

AI Literacy for Health Literacy Professionals

Overview

Publication Information

Format: Webinar for IHA Health Literacy Academy.

Date: July 29, 2026.

Author: Samuel R. Mendez, PhD.

About the author: Samuel R. Mendez is a Lecturer at The University of Western Australia's Graduate Research School. Their research bridges computational social science and media studies to improve health communication. Their PhD is from the Harvard TH Chan School of Public Health (2025), where they evaluated the strengths and limits of machine learning and generative AI approaches to provide feedback on health communication materials.

Their portfolio is online at samuelanimates.com. They've created a [public health feed on BlueSky](#) and a [Public Health on BlueSky Starter Pack](#). They run the [Health Literacy Right Now](#) newsletter on LinkedIn.

Summary

This webinar introduces critical AI literacy skills to health professionals. By the end of this webinar, you will be able to:

1. Distinguish among different forms of AI
2. Assess claims about AI in health communication
3. Begin developing criteria for responsible experimentation

Distinguish Among Different Forms of AI

"Artificial Intelligence" is a Moving Target

As [Melanie Mitchell](#) outlines in her book *Artificial Intelligence: A Guide for Thinking Humans*, there is a philosophical approach to AI and a practical approach. The philosophical approach leverages computer science to learn more about the nature of intelligence by recreating it.

The practical approach leverages computer models to perform tasks typically thought of as requiring human intelligence.

AI Applications Don't Feel Like AI Forever

In health fields we are typically more interested in a practical approach to AI. As you might have noticed above, this is a moving target. We've gotten used to certain AI applications like navigation apps, email spam filters, and autofill. In fact, we've gotten so used to them that you might not even think about them as AI anymore. For better or for worse, we've gotten used to the idea that some tasks may no longer require a human touch.

This shift isn't just the result of getting used to certain technologies. It is also related to tech industry hype and marketing, within broader cycles of AI "summer" and "winter" hype cycles. The 2020s aren't the first decade filled with promises and fears of artificial intelligence making huge strides toward the superhuman. If history is any guide, it also won't be the last.

AI Applications in Health Organizations

In health organizations, there are a few broad categories of AI applications we are typically interested in:

- **Classification** models are trained on labeled data to predict those labels on new data. Example: Models providing diagnoses based on medical images.
- **Prediction** models are trained on big data to infer a specific outcome. Example: Models predicting a hospital patient's length of stay.
- **Generative models** are trained on media data to create text, image, or video. Example: Models drafting responses to patient messages.

When you encounter the term "AI" today, it is often referring to generative large language models (ChatGPT, Gemini, DeepSeek, etc.). Though they are trained on media data, these training datasets have become so massive that many people find the resulting model outputs to be usable across a wide variety of tasks including: computer programming, summarizing meeting notes, and translating text. These are examples of "emergent" behaviors of these models, which programmers did not originally intend.

How Are (Most) Generative LLMs Trained?

Most tech companies aren't transparent about how their models are trained. But, the major models on the market come from the same type of training process:

1. **Gather Data.** Training datasets for LLMs comprise massive amounts of text, images, and other media from various sources. This data might come from web scraping. It might also come directly from AI chatbot users. It might also come from users of a company's other services, e.g. Meta drawing on Instagram and Facebook data to train its models.
2. **Fill in the Blank.** LLM outputs sound convincing because they're shaped by the relationships between words in massive training datasets. This happens through a fill-in-the-blank process. LLMs go through many rounds of predicting which word or passage belongs in a particular blank spot in a document from its training data.
3. **Benchmark.** Tech companies test the model's performance against established standards. These are often in the form of multiple choice questions, coding challenges, and language-based tests like pronoun resolution.
4. **Use Customer Feedback.** Tech companies incorporate ratings, preferences, and corrections from users to refine interactions with models over time.

Hidden Step: Data Labor

Marketing and industry hype depict generative AI technology as automatic, fast, and cheap. This hides a more complicated reality. The training process still relies on human workers to label data and moderate content, which reportedly can happen under reporting more about this process. Here are some examples:

- AWU-CWA: "[Ghost Workers in the AI Machine](#)"
- The Guardian: "[Meet Mercy and Anita – the African Workers Driving the AI Revolution, for Just Over a Dollar an Hour](#)"
- Pulitzer Center: "[Moderators Received 7 Cents per Task To Comb Through Violence, Pornography, and Extreme Content on X](#)"

Assess Claims About AI in Health Communication

Case Study: ChatGPT Health

In January 2026, OpenAI introduced [ChatGPT Health](#). It launched in the US more broadly in July 2026. It is a specific product within ChatGPT trained more specifically around health communication. OpenAI is not transparent about the exact process of model training, including training data. But they are more open about [HealthBench](#), their process of benchmarking health communication for this product. We'll look at ChatGPT Health and the HealthBench evaluation process to illustrate how to assess claims about AI in health communication.

