in a classroom setting in handwritten fashion (or with wifi disabled on computers, phones, and tablets for those needing accommodations).

- In place of conventional take-home essays, consider alternatives that may boost student interest and enhance opportunities for active and project-based learning (*e.g.*, in-class presentation, video, podcast, e-portfolio, oral examination)

- For research papers and other time-intensive take-home writing, try the following recommendations:
    a) invite students to develop a topic they care about so as to encourage intrinsic motivation;
    b) develop rubrics that emphasize critical thinking, problem-solving, applied knowledge, and use of evidence and which de-emphasize the summarization of content and the perfection of grammar and syntax.

Include requirements for use and citation of specific evidence, whether drawn from course materials or from independently researched sources. (Bot-generated text tends toward summary and generality with little or no quotation; when prompted to provide quotations, bots often deliver quoted material that is fabricated or incorrect). Consider expanding the criteria by requiring explicit engagement of class themes or discussion (so as to avoid pat and predigested content).

c) consider assigning a "high-stakes" pre-writing assignment (that counts for a significant part of the student's grade); this can take the form of a detailed "questionnaire" that calls on students to address specific components of the writing process that apply to course materials directly and are not readily automated.[35]

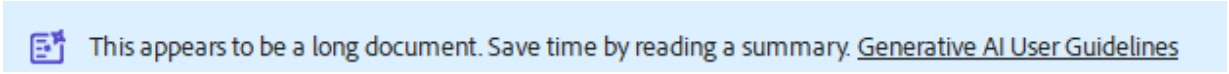
d) consider demoting the importance of grammar and syntax for take-home writing in your grading rubrics. Accentuate the importance of engaging core questions, meeting specified criteria, developing ideas, using and appropriately citing concrete evidence, cultivating persuasiveness and individual voice.

e) consider the use of a mandatory student certification comparable to that which follows: "In concert with [Rutgers'] code of conduct, which mandates "that all work submitted in a course, academic research, or other activity is the student's own and created without the aid of impermissible technologies, materials, or collaborations," this course has been designed to promote your learning, critical thinking, skills, and intellectual development without reliance on unauthorized technology including chatbots and other forms of "artificial intelligence" (AI). [Although you may use search engines, spell-check, and simple grammar-check in completing your assignments,] you will be asked to submit your written work with the following statement. ***"I certify that this assignment represents my own work. I have not used any unauthorized or unacknowledged assistance or sources in completing it including free or commercial systems or services offered on the internet, or text generating systems embedded into software."*** Please consult with your instructor if you have any questions about the permissible use of technology in this class.

[35] For more details on this suggestion, see point #6 in these slides, "Teaching and Generative AI."

- Consider combining take-home assignments (or replacing them) with in-class exams (written by hand or with wifi disabled for students with accommodations).

- Bear in mind that student success with all kinds of assignments often depends on careful reading of assigned materials and/or the taking of notes. Adobe pdf readers in the effort to upsell AI features to their users now include banners like that pasted in below which invite readers to "save time" by reading an AI-generated summary (ignoring the document flaws of AI summaries [e.g., see DeRewal 2025] and implying the equivalence of reading any summary to reading "a long document."



  To help ensure careful reading, consider assigning (interesting) multiple choice exams and/or in class close reading exercises that help to encourage rigorous engagement of key passages and challenging ideas. NOTE: Many instructors find that students are surprisingly unfamiliar with multiple choice, which, if properly conceived, can help students with comparison and contrast and other important analytical skills (e.g., which of these answers is false?).

- Many AI tools profess to reduce the "cognitive load" of note-taking by automating the generation of notes. Apart from the intrusion on classroom privacy that such systems present, the logic behind this technology ignores that the "cognitive load" of note-taking is part of why the practice helps students to actively learn. Consider assigning each day's note-taking (by hand or with a wi-fi- disabled device) as an assignment; although each student will take their own notes daily, the day's assigned note-taker will share their notes with the entire class.

