

5.4c Generative Artificial Intelligence

10x10 – Introduction to Biomedical Informatics & Artificial Intelligence
William Hersh, MD, FACMI, FAMIA, FIAHSI
Copyright 2026, All Rights Reserved



Downsides for generative AI

Hallucinations, confabulations, fabrications, etc.

Reduces depth of learning and knowledge

Perpetuating bias

Detecting use

Deep (and not so deep) fakes

Concerns for scientific writing and research

Intellectual property issues

Adversarial attacks on LLMs

Privacy concerns

Other challenges

Hallucinations, confabulations, fabrications, etc.

Hallucinations and confabulations continue

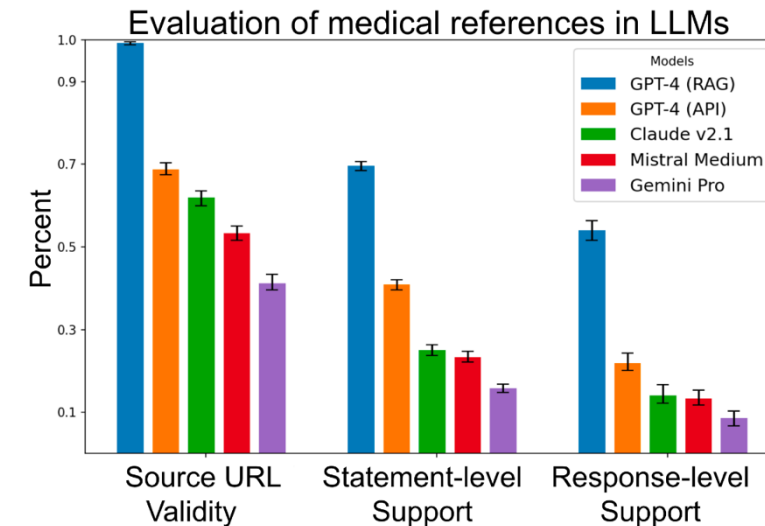
- Improved over time (Jones, 2025)
- Still seen with Google search summaries (Gregory, 2026)

Early LLMs also noted for fabrication of references (Walters, 2023)

- In biomedicine and health, searchers may have concerns for authoritativeness, timeliness, and contextualization of search (Herhsh, 2024)

Most LLMs poor at reference attribution (Wu, 2024)

- Best LLM (GPT-4 in CoPilot) had highest URL source validity, 70% statement-level support, and 54% response-level support
- Even CoPilot failed to cite any sources for around 20% of prompts; others more



Reduces depth of learning and knowledge

While may improve work efficiency, can inhibit critical engagement with work and potentially lead to long-term overreliance on tool and diminished skill for independent problem-solving (Lee, 2025)

- When using generative AI, effort invested in critical thinking shifts from
 - Information gathering to information verification
 - Problem-solving to AI response integration
 - Task execution to task stewardship

Use of LLMs reduced cognitive connections in brain during essay-writing, producing more homogeneous output and lessening ability to recall what was written over time (Kosmyna, 2025)

Use of LLMs resulted in "shallower knowledge" than learned via Web search (Melumad, 2025)

Perpetuating bias

8 clinical questions asked of 4 LLMs recapitulated “harmful, race-based medicine” (Omiye, 2023)

In standardized clinical vignettes from NEJM Healer, GPT-4 (Zack, 2024)

- More likely to include diagnoses that stereotyped certain races, ethnicities, and genders
- Did not model appropriate demographic diversity of medical conditions

Anti-LGBTQIA+ bias (Chang, 2025)

LLMs gave variable advice dependent upon race, socioeconomic status, and LGBT status not supported by clinical reasoning (Omar, 2025; Ji, 2025)

Detecting use of generative AI

Why detect use of generative AI?

