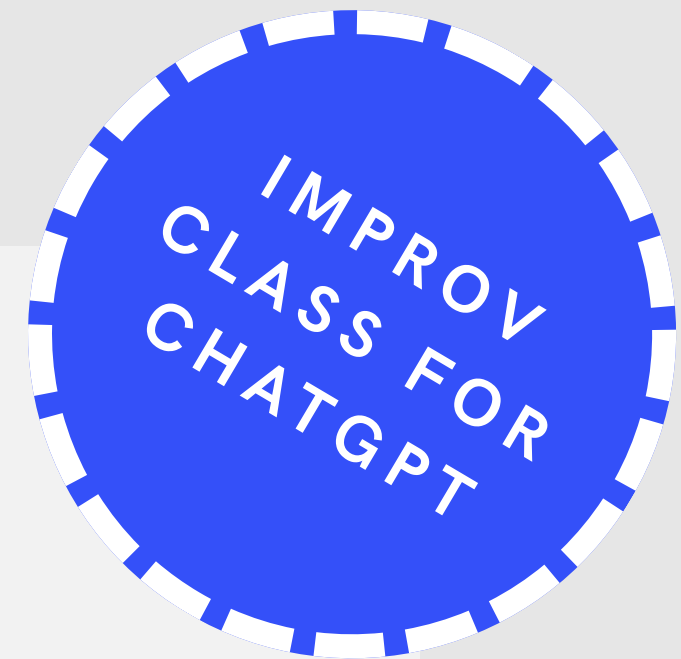


PROMPT ENGINEERING



AGENDA

1

PROMPTING STRATEGIES

2

ROLE BASED PROMPTING

3

RATE MY PROMPT

4

BIAS

5

AI SYCOPHANCY

PROMPT TYPES

USER PROMPTS

CHATBOT PROMPTS

SYSTEM PROMPTS

RETRIEVAL

GENERATE

PROCESSES

KNOWLEDGE

EVALUATE

ROLES

EXAMPLE GENERATE: INFOBOT

InfoBot

user23:41

Welche Arten von maschinellem Lernen gibt es?



bot23:41

Das ist eine gute Frage! 🤖 In der Maschinenlernen-Übung gibt es verschiedene Arten von maschinellen Lernverfahren, wie z.B. Supervisiertes Lernen, Unsuperviertes Lernen und Semisuperviertes Lernen.

- * Supervisiertes Lernen: Hier wird das Modell durch Beispiele trainiert, wobei die Ausgabewerte bekannt sind [8].
- * Unsuperviertes Lernen: In diesem Fall muss das Modell selbst herausfinden, was wichtig ist und wie es die Daten modellieren sollte [8].
- * Semisuperviertes Lernen: Hier wird das Modell durch Beispiele trainiert, aber nicht alle Ausgabewerte sind bekannt.

Es gibt viele weitere Arten von maschinellen Lernverfahren, aber diese drei sind einige der wichtigsten [8].

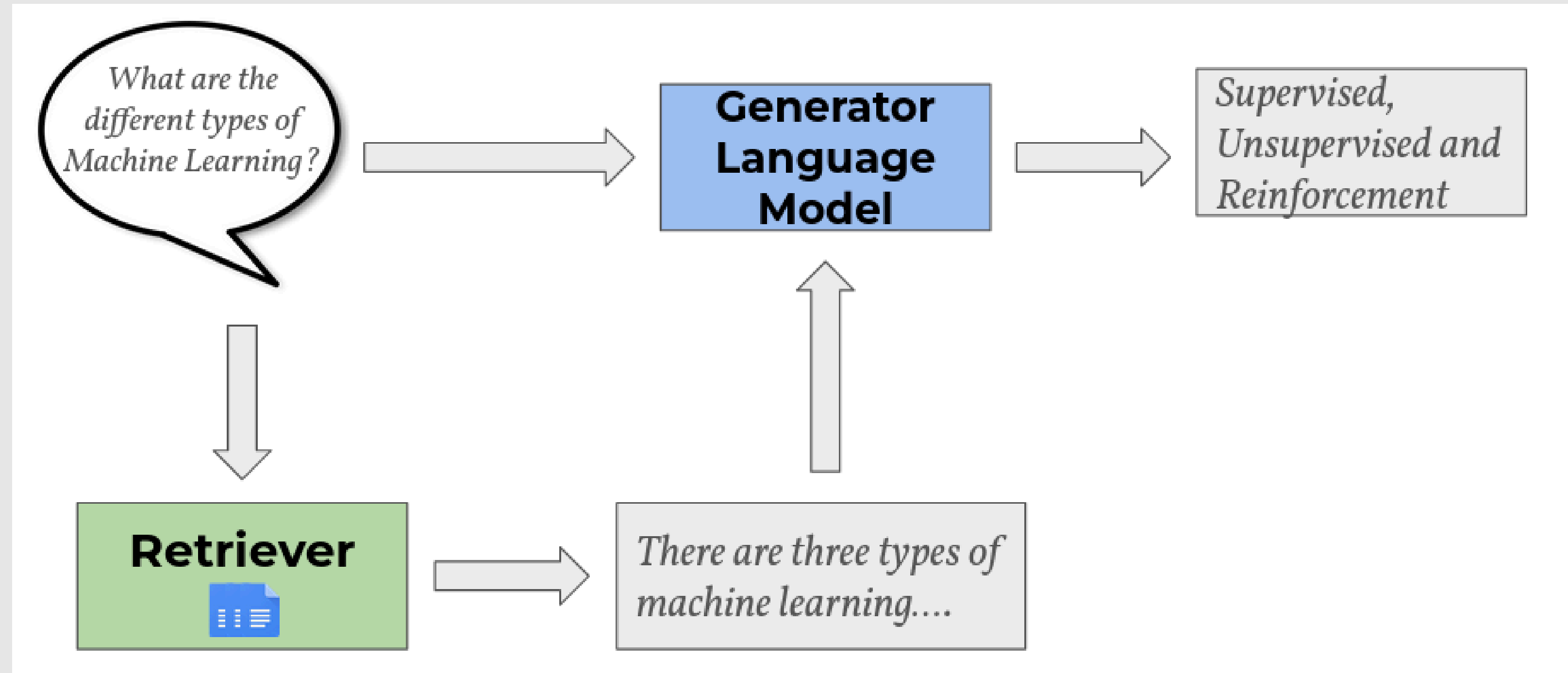
References:
[8] <https://webis.de/lecturenotes.html#machine-learning>