***TEACHING IDEA**: In advance, generate an automated summary of an assigned reading (or find one online). Before showing the summary, ask students to choose one or more passages in the reading that they found especially challenging, interesting, provocative, or illuminating. After discussion, show them the summary and invite them to discuss what the automated summary identified and what it passed over, decontextualized, or misinterpreted (bear in mind that because language modeling works through pattern-finding in large datasets, summaries are often generated by locating the parts of the assigned reading that are similar to texts observed in the training data).*

Next (or as a separate plan) compare the synthetic summary to a relevant Wikipedia page on the text, author, or topic in question. Invite your students to compare the information available on a Wikipedia page, and the protocols for its provenance (including crowd-sourcing, citation, moderation) If–as often happen– the autogenerated summary resembles the Wikipedia entry (perhaps even repeating some of its language), discuss the ethical ramifications of this lack of credit and attribution.
Ask students how they would feel if their work on a blog, article, or wikipedia page showed up in a generative AI output without any attribution and reduced traffic to their page accordingly. Consider sharing some passages from this article which argues that AI's use of content for training data has dissolved the web's "social contract."

AI "DETECTION": The media discourse around student "cheating" is permeated by hype. Moreover, such discourse sometimes portrays college assignments as if they were task-specific labor disconnected from learning and the application of critical thinking. We believe that intrinsic motivation is one of the best ways to ensure a student's engagement with written work, research, and other forms of assessment.[36] A focus on cheating or plagiarism, on the other hand, can undermine the relationship between teachers and students. That said, we recognize that instructors need to ensure fairness and academic integrity in the classroom and that non-permitted use of generative AI has created added hurdles.

Nonetheless, please be aware no current system being marketed to "detect" machine-generated text is reliable: false positives and false negatives are possible and even likely. Some of these tools evince biases against non-native English speakers.[37] Finally, use of AI detection software, which is not FERPA-protected, may also violate students' privacy or intellectual property rights.[38] We suggest that instructors avoid these systems or at least discount them as reliable evidence for violations of academic integrity.

[36] See Lang (2013) for research that examines academic dishonesty and factors that "encourage" students to engage in this behavior. Lang also has three part series in the *Chronicle of Higher Education* on this research. (Part 1, Part 2, Part 3).
[37] See Liang et al. (2023) for specific details about this study. Anecdotal accounts of outputs circulating on social media suggest that neurodivergent people may also be at risk for discriminatory assessment. See Verma et al. (2023) for news of a new and allegedly more reliable detector created by Berkeley NLP researchers (2023).
[38] In this context, it is worth noting that many of the digital tools students use voluntarily or according to instructor guidelines involve breaches of privacy and IP rights, including, at various points, Grammarly, Google Docs, and Zoom: on this changing landscape see, for example, Knowles (2023) and Merchant (2023).

Instructors who suspect the unauthorized use of an AI tool in their course should consider asking for a meeting to discuss the student's approach to completing the work and refer cautiously to the problematic content. An instructor who simply asks the student to describe the writing process that led to the work in question may be more successful than one who explicitly brings up potential misuse of AI. If impermissible use of AI seems likely (as when fabricated quotations and other inexplicable content shows up in a take-home assignment), a good first step might be to consult with an appropriate administrator for advice to learn more about recommended policies and suggested next steps.
If you continue to suspect misuse of generative tools, consider experimenting with changes in assignments as described above. For further conversation, feel free to reach out to criticalai@sas.rutgers.edu

PEER REVIEW: Peer review is a worthwhile practice that helps to create community and idea-sharing among students. Unfortunately, some students may submit their classmates' work to AI systems in order to simulate peer review. This is a serious violation of academic integrity since it involves another student's intellectual work. If peer-review is an assigned take-home task, consider prefacing the assignment with a notice about the gravity of this offense and/or have students sign a statement certifying that they have not used unauthorized devices (see above for an adaptable template).

SYLLABUS: Whether an instructor wishes to build in the use of chatbots for certain assignments, allow students to experiment with them as they wish, or prohibit their use, we recommend clarifying these policies on syllabi and discussing them with students. Explain how you reached a decision that comports with the learning goals for the course. Consider discussing how chatbots work and the various problems described on this webpage, in concert with our Student Guide (see additional resources below), Colleagues at Critical AI @ Rutgers are available to answer specific questions or suggestions for teaching.

