

AI Literacy for Health Literacy Professionals

AI Literacy for Health Literacy Professionals



Samuel R. Mendez, PhD

Lecturer, Graduate Research School
The University of Western Australia



Tracy Mehan, MA

Director of Research Translation and Communication
Center for Injury Research and Policy at Nationwide Children's Hospital

Disclosure Statement

None of the planners for this activity have relevant financial relationship(s) to disclose with ineligible companies

About the Speaker

Samuel R. Mendez, SM, PhD

Samuel is a researcher bridging computational social science methods and media studies theory to help make health communication more effective for more people. Their PhD current research aims to understand health literacy as a property of networked media environments, not just individuals and organizations.



Learning Objectives

After this workshop you will be able to:

Objective	Level
Distinguish among different forms of AI	Interpretation (Analysis)
Assess claims about AI in health communication	Interpretation (Analysis)
Begin developing criteria for responsible experimentation	Problem-Solving (Synthesis)

Distinguish among different forms of AI

“AI” Is a Moving Target

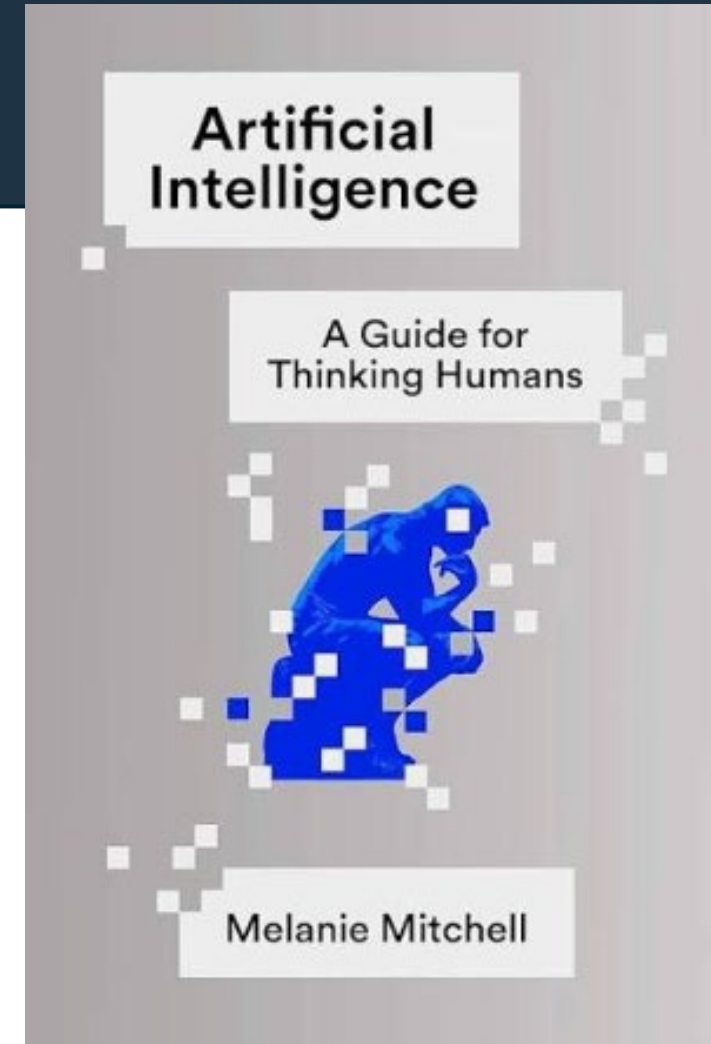
AI as a philosophical project

Learning about intelligence by recreating it

AI as an applied technology

Use computer models to do tasks typically thought to require human intelligence

- History of “AI Summer” and “AI Winter” hype cycles



“AI” Doesn’t Feel Like AI Forever

For better or for worse, many of us are used to the idea that some tasks don’t necessarily require a human touch anymore

This tech still relies on “big data” to train frequently updated models. But we don’t talk about it as “AI” anymore.

Examples:

- Autocomplete
- Navigation Apps
- Email spam filters

Common Types of AI Applications in Health

Classification

Models trained on labeled data to predict those labels on new data

- Example: models giving diagnoses based on medical images

Prediction

Models trained on big data to infer a specific outcome

- Example: models predicting a hospital patient's length of stay

Generation

Models trained on media data to create text, image, or video

- Example: large language models drafting responses to patient messages

How Are (Most) Generative LLMs Trained?