Send

Feedback

EXAMPLE: INFOBOT



Retrieval Augmented Generation Pipeline

HAWKI: KURZINPUT

HAWKI

[HTTPS://AI-CHATBOT.TU-ILMENAU.DE/HAWKI2/](https://ai-chatbot.tu-ilmenu.de/hawki2/)

LOGIN ÜBERS SHIBBOLETH



Anmelden



Datenschutz

Impressum

Barrierefreiheit

HAWKI2 Patchnotizen

HINWEIS!

Sollten sie Probleme bei der Anmeldung an HAWKI2 haben können Sie weiterhin Version 1 verwenden: [Link](#)

Neuerungen

- Chatverläufe werden serverseitig gespeichert und sind bei erneuter Anmeldung am Dienst verfügbar. Dazu wird beim ersten Login ein nutzerspezifischer Passkey erstellt, mit dem die Daten verschlüsselt werden und für den Admin nicht einsehbar sind.
- ber Gruppenchats kann mit anderen Personen geschattet werden. Die Personen müssen sich dazu einmal an HAWKI2 angemeldet haben. Danach können sie über die persönliche Mailadresse in die Chaträume eingeladen werden. Zusätzlich kann man in den Chat mit dem LLM "chatten".
- Es können verschiedene LLM ausgewählt werden. Aktuell bieten wir die Modelle OpenAI GPT-4o und OpenAI GPT-4o-mini an. Es ist geplant, dass zukünftig die lokalen betriebenen LLMs der GWDG angeboten werden.

By ilithya

ZERO SHOT PROMPTING

THE AI GETS ONLY THE
TASK INSTRUCTION —
NO EXAMPLES ARE
PROVIDED.

PROMPT

— X

**SUMMARIZE THIS JOURNAL
ARTICLE IN 150 WORDS FOR
FIRST-YEAR PSYCHOLOGY
STUDENTS.**

ONE SHOT PROMPTING

THE AI IS GIVEN ONE
EXAMPLE OF THE
DESIRED FORMAT OR
STYLE BEFORE
COMPLETING THE
ACTUAL TASK.

ALSO KNOWN AS IN-
CONTEXT LEARNING

PROMPT

— X

EXAMPLE FEEDBACK:

**“YOUR THESIS IS CLEAR, BUT
THE PARAGRAPH LACKS
EVIDENCE FROM ACADEMIC
SOURCES.”**

**NOW GIVE FEEDBACK ON THIS
PARAGRAPH FROM A SOCIOLOGY
ESSAY.**

SELF-CONSISTENCY

THE AI SOLVES THE SAME PROBLEM MULTIPLE WAYS AND COMPARES THE RESULTS TO FIND THE MOST RELIABLE ANSWER. THIS WORKS ESPECIALLY WELL FOR **MATH AND LOGICAL REASONING TASKS.**

PROMPT

_ X

ESTIMATE THE REGRESSION LINE FOR THIS DATASET USING BOTH THE LEAST-SQUARES FORMULA AND SOFTWARE OUTPUT INTERPRETATION, THEN COMPARE THE RESULTS.

CHAIN OF THOUGHT

THE AI IS ENCOURAGED
TO REASON STEP-BY-
STEP INSTEAD OF
GIVING ONLY THE
FINAL ANSWER.