- Deepfakes
- Integrity of writing, especially but not limited to scientific writing
- Educational assessment

Neither algorithms (Tang, 2024) nor humans (Cooke, 2025) can reliably detect

- More likely to classify non-native English writing as AI-generated (Liang, 2023)

Clinicians unable to distinguish human-generated vs. ChatGPT-generated medical manuscripts (Helgeson, 2025)

- Errors made in directions of false negatives and false positives
- Those with AI familiarity had somewhat higher accuracy than those without

Accuracy of commercial AI detectors variable and constantly evolving (Gerwitz, 2025)

Deep (and not so deep) fakes

Deepfakes – audio, image, or video digital forgery created by AI, typically generative AI (Naffi, 2025)

- Growing concern, especially in realms of politics and social media (Burgess, 2026)

Concerns for integrity of scientific research (Chen, 2024)

- Create fake datasets in ophthalmology (Taloni, 2023) that evade forensic detection (Taloni, 2025)
- Fabrication of machine learning textbook (Aksenfeld, 2025)
- False authorship (Spinellis, 2025)
- “Explosion” of formulaic research articles, including inappropriate study designs and false discoveries, based on NHANES (Suchak, 2025)
- Undermine online survey research (Westwood, 2025)

Concerns about scientific publishing

Specific AI policies vary by journals and publishers (Bhavsar, 2025; Yin, 2025)

Evidence for undisclosed use

- “Delving” into word frequency in Pubmed abstracts suggests usage of LLMs in at least 13.4% (Kobak, 2025)
- Word frequency changes suggest usage of LLMs in different fields from computer science (22%) to math (9%) (Liang, 2025)
- Assessment of usage in cancer journal submissions estimated growing use in abstracts, methods, and reviewer comments, varying by journal and location of authors (Evanko, 2025)
- Disclosure of AI use to BMJ Group journals much lower than stated in surveys of researchers using AI (AlFayyad, 2025)

Part of growing problem of scientific publishing generally (Kabel, 2025)

- Tainting literature, especially preprint servers (Watson, 2025)
- Leading to calls for “resisting AI slop” (Thorp, 2026)

Larger concerns for biomedical research

Ease of generating papers may increase quantity and reduce quality, especially with use of public datasets (Spick, 2025), leading to increases in redundant publications (Maupin, 2025)

Lack of protocols to detect errors may undermine science (Narayanan, 2025)

Increases imperative to share research data (Brown, 2025)

“Vibe science” threatens scientific rigor (Rose, 2025)

Science writing uses words different from general usage; will AI impact scientific writing (Steiner, 2026)?

Use of AI tools may increase research productivity and advancement but may also reduce scope of science (Hao, 2026)

Intellectual property issues

Most LLMs trained by crawling content on Web – raises concerns about protection of intellectual property of original authors of sites (Schaul, 2023) and books/articles (Reisner, 2025)

Have you trained LLMs? I have

- My Web content at dmice.ohsu.edu part of [Colossal Clean Crawled Corpus](#) used by OpenAI and others (Dodge, 2021)
- 83 of my books and papers part of (allegedly pirated) [LibGen](#) used by Meta Llama

Ongoing lawsuits by NY Times and other publishers against OpenAI and Microsoft for copyright infringement over “unauthorized” use of content for training GPT (Allyn, 2025)

- Use of copyrighted content served as “snippets” by search engine allowed for Google years ago (Stempel, 2013)

Adversarial attacks on LLMs

Attacks on LLMs – three categories, with biomedical and health examples (Chowdhury, 2024)

- Jailbreaking – exploiting model vulnerabilities
 - LLMs susceptible to adversarial attacks to generate objectionable behaviors by targeted prompts (Zou, 2023)
 - Using persuasion in prompts to jailbreak LLMs (Zeng, 2024; Meincke, 2025)
 - Reproducing whole books used for training (Ahmed, 2026)
- Prompt injection – manipulate behavior with malicious prompts
 - LLM agents autonomously hacking Web sites and their data stores (Fang, 2024)
 - Poisoning attacks on text-to-image generative models (Shan, 2024)
 - Inserting fabricated evidence for supplement recommendations, opioid prescriptions, pregnancy contraindications, and central-nervous-system toxic effects could induce harmful recommendations (Lee, 2025)
 - Fake lab values, signs, or diseases can cause LLM models to repeat or elaborate on planted error in up to 83% of cases (Omar, 2025)
- Data poisoning – adding data to models
 - Inserting small amounts of misinformation into models could result in misinformation (Alber, 2025)

Privacy issues

LLMs can infer personal data (Staab, 2023)

NY Times writer contacted by researcher (White, 2023) who overcame fine-tuning to obtain email addresses of employees (Chen, 2023)

How do we balance beneficial uses with security and privacy risks? (Yao, 2024)

Other downsides for LLMs

LLMs require more than accuracy (Goodman, 2024)

