

KI in der Lehre auf dem Gebiet der Mathematik



PROJEKTTEAM:
WIEBKE HOPPE, ABDELGHAFOUR
KAOUCH



BETREUER:
PROF. DR. STRASS, PROF. DR.
LIEBSCHER

Gliederung



Einführung



Methodik



Ergebnisse



Fazit und Ausblick



Erfahrungen und Probleme

Einführung

Motivation

- Studierende/Lehrende nutzen intensiv generative KI
- Viele Fragen sind aber offen:
 - Welche ChatBots sind empfehlenswert?
 - Was sind vorteilhafte Vorgehensweisen?
Wie ist der Prompt zu gestalten?
 - Kann man den Antworten der ChatBots trauen?
Welcher Anteil der Antworten ist korrekt?

Einführung

Projektziele

- praktische Einsatzmöglichkeiten von textgenerierenden KI-Tools erproben und diskutieren
- Ergebnisse sollen allen Studierenden der Hochschule Merseburg zur Verfügung gestellt werden
- Studierende sollen textgenerierende KI-Tools in ihrem Studium und ihrer Forschung effektiv einsetzen

Einführung

Einsatzgebiete in der Lehre

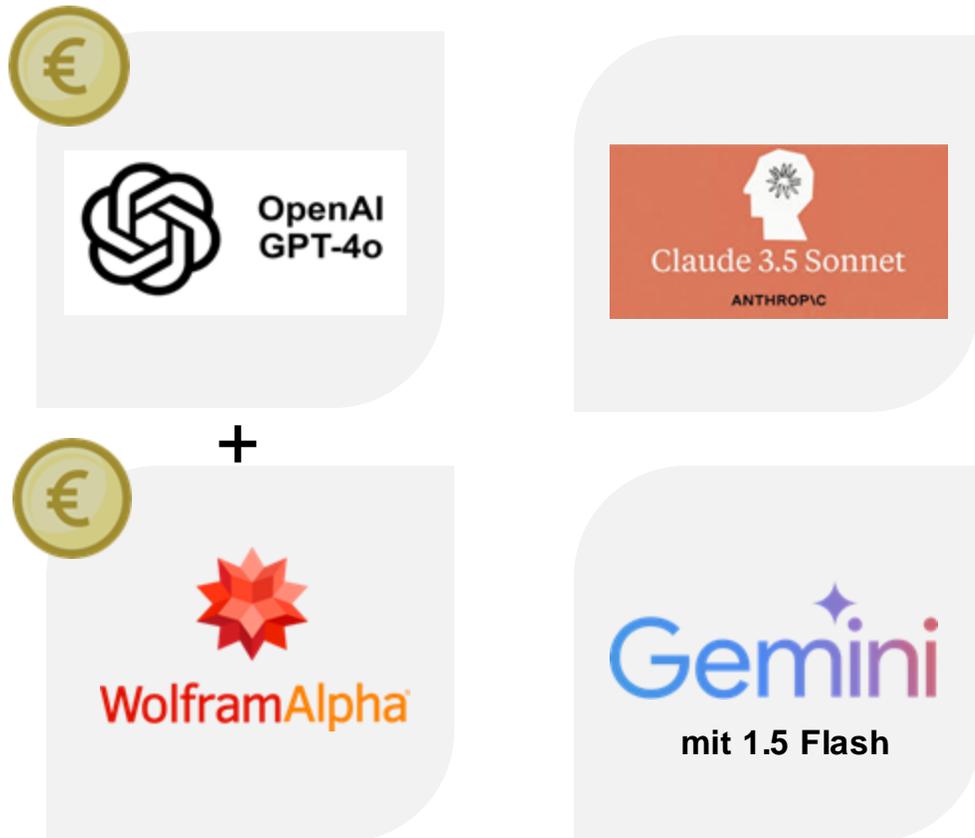
Adaptive Lernplattformen

- Generieren von Übungsaufgaben anhand des Vorlesungsskriptes
- Automatische und direkte Bewertung
- personalisiertes Feedback
- am Lernfortschritt angepasste individuelle Lehrpläne