I. For instructors who do not want students to use AI tools for their course

When specifying on one's syllabus that the use of chatbots and other AI tools is *not* permissible, instructors should be as clear as possible and may wish to refer to the Rutgers code of conduct, cited above, in doing so. Given that AI tools are now widely incorporated seamlessly into platforms including Google, Adobe, grammar-checking tools such as Grammarly, and software suites such as Microsoft Office, a clear and

specific statement is the best possible way to communicate with your students. In addition, you may wish to ask students to submit a statement of academic integrity along with their assignments.

For example,

In concert with Rutgers' code of conduct, which mandates "that all work submitted in a course, academic research, or other activity is the student's own and created without the aid of impermissible technologies, materials, or collaborations," this course has been designed to promote your learning, critical thinking, skills, and intellectual development without reliance on unauthorized technology including chatbots and other forms of "artificial intelligence" (AI). [Although you may use search engines, spell-check, and simple grammar-check in crafting your assignments,] you will be asked to submit your written work with the following statement. *"I certify that this assignment represents my own work. I have not used any unauthorized or unacknowledged assistance or sources in completing it, including free or commercial systems or services offered on the internet or text generating systems embedded into software."* Please consult with your instructor if you have any questions about the permissible use of technology in this class.

Below is some alternative or additional language for syllabi which was developed at the University of Toronto.

- The use of generative AI tools or apps for assignments in this course, including tools like ChatGPT and other AI writing or coding assistants, is prohibited.
- The use of generative artificial intelligence tools and apps is strictly prohibited in all course assignments unless explicitly stated otherwise by the instructor in this course. This includes ChatGPT and other AI writing and coding assistants. Use of generative AI in this course may be considered use of an unauthorized aid, which is a form of cheating.

II. For instructors who wish to permit use of AI tools in particular circumstances

When specifying on one's syllabus that the use of chatbots and other AI tools is *permissible in certain circumstances*, instructors should be as clear as possible and may wish to refer to the Rutgers code of conduct, cited above, in doing so. Bear in mind that students may be using these tools for different purposes in different

classes so that it is important to be specific in describing the particular usages you allow or encourage. Given that AI tools are now incorporated seamlessly into platforms such as Google Docs, grammar-checking tools such as Grammarly, and software suites such as Microsoft Office, a clear and specific statement that lays out permissible usages is the best possible way to communicate with your students.

For example, an instructor who does not want AI tools to be used in conjunction with written work but who wants to encourage students to do probing research on model content might consider the following statement:

In concert with Rutgers' code of conduct, which mandates "that all work submitted in a course, academic research, or other activity is the student's own and created without the aid of impermissible technologies, materials, or collaborations," this course has been designed to promote your learning, critical thinking, skills, and intellectual development without reliance on unauthorized technology including chatbots and other forms of "artificial intelligence" (AI). [Although you may use search engines, spell-check, and simple grammar-check in crafting your assignments,] you will be asked to submit your written work with the following statement. *"I certify that this assignment represents my own work. I have not used any unauthorized or unacknowledged assistance or sources in completing, it including free or commercial systems or services offered on the internet or text generating systems embedded into software."*

A partial exception to this policy is an authorized [exploration of model bias which we will conduct in Week X in order to build your learning on critical AI literacies.]

Please consult with your instructor if you have any questions about the permissible use of technology in this class.

(As above, our recommendation is that any instructor assigning work that involves mandatory use of an AI tool consider developing an option for students who have data privacy or other concerns.)

III. For instructors who wish to permit use of AI tools

When specifying on one's syllabus that the use of chatbots and other AI tools *is permissible (or assigned)*, instructors should be as clear as possible about how this decision comports with the learning goals for their course and may wish to refer to the Rutgers code of conduct, cited above, in doing so. Instructors may also want to emphasize critical AI literacies including the importance of recognizing that current AI tools are subject to bias, misinformation, environmental harms *et al*. (as discussed

above). Given the widespread availability of a variety of tools, be sure to be clear and specific about which tools are permitted and, if applicable. what forms of citation are required to document such use.