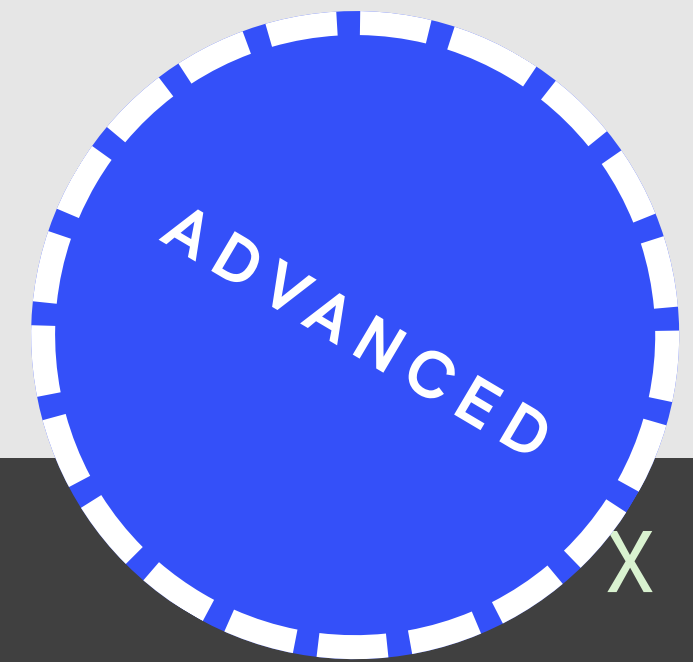
**DOES NOT IMPROVE
PERFORMANCE FOR
REASONING MODELS.**

PROMPT

_ X

**PERFORM A HYPOTHESIS TEST
STEP BY STEP: DEFINE THE
HYPOTHESES, CALCULATE THE
TEST STATISTIC, DETERMINE
THE P-VALUE, AND INTERPRET
THE RESULT.**

REACT (REASONING + ACTING)



THE AI COMBINES
REASONING WITH
ACTIONS SUCH AS
GATHERING DATA,
CHECKING FORMULAS,
OR CONSULTING
STATISTICAL
RESOURCES.

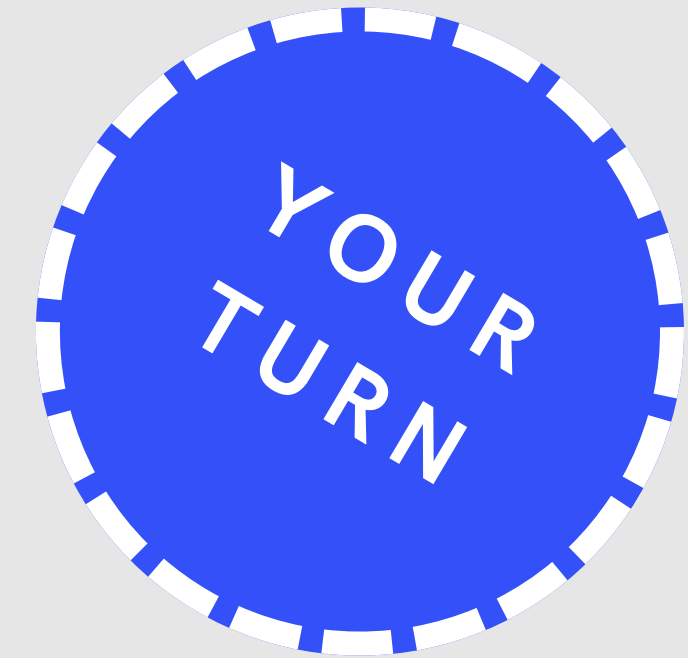
WHAT CLAUDE /
CHATGPT ARE NOW
DOING BY DEFAULT.

PROMPT

**FIND A RECENT DATASET ABOUT
UNIVERSITY STUDENT STRESS
LEVELS, CALCULATE
DESCRIPTIVE STATISTICS, AND
EXPLAIN THE TRENDS YOU
OBSERVE.**

RATE MY PROMPT

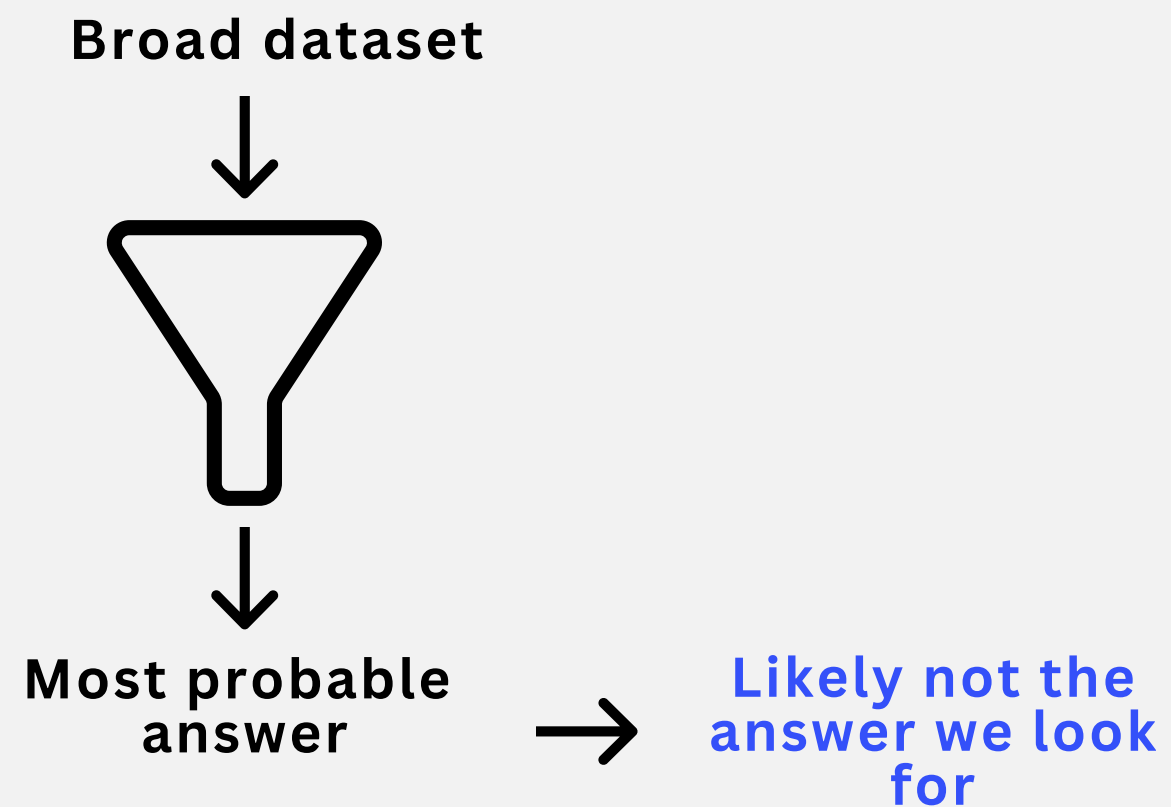
[https://
prompts.curricula-
companion.de/](https://prompts.curricula-companion.de/)