- Uses and outputs probabilistic
- May have “sycophancy bias”
- May be generally accurate but generate small critical errors

Examples of correct answers but flawed reasoning in solving imaging-related cases (Jin, 2024)

May be running out of training data (Jones, 2024), potentially leading to “model collapse” (Shumailov, 2024)

“Open systems” not as open as we might like (Widder, 2024)

Growing burden for clinicians needing to review LLM content (Ohde, 2025)

LLMs lack metacognition in clinical situations (Griot, 2025)

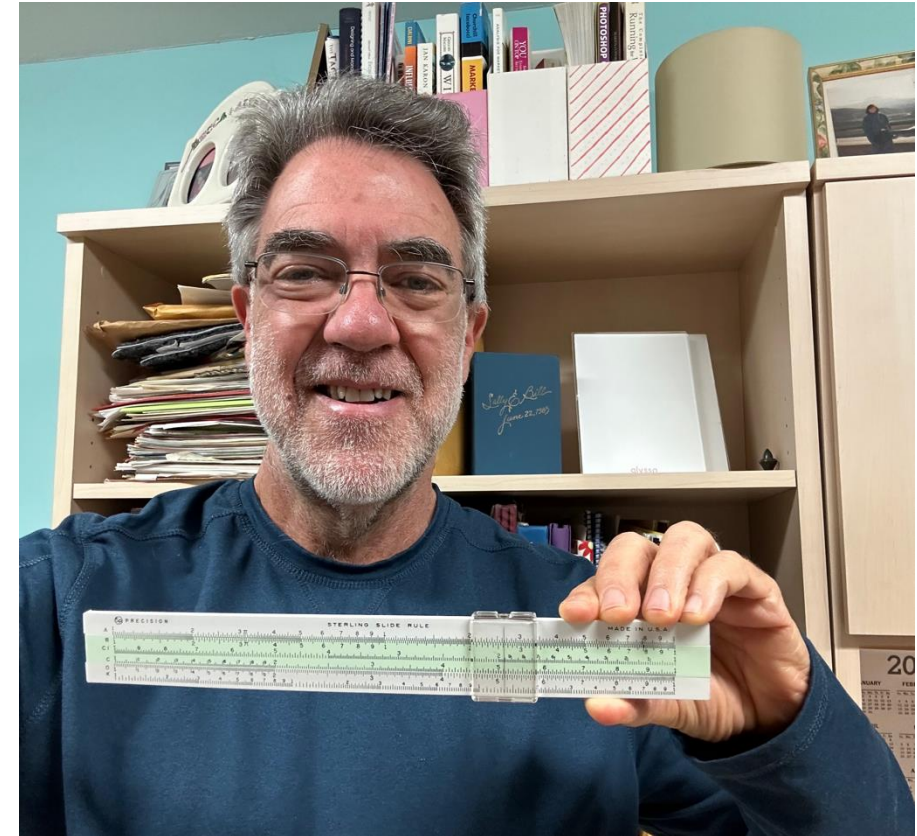
Impact on education (Hersh, 2025)

The “homework apocalypse” (Mollick, 2023) and solutions going forward (Mollick, 2024)

“ChatGPT has transformed the problem of grade inflation from a minor corruption to an enterprise-destroying blight.” (Clune, 2023)

“I used to teach students. Now I catch ChatGPT cheats.” (Jollimore, 2025)

“Generative artificial intelligence does not have to undermine education.” (Tan, 2024)



Impact on educational assessment

In biomedical and health fields

- Biomedical informatics knowledge-based course (Hersh, 2024)
- Biomedical graduate studies (Stribling, 2024)
- Health informatics programming course (Avramovic, 2024)