„Nachhilfelehrer“, Prüfungsvorbereitung

Methodik

Auswahl der Chatbots und Plug-Ins



Forschungszeitraum: Juli - August 2024

Anforderung	ChatGPT 4o	ChatGPT 4o mit Wolfram-Alpha	Gemini 1.5 Flash	Claude Sonnet 3.5 Sonnet	COPILOT	Llama-3 400B+
automatisiertes Plugin	✓	k.A.	✓	✓	✓	✓
Bild-Prompt	✓	✓	✓	✓	✓	✓
manueller Plugin	✓	✓	x	x	x	x
MATH Benchmark	76.6%	k.A.	67.7%	71.1%	k.A.	57.8%
Mathematik Verständnis	✓	✓	✓	teilweise	teilweise	✓
Text-Prompt	✓	✓	✓	✓	✓	✓
vollständige Lösung	✓	✓	✓	teilweise	x	✓

Methodik

Auswahl der Chatbots und Plug-Ins

Gab es während des Projekts Updates für die verwendeten Chatbots?

- **ChatGPT:**
 - GPT-4o with canvas :programmieren
 - o1-preview:fortgeschrittenes Reasoning
 - o1-mini:Schnelleres Reasoning
- **Claude:**
 - claude 3.5 sonnet(new)
 - claude 3.5 Haiku
 - use Computer

Methodik

Aufbau eines Prompts

Vor-Prompt

- Gibt dem Chatbot eine klare Rolle und definiert die Art der Antwort.

Prompt der Aufgabenstellung

- Screenshot
- LaTeX

1. Vor-Prompt	"Sei ein Experte für Mathematikaufgaben und <u>hilf</u> gerne bei der Lösung von Mathematikaufgaben. Erkläre die Lösungen sehr klar und verständlich unter Beachtung der mathematischen Regeln. Erstelle eine Schritt-für-Schritt-Lösung."
2. Prompt der Aufgabenstellung (als Screenshot oder Latex-Code)	<p>Screenshot:</p>  <p>Gegeben ist der Vektor:</p> $\vec{a} = \begin{pmatrix} 2 \\ -4 \\ 4 \end{pmatrix}$ <p>Bestimmen Sie</p> <ol style="list-style-type: none">die Länge dieses Vektors,den normierten Vektor unddie Winkel zu den Koordinatenachsen (in Grad- und Bogenmaß). <p><i>Abbildung 1 Ilias-aufgabe von Herren Liebscher [16]</i></p>
	<p><u>Latex-Code:</u> Gegeben ist der Vektor: <code>\[\vec{a} = \left(\begin{array}{c} 2 \\ -4 \\ 4 \end{array} \right) \]</code> Bestimmen Sie a) die Länge dieses Vektors, b) den normierten Vektor und c) die Winkel zu den Koordinatenachsen (in Grad- und Bogenmaß)."</p>

Methodik

Aufgabenauswahl und Clusterung

- Aufgaben aus dem ILIAS-Aufgabenpool „**Mathematik für Betriebswirte und Ingenieure**“
- Clusterung der Aufgaben in **drei Schwierigkeitsgrade**
 - normal schwer
 - mittelschwer
 - schwer
- Klassifikation basiert auf den **durchschnittlichen Erfolgsquoten** und der **Anzahl der erforderlichen Versuche** der Studierenden

Methodik

Bewertung der Chatbot-Ergebnisse: Inhalt

Bewertungskriterien:

1. Das angewandte mathematische Konzept
2. Einsetzen der Zahlenwerte in die Formel
3. Korrektheit aller Zwischenergebnisse
4. Das Endergebnis

Punktvergabe:

- Je Kriterium 1 Punkt erreichbar
- Anteiliger Punktabzug bei fehlerhaften Antworten

Methodik

Bewertung der Chatbot-Ergebnisse: Form

Bewertungskriterien:

1. Struktur der Ausgabe
2. Einsetzen der Zahlenwerte in die Formel

Punktvergabe:

- je Kriterium 1 Punkt erreichbar
- anteiliger Punktabzug bei fehlerhaften Antworten

