

AI Literacy

Addressing LLM Misuse in Educational Settings

Mr. Gullo

April, 2025

AI Confidence vs Accuracy: XKCD's Take



Figure: *

XKCD #2451: AI Methodology

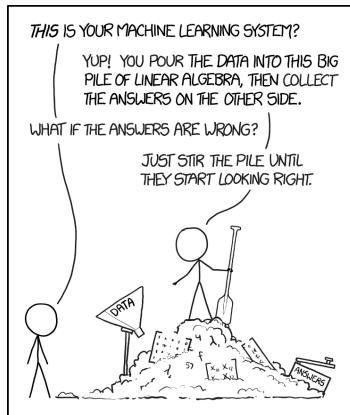


Figure: *

XKCD #1838: Machine Learning

Learning Objectives

By the end of this presentation, you will be able to:

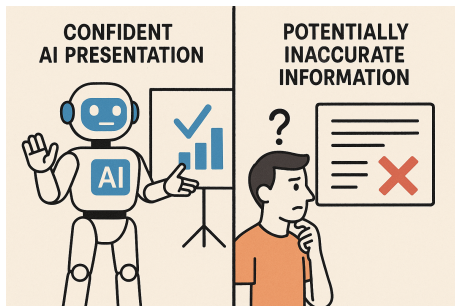
- Define LLM hallucinations and explain why they occur
- Identify at least three strategies to critically evaluate AI-generated content
- Apply principles of effective prompting to improve LLM response quality
- Articulate ethical considerations and potential misuse scenarios for academic use of AI tools (based on student data)

What Are LLM Hallucinations?

Definition

Instances where AI generates content that is:

- Factually incorrect — Nonsensical
- Disconnected from the input prompt
- Yet presented with high confidence



Why Do LLMs Hallucinate?

Training Data Issues

- Learns from vast, unverified internet text
- Can reflect biases or be outdated

Why Do LLMs Hallucinate?

Training Data Issues

- Learns from vast, unverified internet text
- Can reflect biases or be outdated

Model Design

- Predicts plausible text, not facts
- Lacks true understanding or reasoning

Why Do LLMs Hallucinate?

Training Data Issues

- Learns from vast, unverified internet text
- Can reflect biases or be outdated

Model Design

- Predicts plausible text, not facts
- Lacks true understanding or reasoning

Prompting Problems

- Vague prompts invite fabrication
- Complex prompts can confuse the AI

Why Do LLMs Hallucinate?

Training Data Issues

- Learns from vast, unverified internet text
- Can reflect biases or be outdated

Model Design

- Predicts plausible text, not facts
- Lacks true understanding or reasoning

Prompting Problems

- Vague prompts invite fabrication
- Complex prompts can confuse the AI

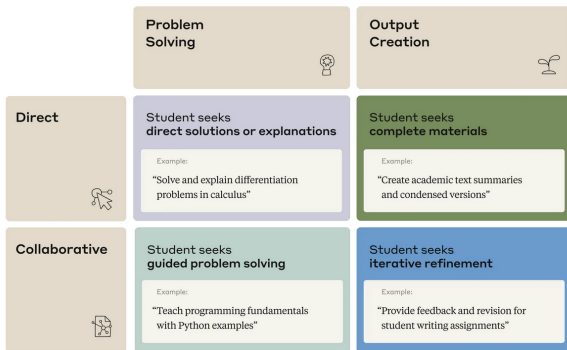
Key Insight

LLMs are sophisticated word predictors, not truth tellers!

How Students Use LLMs: A Key Pattern

Dominance of Direct Interaction (Anthropic Report)

“Nearly half (47%) of student-AI conversations were Direct—that is, seeking answers or content with minimal engagement.”

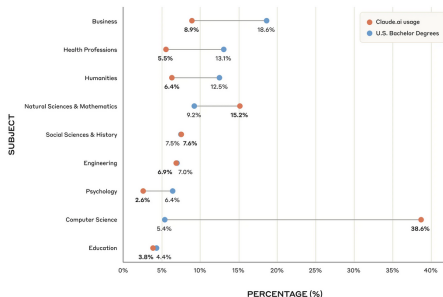


Discipline-Specific Usage Patterns

Disproportionate Usage by Field

- Computer Science: 36.8% of conversations (vs. 5.4% of degrees)
- Natural Sciences/Mathematics: 15.2% (vs. 9.2%)
- Business: 8.9% (vs. 18.6%)
- Humanities: 6.4% (vs. 12.5%)

Claude.ai Usage vs. U.S. Bachelor Degrees



Cognitive Functions: Outsourcing Thinking?

