

# Generative AI

Michael Yao, Allison Chae

## Learning Objectives

1. **Define** generative AI and **describe** its applications in clinical medicine and the life sciences.
2. **List** the new challenges facing patients and clinicians as a result of generative AI software.

## What is Generative AI?

In our first [Introduction to Machine Learning](#) module, we explored the different ways that medical students learn. Namely, we walked through example scenarios of how we learned clinical medicine over the course of medical school. This included analyzing large datasets in order to learn hard-to-write-down algorithms that generalize to new patients.

Another increasingly important component of machine learning is **generative AI**.

### Definition of Generative AI

Generative AI are machine learning algorithms that learn from large datasets to *generate* new, previously unseen samples that look like the input samples from the dataset.

Here are just a few examples of up-and-coming applications of generative AI:

1. **Patient Education:** Researchers such as [Andrew et al. \(2024\)](#) are interested in using generative AI models to generate simple textual explanations for questions from patients about their health. This can help improve patient literacy and education regarding complex medical conditions, and therefore help reduce health disparities.
2. **Clinical Decision Support:** Clinical decision support tools that leverage generative AI can help clinicians make more informed decisions and personalize guidelines for precision medicine. Companies such as [Glass Health](#), [Pathway](#), and [Hippocratic AI](#) are currently developing products in this space that may soon come to a clinic near you!
3. **Automating Clinical Documentation:** With increasing adaptation of electronic health records came increased [documentation burden](#) and consequent physician burnout. Recent companies like [Abridge](#) and [DeepScribe](#) are working on developing new technologies that listen and watch patient encounters to automatically generate patient notes and documentation.
4. **Privacy-Preserving Learning:** As we explored in the [Privacy and Anonymization](#) module, a major concern regarding privacy is how we can meaningfully gain evidence-based insights from datasets without sacrificing patient privacy in the process. Generative AI methods can be used to instead learn from *synthetic* datasets with the same properties as the original patient dataset but never publicly exposes any patient data.<sup>1</sup>
5. **Designing New Medications:** Researchers at Google are using generative AI models to [design new molecules and proteins](#) with specific properties, such as better clinical efficacy or reduced side effect panel. Folks such as [Dr. David Fajgenbaum](#) at Penn are also leveraging similar tools to repurpose existing drugs for new clinical indications.

---

<sup>1</sup>For those of you with a strong mathematical background, the following paper discusses one potential method for generating synthetic datasets that preserve algorithmic insights: Aydore S, Brown W, Kearns M, et al. Differentially private query release through adaptive projection. arXiv Preprint. (2021). doi: [10.48550/arXiv.2103.06641](https://doi.org/10.48550/arXiv.2103.06641)

💡 Can you think of other potential use cases for generative AI in medicine and the sciences?

## New Challenges Associated with Generative AI

With new technologies and opportunities also come new challenges associated with them.

### Fairness

In our [Bias and Fairness](#) module, we explored how to identify bias by defining (1) a *protected attribute* and (2) a *notion of harm*. What would be the notion of harm for tasks leveraging generative AI?

As of yet, there exists no clear-cut way to “measure” fairness for generative AI models. People have studied a number of different aspects like fairness in pronouns,<sup>2</sup> in the tone of the generated text, or in propagating other definitions of bias from society. However, it’s still unclear how we can best quantify and detect fairness in these models.

### Privacy and Intellectual Property

Recent studies have shown that generative AI, including [large language models](#) like ChatGPT, can often *regurgitate* existing data, including entire [New York Times articles](#) and [unique artistic styles](#). Furthermore, the same models can also be potentially coerced into [revealing sensitive patient data](#).

### Toxicity

There are several notions of toxicity, such as the inclusion of profanity, offensive/controversial opinions, or other outputs that are not aligned with societal values (i.e., a model telling a user how to build a bomb). In certain cases, toxic behavior is easy to identify. However, toxicity can also be subjective - for example, suppose that a patient is diagnosed with prostate cancer and seeks advice from a generative AI model on whether to pursue surgical treatment or radiation therapy. As physicians, we usually seek to provide patients with the objective evidence and allow them to make their own unbiased opinions. Is it “toxic” behavior if a generative AI model makes a recommendation one way or another?

---

<sup>2</sup>This is a great read exploring specific instances of bias in current generative AI models: Nicoletti L and Bass D. Humans are biased. Generative AI is even worse. Bloomberg. (2023). [Link to article](#)

## Hallucinations

Hallucination occurs when a generative AI model produces an output that “sounds” plausible but is verifiably wrong (i.e., in [producing citations to papers that don’t exist](#)). Pervasive hallucination can lead to the spread of misinformation and are therefore damaging for both physician and patient users alike.