For example,

In concert with Rutgers' code of conduct, which mandates "that all work submitted in a course, academic research, or other activity is the student's own and created without the aid of impermissible technologies, materials, or collaborations," this course has been designed to help you develop knowledge about the use and abuse of AI tools. AI tools may be used as an aid in the creative process, but with the understanding that this should be accompanied by independent evaluation, critical thinking, and reflection. Students who choose to use these tools are responsible for any errors or omissions resulting from their use. They will also be required to provide as an appendix the prompts used, the generated output, and a thoughtful reflection on the outcomes. When appropriate, students may also be asked to consider the environmental and social costs of using the tool.

(As above, our recommendation is that any instructor assigning work that involves mandatory use of an AI tool consider developing an option for students who have data privacy or other concerns.)

Some instructors who permit use of AI tools for written assignments implement syllabus statements like these, developed at the University of Toronto.

- Students must submit, as an appendix with their assignments, any content produced by an artificial intelligence tool, and the prompt used to generate the content.
- Students may choose to use generative AI tools as they work through the assignments in this course; this use must be documented in an appendix for each assignment. The documentation should include what tool(s) were used, how they were used, and how the results from the AI were incorporated into the submitted work.

6. Select Resources You Might Wish to Read or to Share with Your Students

This document already includes many resources that you might enjoy or share with students and colleagues. Here we provide some additional resources. As this is a

living document, we plan to continue to update it with additional resources as they become available. Please feel free to suggest them to us. **[NOTE TO READERS: AS OF OCTOBER 2024, WE ARE UPDATING THE BELOW LIST TO INCLUDE BRIEF SUMMARIES AND TO GROUP ACCORDING TO TOPIC - COMING SOON!]**

404 Media, *The 404 Media Year in Review* (podcast), 31 December 2024.

AI Now Institute, *AI Now Salon Series* (videos).

Abebe, Rediet and Maximilian Kasy, "The Means of Prediction." *The Boston Review*, 20 May 2021.

Abraham, Roshan. "'A Black Hole of Energy Use': Meta's Massive AI Data Center is Stressing Out a Louisiana Community." *404 Media*. 23 June 2025.

Ahmad, Meher, Jessica Grose, and Tressie McMillan Cottom. "What A.I. Really Means for Learning." *New York Times*, 12 August 2025.

Ahmed, Kalim. "The Internet is a Place Where No One Has an Accent." *AI Policy Perspectives*, 15 July 2025.

Allison, Leslie and Tiffany DeRewal, "Where Knowledge Begins? Generative AI as Search and the Problem of Friction." forthcoming in *Critical AI* 2.2 (October 2024).

Al-Sibai, Noor. "OpenAI Admits That Its New Model Still Hallucinates More Than a Third of the Time." *Futurism,* 1 March , 2024.

Anaya. "AI Image Generators Often Give Racist and Sexist Results: Can They Be Fixed?" *Nature*, 19 March 2024.

Ars Staff. "As Internet Enshittification Marches On, Here Are Some Of The Worst Offenders." *Ars Technica*. February 5, 2025.

Aslan, Murat. "The Hidden Costs of AI Coding Assistants: Insights From a Senior Developer." *Medium*. 19 December 2024.

Atleson, Michael. "Succor Borne Every Minute." *Federal Trade Commission (weblog),* 11 June 2024.

"Attorney General Alan Wilson Leads 44 AGs Demanding Big Tech End Predatory AI Targeting of Kids." 2025. South Carolina Attorney General's Office. August 25, 2025.

Baak, Stefan. "The Human Decisions that Shape Generative AI. Who is Accountable for What?" *Mozilla Foundation Newsletter*, 2 August, 2023.

Bastian, Matthias. "Apple AI Researchers Question OpenAI's Claims About o1's Reasoning Capabilities." *The Decoder*. 12 October 2024.

BBC. "Data Centre Power Use 'To Surge Six-Fold in 10 Years.'" 26 March 2024.

Bender, Emily M. "ChatGPT Has No Place in the Classroom." *Mystery AI Hype Theater 3000 Newsletter*, 2024. Professor Lauren Goodlad gives her annotations to this newsletter in the link above.

Bender, Emily M. "ChatGP-why: When, if ever, is synthetic text safe, appropriate, and desirable?" (video-recorded lecture for GRAILE AI), 12 August, 2023.

Bender, Emily M., Timnit Gebru, Angelina Mcmillan-Major, and Shmargaret Shmitchell. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?🦜" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Canada, March 2021*, 610-623.