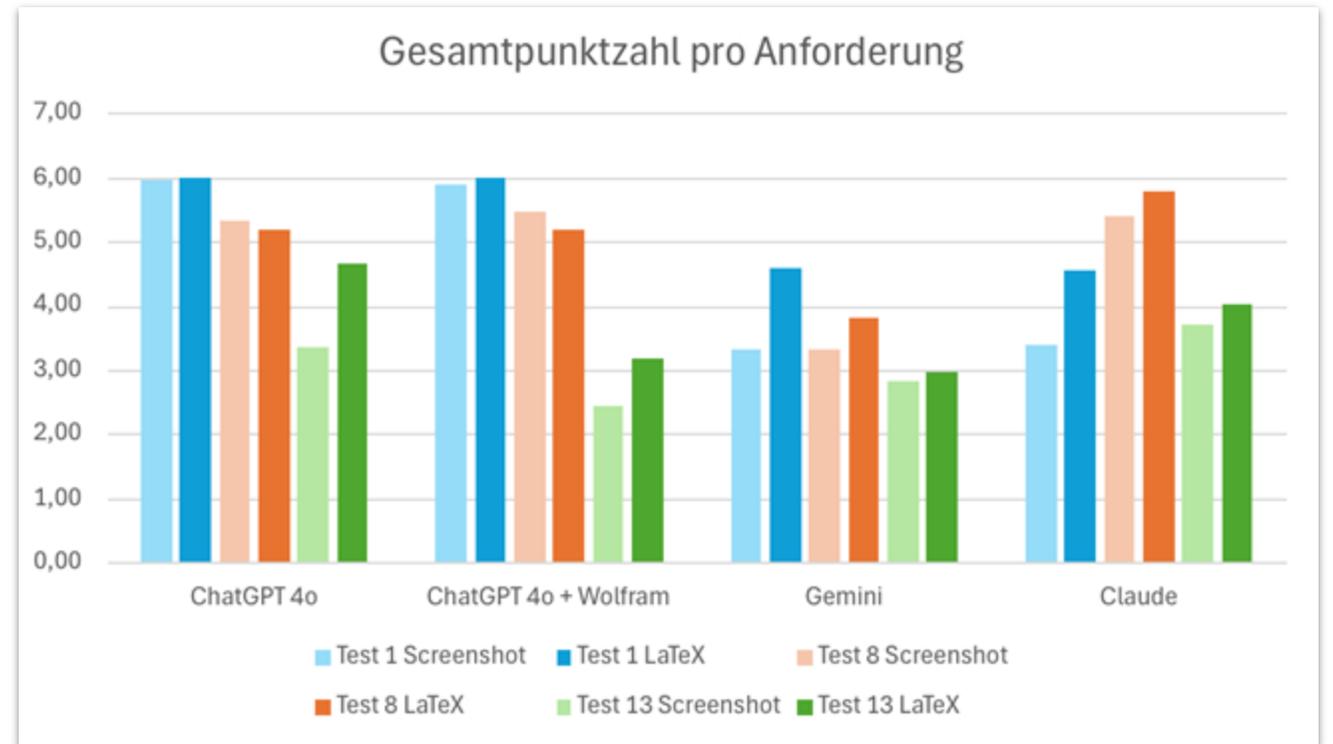
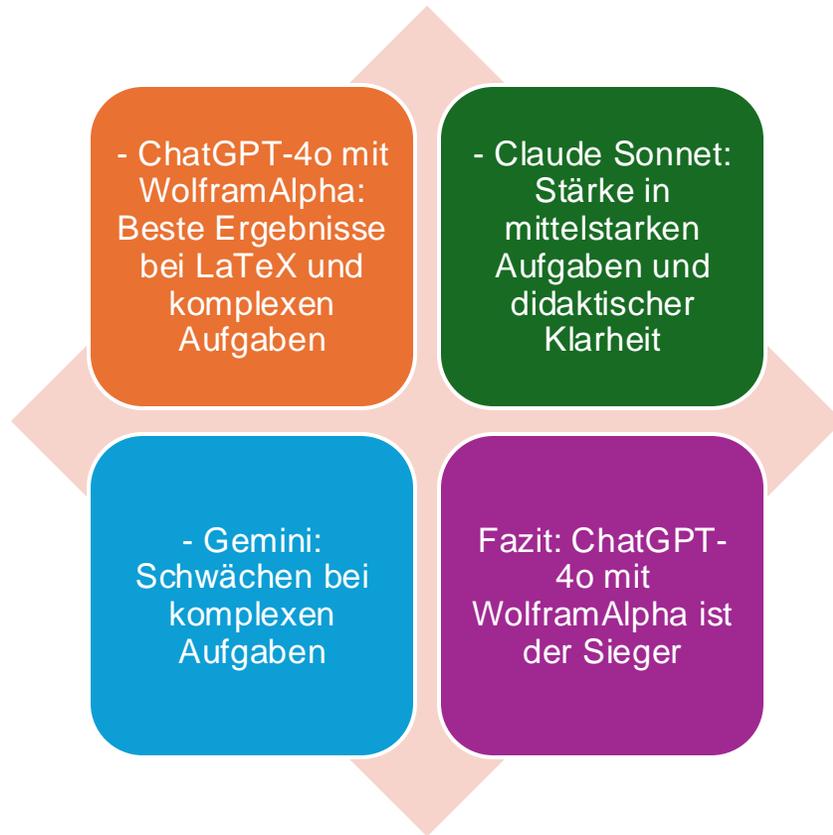
Methodik

