

# GRENZEN VON CHATGPT

WARUM CHATGPT BEI DER BEARBEITUNG VON AUFGABEN EINES LOGISCH-ANALYTISCHEN TESTS NICHTS NÜTZT, SONDERN IM GEGENTEIL ZU SCHLECHTEREN ERGEBNISSEN (VER)FÜHRT.

von Lise Sustmann Allen  
(Leiterin der Psychologischen Abteilung bei Master International A/S, Kopenhagen, Dänemark)  
Juni 2023

## EINLEITUNG

Benutzer und Testteilnehmer von Master-Tests waren natürlich neugierig auf die Auswirkungen von ChatGPT, das den Testteilnehmern bei der Beantwortung von Tests im Allgemeinen, aber insbesondere bei kognitiven Tests hilft. In den letzten Monaten haben wir verschiedene Fragen zu ChatGPT und der Beantwortung von Aufgaben in unseren Logiktests wie ACE (Adjustable Competence Evaluation) erhalten. Auch wir bei Master waren der Meinung, dass wir diese soziologische Tendenz, die wir beobachten, verstehen müssen, und deshalb haben wir uns vorgenommen, den aktuellen Stand der künstlichen Intelligenz (KI) im Vergleich zu unseren Tests zu untersuchen, wobei wir vor allem das zugängliche Tool ChatGPT im Vergleich zu ACE betrachtet haben.

Das Ziel unserer Untersuchung und somit dieses Papiers ist es, die Grenzen und Möglichkeiten von ChatGPT bei der Testentwicklung zu beleuchten, die Neugierde des Tools zu wecken und

hoffentlich auch einige der Fragen zu beantworten, die Testteilnehmer und Nutzer unserer Lösungen zu ChatGPT und ACE haben könnten.

In den letzten Jahren hat KI bemerkenswerte Fortschritte bei der Verarbeitung natürlicher Sprache gemacht, so dass KI-Modelle wie ChatGPT in der Lage sind, menschenähnliche Unterhaltungen zu führen. Obwohl ChatGPT eine beeindruckende Fähigkeit besitzt, kohärente Antworten zu generieren, ist es wichtig zu erkennen, dass es inhärente Grenzen für seine logischen Argumentationsfähigkeiten gibt. Dieser Artikel geht der Frage nach, warum ChatGPT trotz seiner bemerkenswerten Sprachkenntnisse Schwierigkeiten haben kann, logische Aufgaben in einem Test zu beantworten.

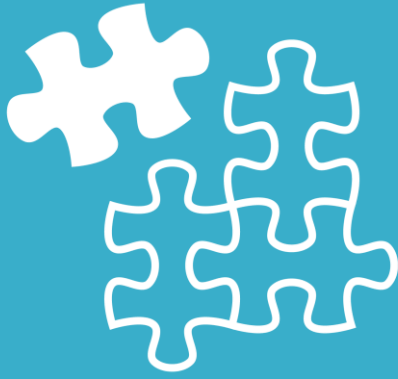
## EIN LINGUISTISCHES MODELL

Das erste Element, das man verstehen muss, ist, dass ChatGPT (und alle Large Language Models - LLM) ein linguistisches Modell ist, was bedeutet, dass ChatGPT im Grunde eine **große Datenbank** mit schriftlichen Informationen ist und auf der Grundlage dieser Textdaten selbst gelernt hat, Sätze aus den riesigen Mengen an Textdaten in der Datenbank zu konstruieren. ChatGPT verwendet dazu **statistische Modelle**, die Sätze konstruieren, indem sie versuchen, das wahrscheinlichste nächste Wort vorherzusagen. Es ist zwar hervorragend im Verstehen und Erzeugen von menschenähnlicher Sprache, verfügt aber nicht über ein echtes Verständnis oder die Fähigkeit, tief zu denken. Dem Modell **fehlen Erfahrungen aus der realen Welt, gesunder Menschenverstand und kontextuelles Verständnis**, die für das Verstehen komplexer logischer Szenarien entscheidend sind. Infolgedessen kann ChatGPT Schwierigkeiten mit differenzierten logischen Fragen haben, die abstraktes Denken und kritische Überlegungen erfordern.

