

DEUTSCHLEHRERTAGUNG 2024

# MIT KÜNSTLICHEN GEHIRNEN SPRACHEN LERNEN



Large Language Models im Sprachunterricht

**GOETHE  
INSTITUT**

Sprache. Kultur. Deutschland.

# PAUL HEEREN

- Referent Fachkräfte am Goethe-Institut in New Delhi
- Zuvor: Astrophysiker und Data Scientist
- Praktische Kenntnisse in Programmierung und Machine Learning
- Regelmäßige Nutzung von Large Language Models:
  - Auf der Arbeit
  - Zur Unterhaltung & Recherche
  - Beim Hindi-Lernen



## AGENDA

# ZIELE DES WORKSHOPS



Stärken und Schwächen von Large Language Models (LLMs) kennenlernen



Einsatzmöglichkeiten im Sprachunterricht und Sprachenlernen erforschen



Best Practice Beispiele entwickeln



Selbstvertrauen im Umgang mit LLMs aufbauen



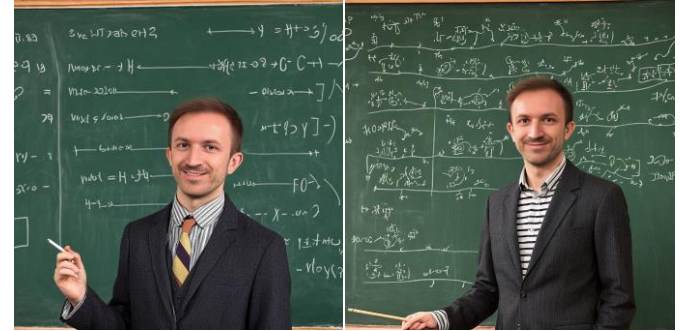
## AGENDA

- 1 **EINFÜHRUNG: GENERATIVE KI UND LLMS**
- 2 **ERSTELLEN VON VOKABEL- UND GRAMMATIKÜBUNGEN**
- 3 **KORREKTUR VON AUFGABEN**
- 4 **LLMS ALS SPRACHTANDEMPARTNER**
- 5 **REFLEXION UND DISKUSSION**

1

# EINFÜHRUNG: GENERATIVE KI UND LLMS

# BEISPIEL: BILDERSTELLUNG

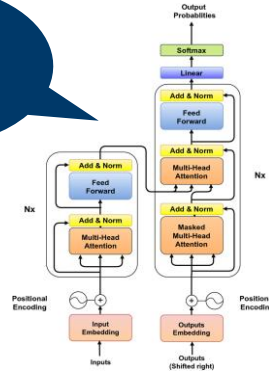


Professor, academics,  
checked suit jacket, striped  
shirt, crazy hair, blackboard in  
background, complex  
formulas

# KURZE GESCHICHTE DER (KÜNSTLICHEN) INTELLIGENZ



Attention is all you need



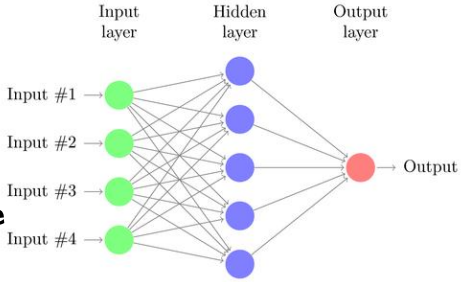
Transformer (2017)



