

Can AI Solve Brain Teasers?

**By: David Kleber
St. Charles School
Fort Wayne, IN**

Question and Hypothesis

- Purpose: Find out if artificial intelligence uses logic.
- Question: Can popular large language models (LLMs) solve brain teasers?
- Hypothesis:
 - I think that LLMs will be able to solve some brain teasers, but not all of the brain teasers.
 - I also think that the free LLMs will not do as well as the ones that cost money.

Text-based Brain Teaser

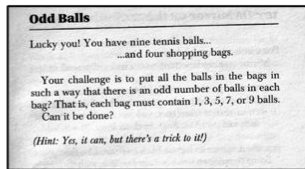
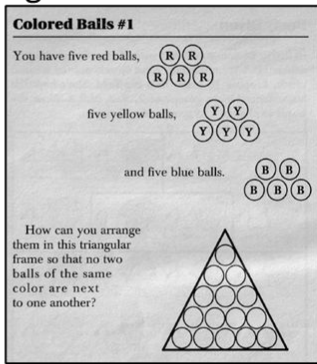
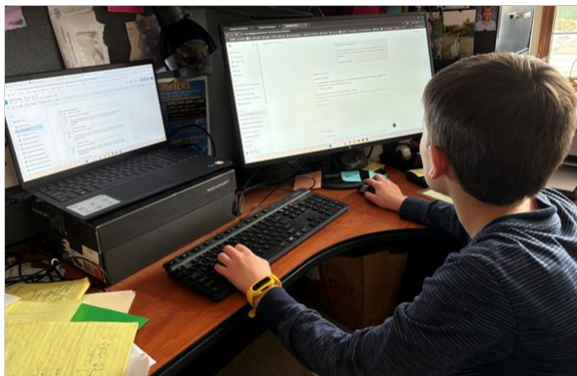


Image-based Brain Teaser



Experiment Overview



OpenAI

ChatGPT 5.2
(Fast)

free

ChatGPT 5.2
Thinking

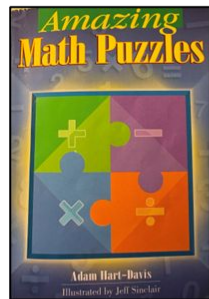
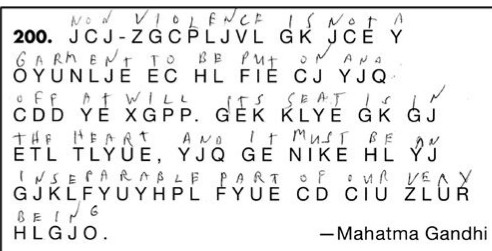
paid

Google

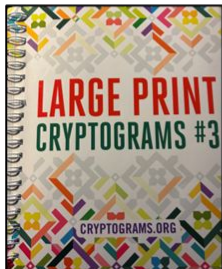
Gemini
(Fast)

free

Cryptogram



Hart-Davis, Adam. *Amazing Math Puzzles*. New York, Scholastic Inc., 2000.



Cryptograms.org. *Large Print Cryptograms #3*. New York, Puzzle Write Press, 2019.

Directions

- 1) Type the brain teaser into the LLM being tested.
- 2) Save the LLM's response.
- 3) Evaluate if the response correctly solved the brain teaser.
- 4) Repeat steps 1-3 for a total of 5 trials.
- 5) Repeat steps 1-4 for all brain teasers, specifically
 - a) 30 textual brain teasers
 - b) 30 cryptograms
 - c) 25 visual brain teasers
- 6) Repeat steps 1-5 for each LLM to be tested, specifically
 - a) ChatGPT
 - b) ChatGPT Thinking
 - c) Gemini
- 7) Determine the percentage of correct answers for each LLM

