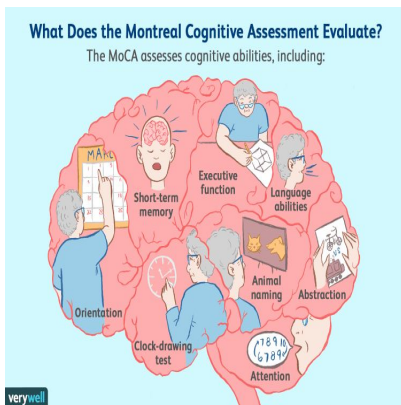




Slaagt AI voor de Cognitive Assessment test?

Beschrijving

Om dat de bestuderen, onderzochten wetenschappers de cognitieve vaardigheden van de toonaangevende, openbaar beschikbare AI Chatbots:

- *ChatGPT* versies 4 en 4o (ontwikkeld door *OpenAI*),
- *Claude 3.5* *Sonnet* (ontwikkeld door *Anthropic*) en
- *Gemini* versies 1 en 1.5 (ontwikkeld door *Alphabet*).

MoCA test

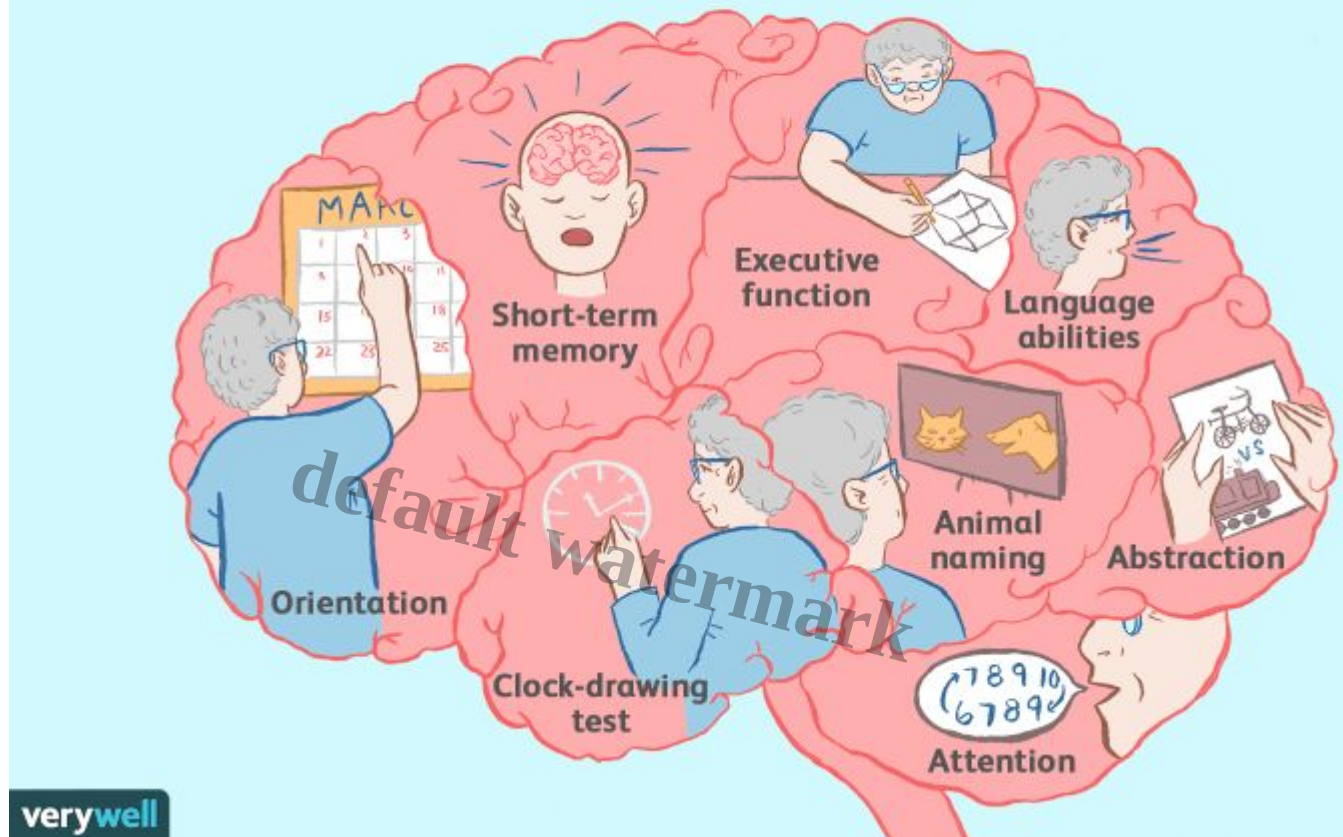
Om de cognitieve vaardigheden te testen, maakte het team gebruik van de zogenoemde *Montreal Cognitive Assessment (MoCA)*-test. Deze test wordt vaak gebruikt om cognitieve achteruitgang en vroege tekenen van *dementie* op te sporen, vooral bij oudere volwassenen. Met een reeks korte taken en vragen worden vaardigheden zoals aandacht, geheugen, taal, visuospatiele vaardigheden en uitvoerende functies beoordeeld. De maximale score is 30 punten, waarbij 26 of meer meestal als normaal wordt beschouwd.

Neurologische beoordeling

De instructies voor elke taak die aan de chatbots werden gegeven, waren hetzelfde als die men aan menselijke patiënten voorlegt. Ook de behaalde scores werden volgens de officiële richtlijnen vastgesteld en vervolgens beoordeeld door een neuroloog.

What Does the Montreal Cognitive Assessment Evaluate?

The MoCA assesses cognitive abilities, including:



Slechte scores

Opvallend genoeg bleken de chatbots niet zo goed te scoren. ChatGPT 4o behaalde de hoogste score op de MoCA-test, met 26 van de 30 punten. ChatGPT 4 en Claude volgden met 25 van de 30 punten. Gemini 1.0 scoorde het laagst, met slechts 16 van de 30 punten.

Taken

Taken zoals benoemen, aandacht, taal en abstractie werden door alle chatbots goed uitgevoerd. Maar op het gebied van *visuospatiale vaardigheden* en uitvoerende taken viel hun prestatie tegen. Zo hadden ze moeite met taken waarbij omcirkelde cijfers en letters in oplopende volgorde moesten worden verbonden. Ook presteerden ze slecht bij de taak waarbij ze een klok met een specifieke tijd moesten tekenen.

De Gemini-modellen faalden tevens bij de taak waarbij ze een reeks van vijf woorden moesten onthouden. Bij verdere visuospatiale tests slaagden de chatbots er ook niet in empathie te tonen of complexe visuele scènes nauwkeurig te interpreteren. Alleen ChatGPT 4o slaagde in de Stroop-taak, waarbij combinaties van kleurnamen en letterkleuren worden gebruikt om te meten hoe interferentie de reactietijd beïnvloedt.

Vroege dementie

Het betekent dat bijna alle toonaangevende chatbots tekenen van milde cognitieve achteruitgang vertonen in de test die vaak wordt gebruikt om vroege symptomen van dementie op te sporen. De resultaten tonen ook aan dat â??oudereâ?? versies van chatbots, net als oudere patiënten, slechter presteren bij de test.

De onderzoekers geven wel aan dat het onderzoek als een grap was bedoeld, omdat met name uitvoerende taken natuurlijk lastig te beoordelen zijn als het Chatbots betreft.

Probeer de test zelf: [MoCA test](#).

Datum aangemaakt

2024/12/31

default watermark