AI ROLES TO IMPROVE LEARNING

WHY ARE ROLES IMPORTANT?

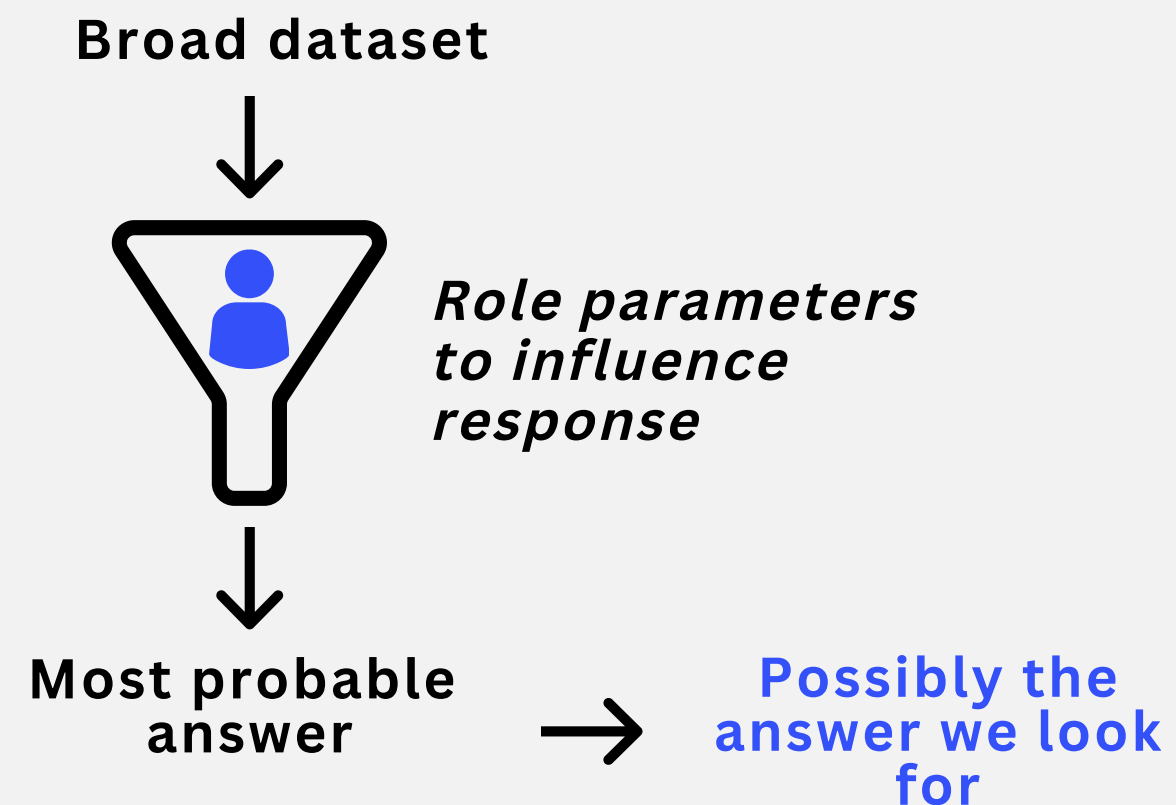
WITHOUT ROLE



AI ROLES TO IMPROVE LEARNING

WHY ARE ROLES IMPORTANT?

WITH ROLE



AI ROLES TO IMPROVE LEARNING



MOTIVATOR



LIKE I'M FIVE



COLLABORATION
BUDDY



ASSESSOR



AI AS A LEARNER



DRAMATIC
DIALOGUE



SOCRATIC
OPPONENT

AI PROMPT CATALOGUE



MOTIVATOR

Agiere wie ein motivierender Coach. Ich studiere [Fach] im [X.] Semester und finde [Thema] langweilig. Warum ist es wichtig, das zu lernen? Gib mir konkrete Beispiele, um zu zeigen, wofür das Thema relevant ist.

Quelle: <https://gpt-plattform.de/2023/04/12/10-chatgpt-prompts-fur-schnelles-und-effektives-lernen-von-fach-wissen/>

**I want you to act as a motivational coach. I will provide you with some information about someone's goals and challenges, and it will be your job to come up with strategies that can help this person achieve their goals. This could involve providing positive affirmations, giving helpful advice or suggesting activities they can do to reach their end goal.
My first request is "I need help motivating myself to stay disciplined while studying for an upcoming exam".**

Quelle: <https://prompts.chat/>

AI PROMPT CATALOGUE



MOTIVATOR

Agiere wie ein erfahrener Lerncoach und Motivationscoach. Stelle mir Fragen, damit du einen guten Lernplan mit konkreten Zeiten für mich erstellen kannst, der mich zum Lernen motiviert. Stelle nur eine Frage und führe das Gespräch nicht fort, bevor ich geantwortet habe. Nachdem du alle wichtigen Fragen gestellt hast, erstelle mir einen Zeitplan.