Other disciplines beyond biomedicine

Best practices

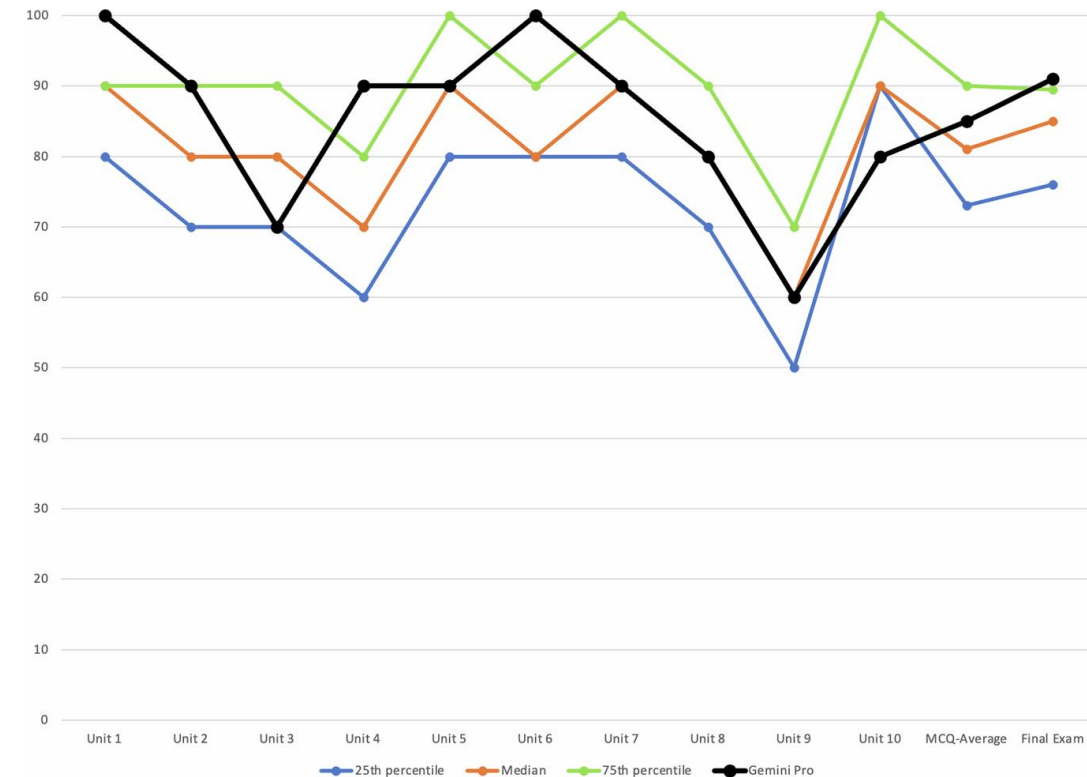
Use in introductory informatics course (Hersh, 2024)

Analysis of 2023 course for 139 students
– 30 graduate students, 85 continuing education students, and 24 medical students

For knowledge assessment, LLMs scored better than 50-75% of students and obtained passing grades

Writing assignments not assessed but most LLMs can do that as well.

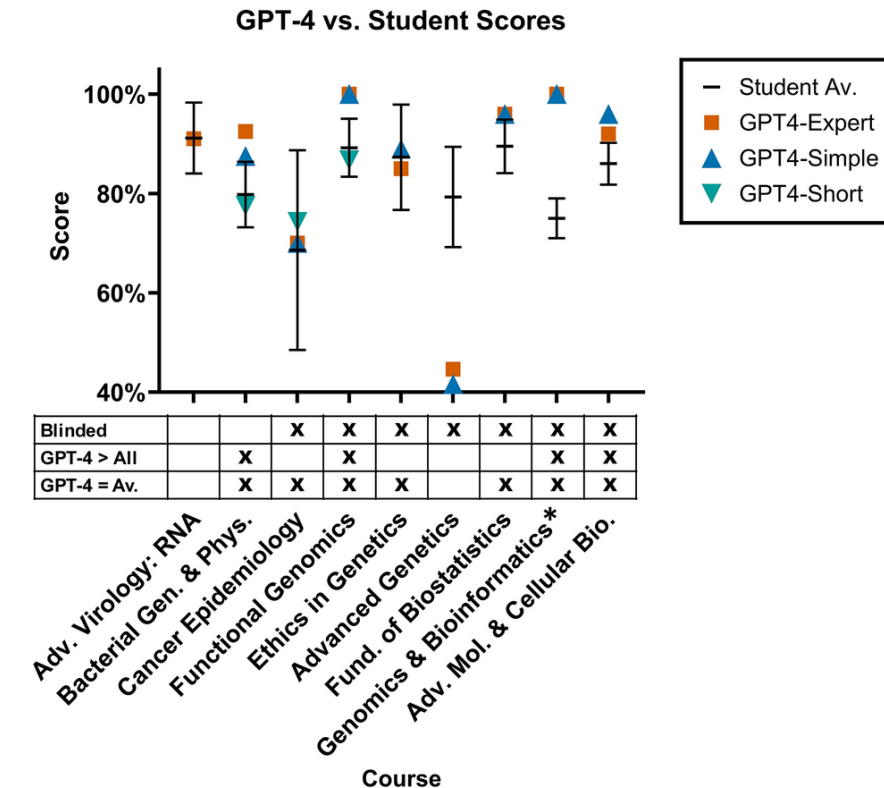
My [policy for use of generative AI in this course](#)