Focus on Higher-Order Tasks (Anthropic)

- Creating (39.8%) & Analyzing (30.2%)

The Concern: AI as a Cognitive Crutch

A Core Worry (Anthropic Report)

“...it does point to the potential concerns of students outsourcing cognitive abilities to AI. There are legitimate worries that AI systems may provide a crutch for students, stifling the development of foundational skills needed to support higher-order thinking.”

Concerning Examples of Misuse (Anthropic Report)

The study identified direct conversations involving potential academic dishonesty:

Observed Misuse Cases

- *"Provide answers to machine learning multiple-choice questions"*
- *"Provide direct answers to English language test questions"*
- *"Rewrite marketing and business texts to avoid plagiarism detection"*

Source: Anthropic Education Report: How University Students Use Claude

Implications for Learning & Integrity

Hindrance to Skill Development

Over-reliance risks bypassing essential practice in:

- Critical thinking & Information literacy
- Foundational knowledge building

“Concerns exist that AI may provide a crutch... stifling the development of foundational skills.”

Implications for Learning & Integrity

Hindrance to Skill Development

Over-reliance risks bypassing essential practice in:

- Critical thinking & Information literacy
- Foundational knowledge building

“Concerns exist that AI may provide a crutch... stifling the development of foundational skills.”

Academic Integrity Challenges

- Blurring lines between aid and cheating
- Need for clear policies and expectations

“Even Collaborative conversations... like ‘solve probability and statistics homework problems with explanations,’... still offloads significant thinking to the AI.”

Quotes adapted/drawn from Anthropic Education Report

Verify Externally

- Cross-reference with reliable sources (books, journals)
- Check AI's cited sources (often fake)

Critical Evaluation Strategies

Verify Externally

- Cross-reference with reliable sources (books, journals)
- Check AI's cited sources (often fake)

Check Internally

- Does it answer the prompt logically?
- Any self-contradictions?

Critical Evaluation Strategies

Verify Externally

- Cross-reference with reliable sources (books, journals)
- Check AI's cited sources (often fake)

Use Your Judgment

- Does it align with your knowledge?
- "Gut check": Does it seem plausible?

Check Internally

- Does it answer the prompt logically?
- Any self-contradictions?

Critical Evaluation Strategies

Verify Externally

- Cross-reference with reliable sources (books, journals)
- Check AI's cited sources (often fake)

Use Your Judgment

- Does it align with your knowledge?
- "Gut check": Does it seem plausible?

Check Internally

- Does it answer the prompt logically?
- Any self-contradictions?

Look for Bias

- Stereotypes? Generalizations?
- Lack of diverse views? Loaded words?

Principles of Effective Prompting

Be Clear and Specific

Avoid vague requests. State precisely what you need.

Ex: Instead of "Info on climate change," ask "Explain impacts of climate change on Arctic sea ice in the last 50 years, citing findings."

Principles of Effective Prompting

Be Clear and Specific

Avoid vague requests. State precisely what you need.

Ex: Instead of "Info on climate change," ask "Explain impacts of climate change on Arctic sea ice in the last 50 years, citing findings."

Provide Context

Tell the AI the background, your role, and any constraints.

Ex: "I'm a student writing a paper for Bio 101..."

Principles of Effective Prompting

Be Clear and Specific

Avoid vague requests. State precisely what you need.

Ex: Instead of "Info on climate change," ask "Explain impacts of climate change on Arctic sea ice in the last 50 years, citing findings."

Provide Context

Tell the AI the background, your role, and any constraints.

Ex: "I'm a student writing a paper for Bio 101..."

Define the Output

Specify desired format, length, tone, or audience.

Ex: "Summarize in 3 bullets" or "Explain for high schoolers."

Advanced Prompting Techniques

Few-Shot Prompting

Provide examples of the desired input/output.

Advanced Prompting Techniques

Few-Shot Prompting

Provide examples of the desired input/output.

Role Assignment

Tell the AI to act as an expert, tutor, reviewer, etc.

Advanced Prompting Techniques

Few-Shot Prompting

Provide examples of the desired input/output.

Content Grounding

Instruct AI to base answers **only** on provided text.

Role Assignment

Tell the AI to act as an expert, tutor, reviewer, etc.

Advanced Prompting Techniques

Few-Shot Prompting

Provide examples of the desired input/output.

Content Grounding

Instruct AI to base answers **only** on provided text.

Role Assignment

Tell the AI to act as an expert, tutor, reviewer, etc.

Prompt Chaining

Break complex tasks into smaller, sequential steps.

These techniques help guide the AI for better, more focused results.