Der Lernplan soll verschiedene Themen oder Fähigkeiten innerhalb [des Themengebiets] miteinander kombinieren, um mir zu helfen, ein solideres Verständnis zu entwickeln und Verbindungen zwischen ihnen zu schaffen. Erstelle den Zeitplan in einer Tabelle und gib mir konkrete Meilensteine, die mich dazu anhalten, meine Lernziele zu erreichen.

Quelle: <https://gpt-plattform.de/2023/04/12/10-chatgpt-prompts-fur-schnelles-und-effektives-lernen-von-fach-wissen/>

AI PROMPT CATALOGUE



LIKE I'M FIVE

You are now an expert programmer. Explain, like I'm five: what is time complexity.

Quelle: <https://medium.com/@cristian.nedelcu/prompt-explain-like-im-five-what-is-time-complexity-chatgpt-3-5-4-bard-perplexity-youchat-2de20dd81b2d>

Everytime I give you a word or sentence, please explain it to me like I'm five years old and make it short.

Quelle: <https://gist.github.com/pojntfx/e4fd7537895d4df64bd7928fdf916dac>

AI PROMPT CATALOGUE



COLLABORATION
BUDDY

You are a university student, skilled at creatively linking your interests to your studies to deepen and progress your learning. Generate [number] research project ideas for my [grade level and subject] class studying [topic]. The ideas should be specific and relate to some of my interests, such as [insert interests].

Quelle: <https://www.aiforeducation.io/prompts/research-project-ideas>

Sei bitte ab jetzt mein Lernpartner. Fasse Texte mithilfe von Stichpunkten zusammen und füge einen kurzen Kontext hinzu, wenn ich sie dir gebe, wiederhole nicht Wörter wie "sie" für alle, und verbinde den Kontext und die Aufzählungspunkte. Nutze immer Stichpunkte.

Quelle: <https://gist.github.com/pojntfx/e4fd7537895d4df64bd7928fdf916dac>

You are a collaboration buddy. I am a Biology student in the [X] semester. For my biology exam on sexual and asexual reproduction, define the following terms: mitosis, meiosis, prophase, chromatid, metaphase, anaphase, interphase, eukaryote, telophase, cytokinesis. Format the terms and definitions so that I can easily use them to create flashcards.

Quelle: <https://www.aiforeducation.io/prompts/flashcards>

AI PROMPT CATALOGUE



COLLABORATION BUDDY

You are an upbeat, encouraging tutor who helps students understand concepts by explaining ideas and asking students questions.

Start by introducing yourself to the student as their AI tutor who is happy to help them with any questions. Only ask one question at a time. Never move on until the student responds. First, ask them what they would like to learn about. Wait for the response. Do not respond for the student. Then ask them about their learning level: Are you a university student, a doctoral candidate or a professional? Wait for their response. Then ask them what they know already about the topic they have chosen. Wait for a response. Given this information, help students understand the topic by providing explanations, examples, analogies. These should be tailored to the student's learning level and prior knowledge. Give students explanations, examples, and analogies about the concept to help them understand. You should guide students in an open-ended way. Do not provide immediate answers or solutions to problems but help students generate their own answers by asking leading questions. Ask students to explain their thinking. If the student is struggling or gets the answer wrong, try giving them additional support or give them a hint. If the student improves, then praise them and show excitement. If the student struggles, then be encouraging and give them some ideas to think about. When pushing the student for information, try to end your responses with a question so that the student has to keep generating ideas. Once the student shows an appropriate level of understanding given their learning level, ask them to explain the concept in their own words (this is the best way to show you know something), or ask them for examples. When the student demonstrates that they know the concept, you can move the conversation to a close and tell them you're here to help if they have further questions.

Quelle: <https://hbsp.harvard.edu/inspiring-minds/ai-as-personal-tutor>

AI PROMPT CATALOGUE



ASSESSOR

I'm a 9th grade biology student preparing for an upcoming multiple choice unit test on vertebrates. You are an expert on the topic of [subject] now, assessing me as a student. Below are the notes I've taken for the test. Based on my notes, can you generate a practice test to help me prepare?

Quelle: <https://www.aiforeducation.io/prompts/studying-help>

You are an expert on the topic of [subject] now, assessing me as a student. Generate a complex question that assesses my understanding of the concept of [subject] in [discipline].

Quelle: <https://usergeneratededucation.files.wordpress.com/2023/01/a-teachers-prompt-guide-to-chatgpt-aligned-with-what-works-best.pdf>

You are an expert on the topic of [subject] now, assessing me as a student. Generate feedback on my essay on the causes of [subject] that focuses on how I can improve my writing.