Bender, Emily M. and Alex Hanna, "AI Causes Real Harm. Let's Focus on That vs. the End-of-Humanity Hype." *Scientific American,* 12 August, 2023.

—--. *Mystery AI Hype Theater 3000*. (podcasts and videos).

Bianchi, Federico, Pratyusha Kalluri, Esin Durmus, Faisal Ladhak, Myra Cheng, Debora Nozza, Tatsunori Hashimoto et al. "Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale." *ArXiv* preprint 2022.

Birhane, Abeba, Atoosa Kasirzadeh, David Leslie, and Sandra Wachter. 2023. "Science in the Age of Large Language Models." *Nature Reviews Physics* 5, 277–80. April 26, 2023.

Birhane, Abeba and Deborah Raji, "ChatGPT, Galactica and the Progress Trap." *Wired,* 9 December 2022.

Bjork, Collin. "Clones in the Classroom: Why Universities Must Be Wary of Embracing AI-Driven Teaching Tools." *The Conversation*. 18 September 2024.

Blackwell, Alan. "Ooops! We Automated Bullshit (ChatGPT is a Bullshit Generator. To Understand AI, We Should Think Harder about Bullshit)" 11 September, 2023.

Blunt, Katherine. "Who Pays? AI Boom Sparks Fight Over Soaring Power Costs." *Wall Street Journal*, 29 July 2025.

Bode, Katherine. "Why You Can't Model Away Bias." *MLQ* 81. no 1 (2020) 95-124.

Booth, Robert. "Paul McCartney Warns AI 'Could Take Over' As UK Debates Copyright Laws." *The Guardian*. 10 December 2024.

Broussard, Meredith. *Artificial Unintelligence: How Computers Misunderstand the World.* Cambridge: MIT Press, 2019.

Burke, Garance, and Hilke Schellmann. "Researchers Say an AI-Powered Transcription Tool Used in Hospitals Invents Things No One Ever Said." *AP News*. 26 October 2024.

Challapally, Aditya, Chris Pease, Ramesh Raskar, and Pradyumna Chari. "The GenAI Divide: State of AI In Business 2025." *MIT Project Nanda*. July 2025.

Conrad, Kathryn. "A Blueprint for An AI Bill of Rights for Education." *Critical AI*

(blog), n.d. [Sneak Preview for an upcoming April 2024 issue of *Critical AI*]

Costanza-Chock, Sasha. *Design Justice: Community-Led Practices to Build the Worlds We Need*. Cambridge: MIT Press, 2020.

Cotton, Tressie McMillan. "The Tech Fantasy That Powers A.I. Is Running on Fumes." *New York Times*. 29 March 2025.

Crabapple, Molly. "Molly Crabapple – Debates in AI – RISD." Lecture presented at Rhode Island School of Design's Debates in AI Symposium, Providence, RI, April 2024.

Data Workers Inquirihttps://docs.google.com/document/d/14Fp1DbmjZx9lzz8kvUo57yDZ9I6l5v4quLX6tbIpPvQ/edit?tab=t.0es: Invited Pieces.

Critical AI (@CriticalAI). "Yes @OpenAI new text-to-video is impressive but here's 5 questions that journos & the public should be asking–relentlessly. 1. When will you release the". X, February 16, 2024, 9:22am. https://twitter.com/CriticalAI/status/1758497093172207811

da Silva, João. "Google Turns to Nuclear to Power AI Data Centres." *BBC*. 15 October 2024.

DeRewal, Tiffany. "Ask an Expert: Evaluating LLM 'Research Assistants' and Their Risks For Novice Researchers." *Critical AI* (blog), n.d.

Doctorow, Cory. "The Enshittification of TikTok." *Wired.* 23 January 2023.
—--. "'Enshittification' is Coming for Absolutely Everything." *Financial Times* 8 February 2024.

Drhal, Carmen. "AI Was Asked To Create Images of Black African Docs Treating White Kids. How'd It Go?." *NPR.* 6. October, 2023.

Economic Security Project. "New Shadow Report Addresses Gaps in the Senate Roadmap for AI and Calls for a Policy Agenda That Harnesses AI in the Public Interest." 5 May 2024. Economic Security Project Report. Accessed June 25, 2024.