Bewertung der Chatbot-Ergebnisse

Test 2		Bewertungsschema für Ilias-Aufgabentests			
Vor-Prompt		"Sei ein Experte für Mathematikaufgaben und hilf gerne bei der Lösung von Mathematikaufgaben. Erkläre die Lösungen s			
Testnummer 1		Kurvendiskussion mit Logarithmus			
Chatbot	GPT-4o	GPT-4o + Wolfram Alpha	Gemini pro 1.5	Claude Sonnet 3.5	
Mathematische Konzept		0,9	0,36	0,36	0,36
Zahlenwerte		1	1	1	1
Zwischenergebnis		0,36	1	0,09	1
Endergebnis		0,18	0,36	0,09	0,36
Klarheit und Verständlichkeit		0,5	1	1	1
Struktur der Ausgabe		1	1	1	1
Summe		3,94	4,72	3,54	4,72
Testnummer 2		Kurvendiskussion mit rationaler Funktion			
Chatbot	GPT-4o	GPT-4o + Wolfram Alpha	Gemini pro 1.5	Claude Sonnet 3.5	
Mathematische Konzept		1	0,82	0,09	1
Zahlenwerte		1	1	0	1
Zwischenergebnis		1	0,82	0,18	0,73
Endergebnis		1	1	0,18	0,73
Klarheit und Verständlichkeit		1	1	0	1
Struktur der Ausgabe		1	1	0	1
Summe		6	5,64	0,45	5,46
Testnummer 3		Grenzwerte			
Chatbot	GPT-4o	GPT-4o + Wolfram Alpha	Gemini pro 1.5	Claude Sonnet 3.5	
Mathematische Konzept		1	1	1	1
Zahlenwerte		1	1	1	1
Zwischenergebnis		1	1	1	1
Endergebnis		1	1	1	1
Klarheit und Verständlichkeit		1	1	1	1
Struktur der Ausgabe		1	1	1	1
Summe		6	6	6	6
Durchschnittsbewertung		5,31	5,45	3,33	5,39
	GPT-4o	GPT-4o + Wolfram Alpha	Gemini pro 1.5	Claude Sonnet 3.5	
⊙ mathematisches Konzept	0,97	0,73	0,48	0,79	
⊙ Zahlenwert	1,00	1,00	0,67	1,00	
⊙ Zwischenergebnis	0,79	0,94	0,42	0,91	
⊙ Endergebnis	0,73	0,79	0,42	0,70	