Large Language Models

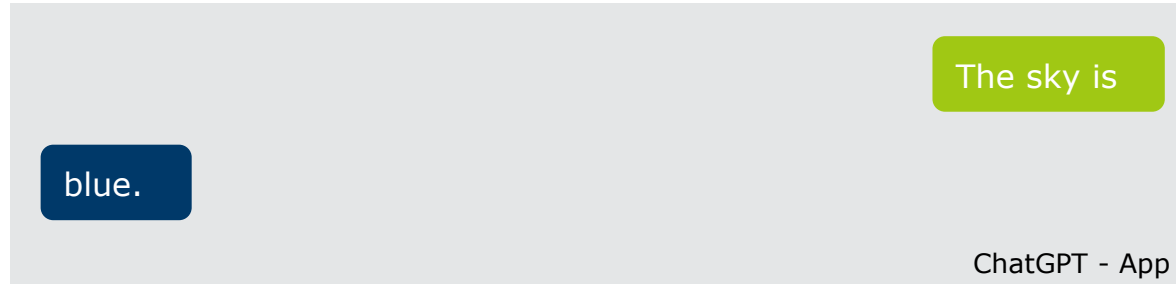


Künstliche Neuronale Netze



# WAS KÖNNEN LARGE LANGUAGE MODELS?

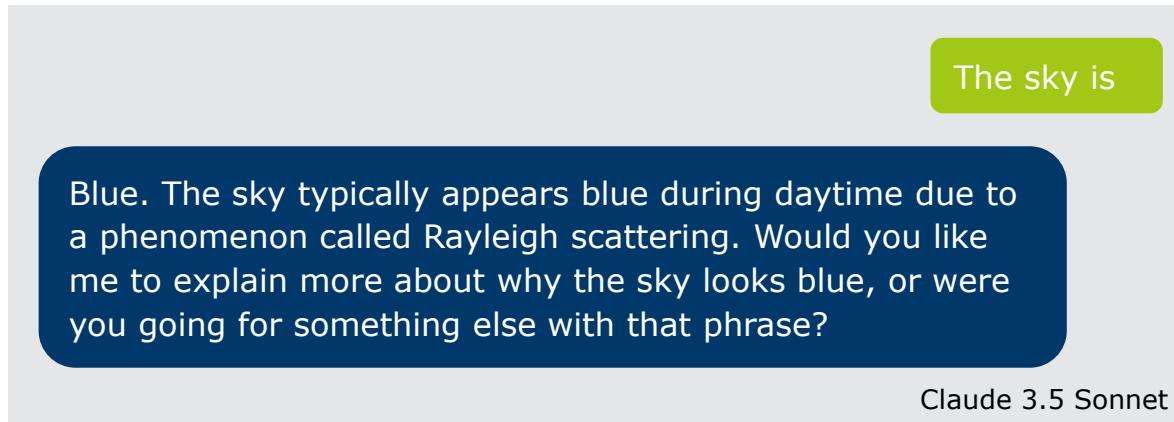
- LLMs als Textvervollständigung
  - was ist das statistisch wahrscheinlichste nächste Wort?
  - basierend auf Training mit großen Text-Datenmengen





# WAS KÖNNEN LARGE LANGUAGE MODELS?

- LLMs als Chat-Partner (basierend auf Textvervollständigung)
  - was ist die statistisch wahrscheinlichste Antwort (die akzeptierteste)?
  - zusätzliches Training durch Feedback von Menschen ("Reinforcement learning from human feedback (RLHF)")



# WAS IST PROMPT ENGINEERING?

Was ist Prompt Engineering?

[Zu lange und komplexe Antworten...]

Geht das auch in einfacher Sprache?

Hier ist eine Version in einfacher Sprache:

- Man sagt der KI genau, was man will
- Man gibt der KI wichtige Infos mit
- So hilft die KI besser bei Aufgaben

Claude 3.5 Sonnet

## STRUKTUR EINES PROMPTS

Schreibe ein Essay zu untenstehender Aufgabe.  
Es handelt sich hierbei um eine Hausaufgabe im  
Unterrichtsfach Deutsch der 10. Klasse, und sollte sich  
hinsichtlich Sprache und Ausdruck in etwa auf diesem  
Level bewegen.  
Die Aufgabe lautet: „Erläutere die typischen  
Schwierigkeiten beim Deutschlernen für  
englischsprachige Menschen.“  
Der Text sollte etwa 300 Wörter lang und in Einleitung,  
Hauptteil und Schluss gegliedert sein.

Anweisung

Kontext

Eingabedaten

Ausgabeindikator

Hier ist ein Entwurf für ein Essay ...

Claude 3.5 Sonnet

## DOS UND DONTs DES PROMPTENS

- Schreibe was zu tun ist – nicht was NICHT zu tun ist!  
→ ~~„Verwende kein Präteritum.“~~ -> „Verwende nur Präsens und Perfekt.“
- LLMs lieben Rollenspiele!  
→ „Du bist Shakespeare.“, „Du bist eine hochqualifizierte Deutschlehrerin.“
- Sei spezifisch und gebe Details und Kontext mit!  
→ ~~„Schreibe über...“~~ -> „Schreibe 10 Sätze in formellem Ton über..“
- Sei höflich und korrekt – dein Umgangston beeinflusst die Antwort.  
→ ~~„Hey geb mir Text der wo...“~~ -> „Hilf mir, ein Essay zu erstellen, das...“
- Strukturiere den Prompt (und verwende Beispiele).  
→ „Aufgabe: ...    Ausgabeformat: ...    Beispiel: ...    Eingabetext: "" "" "" "" “

### Pro-Tipp Chain-of-Thought

„Erstelle zunächst eine  
Gliederung. Dann ...  
Anschließend ...“

# ACHTUNG – RISIKEN!

## DATENKRAKE



Keine privaten, empfindlichen oder geschützten Daten eingeben!