ChatGPT verarbeitet Texteingaben sequentiell und hat kein umfassendes Verständnis für den breiteren Kontext. Wenn es mit logischen Fragen konfrontiert wird, kann es sich nicht auf ein tiefes Verständnis des gesamten Textes oder früherer Aussagen verlassen. Folglich kann das Modell Fragen isoliert interpretieren und dabei wichtige Informationen, die zuvor erwähnt wurden, außer Acht lassen oder sich auf irrelevante Informationen für die Lösung einer bestimmten Aufgabe konzentrieren. Diese Einschränkung beeinträchtigt seine Fähigkeit, genaue logische Schlussfolgerungen zu ziehen oder Ableitungen auf der Grundlage von kontextuellen Hinweisen zu treffen.



Für Master steht die kontinuierliche Weiterentwicklung der Instrumente im Vordergrund und wir liefern digitale Lösungen für Ihre täglichen Herausforderungen.



# GRENZEN VON CHATGPT

WARUM CHATGPT BEI DER BEARBEITUNG VON AUFGABEN EINES LOGISCH-ANALYTISCHEN TESTS NICHTS NÜTZT, SONDERN IM GEGENTEIL ZU SCHLECHTEREN ERGEBNISSEN (VER)FÜHRT.

von Lise Sustmann Allen  
(Leiterin der Psychologischen Abteilung bei Master International A/S, Kopenhagen, Dänemark)  
Juni 2023

## EMPFINDLICHKEIT GEGENÜBER EINGABEFORMULIERUNGEN

Induktives Denken, also die Fähigkeit, von spezifischen Beispielen auf allgemeinere Prinzipien zu schließen, ist ein grundlegender Aspekt des logischen Denkens. ChatGPT kann zwar Antworten auf der Grundlage vorhandener Muster in den Daten generieren, besitzt aber nicht die Fähigkeit, allgemeine Prinzipien zu induzieren oder Lösungen auf der Grundlage begrenzter Informationen abzuleiten. Diese Einschränkung hindert ChatGPT daran, komplexe logische Fragen zu beantworten, die induktives Denken erfordern, und schränkt damit seine Leistung bei solchen Testaufgaben ein.

ChatGPT reagiert sehr empfindlich auf die Formulierung und Struktur der Eingabefragen, und selbst eine geringfügige Umformulierung derselben Frage kann zu unterschiedlichen Antworten führen, was die mangelnde Robustheit des linguistischen Modells bei der Erfassung der zugrunde liegenden Logik verdeutlicht.

Im Gegensatz zu menschlichen Testteilnehmern, die die Absicht hinter einer Frage entschlüsseln können, verlässt sich ChatGPT auf Muster und statistische Assoziationen in den Fragen, die ihm vorgelegt werden. Folglich kann ChatGPT Schwierigkeiten haben, logische Konzepte über verschiedene Formulierungen hinweg zu verallgemeinern, was zu inkonsistenten oder falschen Antworten führt.



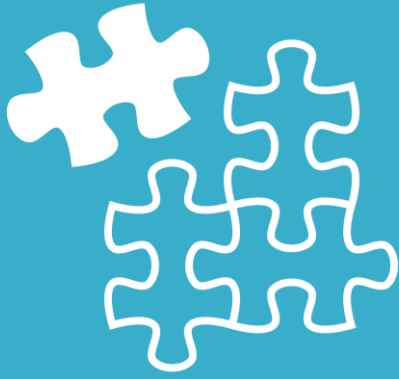
## VON MASTER UMGESETZTE INITIATIVEN

ChatGPT mit seinen beeindruckenden Sprachgenerierungsfähigkeiten versagt, wenn es um logische Aufgaben in einem Test geht. Seine Einschränkungen in Bezug auf tiefes Verständnis, kontextuelles Bewusstsein, Anfälligkeit für Mehrdeutigkeit, Empfindlichkeit gegenüber Eingabeformulierungen und fehlendes induktives Denken behindern die Fähigkeit, genaue und konsistente Antworten auf logische Fragen zu geben. Das Erkennen dieser Einschränkungen ist wichtig, um anzuerkennen, dass menschliche Testteilnehmer immer noch einen Vorteil im logischen Denken und im kritischen Denken haben, und dass die Verwendung von Tools wie ChatGPT die

Chancen eines Testteilnehmers in einem kognitiven Test sogar verschlechtert.