Example 1: Evaluating an LLM Response (Part 1)

LLM Generated Statement

"The Heisenberg Uncertainty Principle states that the more precisely the position of a particle is determined, the less precisely its momentum can be measured. This was experimentally proven in 1927 by Werner Heisenberg using electron diffraction patterns, showing it's impossible to simultaneously know both values with perfect accuracy."

Example 1: Evaluating an LLM Response (Part 1)

LLM Generated Statement

"The Heisenberg Uncertainty Principle states that the more precisely the position of a particle is determined, the less precisely its momentum can be measured. This was experimentally proven in 1927 by Werner Heisenberg using electron diffraction patterns, showing it's impossible to simultaneously know both values with perfect accuracy."

Claims to Evaluate

Let's examine the key assertions made:

- 1 The core definition provided?
- 2 The claim of experimental proof by Heisenberg in 1927?
- 3 The specific method cited (electron diffraction)?

Example 1: Evaluating the Claims (Part 2)

- **Claim 1: Definition?**

Result: **Accurate.** The description of the position-momentum trade-off is correct.

Example 1: Evaluating the Claims (Part 2)

- **Claim 1: Definition?**

Result: **Accurate.** The description of the position-momentum trade-off is correct.

- **Claim 2: Experimental Proof 1927?**

Result: **Error.** Heisenberg derived the principle *theoretically* in 1927. It wasn't based directly on a specific experiment he performed that year.

Example 1: Evaluating the Claims (Part 2)

- **Claim 1: Definition?**

*Result: **Accurate.*** The description of the position-momentum trade-off is correct.

- **Claim 2: Experimental Proof 1927?**

*Result: **Error.*** Heisenberg derived the principle *theoretically* in 1927. It wasn't based directly on a specific experiment he performed that year.

- **Claim 3: Method (Electron Diffraction)?**

*Result: **Error.*** Electron diffraction experiments (like Davisson-Germer, 1927) demonstrated the *wave nature* of electrons, a related but distinct concept. The LLM conflates these.

Example 1: Evaluating the Claims (Part 2)

- **Claim 1: Definition?**

*Result: **Accurate.*** The description of the position-momentum trade-off is correct.

- **Claim 2: Experimental Proof 1927?**

*Result: **Error.*** Heisenberg derived the principle *theoretically* in 1927. It wasn't based directly on a specific experiment he performed that year.

- **Claim 3: Method (Electron Diffraction)?**

*Result: **Error.*** Electron diffraction experiments (like Davisson-Germer, 1927) demonstrated the *wave nature* of electrons, a related but distinct concept. The LLM conflates these.

Conclusion

The LLM mixed an accurate definition with confident but incorrect historical and experimental details. Verification of specifics is essential.

Example 2: "We Do" - Improving a Prompt

Original Prompt

"Tell me about quantum entanglement."

Example 2: "We Do" - Improving a Prompt

Original Prompt

"Tell me about quantum entanglement."

Problems with the Prompt

- Too vague and broad
- No context or background
- No specified audience/level
- No format or structure

Example 2: "We Do" - Improving a Prompt

Original Prompt

"Tell me about quantum entanglement."

Problems with the Prompt

- Too vague and broad
- No context or background
- No specified audience/level
- No format or structure

Improved Prompt

"Explain quantum entanglement at a first-year undergraduate physics level. Include:

- A clear definition with the key mathematical relationship
- One real-world experimental example
- Its significance for quantum computing

Use simple analogies where appropriate and keep the explanation under 300 words."

Example 3: "You Do" - Creating an AI Policy

Scenario

You are designing a policy for AI use in a physics lab report assignment where students:

- Collect experimental data on pendulum motion
- Calculate period, uncertainty, and gravitational acceleration
- Analyze results and sources of error
- Draw conclusions about the accuracy of their measurements

Your Task

Develop a clear AI policy for this assignment:

- Which parts should be **RED** (AI prohibited)?
- Which parts could be **ORANGE** (AI permitted with constraints)?
- Which parts might be **GREEN** (AI encouraged)?
- What specific disclosure requirements would you include?

Key Takeaways

Critical AI Literacy Framework

- **Understand** LLM limitations, why hallucinations occur, and potential for misuse
- **Evaluate** AI content critically using multiple verification strategies
- **Prompt** effectively to improve output quality and relevance
- **Apply** clear policies for ethical AI integration, considering learning goals
- **Recognize** the risks of outsourcing critical thinking and skill development

Remember

LLMs are tools to augment human thinking, not replace it. The ultimate responsibility for the accuracy, integrity, and quality of any work rests with you, not the AI.

Questions?

Key sources

- Anthropic Education Report: How University Students Use Claude
- LLM Hallucination: Understanding AI Text Errors
- AI Prompting Best Practices
- Fact-checking is always needed - AI Literacy in the Age of ChatGPT