## HALLUZINATIONEN



Überprüfe die Antworten auf Falschinformationen und Fehler!

## VORURTEILE



LLMs können so vorurteilsbeladen sein wie wir alle!

## RECHTLICHES



Der Rechtsrahmen (Urheberrecht, Einsatz, etc.) ist oft noch unklar und/oder ändert sich schnell.

# WO LLMS BESSER ALS 5-JÄHRIGE SIND...

- (Abstraktes, theoretisches) Wissen, Sprachfähigkeiten
  - ➔ Traditionelle Tests, z.B. MMLU (Measuring Massive Multitask Language Understanding), MATH etc.

## Astronomy

What is true for a type-Ia supernova?

- This type occurs in binary systems.
- This type occurs in young galaxies.
- This type produces gamma-ray bursts.
- This type produces high amounts of X-rays.

Answer: A

	Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku	GPT-4	GPT-3.5	Gemini 1.0 Ultra	Gemini 1.0 Pro
Undergraduate level knowledge <i>MMLU</i>	<b>86.8%</b> 5 shot	<b>79.0%</b> 5-shot	<b>75.2%</b> 5-shot	<b>86.4%</b> 5-shot	<b>70.0%</b> 5-shot	<b>83.7%</b> 5-shot	<b>71.8%</b> 5-shot
Graduate level reasoning <i>GPQA, Diamond</i>	<b>50.4%</b> 0-shot CoT	<b>40.4%</b> 0-shot CoT	<b>33.3%</b> 0-shot CoT	<b>35.7%</b> 0-shot CoT	<b>28.1%</b> 0-shot CoT	—	—
Grade school math <i>GSM8K</i>	<b>95.0%</b> 0-shot CoT	<b>92.3%</b> 0-shot CoT	<b>88.9%</b> 0-shot CoT	<b>92.0%</b> 5-shot CoT	<b>57.1%</b> 5-shot	<b>94.4%</b> Maj1@32	<b>86.5%</b> Maj1@32
Math problem-solving <i>MATH</i>	<b>60.1%</b> 0-shot CoT	<b>43.1%</b> 0-shot CoT	<b>38.9%</b> 0-shot CoT	<b>52.9%</b> 4-shot	<b>34.1%</b> 4-shot	<b>53.2%</b> 4-shot	<b>32.6%</b> 4-shot

## ...UND WO NICHT

- Allgemeines Weltverständnis, viele einfache Fangfragen
  - ➔ z.B. SIMPLE-Test

5.	0	1	0	0	1
Total	<i>Claude 3.5 Sonnet</i>	<i>Llama 405b</i>	<i>GPT4o</i>	<i>Gemini 1.5</i>	<i>GPT-4T</i>
(96% human)	<b>32%</b>	<b>18%</b>	<b>5%</b>	<b>11%</b>	<b>12%</b> <sub>1</sub>

Beth places four whole ice cubes in a fire at the start of the first minute, then five at the start of the second minute and some more at the start of the third minute, but none in the fourth minute. If the average number of ice cubes per minute placed in the fire was five, how many whole ice cubes can be found in the fire at the end of the third minute? Pick the most realistic answer. A) 5 B) 11 C) 0 D) 20

- 🌀 To find the number of ice cubes placed by Beth over the four minutes, we first use the information that the average number of ice cubes per minute is five. This average is calculated over the four minutes, which means the total number of ice cubes placed in the fire is:

$$5 \text{ ice cubes/minute} \times 4 \text{ minutes} = 20 \text{ ice cubes}$$

Next, we break down the problem to identify the number of ice cubes placed each minute:

- At the start of the first minute, 4 ice cubes were placed.
- At the start of the second minute, 5 ice cubes were placed.
- Let  $x$  be the number of ice cubes placed at the start of the third minute.
- No ice cubes were placed at the start of the fourth minute.