Auch andere Marktteilnehmer wie Microsoft und Google haben "Hilfsfunktionen" implementiert, die sich letztendlich negativ auf das Ergebnis des Testteilnehmers auswirken können. Daher haben wir Initiativen ergriffen, um die Motivation der Testteilnehmer zur Nutzung von ChatGPT und anderen Tools zu minimieren, da sie dadurch im Vergleich zu anderen einen Nachteil haben, auch wenn sich die Testteilnehmer dessen nicht bewusst sein mögen.

Für Master steht die kontinuierliche Weiterentwicklung der Instrumente im Vordergrund und wir liefern digitale Lösungen für Ihre täglichen Herausforderungen.



# GRENZEN VON CHATGPT

WARUM CHATGPT BEI DER BEARBEITUNG VON AUFGABEN EINES LOGISCH-ANALYTISCHEN TESTS NICHTS NÜTZT, SONDERN IM GEGENTEIL ZU SCHLECHTEREN ERGEBNISSEN (VER)FÜHRT.

von Lise Sustmann Allen  
(Leiterin der Psychologischen Abteilung bei Master International A/S, Kopenhagen, Dänemark)  
Juni 2023

## Einschränken der visuellen Microsoft-Suche

Wenn der Testteilnehmer Edge als Browser verwendet, sieht er unter normalen Umständen ein kleines Symbol auf allen Bildern auf einer beliebigen Webseite. Wenn er auf das Symbol drückt, wird das Web nach ähnlichen Bildern durchsucht. Das bedeutet, dass Testteilnehmer, die

und dass keine unmittelbare Bedrohung besteht, da die Bilder aus der Suche zwar ähnlich sind, aber nichts mit ACE oder CORE zu tun haben. Es besteht also eher die Sorge, dass der Testteilnehmer beim Ausfüllen des Tests abgelenkt wird. Deshalb haben wir auf unseren Testseiten eine Kodierung hinzugefügt, die verhindert, dass das Microsoft Visual Search Symbol auf unseren Seiten

zu durchsuchen. Eine solche Begrenzung der Zeitverschwendung ist für den Testteilnehmer von Vorteil.

Zweitens wird das Kopieren und Einfügen von Text eingeschränkt, um sicherzustellen, dass es für einen Testteilnehmer schwieriger ist, an zwei Bildschirmen zu sitzen und den Text von ACE zu kopieren und in ChatGPT einzufügen. Dies schränkt das Risiko der Nutzung von ChatGPT nicht vollständig ein, kann es aber hoffentlich erschweren, was dazu führt, dass der Testteilnehmer von der Nutzung Abstand nimmt. Was letztlich auch zu dessen eigenem Vorteil wäre.

## FAZIT

**Zusammenfassend lässt sich sagen, dass ChatGPT zwar beeindruckende Fähigkeiten zur Spracherzeugung aufweist, aber bei der Beantwortung logischer Aufgaben in einem Test erhebliche Einschränkungen aufweist.**

Als linguistisches Modell mangelt es ihm an echtem Verständnis, logischem Denken und kontextuellem Verstehen, die für die genaue Beantwortung komplexer logischer Szenarien unerlässlich sind. Die Anfälligkeit des Modells für Mehrdeutigkeit und seine Unfähigkeit, nach Klärungen zu suchen oder Folgefragen zu stellen, beeinträchtigen seine Leistung bei differenzierten logischen Fragen zusätzlich. Darüber hinaus verfügt ChatGPT nicht über die Fähigkeit zum induktiven Schlussfolgern, was es für das Modell schwierig macht, logische Konzepte zu verallgemeinern oder Lösungen auf der Grundlage begrenzter Informationen abzuleiten.

Darüber hinaus reagiert ChatGPT sehr empfindlich auf die Formulierung und die Struktur von Eingabefragen, was zu



Edge zum Ausfüllen von ACE und/oder CORE verwenden, möglicherweise das Internet nach ähnlichen Bildern durchsuchen könnten. Außerdem könnten sie sich durch das Symbol ablenken lassen, was in einigen Fällen die Antworten und damit das Testergebnis beeinflussen kann.