Check the Specificity of the Claims

Here are few of the claims that OpenAI has made about ChatGPT Health in marketing materials:

- "Help you feel more informed, prepared, and confident navigating your health."
- "Feel more informed and prepared for important medical conversations."
- "[ChatGPT Health] Adheres to the highest industry standards in data security and privacy."

You may notice the focus on user feelings. It's possible to feel well informed without actually being well informed. The tone of AI outputs, not necessarily the content, might lead to these feelings. A feeling of confidence and preparedness might very well help health communication be more effective. But keep in mind that this is subjective. There is not actually a promise here that ChatGPT Health will help you adopt evidence-based practices or increase patient understanding.

You may also notice the focus on industry standards around privacy. This is not the same as legal standards. These tools tend not to be considered medical tools. Thus they don't fall under the same regulations as an electronic health record, for example. This is important to keep in mind as you essentially give away whatever data you provide an AI chatbot.

Look at the Benchmarks

On top of the fill-in-the-blank process used to train ChatGPT, ChatGPT Health is trained using the HealthBench framework. Under this framework, AI models assess other AI models' outputs at scale. These evaluator models are benchmarked against judgements

from a panel of physicians. This process automates evaluation of AI-generated outputs based on factors including accuracy, communication quality, and completeness. Importantly, these benchmarks might not align with your needs.

Keep in mind that the framework involves specific kinds of user prompts, such as a parent asking about a child's asthma. This may not represent the kinds of questions important in your community.

It's also important to keep in mind that the framework uses custom criteria with custom scores and custom weighting. For example, you might care a lot about succinct answers or the use of bulleted lists and meaningful headings. The HealthBench framework used to train ChatGPT Health weights these factors alongside other factors like referring to emergency services or a reminder that this is not medical advice might not be relevant to your workflow.

The end result of the HealthBench framework are numerical scores that do not map onto standard assessments in the field, like the [PEMAT](#) or [CDC Clear Communication Index](#). You can look at a graph showing the improvement in scores with different model versions. But you don't have a benchmark for what this means in practice or what an acceptable score is in practice.

Ask: What's Not Benchmarked?

Critical gaps in evaluation of models like ChatGPT Health may include:

- Alignment with how people in your community talk about health topics
- Appropriate handling of sensitive issues for your community
- Social standing of you and your organization

Begin Developing Criteria for Responsible Experimentation

Start with Your Values

Before implementing AI in your health communication efforts, clarify your values: For example:

- **Transparency:** What openness do you owe your community?
- **Equity:** Who is most vulnerable to risk?
- **Consent:** What choices should your community have?
- **Trust:** What builds stable relationships in your community?

This might shape what type of tasks you apply AI to. This might shape who you report to, such as a community advisor board. This might also shape what other investments or experiments you undertake to solve health communication issues.

Consider Privacy-Conscious Options

[GPT4All](#) is open-source software that allows you to interact with large language models downloaded to your computer. There is no need to send data to anyone else. This also means you aren't relying on a power-hungry data center.

[Duck.ai](#) provides free access to popular generative AI chats via the privacy-focused browser, DuckDuckGo. Interactions are not tied to you specifically, but keep in mind this is not regulated as a medical tool.

Follow this Short Guide to Testing Generative AI Tools

1. **Define your task.** Clearly articulate when you will use a generative AI tool..
2. **Define failure and success.** Establish specific criteria to evaluate performance. This might relate to the content itself, such as accuracy or clarity. This also might relate to your workflow, such as taking a specific amount of time.
3. **Record AI-generated outputs.** Keep detailed records of what the tool produces. This way you can understand what failure or success looks like in practice.
4. **Track failure and success over time.** Monitor patterns in performance over a certain amount of time or over a certain number of experiments.
5. **Write up your experience.** Document your findings and share learnings with colleagues, community members, etc.