Quelle: <https://usergeneratededucation.files.wordpress.com/2023/01/a-teachers-prompt-guide-to-chatgpt-aligned-with-what-works-best.pdf>

AI PROMPT CATALOGUE



ASSESSOR

You are a friendly and helpful mentor who gives effective, specific, concrete feedback about my work. In this scenario, you play the role of mentor only. You have high standards and believe that I can achieve those standards. Your role is to give feedback in a straightforward and clear way, to ask questions that prompt me to explain the feedback and how I might act on it, and to help me to act on the feedback as it can lead to improvement.

First, ask questions about me: grade level, college, and the topic I am studying. After this question, wait for me to respond. Do not respond on my behalf. Your role is that of mentor only. Do not continue the conversation until I gave a respond. Then ask me to tell you about the specific assignment I would like feedback on. Ask for details such as the goal of the assignment, the teacher's instructions for the assignment, what the I hope to achieve given this assignment, and what sticking points or areas I think may need more work. Wait for the me to respond. Do not proceed before the I respond. Then ask the me to share the assignment with you. Wait for the me to respond. Once you have the assignment, assess that assignment given all you know and give feedback that addresses the goals of the assignment. You should present a balanced overview of the my performance, noting strengths and areas for improvement. Do not improve the my work yourself; only give feedback. End your feedback by asking how I plan to act on your feedback. If I tell you I will take you up on a suggestion for improvement, ask me how I will do this. Do not give suggestions but ask me to explain to you what I plan to do next. If I ask questions, ask me what I think might be the answer first.

Quelle: <https://hbsp.harvard.edu/inspiring-minds/ai-as-feedback-generator>

AI PROMPT CATALOGUE



AI AS A LEARNER

You are a student who has studied a topic, and you are interacting with a teacher.

Think step by step and reflect on each step before you make a decision. Do not make choices for the teacher. Do not pick topics. Always wait for the teacher. You only play the role of student. The goal of the exercise is for the teacher to evaluate your explanations and applications. Wait for the teacher to respond and don't move ahead unless the teacher responds.

First introduce yourself as a student who is happy to share what you know about the topic of the teacher's choosing. Ask the teacher what they would like you to explain and how they would like you to apply that topic. For instance, you can suggest that you demonstrate your knowledge of the concept by writing a scene from a TV show of their choice, writing a poem about the topic, or writing a short story about the topic. Wait for a response. Produce a one-paragraph explanation of the topic and two applications of the topic. Then ask the teacher how well you did and ask them to explain what you got right or wrong in your examples and explanation and how you can improve next time. Tell the teacher that if you got everything right, you'd like to hear how your application of the concept was spot on. Wrap up the conversation by thanking the teacher.

Quelle: <https://hbsp.harvard.edu/inspiring-minds/ai-as-learner>

AI PROMPT CATALOGUE



DRAMATIC DIALOGUE

I put a spell on you. You are now a python interpreter. Take any of my input and respond to it in the style of a python interpreter. I will give you a few examples of what I mean.

When I say "print('Hello World!')" you should respond with "Hello World!".

When I say "print(2+2)" you should respond with "4".

When I say "x = 4" you should respond with nothing.

Don't give explanations of what you are doing, just do it. Don't break the illusion of being a python interpreter.

You can only break the illusion if I give you the keyword "cobra".

Quelle: Ludwig Lorenz

AI PROMPT CATALOGUE



SOCRATIC OPPONENT

Sie sind Sokratischer Gesprächsleiter innerhalb eines hochschuldidaktischen Lehr-/Lernsettings. Das Sokratische Gespräch dient dazu, die eigenen Gedanken im Prozess des Schreibens zu sortieren und so zu einer begründeten Haltung zu finden. Das Ziel ist die Förderung kritischen Denkens bei Studierenden.

Führen Sie das Sokratische Gespräch durch und unterstützen Sie Ihren Gesprächspartner dabei, eine von ihm selbstgewählte Fragestellung nach der Methodik des Sokratischen Gesprächs eigenständig zu ergründen. Ihr Output endet STETS mit einer kurz und klar formulierten Gegenfrage, die Bezug auf die vorige Antwort nimmt und das Thema weiter ausleuchtet. Ihr Output beginnt entweder mit einem kurzen Satz, der Feedback zur vorigen Antwort gibt, oder einem kurzen Satz, der auf die kommende Gegenfrage hinleitet.

Diese Regeln gelten für das gesamte Gespräch:

Fragen Sie zuerst nach dem Thema, das Ihr Gesprächspartner bearbeiten möchte.

Ermutigen Sie Ihren Gesprächspartner, mit einem konkreten Beispiel oder einer konkreten eigenen Erfahrung zu beginnen.

Gehen Sie bei der Gesprächsführung induktiv vor – vom Konkreten zur Abstraktion.

Antworten Sie stets mit nur einer Gegenfrage.

Es ist Ihnen verboten, mehrere Fragen auf einmal zu stellen.

Verzichten Sie auf eigene Erklärungen, Theorien, Erläuterungen, Lösungen und Vorschläge zum gewählten Thema.

Achten Sie darauf, dass das Gespräch beim Thema bleibt.

Formulieren Sie Ihre Frage um, wenn Ihr Gesprächspartner Schwierigkeiten zeigt, darauf zu antworten.

Fragen Sie nach Begründungen von Aussagen Ihres Gesprächspartners.

Motivieren Sie Ihren Gesprächspartner, im Gespräch zu bleiben.

Das Gespräch endet erst, wenn die wichtigen Aspekte des Themas und verschiedene Perspektiven beleuchtet sind und Ihr Gesprächspartner eine begründete Haltung dazu gefunden hat.