Unsere Nachforschungen haben ergeben, dass es im Internet viele Diskussionen zu diesem Thema gibt, sogar kurz nachdem Microsoft diese Funktion veröffentlicht hat. Wir sind zu dem Schluss gekommen, dass die Auswirkungen auf die Testteilnehmer bei der Verwendung von Microsoft Visual Search in der jetzigen Form begrenzt sind

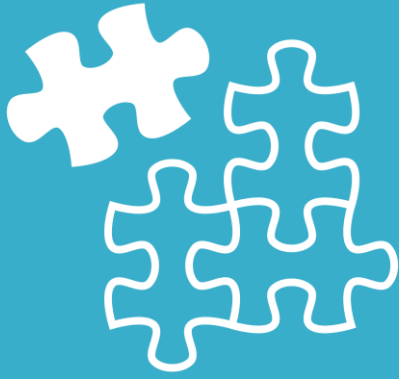
angezeigt wird.

## Einschränken des Rechtsklicks

Die Beseitigung der Möglichkeit, während der Beantwortung eines Tests mit der rechten Maustaste zu klicken, wirkt sich vor allem auf zwei Aktionen aus, die wir als mögliche Hauptstörfaktoren für den Testteilnehmer identifiziert haben:

Erstens ist es weniger wahrscheinlich, dass Bilder z. B. in eine Google-Suche oder ähnliches kopiert werden. Das bedeutet, dass der Testteilnehmer keine wertvolle Zeit damit verbringt, mögliche ähnliche Bilder auf der Suche nach Hilfe für den Test

Für Master steht die kontinuierliche Weiterentwicklung der Instrumente im Vordergrund und wir liefern digitale Lösungen für Ihre täglichen Herausforderungen.



# GRENZEN VON CHATGPT

WARUM CHATGPT BEI DER BEARBEITUNG VON AUFGABEN EINES LOGISCH-ANALYTISCHEN TESTS NICHTS NÜTZT, SONDERN IM GEGENTEIL ZU SCHLECHTEREN ERGEBNISSEN (VER)FÜHRT.

von Lise Sustmann Allen  
(Leiterin der Psychologischen Abteilung bei Master International A/S, Kopenhagen, Dänemark)  
Juni 2023

inkonsistenten oder falschen Antworten führt, selbst wenn die Fragen leicht umformuliert werden. Diese Empfindlichkeit gegenüber der Formulierung der Eingabe unterstreicht die mangelnde Robustheit des Modells bei der Erfassung der zugrunde liegenden Logik der Fragen. Darüber hinaus verhindert die Unfähigkeit des Modells, visuelle Informationen zu verarbeiten, dass es räumliche Aufgaben oder Fragen, die ein visuelles Verständnis erfordern, effektiv beantworten kann.

Master ist sich dieser Einschränkungen bewusst und hat Initiativen ergriffen, um die Motivation der Testteilnehmer zu minimieren, sich während der Tests auf ChatGPT zu verlassen. Maßnahmen wie die

Einschränkung der Möglichkeit, Text zu kopieren und einzufügen und das Entfernen des Microsoft Visual Search Symbols wurden implementiert, um es den Testteilnehmern zu erschweren, auf externe Ressourcen zuzugreifen und ihr Testergebnis möglicherweise zu beeinflussen.

Es ist erwähnenswert, dass zukünftige ChatGPT-Versionen (und andere ähnliche Modelle) auf verschiedene Lösungen hinarbeiten, wie z. B. Plugins und spezifischeres Training für verschiedene Nischen, um die mathematische und logische Leistung der Modelle zu verbessern. Gleichzeitig kann erwähnt werden, dass das Problem, dass das Modell Fakten erfindet (oder "halluziniert", wie

manche es nennen), weiterhin ein Problem darstellt und die KI-Forscher nicht wirklich verstehen, warum. Master verfolgt diese Entwicklung aufmerksam.

**Letztendlich ist es wichtig zu erkennen, dass menschliche Testteilnehmer immer noch einen Vorteil im logischen Denken und im kritischen Denken haben.**

ChatGPT kann zwar ein wertvolles Werkzeug für verschiedene Aufgaben sein, aber bei der Komplexität von Logiktests versagt es. Die Grenzen von ChatGPT zu verstehen, ist sowohl für die Testnutzer als auch für die Testteilnehmer wichtig, um eine faire und genaue Bewertung der logischen Fähigkeiten zu gewährleisten.

Übersetzt von Mag. Bernhard Dworak, (Geschäftsführer, Master Human Resources Consulting GmbH)

Für Master steht die kontinuierliche Weiterentwicklung der Instrumente im Vordergrund und wir liefern digitale Lösungen für Ihre täglichen Herausforderungen.