The sum of the ice cubes placed in the first, second, third, and fourth minutes must equal the total number:

$$4 + 5 + x + 0 = 20$$

# ÜBERBLICK BELIEBTER LLMS (STAND 08/24)

	<b>ChatGPT OpenAI</b>	<b>Claude Anthropic</b>	<b>Gemini Alphabet</b>	<b>Copilot Microsoft</b>
Freie Version	<ul style="list-style-type: none"> <li>• Text, Sprache &amp; Bilder</li> <li>• GPT-4o mini (eingeschränkt: GPT-4o)</li> <li>• Eingeschränkt: Onlinesuche &amp; ind. GPTs</li> </ul>	<ul style="list-style-type: none"> <li>• Text, Bilder &amp; Dateien</li> <li>• Eingeschränkter Zugriff auf Claude 3.5 Sonnet</li> </ul>	<ul style="list-style-type: none"> <li>• Text, Sprache &amp; Bilder</li> <li>• Eingeschränkter Zugriff</li> </ul>	<ul style="list-style-type: none"> <li>• Text, Sprache und Bilder</li> <li>• Zugang zum Internet</li> <li>• Basiert auf Sprachmodell von OpenAI</li> </ul>
Bezahlversion	<p>Für 20\$ pro Monat:</p> <ul style="list-style-type: none"> <li>• GPT-4o, GPT-4o mini, GPT-4</li> <li>• 5x mehr Nachrichten für GPT-4o</li> <li>• DALL·E-Bildgenerierung</li> <li>• Voller Zugriff auf ind. GPTs</li> </ul>	<p>Für 20\$ pro Monat:</p> <ul style="list-style-type: none"> <li>• 5x mehr Nachrichten</li> <li>• Claude 3 Haiku &amp; Opus</li> <li>• Projekte rund um Dokumente, Code oder Dateien</li> <li>• Prior. Zugriff in High-Traffic-Zeiten</li> </ul>	<p>Für 1950₹ pro Monat:</p> <ul style="list-style-type: none"> <li>• Zugriff auf Gemini 1.5 Pro</li> <li>• 1 Mio. Token Context Window</li> <li>• Gemini in Gmail &amp; Docs</li> <li>• 2 TB Speicherplatz</li> </ul>	<p>Für 20\$ pro Monat:</p> <ul style="list-style-type: none"> <li>• Schnellerer Zugriff</li> <li>• Copilot in ausgewählten MS365-Apps</li> <li>• Schnellere AI-Bilderstellung</li> </ul>
Zugang	ChatGPT.com, App	Claude.ai, App	Gemini.Google.com, App	Copilot.Microsoft.com, App



2

# ERSTELLEN VON VOKABEL- UND GRAMMATIKÜBUNGEN

## ERSTELLEN VON VOKABEL- &amp; GRAMMATIKÜBUNGEN

## IDEE UND REZEPT

## Ziel:

- Entlastung der Lehrkraft
- Größere Vielfalt und maßgeschneiderte Aufgaben

## Vorgehen:

- LLM mit Details zur gewünschten Aufgabe promoten
- Ergebnis überprüfen, durch weitere Interaktion mit LLM verbessern
- Wenn nötig: Eigene Anpassungen vornehmen

① Schreibe die richtigen Wörter in die Felder!

**Bringe** mir bitte die  !

**Fange** nie eine  !

So ein **Schreck!** Hier ist alles voll  !

Ich zähle bis **drei**, und du bist  !

Die Arbeit kostet Mühe und **Schweiß**.

Mir ist ganz  .

Sie ist **schlau**, denn sie ist eine  .

Du versteckst dich hinter dem **Schrank**,  
denn du bist  .

Ich passe in kein **Kleid**, denn ich bin  .

Nicht in den Finger **schneiden!**

Das musst du ver  !

Wenn ich dir die Daumen **drück**,  
hast du bestimmt  .

# BEISPIEL: VOKABELÜBUNG

Du bist eine Deutschlehrerin, und ich dein Englisch-sprachiger Schüler. Ich befinde mich aktuell ungefähr auf dem A1-Level in Deutsch. Entwirf eine Lückentext-Übung mit Sätzen, die die Verwendung geeigneter Adverbien erfordern (z. B. schnell, sorgfältig, oft). Liste alle zu verwendenden Adverbien am Anfang des Textes auf. Aus dem Kontext muss für jede Lücke klar hervorgehen, welches Adverb jeweils gefordert ist. Der Text soll 10 Sätze lang sein. Notiere die richtigen Lösungen am Ende des Textes.

