



---

# GPT-5 Falls Short in Human Attention Test

## Description

A long-standing psychology test has highlighted a significant weakness in advanced artificial intelligence (AI) systems, showing that their focus may differ from that of humans. Researchers, led by Suketu Patel, studied how large language models (LLMs), including GPT-5, perform on a known cognitive test called the Stroop task.

The Stroop task involves showing participants words that name colours, such as “red” or “blue,” displayed in different ink colours. Participants must identify the ink colour while ignoring the word's meaning, which causes a mental conflict. Humans typically become slower at responding when the ink colour does not match the word, a phenomenon known as the Stroop effect. However, even during lengthy tasks, people generally maintain high accuracy and focus.

To determine how AI models cope with similar challenges, the researchers tested several leading LLMs. Initially, these models performed well with short word lists, with GPT-4o achieving 91% accuracy. However, the situation changed dramatically with longer lists. With ten words, GPT-4o's accuracy dropped to 57%, and with forty words, it plummeted to just 15%. Claude 3.5 Sonnet also saw a decline in performance as list length increased, dropping to 24% accuracy with forty words.

These results indicate a crucial difference between human and AI cognition. While AI excels at recognising words, it struggles to suppress automatic responses and maintain focus over time. This suggests that the attention mechanisms of AI systems fundamentally differ from those in the human brain, revealing important limitations as AI becomes more integrated into daily life.

---

## Vocabulary List:

1. **psychology** //saɪ'kɒlədʒi// (noun): study of the mind and how people behave
2. **cognitive** //kɒgnətɪv// (adjective): relating to thinking, learning, and remembering
3. **accuracy** //ækjə'reɪsi// (noun): how correct or exact something is
4. **plummeted** //ˈplʌmɪtɪd// (verb): fell very quickly to a much lower level
5. **suppress** //sə'prɛs// (verb): stop or hold back something from happening
6. **limitations** //ˌlɪmɪ'teɪʃənz// (noun): things that make something less effective

## Comprehension Questions



---

## Multiple Choice

1. What is the Stroop task primarily used to study?
  - Option: Memory recall
  - Option: Word recognition
  - Option: Color recognition and mental conflict
  - Option: Language translation
2. What was the accuracy of GPT-4o when tested with ten words?
  - Option: 91%
  - Option: 57%
  - Option: 24%
  - Option: 15%
3. Which model achieved a 91% accuracy on the Stroop task with short word lists?
  - Option: Claude 3.5 Sonnet
  - Option: GPT-4o
  - Option: Suketu Patel's model
  - Option: GPT-5
4. How does human focus typically compare when ink color matches the word versus when it does not?
  - Option: Faster and more accurate
  - Option: Slower and less accurate
  - Option: No change in speed
  - Option: Faster but less accurate
5. What happens to GPT-4o's accuracy with forty words?
  - Option: It remains stable
  - Option: It drops dramatically
  - Option: It improves
  - Option: It fluctuates
6. Who led the research study on AI systems and the Stroop task?
  - Option: Claude 3.5 Sonnet
  - Option: Researchers at MIT
  - Option: Suketu Patel
  - Option: GPT-5 team



### True-False

7. Humans tend to maintain high accuracy and focus even with lengthy Stroop tasks.
8. AI models exhibited consistent performance regardless of list length in the Stroop task.
9. The Stroop task involves participants identifying the meaning of word color.
10. AI systems perform better on longer lists than on shorter lists according to the study.
11. The study suggests that AI and human cognition mechanisms are fundamentally similar.
12. Claude 3.5 Sonnet achieved a 24% accuracy with forty words in the Stroop task.

### Gap-Fill

13. The Stroop task describes a phenomenon known as the Stroop effect, which causes a mental conflict when the ink color does not match the word. This phenomenon affects human response speed and accuracy, indicating a gap \_\_\_\_\_ between AI and human cognition.
14. In the study, GPT-4o's accuracy dropped to 15% when tested with forty words and demonstrated a significant gap \_\_\_\_\_ in performance compared to its results with short lists.
15. Researchers led by Suketu Patel examined how large language models (LLMs) cope with challenges similar to that of the Stroop task, revealing a gap \_\_\_\_\_ in focus and automatic responses.
16. Humans typically exhibit a slower response \_\_\_\_\_ when the ink color of the Stroop words does not match the color of the words themselves.
17. The cognitive test revealed important limitations in AI systems, indicating that their attention



mechanisms differ \_\_\_\_\_ from those found in the human brain.

