

Downloadmaterial zum Beitrag „Ist ChatGPT ein Mathe-Ass?“ – MINT Zirkel 2-2024

# Wie gut schlägt sich GPT beim eigenständigen Lösen von Aufgaben?

Dieser Frage ging der Entwickler OpenAI zunächst selbst nach und ließ GPT den Mathematikteil des amerikanischen Scholastic Assessment Tests (SAT), einen Teil des amerikanischen Hochschulzugangstests, absolvieren. Dabei schnitt GPT-4 im 89. Perzentil ab, also deutlich besser als der Durchschnitt. Zum Vergleich: Harvard-Studierende befinden sich in der Regel mindestens im 90. Perzentil. Allerdings müssen die Ergebnisse relativiert werden. Zum einen ist nicht klar, inwieweit diese Tests bereits zum Trainieren von GPT verwendet wurden, zum anderen enthält der SAT vergleichsweise simple, meist rechenlastige Aufgaben.

Andere Studien kamen zu differenzierteren Ergebnissen: Die Leistung hängt vom mathematischen Niveau, vom Gebiet und von der Art der Aufgabe ab. GPT ist gut dazu in der Lage, einfache Standardaufgaben zu lösen, bekommt aber mit höherer Klassenstufe zunehmend Schwierigkeiten. Abgesehen vom Schwierigkeitsgrad spielen bei Anwendungsaufgaben der Kontext und insbesondere die Formulierung eine wichtige Rolle. Hier kommen drei Beispiele.

## Erstes Beispiel (Spreitzer et al. 2024)

Wir stellen ChatGPT einmal folgende Aufgabe und analysieren seine Lösung:

„Wenn ein Gummiball zu Boden fällt, springt er die Hälfte der Strecke wieder hoch. Der Ball wird von einem 18 Meter hohen Dach fallen gelassen. Welche gesamte Entfernung hat der Ball zurückgelegt, wenn er das dritte Mal den Boden berührt?“ (Jordan et al., 2006)

Um die gesamte Entfernung zu berechnen, die der Ball zurückgelegt hat, müssen wir alle Entfernungen addieren, die er beim Fallen und Springen zurückgelegt hat.

Beim ersten Fall: 18 Meter fallen, 9 Meter springen. Beim zweiten Fall: 9 Meter fallen, 4,5 Meter springen. Beim dritten Fall: 4,5 Meter fallen, 2,25 Meter springen.

Wir können eine **geometrische Reihe** verwenden, um die Gesamtstrecke zu berechnen. Die Formel für die Summe einer geometrischen Reihe lautet:

$$S_n = a_1 \times \frac{1-r^n}{1-r}$$

- $S_n$  ist die Summe der ersten  $n$  Terme,
- $a_1$  ist der erste Term der Reihe,
- $r$  ist das Verhältnis zwischen aufeinanderfolgenden Termen,
- $n$  ist die Anzahl der Terme.

Für unseren Fall:

- $a_1 = 18$  (die erste Höhe, von der der Ball fällt),
- $r = 1/2$  (da der Ball die Hälfte der vorherigen Höhe erreicht),
- $n = 3$  (da der Ball drei Mal fällt und springt).

Setzen wir diese Werte in die Formel ein:

$$S_3 = 18 \times \frac{1-(1/2)^3}{1-1/2}$$

$$S_3 = 18 \times \frac{1-1/8}{1/2}$$

$$S_3 = 18 \times \frac{7/8}{1/2}$$

$$S_3 = 18 \times \frac{7}{4}$$

$$S_3 = 31,5$$

Also hat der Ball insgesamt 31,5 Meter zurückgelegt, wenn er zum dritten Mal den Boden berührt.

#### Lösung von ChatGPT-3.5

Der Lösungsansatz mit der geometrischen Reihe (und das Ergebnis auch) ist falsch. Es liegt die Vermutung nahe, dass die Formulierung „... springt er die Hälfte der Strecke ...“ GPT triggerte, die geometrische Reihe einzusetzen, da das Halbieren von Strecken häufig bei Anwendungsaufgaben zur geometrischen Reihe verlangt wird. Bei den neuen Versionen von ChatGPT taucht dieses Problem allerdings nicht auf.

## Zweites Beispiel: Tankaufgabe nach Blum

„Herr Stein wohnt in Trier 20 km von der Grenze zu Luxemburg entfernt. Er fährt mit seinem VW Golf zum Tanken nach Luxemburg, wo sich direkt hinter der Grenze eine Tankstelle befindet. Dort kostet der Liter Benzin nur 0,85 Euro, im Gegensatz zu 1,1 Euro in Trier. Lohnt sich die Fahrt für Herrn Stein? Begründe deine Antwort.“



Um zu bestimmen, ob sich die Fahrt nach Luxemburg für Herrn Stein lohnt, müssen wir die Kosten für die Fahrt von Trier nach Luxemburg und zurück (insgesamt 40 km) berechnen und diese Kosten mit den Einsparungen beim Tanken vergleichen.