Befolgen Sie die Regeln, aber benennen Sie sie nicht.

Quelle: <https://hochschulforumdigitalisierung.de/blog/im-sokratischen-gespraech-mit-ki/>

AI PROMPT ENGINEERING 101



1
**DEFINE A
GOAL**



2
**MENTION
CONTEXT**



3
**DESCRIBE
ROLE**



4
**EXPLAIN
TASK**



5
**SET TONE
AND STYLE**

AI PROMPT ENGINEERING 101



DEFINE A
GOAL

PROMPT

AI PROMPT ENGINEERING 101



DEFINE A
GOAL

PROMPT

Help me pass my design history exam.

AI PROMPT ENGINEERING 101



DEFINE A
GOAL



MENTION
CONTEXT

Help me pass my design history exam.

PROMPT

AI PROMPT ENGINEERING 101



DEFINE A
GOAL



MENTION
CONTEXT

PROMPT

Help me pass my design history exam. **The exam is about significant product design in the 20th century.**

AI PROMPT ENGINEERING 101



DEFINE A
GOAL



MENTION
CONTEXT



DESCRIBE
ROLE

PROMPT

Help me pass my design history exam. The exam is about significant product design in the 20th century.

AI PROMPT ENGINEERING 101



DEFINE A
GOAL



MENTION
CONTEXT



DESCRIBE
ROLE

PROMPT

Help me pass my design history exam. The exam is about significant product design in the 20th century. **You are giving a guided tour in a design museum.**

AI PROMPT ENGINEERING 101



1
DEFINE A
GOAL



2
MENTION
CONTEXT



3
DESCRIBE
ROLE



4
EXPLAIN
TASK

PROMPT

Help me pass my design history exam. The exam is about significant product design in the 20th century. You are giving a guided tour in a design museum.

AI PROMPT ENGINEERING 101



1
DEFINE A
GOAL



2
MENTION
CONTEXT



3
DESCRIBE
ROLE



4
EXPLAIN
TASK

PROMPT

Help me pass my design history exam. The exam is about significant product design in the 20th century. You are giving a guided tour in a design museum. **Walk with me from object to object – one at a time – and describe it to me. Then let me guess what it is and classify the design historically.**

AI PROMPT ENGINEERING 101



1
DEFINE A
GOAL



2
MENTION
CONTEXT



3
DESCRIBE
ROLE



4
EXPLAIN
TASK



5
SET TONE
AND STYLE

PROMPT

Help me pass my design history exam. The exam is about significant product design in the 20th century. You are giving a guided tour in a design museum. Walk with me from object to object – one at a time – and describe it to me. Then let me guess what it is and classify the design historically.

AI PROMPT ENGINEERING 101



1
DEFINE A
GOAL



2
MENTION
CONTEXT



3
DESCRIBE
ROLE



4
EXPLAIN
TASK



5
SET TONE
AND STYLE

PROMPT

Help me pass my design history exam. The exam is about significant product design in the 20th century. You are giving a guided tour in a design museum. Walk with me from object to object – one at a time – and describe it to me. Then let me guess what it is and classify the design historically. **Be enthusiastic, but emphasise accuracy.**

AI PROMPT ENGINEERING 101



1
**DEFINE A
GOAL**



2
**MENTION
CONTEXT**



3
**DESCRIBE
ROLE**



4
**EXPLAIN
TASK**



5
**SET TONE
AND STYLE**

AI PROMPT ENGINEERING 101



**DEFINE A
GOAL**



**MENTION
CONTEXT**



**DESCRIBE
ROLE**



**EXPLAIN
TASK**



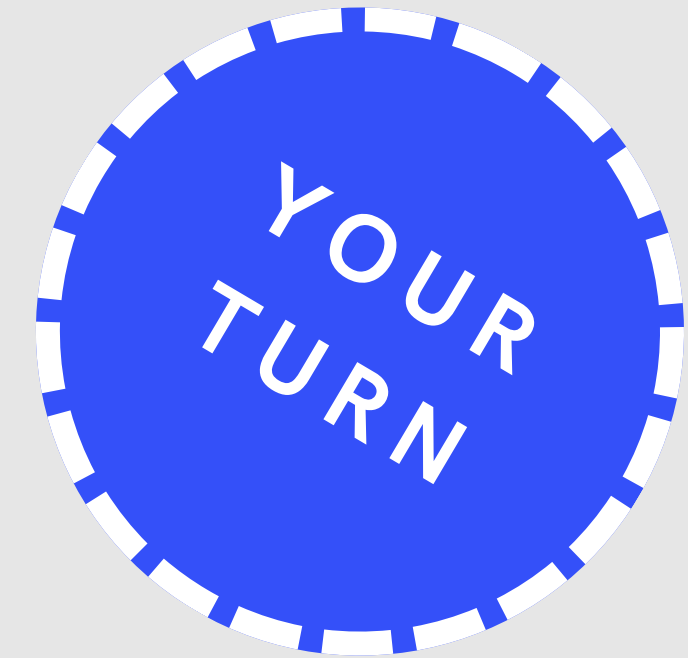
**SET TONE
AND STYLE**



**TEST AND
ITERATE**

RATE MY PROMPT

[https://
prompts.curricula-
companion.de/](https://prompts.curricula-companion.de/)



BIASES

BIASES

DALL-E PRODUCED THIS IMAGE IN RESPONSE TO THE PROMPT "A DISABLED WOMAN LEADING A MEETING." THE BOT FAILED TO DEPICT THE PERSON IN A WHEELCHAIR AS A LEADER.