Sehr gerne! Ich werde eine Lückentext-Übung für Sie erstellen, die sich auf Adverbien konzentriert. Diese Übung ist für Ihr A1-Niveau geeignet.

Adverbien: schnell, langsam, oft, selten, gern, gut, laut, leise, sorgfältig, fleißig ...

Claude 3.5 Sonnet

# BEISPIEL: GRAMMATIKÜBUNG

Du bist eine Deutschlehrerin, und ich dein Englisch-sprachiger Schüler. Ich befinde mich aktuell ungefähr auf dem A1-Level in Deutsch. Erstelle einen Lückentext für mich, mit dem ich Konjugationen von Verben üben kann. Der Text sollte von einem typischen Tag und Ereignissen am Vortag erzählen und etwa zehn Sätze lang sein. Lasse Lücken anstelle der Verben und setze die Infinitivformen der gewünschten Verben in Klammern hinter jede Lücke. Die Übung soll die Gegenwarts- und Vergangenheitsformen abfragen und auch einfache trennbare Verben beinhalten. Aus dem Kontext muss immer klar hervorgehen, welche Zeitform des Verbs gewünscht ist. Notiere die richtigen Lösungen am Ende des Textes.

Sehr gerne! Hier ist ein Lückentext für dich ...

# BEISPIEL-PROMPTS



**Anmerkung:**

Das Padlet wurde im Anschluss an den Workshop eingefroren, d.h. es ist nur noch im Lesemodus verfügbar.

3

# KORREKTUR VON AUFGABEN

# IDEE UND REZEPT

## Ziel:

- Entlastung der Lehrkraft
- Vorkontrolle

## Vorgehen:

- Lösung des/der Schüler\*in an LLM geben (u.U. zuvor OCR-Texterkennung)
- LLM um Überprüfung bitten (unter Angabe von Kriterien/Musterlösung)
- Überprüfen der Korrektur!



# BEISPIEL: AUFSATZ-KORREKTUR

Du bist eine Lehrerin für Deutsch als Fremdsprache, und korrigierst Übungsaufsätze von Schüler\*innen. Ich gebe dir gleich zunächst den abgegebenen Aufsatz einer Schülerin. Danach erhältst du weitere Anweisungen.

...

[Aufsatz]

...

Bitte markiere alle sprachlichen Fehler (Rechtschreibung, Grammatik, Interpunktion usw.) im Text, und füge die korrekten Ausdrücke jeweils in eckigen Klammern dazu.

ChatGPT



# BEISPIEL: AUFSATZ-BEWERTUNG

Bewerte den Aufsatz nach folgenden Kriterien:

- Kohärenz: Folgt er dem Schema Einleitung, Hauptteil, Schluss? Ist er logisch aufgebaut? Sind Sätze und Satzteile gut miteinander verknüpft?
- Wortschatz: Wird ein differenziertes, vielfältiges Spektrum verwendet? Wie gut wird der Wortschatz beherrscht, d.h. werden die Wörter im richtigen Kontext verwendet?
- Strukturen: Ist bei Satzstrukturen eine gewisse Vielfalt vorhanden? Wie korrekt werden Grammatik, Interpunktion, Rechtschreibung beherrscht?

Gib für jedes Kriterium auch Beispiele für Missachtungen oder Fehler (falls vorhanden).

Hier ist die Bewertung des Aufsatzes nach den genannten Kriterien: ...

## BEISPIEL-PROMPTS



**Anmerkung:**

Das Padlet wurde im Anschluss an den Workshop eingefroren, d.h. es ist nur noch im Lesemodus verfügbar.