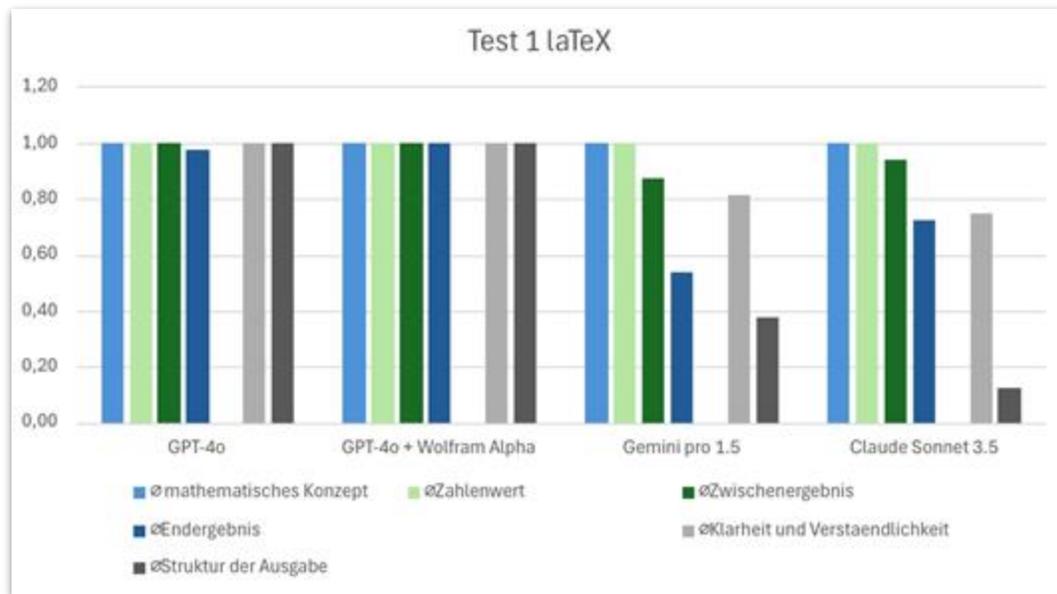
Bewertungsmatrix Test 8 Screenshot

Auswertung: Welcher Chatbot ist am besten?

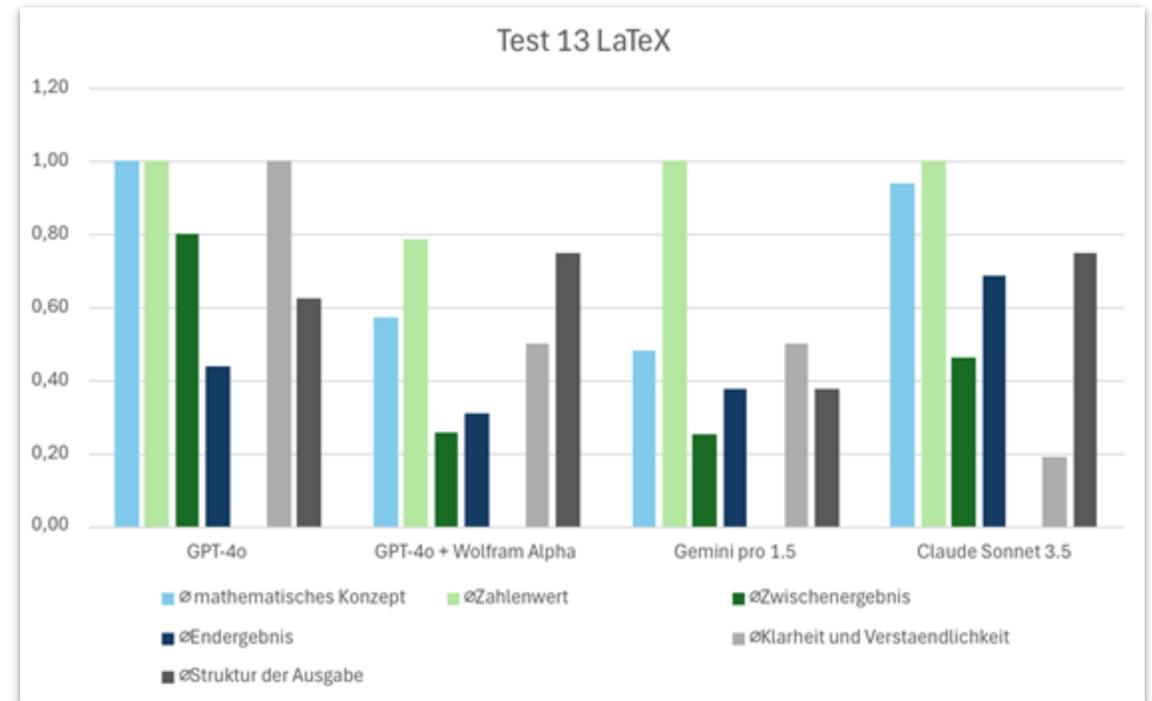


Gesamtpunktzahl pro Anforderung -
Durchschnittliche Bewertung der Anforderungen pro Chatbot

Auswertung: Welcher Chatbot ist am besten?