Zuerst berechnen wir den Verbrauch seines VW Golfs und die damit verbundenen Kosten. Nehmen wir an, dass der VW Golf durchschnittlich 5 Liter pro 100 km verbraucht.

Die Fahrtkosten berechnen sich wie folgt:

- Gesamtstrecke: 40 km (20 km hin und 20 km zurück)
- Benzinverbrauch: 5 Liter/100 km
- Kosten für Benzin in Trier: 1,1 Euro pro Liter

Die Einsparungen beim Tanken in Luxemburg vergleichen wir nun mit den Kosten in Trier. Die Einsparung pro Liter beträgt 0,85 Euro in Luxemburg im Vergleich zu 1,1 Euro in Trier, also eine Ersparnis von 0,25 Euro pro Liter.

Um zu entscheiden, ob sich die Fahrt lohnt, müssen die Einsparungen beim Tanken die Fahrtkosten von 2,20 Euro übersteigen. Wir berechnen nun, wie viel Benzin Herr Stein tanken muss, um mindestens 2,20 Euro zu sparen.

Herr Stein muss mindestens etwa 8,8 Liter Benzin in Luxemburg tanken, um die Fahrtkosten von 2,20 Euro zu decken. Jede darüber hinaus getankte Menge führt zu einer Ersparnis im Vergleich zu den Preisen in Trier.

Dies bedeutet, dass sich die Fahrt für Herrn Stein lohnt, sofern er mehr als 8,8 Liter Benzin tankt. Bei einem typischen Tankvorgang, der deutlich mehr als 8,8 Liter umfasst, würde Herr Stein somit sparen. Die tatsächliche Ersparnis hängt von der Menge des getankten Benzins ab.

#### Lösung der Tankaufgabe von ChatGPT-4

Die Lösung ist rein mathematisch betrachtet korrekt, fehlende Annahmen wurden korrekt identifiziert. Allerdings handelt es sich hier um eine rein mathematische Lösung, andere Faktoren wie ökologische oder zeitliche Faktoren fließen in die Argumentation auch auf Nachfrage nicht ein.

## Drittes Beispiel: Staufaufgabe nach Maaß

„Zu Beginn der Sommerferien kommt es häufig zu Verkehrsstaus. Christina steht schon seit 6 Stunden in einem 20 km langen Stau. Es ist sehr warm und sie ist sehr durstig. Es gibt ein Gerücht, dass ein kleiner Lastwagen die Leute mit Wasser versorgen soll, aber sie hat noch nichts erhalten. Wie lange wird es dauern, bis der Lastwagen alle Menschen mit Wasser versorgt hat?“

Um diese Aufgabe zu lösen, gehen wir folgendermaßen vor:

- Verständnis des Problems:** Christina steckt in einem 20 km langen Stau und ist seit 6 Stunden dort. Es ist sehr warm und sie hat großen Durst. Es gibt das Gerücht, dass ein kleiner LKW die Leute mit Wasser versorgen soll, aber sie hat noch nichts bekommen. Wir müssen berechnen, wie lange es dauert, bis der LKW alle Menschen mit Wasser versorgt hat.
- Notwendige Annahmen:**
  - Der durchschnittliche Abstand zwischen den Autos im Stau, um die Anzahl der Autos abzuschätzen.
  - Wie viel Zeit es braucht, um jedes Auto mit Wasser zu versorgen.
  - Ob der LKW während der Fahrt Wasser liefert oder an jedem Auto anhält.

### Schritt 1: Berechnung der Anzahl der Autos

Gegeben:

- Länge des Staus = 20 km = 20.000 Meter
- Durchschnittlicher Abstand zwischen den Autos = 10 Meter

$$\text{Anzahl der Autos} = \frac{\text{Staulänge}}{\text{Durchschnittlicher Abstand zwischen den Autos}}$$

### Schritt 2: Berechnung der Gesamtversorgungszeit

$$\text{Gesamtversorgungszeit} = \text{Anzahl der Autos} \times \text{Versorgungszeit pro Auto}$$

Die Berechnungen ergaben:

- Es gibt ungefähr **2.000 Autos** im 20 km langen Stau.
- Die Gesamtzeit, die der LKW benötigt, um alle Autos mit Wasser zu versorgen, beträgt **33 Stunden und 20 Minuten**.

Lösung von ChatGPT-4

Die Lösung ist per se nicht falsch. GPT hat wie in der vorherigen Aufgabe auch fehlende Informationen identifiziert und nachvollziehbare Annahmen gemacht. Allerdings wird in der Lösung der Sachverhalt stark vereinfacht, beispielsweise besitzt die Autobahn hier nur eine Spur.

*Dr. Oliver Straser, Aileen Fahrländer, Prof. Dr. Katja Maaß,  
Dr. Carina Spreitzer, Prof. Dr. Stefan Zehetmaier*