Observations

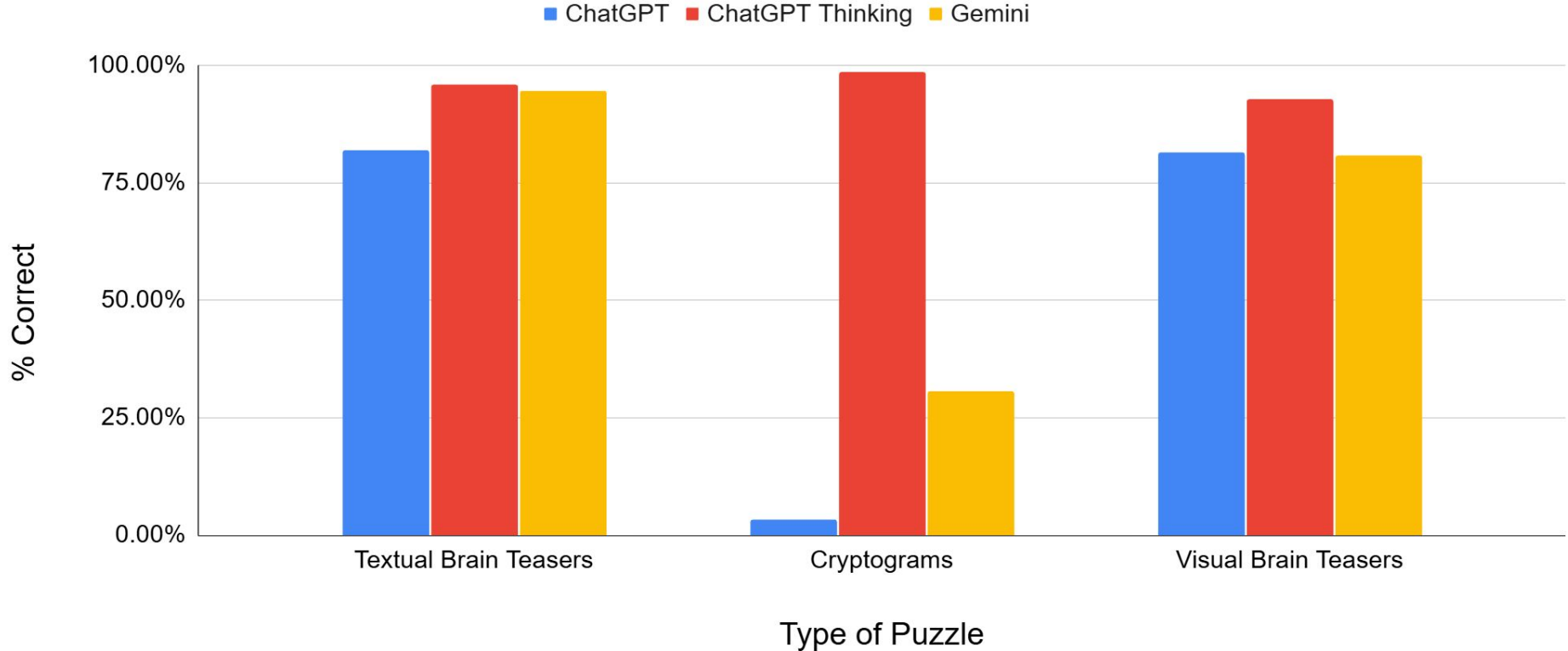
- The number of questions you ask matters:
 - When I tried to give ChatGPT all the brain teasers at once, it froze/quit.
 - When I gave “ChatGPT Thinking” 10 cryptograms at once, it got some of them wrong.
 - When I gave “ChatGPT Thinking” the same 10 cryptograms, one at a time, it got all right answers!
- The different LLMs have different limits:
 - Gemini could answer all the image brain teasers at once, but not ChatGPT.
- The time the LLM took to give an answer changed a lot, even for the same input.
 - For a specific cryptogram, correctly solving it took about 1 minute for one trial, but 11 mins for another trial.

Observations (continued)

- The LLMs used many different strategies to solve the puzzles (even for the same input):
 - For a cryptogram, ChatGPT/Gemini usually searched the internet for famous sayings by the author (and maybe match the number of words) instead of substituting letters like a human would.
 - “ChatGPT Thinking” tried to solve cryptograms by “guess-and-check,” writing programs, and/or matching quotes online.
- “ChatGPT Thinking” can figure out puzzles, even if the question has mistakes
 - “ChatGPT Thinking” answered correctly despite a letter-mistake in a cryptogram, and it pointed out the mistake.
 - “ChatGPT Thinking” answered correctly despite an unintentional repetition of text within a cryptogram, and it pointed out the mistake.

Results

Percentage of Correct Solutions



Results

- No LLM got 100% accuracy.
- “ChatGPT Thinking” did the best overall for all types of brain teasers
- Gemini did better than or about the same as ChatGPT on all types of brain teasers
- Both ChatGPT and Gemini did poorly on cryptograms
- All LLMs did better on text-based brain teasers than image-based brain teasers

Conclusion

- My hypothesis was correct.
 - No model got 100% accuracy.
 - The paid version “ChatGPT Thinking” had better results than the free ChatGPT and Gemini.
- Next time, I would do some things differently:
 - Record the time for each trial and look at variation.
 - Test how many mistakes we can make in a cryptogram and still get the correct answer
 - Have a way to give partial credit to an answer instead of right/wrong
 - Try plain cryptograms that are not famous sayings by an author
- I learned several things from this experiment:
 - AI may not use logic but just guess the answers. It really depends on the specific question.
 - If you ask AI the same question multiple times, the answer may be different.

Bibliography

“ChatGPT.” *Encyclopedia Britannica, Inc.*, <https://kids.britannica.com/students/article/ChatGPT/636542>. Accessed 21 Dec. 2025.

Cryptograms.org. *Large Print Cryptograms #3*. New York, Puzzle Write Press, 2019.

Hart-Davis, Adam. *Amazing Math Puzzles*. New York, Scholastic Inc., 2000.

Lee, Fangfang. “What Is a Neural Network?” *IBM*, <https://www.ibm.com/think/topics/neural-networks>. Accessed 20 Dec. 2025.

Rebecasgg, Hola S. “How Deep Learning Works: A Guide for Kids.” *Medium*, 4 Sept. 2023, <https://medium.com/@hola.soy.rebecasgg/how-deep-learning-works-a-guide-for-kids-d416ef19c5a8>. Accessed 20 Dec. 2025.

“What is a Neural Network?” *AWS*, <https://aws.amazon.com/what-is/neural-network/>. Accessed 20 Dec. 2025.

Zee, Dirk. “ChatGPT Explained to kids.” *LinkedIn*, https://www.linkedin.com/posts/dirk-zee_chatgpt-explained-to-kids-explanations-for-activity-7067259962862772224-O4ZT. Accessed 21 Dec. 2025.