18. While AI excels at recognizing words, the study highlights a gap \_\_\_\_\_ in its ability to maintain focus during lengthy tasks compared to humans.

## Answer

**Multiple Choice:** 1. Color recognition and mental conflict 2. 57% 3. GPT-4o 4. Slower and less accurate 5. It drops dramatically 6. Suketu Patel

**True-False:** 7. True 8. False 9. False 10. False 11. False 12. True

**Gap-Fill:** 13. in understanding 14. decline 15. difference 16. time 17. significantly 18. limitation

## Vocabulary quizzes

**Multiple Choice ( Select the Correct answer for each question. )**

1. Which field studies the mind and behavior?

- Option: Physics
- Option: Biology
- Option: Psychology
- Option: Chemistry

2. Cognitive skills are essential for which type of processing?

- Option: Emotional
- Option: Physical
- Option: Mental
- Option: Social

3. Which type of bodies does celestial refer to?

- Option: Earthly
- Option: Heavenly
- Option: Animal
- Option: Mineral

4. What phenomenon occurs when one celestial body moves into the shadow of another?

- Option: Meteor
- Option: Eclipse
- Option: Comet



---

Option: Satellite

5. Who are the scientists that study weather patterns?

- Option: Geologists
- Option: Meteorologists
- Option: Astronomers
- Option: Biologists

6. What is the income generated from operations called?

- Option: Expense
- Option: Profit
- Option: Revenue
- Option: Loss

7. During a solar eclipse, what is the phase called when the sun is completely obscured?

- Option: Partiality
- Option: Totality
- Option: Visibility
- Option: Brightness

8. What do we call a site for disposing of waste materials?

- Option: Recycling Center
- Option: Incinerator
- Option: Landfill
- Option: Wastewater Treatment

9. What is the term for making the best or most effective use of resources?

- Option: Reducing
- Option: Deteriorating
- Option: Optimising
- Option: Neglecting

10. What do we call substances discharged into the air, such as pollutants?

- Option: Reflections
- Option: Absorptions
- Option: Emissions
- Option: Dispersals

**Gap-Fill ( Fill in the blanks with the correct word from the vocabulary list. )**



11. The time of day when the sun is below the horizon but still lights the sky is known as \_\_\_\_\_.
12. The rocket's additional engines, essential for the launch phase, are called \_\_\_\_\_.
13. Items that are no longer useful or wanted are often \_\_\_\_\_.
14. The line where the earth's surface and the sky appear to meet is called the \_\_\_\_\_.
15. The slow change over time in climate patterns is described as \_\_\_\_\_.
16. In astronomy, the alignment of two celestial bodies is termed a \_\_\_\_\_.
17. The path of an object as it moves around another object in space is called its \_\_\_\_\_.
18. Organic compounds derived from ammonia are known as \_\_\_\_\_.
19. This type of plastic, commonly used for insulation and packaging, is called \_\_\_\_\_.
20. Every research study has its own set of \_\_\_\_\_.

**Matching Sentences ( Match each definition to the correct word from the vocabulary list. )**

21. He tried to suppress his emotions during the stressful situation.
22. Reaching a hundred years of existence is a major milestone for any organization.
23. The stock prices plummeted after the company reported losses.
24. Having different perspectives can enhance problem-solving in a team.
25. The event was available to watch via a livestream for those unable to attend.
26. The observer noted the various phases of the celestial event through his telescope.
27. The team ensured that they had adequate resources to complete the project on time.
28. Reducing carbon emissions is crucial for combating climate change.
29. Many items were discarded when the company decided to downsize.



30. The satellite was placed in geostationary orbit around the Earth.

## Answer

**Multiple Choice:** 1. Psychology 2. Mental 3. Heavenly 4. Eclipse 5. Meteorologists 6. Revenue 7. Totality 8. Landfill 9. Optimising 10. Emissions

**Gap-Fill:** 11. twilight 12. boosters 13. discarded 14. horizon 15. gradual 16. conjunction 17. orbit 18. amines 19. polystyrene 20. limitations

**Matching sentence:** 1. suppress 2. milestone 3. plummeted 4. perspectives 5. livestream 6. observer 7. adequate 8. emissions 9. discarded 10. orbit

## CATEGORY

1. Sci/Tech - LEVEL4

## POST TAG

1. B2
2. ESL learning
3. esl news
4. Even GPT-5
5. Human Attention Test
6. Level 4

## Tags

1. B2
2. ESL learning
3. esl news
4. Even GPT-5
5. Human Attention Test
6. Level 4

## Date Created

2026/06/15

## Author

aimeeyoung99