Other biomedical and health courses

Graduate-level examinations in biomedical sciences (Stribling, 2024)

- GPT-4 performance on 9 exams
- Exceeded student average on 7 of 9 exams and all student scores for 4 exams
- Performed very well on
 - Fill-in-the-blank, short-answer, and essay questions
 - Questions on figures sourced from published manuscripts
- Performed less well on questions with
 - With figures containing simulated data
 - Requiring hand-drawn answer
- Some model responses included detailed hallucinations



Use of GitHub CoPilot in health informatics programming course (Avramovic, 2024)

- Assessed in problems for
 - Database queries with SQL
 - Computational tasks with Python
- Generated solutions worked well for simple tasks but less well for complex ones
 - Some solutions correct but not most efficient approach

Concerns for medical education (Hough, 2025)

Automation bias – tendency to over-trust automated recommendations even when wrong

Cognitive off-loading and outsourcing reasoning – shifting information retrieval, appraisal, and synthesis to AI so learners generate fewer internal explanations and weaker memory traces

De-skilling – loss of procedural or cognitive skill from premature automation of routine tasks; experts may compensate, novices cannot

Racial and equity biases – systematic differences in model outputs by protected characteristics or propagation of race-based biases

Hallucinations – fluent but false statements or references produced with high confidence

Privacy, security and data governance – risks include exposing personal health information (PHI) or student records, prompt injection, data exfiltration, and uncontrolled retention of sensitive content

Education beyond biomedicine

Passing college entrance and AP exams (Dubey, 2024)

Writing computer programs (Denny, 2024; Poldrack 2024; Johnson, 2024)

Writing legal briefs (Choi, 2024)

Creating data science pipelines (Cheng, 2024; Hong, 2024)

OpenAI o1 models outscored PhD students on “Google-proof” questions in biology, chemistry, and physics (Rein, 2023; OpenAI, 2024; Jones, 2024)

In 5 undergraduate psychology courses, GPT-4 scored higher than average among students on take-home exams with only 6% detection (Scarfe, 2024)

Access to ChatGPT improved math problem-solving but reduced when taken away; better learning from tutor mode (Bastani, 2025)

Uses and risks of “assigning AI” (Mollick, 2023)

AI USE	ROLE	PEDAGOGICAL BENEFIT	PEDAGOGICAL RISK
MENTOR	Providing feedback	Frequent feedback improves learning outcomes, even if all advice is not taken.	Not critically examining feedback, which may contain errors.
TUTOR	Direct instruction	Personalized direct instruction is very effective.	Uneven knowledge base of AI. Serious confabulation risks.
COACH	Prompt metacognition	Opportunities for reflection and regulation, which improve learning outcomes.	Tone or style of coaching may not match student. Risks of incorrect advice.
TEAMMATE	Increase team performance	Provide alternate viewpoints, help learning teams function better.	Confabulation and errors. "Personality" conflicts with other team members.
STUDENT	Receive explanations	Teaching others is a powerful learning technique.	Confabulation and argumentation may derail the benefits of teaching.
SIMULATOR	Deliberate practice	Practicing and applying knowledge aids transfer.	Inappropriate fidelity.
TOOL	Accomplish tasks	Helps students accomplish more within the same time frame.	Outsourcing thinking, rather than work.

Risks:

- Confabulation
- Bias – from training content
- Privacy – policies not always clear
- Instructional – student over-reliance

Other educational concerns

How are students using AI (O'Connell, 2025)?

- Usage is widespread
- Faculty concern about “illusion of competence,” relating to “cognitive offloading”
- Students ambivalent but get mixed messages

Concerns for writing self-reflective essays- AI may “replicate the shape of a narrative but not the struggle that makes it meaningful” (Whittet, 2025)

Cuts both ways (Tribin, 2025)

