AI Literacy Framework

Developing AI literacy skills entails thinking critically about how to effectively and responsibly use AI and information it can generate. To facilitate critical reasoning and inquiry, we can critically evaluate AI regarding the following areas:

How to use this framework:

1. Identify your task.
2. Consider stakeholder relevant issues/concerns.
3. Critically evaluate AI use using questions from each category
4. Verify information and revise approach as appropriate.

Technical Evaluation

Model Design

- Where did the training data for the model come from and how reliable is it?
- What background assumptions were the developers working with?
- Are there major biases in the model?
- Is the AI likely to be overly complimentary and uncritical?

Model Functionality

- What AI models are available and what are their relative strengths and weaknesses?
- For what purposes is this AI model designed for?
- What are good practices/what prompts will effectively provide desired feedback?
- What are the different features/settings of an LLM (i.e. thinking mode, individual training) that could help research?

Privacy & Trust

- What information, inputs am I permitted to provide to an LLM?
- What happens to any information I provide?
- Is the LLM designed with my best interests in mind (ex. mental health) or is it a product designed primarily in the interests of the developer/corporation?
- How is this product tempting me to further use it?

Informational Evaluation

Accuracy

- How generally accurate is the model?
- How likely is it to hallucinate?
- What is the probability of false positives?
- Is the model overly confident and am I treating it as overly authoritative?
- If I suspect the accuracy of an output, how can I independently verify information?
- How will the model handle boundary cases?
- How do different LLMs compare in terms of accuracy?

Discipline/Field Relevance

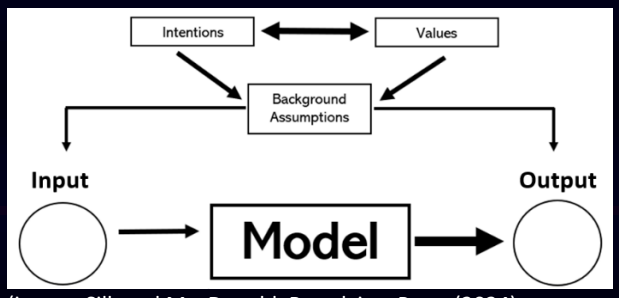
- To what extent is the model familiar with my academic discipline and how deep does that knowledge go? (ex. Continental vs. analytic philosophy)
- Are there disciplinary norms or standards that the model isn't aware of?
- Are there biases/controversies in the discipline that could screw outputs or make the model less usable?
- Does the AI reflect interdisciplinary knowledge?

Authorship

- In what ways am I retaining authorship of research?
- If I am collaborating with AI to develop ideas, what checks and balances are in place to ensure integrity?
- Am I at risk of cognitively depending/deferring to the model? AI psychosis?
- Can I demonstrate and apply knowledge without AI assistance?

Framework Background

Understanding Large Language Models



(image: Silk and MacDonald, Broadview Press (2024)

Intentions, values, and **background assumptions** inform judgments about what training data should be used and why certain outputs are appropriate given certain inputs. This can introduce bias.

Large Language Models (LLMs) split text into small pieces called tokens and learn how those pieces tend to occur together based on examples in their training data. They generate text by statistically predicting what tokens are most likely to come next. Biases and limitations in the training data will be reflected in the model. Unstated or implicit background assumptions by the developer influence judgments about training data appropriateness, what the model will be used for, and what is an appropriate model output for a given input. Understanding LLMs helps evaluate their use.

For example, a developer may choose to omit violent, sexist, or dangerous content from a model and those values will inform assumptions about what constitutes violent, sexist, or dangerous language. Critically evaluating a model involves understanding the developer's assumptions and the model's limitations.

Thinking Critically about AI Outputs

LLMs produce outputs from learned patterns, not intent or true understanding. This makes them well-suited for exploratory tasks (brainstorming, connecting ideas) but less reliable for precise factual claims or exact reproduction of text. It's best to treat LLM outputs like anonymous rumor or gossip. That doesn't mean that it is false, it just means that it requires independent human verification.

Seek out independent reliable sources to corroborate AI-generated claims and be prepared to cite them. AI can stimulate thinking, but you are responsible for verifying that AI-generated information is true. Overly complimentary AI can contribute to confirmation bias.

Epistemic health:

Our capacity to reliably form, revise, and coordinate beliefs that are responsive to evidence, open to correction, and action-guiding in ways remain accountable to the facts and defensible to others.

Protecting Skill Development

Because AI can replace human labor in certain cases, there is a risk that it can inhibit skill development, cause developed skills to atrophy, or give us a false sense of our own knowledge, skills, and abilities. Poor AI habits can undermine epistemic health, constraining our ability to form and critically evaluate beliefs independently.

It is important to maintain human-to-human interaction, including communication, collaboration, and demonstrating our ability to think and reason with other humans and without AI assistance. One must be conscientious about how often and why they may be deferring to what AI generates.

EAISW