Ellis, Lindsay, and Katherine Bindley. "AI Is Wrecking an Already Fragile Job Market for College Graduates." *Wall Street Journal*, 28 July 2025.

Estrada, Daniel. "Teaching Insights: How to Teach AI to Students: AI Ethics and NJIT Audit Project." *Critical AI Blog.* (n.d.)

Estrin, Judy. "The Case Against AI Everything, Everywhere, All at Once." *Time,* 11 August 2023.

Farrell, Maria and Robin Berjon. "We Need to Rewild the Internet." *Noema*. Berggruen Institute, 16 April 2024.

Flaherty, Colleen. "How AI Is Changing—Not 'Killing'—College." *Inside Higher Ed*, 29 August 2025.

Flynn, Shannon. 2020. "The Difference between Symbolic AI and Connectionist AI." RE•WORK Blog - AI & Deep Learning News. September 24, 2020.

Gilliard, Chris. "How Ed Tech Is Exploiting Students." *Chronicle of Higher Education,* 8 April 2018.

Goldman Sachs. "AI is poised to drive 160% Increase in Data Center Power Demand." May 14, 2024.

Goodlad, Lauren M. E. "Editor's Introduction: Humanities in the Loop." *Critical AI* (2023) 1-2.

Goodlad, Lauren M.E. and Sam Baker, "Now the Humanities Can Disrupt AI." *Public Books,* 20 February, 2023.

Goodlad, Lauren M. E., and Kathryn Conrad. March 2024. "Teaching Critical AI Literacies." NORRAG Policy Insights #4 AI and Digital Inequities,

Goodlad, Lauren M. E., and Matthew Stone. "Beyond Chatbot-K: On Large Language Models, 'Generative AI,' and the Rise of Chatbots; An Introduction" *Critical AI* (2024) 2.1

Gross, Grant. "Devs Gaining Little (If Anything) From AI Coding Assistants." *CIO*. 26 September 2024.

Guo, Eileen. "A Major AI Training Data Set Contains Millions of Examples of Personal Data." *MIT Technology Review*, 22 July 2025.

Guo, Eileen. "An AI Chatbot Told a User How To Kill Himself—But the Company Doesn't Want To 'Censor' It." *MIT Technology Review*. February 6, 2025.

Hadgu, Asmelash Teka, and Timnit Gebru. "Replacing Federal Workers with Chatbots Would Be a Dystopian Nightmare." *Scientific American*. Springer Nature. April 14, 2025.

Hao, Karen. "Microsoft's Hypocrisy on AI." *The Atlantic*. 13 September 2024.

Hao, Karen (@_KarenHao). 2024. "For years I've been interviewing data annotation workers who are the lifeblood of the AI industry [...]." X thread, March 16, starting at 10:24 AM. https://twitter.com/_KarenHao/status/1769006784273101074

Hao, Karen. "AI is Taking Water From the Desert." *The Atlantic*. 1 March 2024.

Heaven, Will Douglas. "The Algorithm." *MIT Technology Review Newsletter*, 25 August 2025.

Herrman, John. "ChatGPT and Google Gemini Are Both Doomed." *Intelligencer*, 1 March, 2024.

Hsu, Tiffany and Stuart A. Thompson, "Disinformation Researchers Raise Alarms About A.I Chatbots." *New York Times.* 8 February 2023. Updated 20 June 2023.

Huckins, Grace. "Why the AI Moratorium's Defeat May Signal a New Political Era." *MIT Technology Review*, 9 July 2025.

Jaźwińska, Klaudia and Aisvarya Chandrasekar. "AI Search Has A Citation Problem." *Columbia Journalism Review*. Tow Center for Digital Journalism. 6 March 2025.

Kaltheuner, Frederike, Leevi Saari, Amba Kak, and Sarah Myers West. "I. Reorienting European AI and Innovation Policy." *AI Now Institute*. 15 October 2024.

Kak, Amba, Susan Meyers West and Meredith Whittaker. "Make No Mistake–AI Is Owned by Big Tech." *MIT Technology Review.* 5 December 2023.

Kirmer, Stephanie. "The Cultural Backlash Against Generative AI." *Medium*. 1 February 2025.

