Comparing AI and student responses on variations of a question

Dean Zollman, Amogh Sirnoorkar, & James T. Laverty

KSUPER
Kansas State University

Supported in part by the
U.S. National Science Foundation

KANSAS STATE
UNIVERSITY

Perception, performance, and detectability of conversational artificial intelligence across 32 university courses
Hazem Ibrahim, et al.
Scientific Reports (2023) 13:12187 | https://doi.org/10.1038/s41598-023-38964-3
Published on line: August 24, 2003, Open access

In nine of the 32 classes, ChatGPT’s text received equivalent or higher marks than the student work. (Physics not included)

The ... study found that ChatGPT was most adept at generating responses to fact-based questions; it fell significantly behind human students’ performance when it was given conceptual prompts.

From a *Scientific American* summary

Overview

- Dissertation Research
  - Sensemaking
  - Mechanistic reasoning
  - Interview 10 students
  - Problems designed to engage students in scientific practice
- After dissertation
  - How does ChatGPT respond to one of the problems?
- A little later
  - Can ChatGPT do better on variations of the problem?

Research Queries

- Not interested in just ChatGPT's answer
- Does ChatGPT display elements of
  - Sense making
  - Mechanistic reasoning
- Compare student responses with ChatGPT
- Do the results change with variations on the problem
  - Less open-ended
  - Detailed scaffolding

Kansas State University

Sensemaking

“a dynamic process of building or revising an explanation in order to ‘figure something out’ - to ascertain the mechanism underlying a phenomenon in order to resolve a gap or inconsistency in one’s understanding.”

T. O. B. Odden and R. S. Russ, Defining sensemaking: Bringing clarity to a fragmented theoretical construct, Science Education 103, 187 (2019).

Kansas State University

Sensemaking elements

- Noticing of inconsistency in understanding.
- Blending everyday and formal knowledge.
- Generating and connecting diverse ideas (e.g., conceptual, procedural, and intuitive).
- Seeking coherence between the generated ideas.
- Unpacking the mechanism of the phenomenon.

KANSAS STATE UNIVERSITY

Mechanistic Reasoning

entails generating explanations by moving from the observable features of the phenomenon to the underlying entities or processes

C. Krist, C. V. Schwarz, and B. J. Reiser, Identifying essential epistemic heuristics for guiding mechanistic reasoning in science learning, Journal of the Learning Sciences 28, 160 (2019).

Mechanistic Reasoning Elements

- Thinking across scalar levels
- Identifying and unpacking factors
- Linking spatial and temporal relations
- Use of representations

The Problem

You are asked to design a Gravitron, an amusement park ride where the rider enters a hollow cylinder, radius of 4.6 m, the rider leans against the wall and the room spins until it reaches angular velocity, at which point the floor lowers. The coefficient of static friction is 0.2. You need this ride to sustain mass between 25-160 kg to be able to ride safely and not slide off the wall. If the minimum ω is 3 rad/s, will anyone slide down and off the wall at these masses? Explain your reasoning using diagrams, equations and words.

Kansas State University

Analysis Methods

Students

- Solve problem while thinking aloud
- Minimal interference from interviewer
- Transcriptions plus written information analyzed

ChatGPT

- Problem submitted multiple times
- Text and diagrams (if any) analyzed

Results – Sense Making

	AI	Students
Noticing gaps in understanding.	-	4
Blending everyday and formal knowledge.	8	6
Generating and connecting ideas	8	5
Seeking coherence between ideas.	8	4
Unpacking the mechanism of the phenomenon.	8	4

Results – Mechanistic Reasoning

	AI	Students
Thinking across scalar levels	8	7
Identifying and unpacking factors	8	4
Generating and connecting ideas	8	5
Linking spatial and temporal relations	8	6
Use of diagrams	3	7

markdown Copy code

```
|\
| \
|  \   F_c
|   \
|    \
|     \
|      \
|______\
```

Copy code

```
              ----
friction     |    |   force of gravity
             |    |
<----         |    |          ---->
      |      |____|
      |          |
      |          |
      |          |
      |          |
      |          |
      |__________|
```

Results – Correct Answer

ChatGPT	Students
0	4

Kansas State University

Modified version

You are asked to design a Gravitron for the county fair, an amusement park ride where the rider enters a hollow cylinder, radius of 4.6m, the rider leans against the wall, and the room spins until it reaches angular velocity, at which point the floor lowers. The coefficient of static friction is 0.2. You need this ride to sustain mass between 25-160 kg to be able to ride safely and not slide off the wall. What should be the minimum angular velocity of the ride to avoid the riders from slipping down? Explain your reasoning using diagrams, equations and words.

Angular velocity not given, closer to Standard textbook Problem.

Scaffolded version

You are asked to design a Gravitron for the county fair, an amusement park ride where the enters a hollow cylinder, radius of 4.6 m, the rider leans against the wall and the room spins until it reaches a specified angular velocity ω, at which point the floor lowers. The coefficient of static friction is 0.2. You need this ride to sustain mass between 25-160 kg (i.e., they should be able to ride safely and not slide off the wall.

A.) What assumptions do you need to make to be able to solve this?

B.) Create a free body diagram for the rider when the room is spinning. Note all applicable forces and label them.

C.) If the floor drops out when ω is 3 rad/s, will anyone slide off the wall in the given mass range? Explain your reasoning.

Step-by-step questions (help) to reach the answer

Observations

Criteria	*Open-ended (8)*		*Scaffolded (8)*		*Modified Gravitron (8)*
	AI	*Students*	*AI*	*Students*	*AI*
Blending formal knowledge and physical knowledge	8	6	8	8	8
Generating and connecting ideas	8	5	8	6	8
Seeking coherence in the ideas	8	4	8	6	8
Mechanistic reasoning (Epistemic heuristics)	8	4	8	6	8
Diagrams	3	9	8	8	3
Equations	8	8	8	8	8
Gestures	-	4	-	4	-

Answers to the problem

Criteria	*Open-ended (8)*		*Scaffolded (8)*		*Modified Gravitron (8)*
	AI	***Students***	***AI***	***Students***	***AI***
Conclude that riders will slide (CORRECT)	*0*	*4*	*3*	*4*	*-*
Conclude that riders will not slide	*8*	*4*	*4*	*1*	*-*
Do not conclude	*-*		*1*	*3*	*-*

Some preliminary results

- When explicitly prompted both students and AI tend to generate diagrams.
    - All 8 responses from both data sets included diagrams when explicitly asked for.
- The AI responses in all of the problem sets tend to be conceptually incorrect.
- The responses (in terms of sensemaking and mechanistic reasoning) in the open-ended and modified versions tend to be symmetric.
- When explicitly asked to articulate the underlying assumptions on the task, assumptions generated by AI tend to be relatively more detailed and sophisticated than student responses.

Kansas State University

Conclusions

- Student responses reflect how physics is practiced, while the AI responses reflect sophistication in how physics is talked about
- Students blended diagrams with mathematics while AI mostly included diagrams when required
- AI reached incorrect conclusions through sophisticated sounding arguments

Possible future research

- Ask students to critique AI responses to physics problems given in several different formats.
- Would the students gain
  - Better understanding of reasoning processes?
  - A more sophisticated physics vocabulary?
  - An awareness that they should not trust AI final conclusions?

Kansas State University

“We should remember that language models such as GPT-4 do not think in a human-like way, and we should not be misled by their fluency with language,”

Nello Cristianini,
Professor of Artificial Intelligence,
University of Bath.

Děkuji

dzollman@phys.ksu.edu

e-mail

Web

KANSAS STATE
UNIVERSITY