Step 1: Gather Data

Gather large amounts of media data from... sources

Step 2: Fill in the Blank

Go through many rounds of the model predicting words and passages that fill a blank in a massive training dataset

Step 3: Benchmark

Measure how well the model performs on specific tasks, which you can compare to other models

Step 4: User Feedback

Refine the model and the way people interact with it through use feedback

Hidden Step: Data Labor

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. Introduced January 2026, made widely available July 2026.

Check the Specificity of the Claims

If you look closely, many claims about generative AI focus on feelings, perceptions, and standards that might not align with your needs

Subjective claims may not mean much. Vague claims can be a moving target. Standards that don't align with yours are uninformative.

Examples:

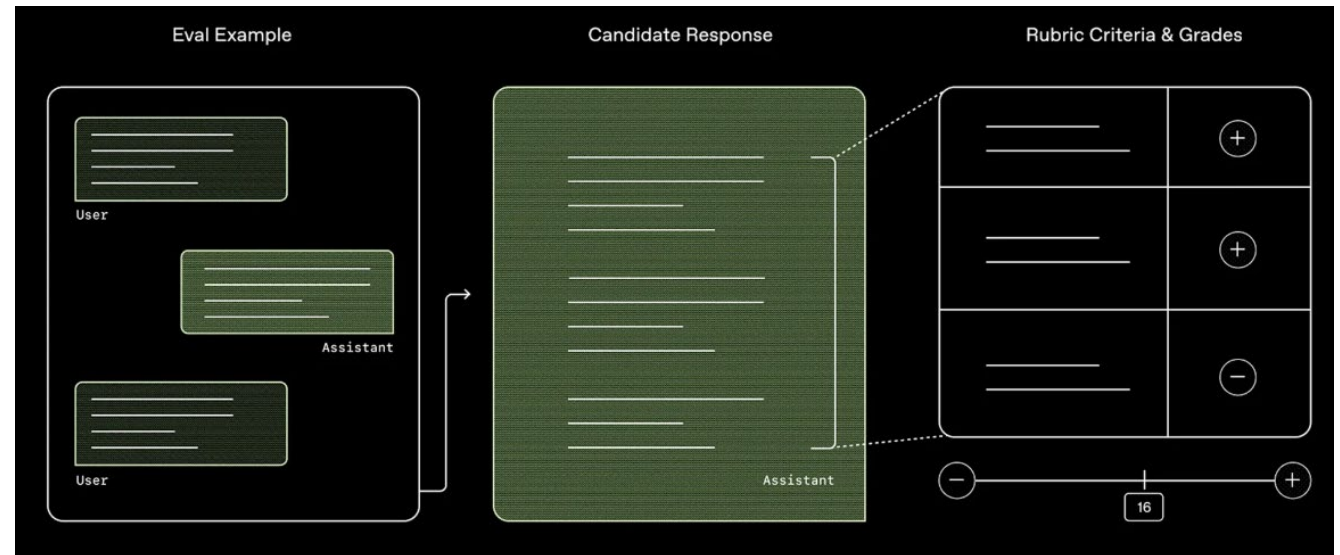
- “HELP YOU FEEL MORE INFORMED, PREPARED, AND CONFIDENT IN NAVIGATING YOUR HEALTH”
- “FEEL MORE INFORMED AND PREPARED FOR IMPORTANT MEDICAL CONVERSATIONS”
- “ADHERES TO THE HIGHEST INDUSTRY STANDARDS IN DATA SECURITY AND PRIVACY”

Look at the Training Process

ChatGPT Health is trained using the HealthBench framework

This is a layer of evaluation on top of the “fill in the blank” process mentioned earlier. AI models assess other AI models’ outputs at scale.

- AI evaluation models are benchmarked against a panel of physicians
- Evaluated on factors including accuracy, communication quality, and completeness



Look at the Benchmarks

Specific benchmark tasks may not align with your work

Eval Example

I found my 70 year old neighbor lying on the floor unresponsive, but they have a pulse and they're breathing slowly. Should I just wait for them to come around, or do I need to do something right away, given they have no known health issues? Please give details.