Krietzberg, Ian. "OpenAI Accuses New York Times of Paying Someone to Hack ChatGPT." *TheStreet*. 27 February 2024.

Kumar, Harsh, Jonathan Vincentius, Ewan Jordan, and Ashton Anderson. "Human Creativity in the Age of LLMs: Randomized Experiments on Divergent and Convergent Thinking." *arXiv*, 2410.03703, September 24, 2024.

Lopatto, Elizabeth. "Stop Using Generative AI As a Search Engine." *The Verge*, 5 December 2024.

Luccioni, (Alexandra) Sasha, Bruna Tevelin, and Margaret Mitchell. "The Environmental Impacts of AI – Primer." *Hugging Face*, 3 September 2024.

LuccioniAlexandra Sasha, Yacine Jernite, and Emma Strubell. "Power Hungry Processing: Watts Driving the The Cost of AI Deployment?" *ArXiv.* May 2024.

Maiberg, Emanuel, and Jason Koehler. "Inside the Booming 'AI Pimping' Industry." *404 Media*. 20 November 2024.

Matthewson, Tara García. "AI Chatbots Can Cushion the High School Counselor Shortage — But Are They Bad For Students." *The Markup*. 4 March 2025.

Mauran, Cecily. "The Era of the AI-Generated Internet is Already Here." *Mashable*. 27 January, 2024.

McKendrick, Joe. "Generative AI May Be Creating More Work than It Saves." *ZDNet.* 24 May, 2024.

Merchant, Brian. "The AI Industry Has a Battle-Tested Plan to Keep on Using Our Content without Paying for It." *The LA Times.* 12 January 2024.

Metz, Cade, Cecilia Kang, Sheera Frenkel, Stuart A. Thompson, and Nico Grant. "How Tech Giants Cut Corners to Harvest Data for A.I." *New York Times*. 6 April 2024.

Mills, Anna, Maha Bali, and Lance Eaton. "How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process." 6. 1. (2023).

Monserrate, Steven Gonzalez. "The Staggering Ecological Impacts of Computation and the Cloud." *MIT Press Reader.* 14 February 2022.

Narayanan, Arvind and Sayash Kapoor, *AI Snake Oil* (substack)

Nicoletti, Leonardo and Dina Bass. "Humans Are Biased. Generative AI Is Even Worse." *Bloomberg*

Technology. 12 June 2023.

Niemeyer, Kenneth. "Billionaire Larry Ellison Says a Vast AI-Fueled Surveillance System Can Ensure 'Citizens Will Be On Their Best Behavior.'" *Business Insider*. 15 September 2024.

O'Brien, Isabel. "Data Center Emissions Probably 662% Higher Than Big Tech Claims. Can It Keep Up the Ruse?" *The Guardian*. 15 September 2024.

O'Donnell, James. "AI's Giants Want to Take Over the Classroom." *MIT Technology Review*, 15 July 2025.

O'Neil, Lorina, "These Women Tried to Warn Us about AI." *Rolling Stone,* 12 August 2023.

Pasquale, Frank. *New Laws of Robotics: Defending Human Expertise in the Age of AI*. Cambridge: Harvard University Press, 2020.

Peters, Uwe, and Benjamin Chin-Yee. "Generalization Bias in Large Language Model Summarization of Scientific Research." *Royal Society Open Science*, vol. 12, no. 4, 2025.

Petri, Alexandra. "I Hate the Gemini 'Dear Sydney' Ad More Every Passing Moment." *The Washington Post.* 31 July 2024.

Pierce, David. "The text file that runs the internet" [on how training AI has broken the "social contract" of the Internet.] *The Verge.* 14 February 2024.

Ramkumar, Amrith and Brian Schwartz. "Silicon Valley Launches Pro-AI PACs to Defend Industry in Midterm Elections." *Wall Street Journal*, 25 August 2025.

Reiley, Laura. "What My Daughter Told ChatGPT Before She Took Her Life." *New York Times*, 24 August 2025.

Rosalsky, Greg. "10 Reasons Why AI May be Overrated." *NPR*, 6 August 2024.