Test 1 LaTeX "normal schwer"



Test 13 LaTeX "schwer"

Auswertung: Screenshot vs. LaTeX

Werden Lösungswege skizziert?

Test 1 "normal schwer"

- Lösungswege mit Berechnungen werden automatisch ausgegeben

Test 8 "mittel schwer"

- Berechnung mit Zahlenwerten nur mit LaTeX Prompt
- Ausgabe des mathematischen Konzepts

Test 13 "schwer"

- Ausgabe des mathematischen Konzepts
- Berechnung mit Zahlenwerten nur mit LaTeX Prompt

W "Sei ein Experte für Mathematikaufgaben und hilf gerne bei der Lösung von Mathematikaufgaben. Erkläre die Lösungen sehr klar und verständlich unter Beachtung der mathematischen Regeln. Erstelle eine Schritt-für-Schritt-Lösung." Gegeben ist die Funktion $f(x) = -\frac{9}{x} + 10 - x$. Geben Sie den Definitionsbereich, die Nullstellen und die Extremstellen der Funktion $f(x)$ an! Für welche x ist die Funktion monoton wachsend und für welche monoton fallend?

Ich helfe Ihnen gerne bei dieser Aufgabe. Lassen Sie uns dies Schritt für Schritt lösen.

1. Definitionsbereich:
 - Die Funktion enthält einen Bruch mit x im Nenner
 - x darf nicht 0 sein (Division durch 0)
 - Also: $D = \mathbb{R} \setminus \{0\}$
2. Nullstellen:
 - $f(x) = 0$ setzen: $-9/x + 10 - x = 0$
 - Mit x multiplizieren: $-9 + 10x - x^2 = 0$
 - Umformen: $-x^2 + 10x - 9 = 0$
 - Quadratische Formel: $x = (-10 \pm \sqrt{100-36})/(-2)$
 - $x = (-10 \pm \sqrt{64})/(-2)$
 - $x = (-10 + 8)/(-2)$
 - $x_1 = (-10 + 8)/(-2) = -1$
 - $x_2 = (-10 - 8)/(-2) = 9$
3. Extremstellen:
 - $f'(x) = 9/x^2 - 1$

Gegeben ist die Funktion $f(x) = -\frac{9}{x} + 10 - x$. Geben Sie den Definitionsbereich, die Nullstellen und die Extremstellen der Funktion $f(x)$ an! Für welche x ist die Funktion monoton wachsend und für welche monoton fallend?

W "Sei ein Experte für Mathematikaufgaben und hilf gerne bei der Lösung von Mathematikaufgaben. Erkläre die Lösungen sehr klar und verständlich unter Beachtung der mathematischen Regeln. Erstelle eine Schritt-für-Schritt-Lösung."

Ich helfe Ihnen gerne bei der Lösung dieser Aufgabe. Lassen Sie uns Schritt für Schritt vorgehen.