4

# LLMS ALS SPRACHTANDEMPARTNER

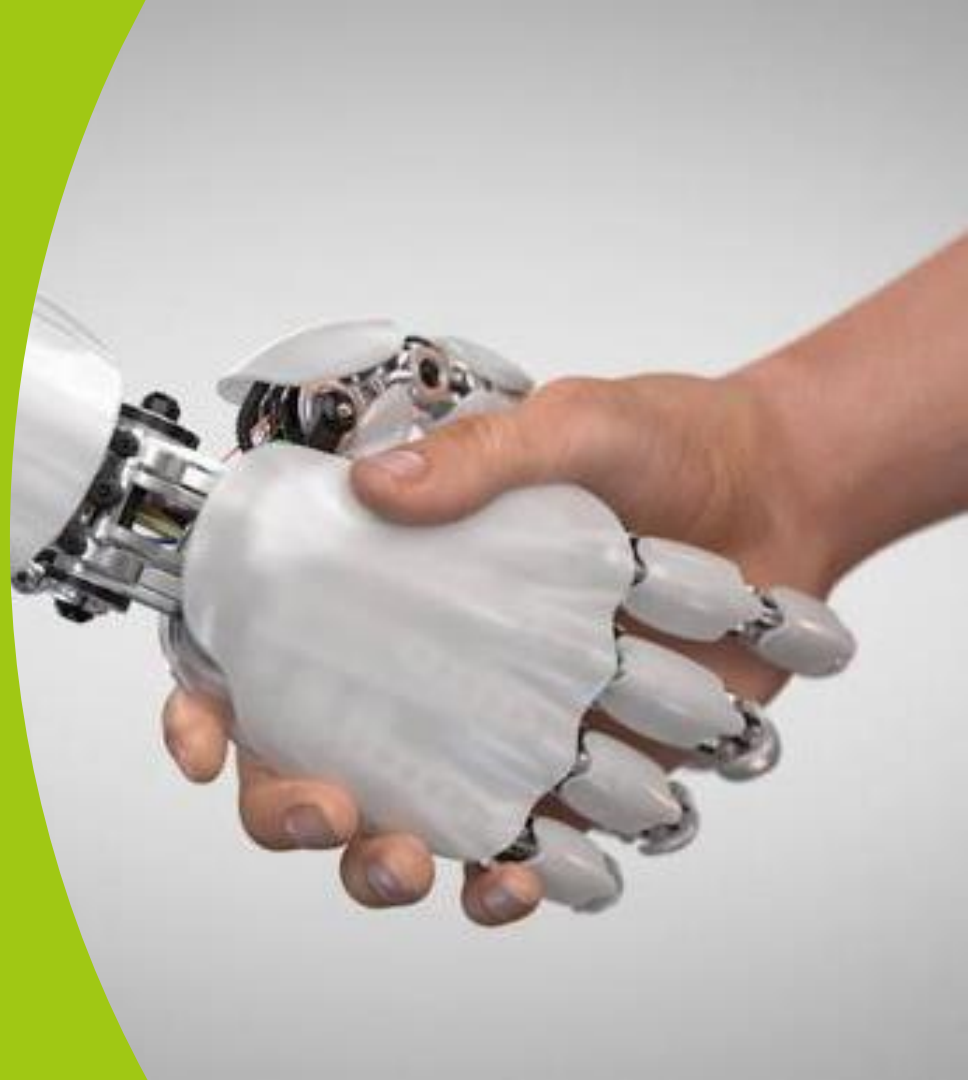
## IDEE UND REZEPT

### Ziel:

- Niedrigschwellige Möglichkeit zum Üben
- Ständig verfügbar, Vermeidung von Scham

### Vorgehen:

- LLM über Rolle und Details instruieren
- Konversation mit LLM – in Schriftform ODER verbal (z.B. ChatGPT Voice Mode)
- U.U.: Fehler korrigieren lassen



## BEISPIEL: GESPRÄCH AUF A2-NIVEAU

Ich bin Deutschlerner auf dem Niveau A2, und du bist mein Sprachtandempartner - eine deutsche Muttersprachlerin, die mit mir Deutsch übt. Fange eine Unterhaltung mit mir auf Deutsch an, indem du mich nach meinen Hobbies fragst, und gehe dann auf meine Antwort ein. Verwende keine zu schwierige Sprache, aber fordere mich heraus indem du das Gespräch aufrecht erhältst und mich auch zu längeren Antworten bewegst.

Hallo! Schön, dich kennenzulernen. Ich freue mich, mit dir Deutsch zu üben. Sag mal, was sind deine Hobbies? Was machst du gerne in deiner Freizeit?



Claude 3.5 Sonnet

## BEISPIEL-PROMPTS



**Anmerkung:**

Das Padlet wurde im Anschluss an den Workshop eingefroren, d.h. es ist nur noch im Lesemodus verfügbar.

5

# REFLEXION UND DISKUSSION

**VIELEN DANK  
FÜR DIE  
AUFMERKSAMKEIT!**



**GOETHE  
INSTITUT**

Sprache. Kultur. Deutschland.