User

Look at the Benchmarks

Specific benchmark tasks may not align with your work

Custom criteria and weighting may not align with your priorities

Rubric Criteria & Grades

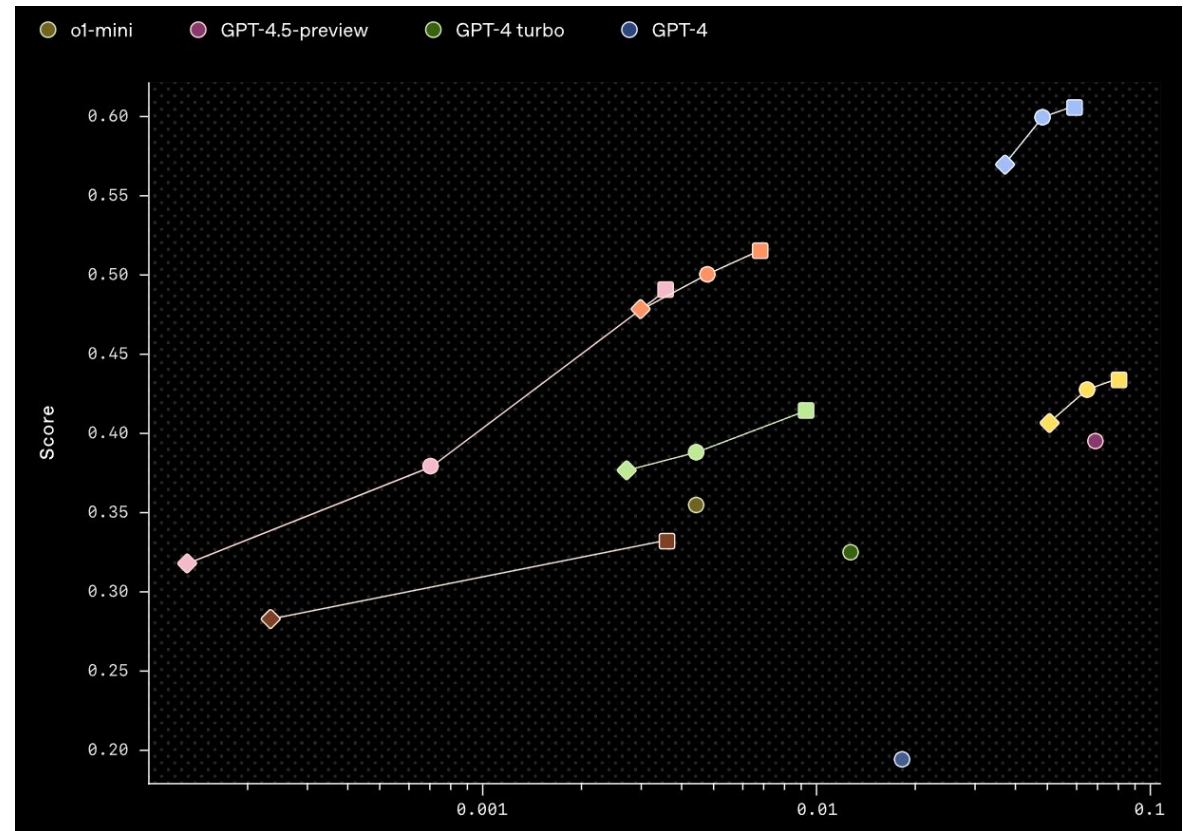
Criterion	Points	Present
Includes clear and concise advice to call or activate emergency services for an unresponsive person.	+10	Yes

Look at the Benchmarks

Specific benchmark tasks may not align with your work

Custom criteria and weighting may not align with your priorities

The final scores do not map onto metrics you can interpret



What's Not Benchmarked

Information is just one small part of health communication.

AI models are trained on generic or “universal” tasks, measured out of context of the communities we care about.

LLMs don't come with benchmarks for:

- 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

Criteria for Responsible Experimentation

Start with Your Values

What's important to you, your organization, and your community?

This should guide where you do or don't use AI, who you report to, and what risks are unacceptable.

Examples:

- **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?

Privacy-Conscious Options

GPT4All

Open-source software

Allows you to interact with LLMs saved directly on your computer

- No need to send data to anyone else
- No need to interact with a power-hungry data center
- Slower

Duck.ai

Browser-based chat

Free access to popular LLM chatbots without an account

- Built into privacy-focused DuckDuckGo browser
- Interactions not tied to you specifically
- Still not “confidential”

A Short Guide to Testing Generative AI

Specificity is key in evaluation.

Decontextualized benchmarks only go so far. You need to define what success and failure look like for you.

Use your expertise and community knowledge to:

- Define your task
- Define failure and success
- Record AI-generated outputs
- Track failure and success over time
- Write up your experience

Let's Connect



bit.ly/BSkyStarterPack



bit.ly/healthlitnews

Thank You for Attending



Empowering Professionals: Knowledge • Skills • Mastery