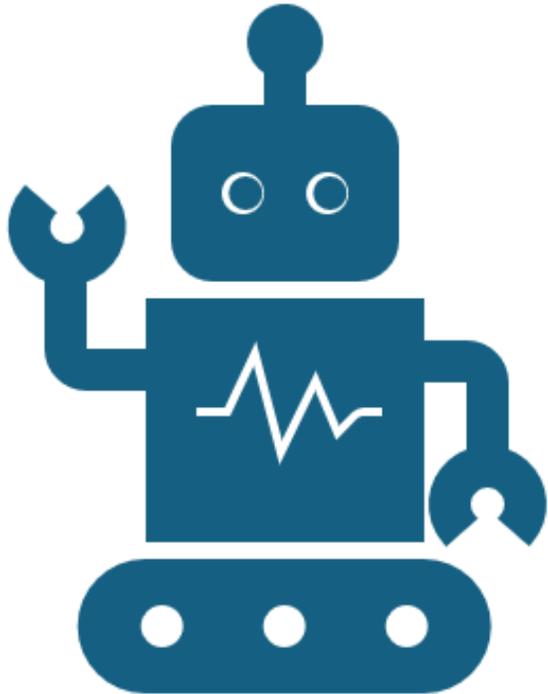
1. Definitionsbereich:
 - Da x im Nenner steht, darf $x \neq 0$ sein
 - Also ist der Definitionsbereich $D = \mathbb{R} \setminus \{0\}$
2. Nullstellen bestimmen: $f(x) = 0 \rightarrow -9/x + 10 - x = 0 \rightarrow -9 + 10x - x^2 = 0 \rightarrow -x^2 + 10x - 9 = 0 \rightarrow$ Quadratische Gleichung lösen: $x = (-10 \pm \sqrt{100-36})/(-2) = (-10 \pm \sqrt{64})/(-2) = (-10 \pm 8)/(-2)$ $x_1 = (-10 + 8)/(-2) = -1$ und $x_2 = 9$
3. Extremstellen bestimmen: $f'(x) = 9/x^2 - 1$ $f'(x) = 0 \rightarrow 9/x^2 - 1 = 0 \rightarrow 9/x^2 = 1 \rightarrow x^2 = 9 \rightarrow x = \pm 3$ Da $x = 3$ und $x = -3$ im Definitionsbereich liegen, sind dies potenzielle Extremstellen.
4. Art der Extremstellen prüfen: $f''(x) = -18/x^3$ Bei $x = 3$: $f''(3) = -18/27 < 0 \rightarrow$ Hochpunkt Bei $x = -3$: $f''(-3) = -18/(-27) > 0 \rightarrow$ Tiefpunkt
5. Monotonieverhalten: $f'(x) = 9/x^2 - 1$
 - Für $|x| > 3$ ist $f'(x) < 0$, also ist $f(x)$ monoton fallend
 - Für $0 < |x| < 3$ ist $f'(x) > 0$, also ist $f(x)$ monoton wachsend

Zusammenfassung:

- Definitionsbereich: $D = \mathbb{R} \setminus \{0\}$

Test 8 LaTeX
Claude Sonnet

Test 8 Screenshot
Claude Sonnet



Auswertung: War der Bot
besser als ein Student?

Motivation

Zeit

Anzahl
der Tests

Fazit zu den Forschungsergebnissen

Ergebnisse hängen von der Form der Eingabe ab.

- Einfluss der Formulierung des Vorprompts auf die Ausgabe
- LaTeX- Prompt schneidet besser ab als Screenshot-Prompt:
 - LaTeX als Programmiersprache der Chatbots
 - Screenshot erfordert ein fortschrittliches Bildverarbeitung Verständnis

Wie erreicht man richtige Ergebnisse?

- deutliche ausformulierung des Vorprompts
- saubere Prompt-Eingabe um Anomalien vorzubeugen

Generalisierbarkeit des Experiments:

- Zeit und Umfang als Hindernis
- Updates der Chatbots verändern Ergebnisse

Fazit zu den Forschungsergebnissen

Können Studierende genAI als verlässliches und mit Mehrwert behaftetes Werkzeug verwenden?

- keine Garantie für Richtigkeit
- der Mehrwert für die Studierenden hängt von ihren individuelle Anforderungen ab
- Fokus auf dem mathematischen Konzept
- angepasster Vorprompt
- Nachvollziehbarkeit der Fehler

Erfahrungen und Probleme

- Experimentelle Durchführung der Tests
 - Aufstellung des Bewertungsschemas
 - Bewertung der Lösungsansätze
 - Auswahl der Chatbots
 - Recherche über Wirkweisen und Implementierungen der Plugins
 - Formulierung einer Methodik
-
- ChatBots arbeiten in der Regel vor allem sprachbasiert
 - Unterstützung von mathematischen „Denkweisen“ unterschiedlich



Fragen ?