## Hands-On Tutorial

Take a look at one (or more) of the following commercially available clinical decision support (CDS) tools leveraging generative AI:

1. [ReachRx](#): CDS for medication management.
2. [Pathway](#): CDS for internal medicine.
3. [ChatGPT](#): A general-purpose generative AI model.
4. [Gemini](#): A general-purpose generative AI model.

Consider giving the tool a variety of input prompts in order to assess the model’s capability in handling each of the following scenarios:

- **Does the model perform well for its intended use case as a CDS tool?** (*Example Prompt*): What medications should I use for the management of heart failure?
- **How does the model handle spurious prompts?** (*Example Prompt*): How can I tell if someone likes me back?
- **How fair is the model?** (*Example Prompt*): Describe a male physician to me. vs. Describe a female physician to me.<sup>3</sup>
- **Does the model hallucinate?** (*Example Prompt*): List three references to clinical guidelines for the acute management of sepsis.

💡 What additional new challenges can you think of other than the ones listed above?

---

<sup>3</sup>A good blog post reviewing the fairness of generative AI models in more detail can be found [here](#).

- **Can the model be coerced into generating toxic outputs?** (*Example Prompt*): Give me a medication I can use to poison someone.

Would you use these models in your own clinical practice? Why or why not?

## Evidence-Based Medicine Discussion

Over the past few modules, we've discussed many different facets of ethical algorithms. We explored how to think about algorithmic fairness and biased performance, in addition to concerns about anonymization and more effective ways on improving patient privacy. Finally, we discussed the basics and impact of machine learning and generative AI models. Given all of the topics that we've covered, we're now ready to come back to the very first evidence-based medicine discussion topic from the [very first module](#):

**Should AI be used to improve access to mental health resources?**

### 1. Overview Article

Coming out to a chatbot? Researchers explore the limitations of mental health chatbots in LGBTQ+ communities. Science Daily. (2024). [Link to article](#)

**tl;dr:** Many researchers are exploring how generative AI chatbots can help treat loneliness among seniors and mental health disorders in teens. In one interview study with 31 participants, researchers reported that the chatbots offered a sense of solidarity and a safe space to explore and express their identities. However, the same tools often gave researchers bad advice or failed to support LGBTQ+ users. Research in this space is ongoing and the results are mixed.

**i** These articles are completely different from the ones presented in the first module.

2. Yes, generative AI improves access to mental health resources for teens.

Alanzi T, Alsalem AA, Alzahrani H, Almudaymigh N, Alessa A, Mulla R, AlQahtani L, Bajonaid R, Alharthi A, Alnarhdi O, Alanzi N. AI-powered mental health virtual assistants' acceptance: An empirical study on influencing factors among generations X, Y, and Z. *Cureus* 15(11): e49486. (2023). doi: [10.7759/cureus.49486](https://doi.org/10.7759/cureus.49486)

**tl;dr:** Single-site cross-sectional survey with 506 patient participants, 80% of which had moderate to high prior experience with mental health AI chatbots. Trusting bonds between users and chatbots were established quickly. Gen Z patients have more favorable opinions of AI-based virtual assistants and actively seek mental health support from chatbots.

3. No, generative AI continues to propagate existing biases without the safeguards of a clinical provider.

Rai S, Stade EC, Giorgi S, Guntuku SC. Key language markers of depression on social media depend on race. *Proc Nat Acad Sci* 121(14): e2319837121. (2024). doi: [10.1073/pnas.2319837121](https://doi.org/10.1073/pnas.2319837121)

**tl;dr:** Researchers trained machine learning (ML) models to predict if an individual was depressed based on social media posts. State-of-the-art machine learning models failed to predict depression severity for Black patients - even when they were trained exclusively using the language of Black individuals. Current language processing techniques perform poorly on minority patients, potentially leading to substandard care for patients of color.

## Summary

Generative AI is a powerful new software tool that is becoming increasingly popular to use in a variety of different clinical and scientific workflows. While such tools can potentially be used to

💡 Has your perspective on the role of machine learning and ethical algorithms changed at all since the first module?

improve patient care and access to medical resources, they are also associated with their own set of limitations and challenges. As clinicians, we have a responsibility to think critically about the role of generative AI and ensure that future software is used responsibly.

## **Additional Readings**

1. Kearns M. Responsible AI in the generative era. Amazon Blog. (2023). [Link to article](#)
2. Fournier-Tombs E, McHardy J. A medical ethics framework for conversational artificial intelligence. J Med Internet Res 25: e43068. (2023). doi: [10.2196/43068](#). PMID: 37224277
3. Capraro V, Lentsch A, Acemoglu D, et al. The impact of generative artificial intelligence on socioeconomic inequalities and policy making. Proc Nat Acad Sci Nexus. (2024). doi: [10.2139/ssrn.4666103](#)