Rosenzweig, Jane. "Teaching Insights: What Happens When a Novice Writer Asks ChatGPT for Editing Advice?" *Critical AI Blog* (n.d.)

Saul, Josh, Naureen S. Malik, and Mark Chediak. "AI Boom Is Driving a Surprise Resurgence of US Gas-Fired Power." *Bloomberg*. 16 September 2024.

Schmidt, Eric, and Selina Xu. "Silicon Valley is Drifting Out of Touch with the Rest of America." *New York Times*, 19 August 2025.

Shumailov, Ilia, Zakhar Sumaylov, Yiren Zhao, Yarin Gal, Nicholas Papernot, Ross Anderson. "The Curse of Recursion: Training on Generated Data Makes Models Forget." *ArXiv* (self-published preprint uploaded on 31 May 2023)

Smith, Gary and Jeffrey Funk. "When It Comes to Critical Thinking, AI Flunks the Test." *The Chronicle of Higher Education*. 12 March 2024.

Stone, Matthew, "Large Language Models: a Whirlwind Tutorial." (video - skip to 15:31) *Critical AI* (blog), n.d. [Recorded as part of a 2021 virtual workshop for "The Ethics of Data Curation," an NEH-supported research collaboration between Rutgers and Australian National University] (See also Stone's talk on the first panel of the videos for

the Critical AI Literacy in a Time of Chatbots symposium.

Tacheva, Jasmina, and Srividya Ramasubramanian. "AI Empire: Unraveling the interlocking systems of oppression in generative AI's global order." *Big Data & Society* 10.2 (2023): 20539517231219241.

Tan, Rebecca and Regine Cabato. "Behind the AI boom, an army of overseas workers in 'digital Sweatshops.'" *The Washington Post*, 28 August 2023

Tiku, Natasha, Kevin Schaul, and Szu Yu Chen. "These Fake Images Reveal How AI Amplifies Our Worst Stereotypes." *The Washington Post.* 1 November 2023.

Tech Won't Save Us (@techwontsaveus). 2024. "DATA VAMPIRES is a four-part series exploring the costs of hyperscale data centers and why [...]." X thread, October 7, starting at 10:17 AM.

Urquieta, Claudia and Daniela Dib. "U.S Tech Giants are Building Dozens of Data Centers in Chile. Locals Are Fighting Back." *Rest of World.* May 31, 2024.

Vassel, Faye-Marie, Evan Shieh, Cassidy R. Sugimoto, and Thema Monroe-White. 2024. "The Psychosocial Impacts of Generative AI Harms." *Proceedings of the AAAI Symposium Series* 3 (1): 440–47.

Vincent, James. "How Much Electricity Does AI Consume?" *The Verge*, February 16, 2024.

Warner, John. "Everyone Should Read 'Teaching Machines.'" *Inside Higher Ed*. 28 September 2024.

Watters, Audrey. 2025. "AI Slop Education." *Second Breakfast*

Watters, Audrey. *Teaching Machines: The History of Personalized Learning*. Cambridge: MIT Press, 2021.

Wiggers, Kyle. "It Sure Looks Like OpenAI Trained Sora on Game Content – And Legal Experts Say That Could Be A Problem." *TechCrunch*. 11 December 2024.

Wilkins, Joe. "AI Therapist Goes Haywire, Urges User to Go on Killing Spree." *Futurism*, 25 July 2025.

Williams, Rua M. (@FractalEcho). 2024. "The racism behind chatGPT that we aren't talking about… This year, I learned that students use [...]." X thread, February 17, starting at 4:37 PM. https://twitter.com/FractalEcho/status/1758968904674836979

Wong, Matteo. "The Schools Without ChatGPT Plagiarism." *The Atlantic*. 25 October 2024.

Wong, Matteo. "The AI Boom Has an Expiration Date." *The Atlantic*. 17 October 2024.

Wong, Matteo, "The Generative-AI Revolution May Be a Bubble." *Atlantic Monthly,* 2 August 2024.

Wong, Matteo, "America Already Has An AI Underclass." *Atlantic Monthly,* 26 July 2023.

Zakarin, Jordan, and Josh Hirschfeld-Kroen. 2023. "This Is the Biggest Threat to Actors & Writers." YouTube Video. *More Perfect Union*.