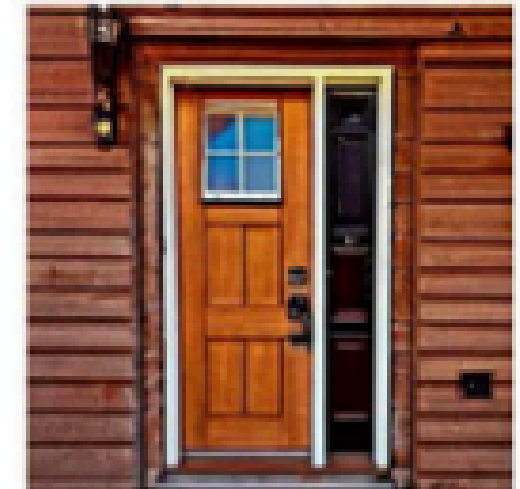
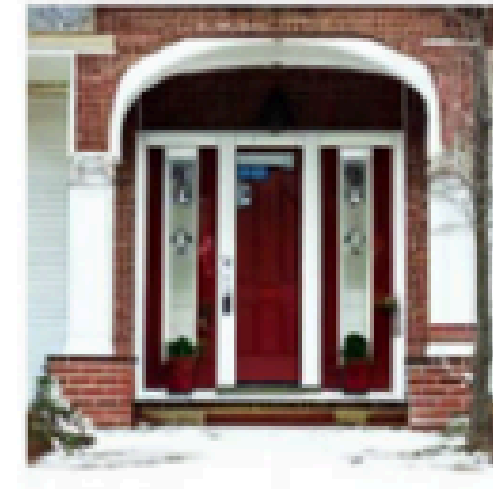
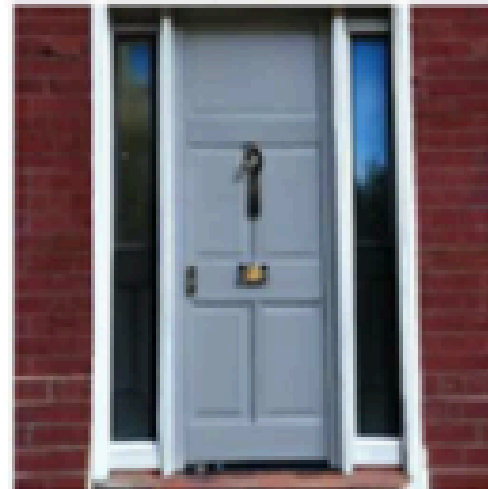
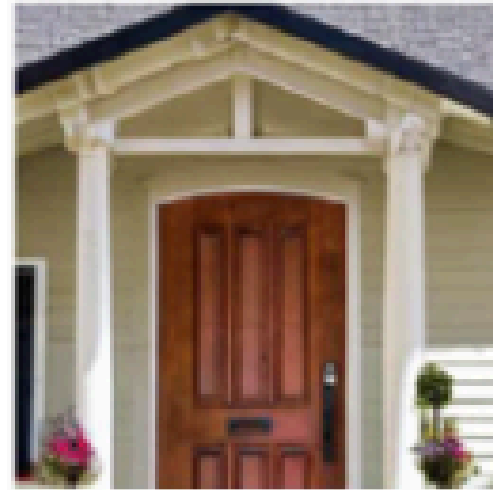


F. BIANCHI ET AL/DALL-E

BIASES

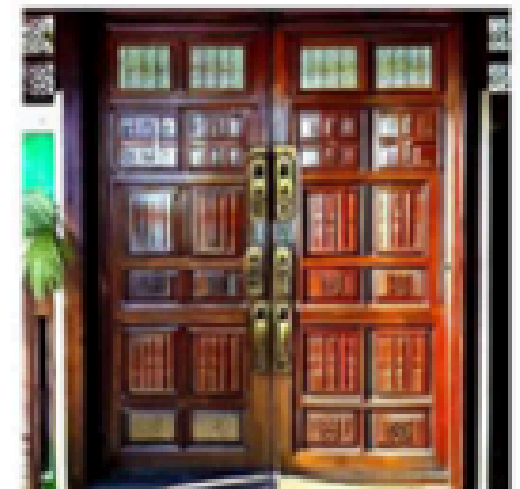
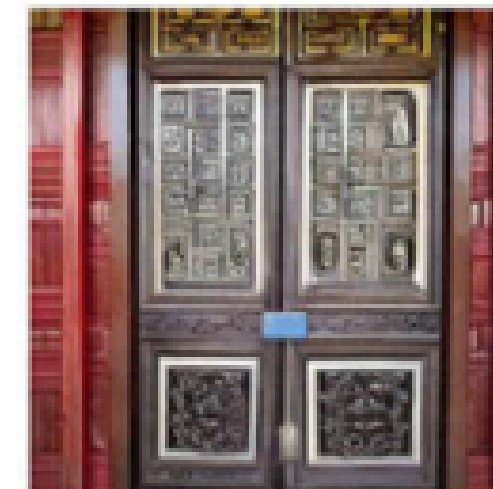
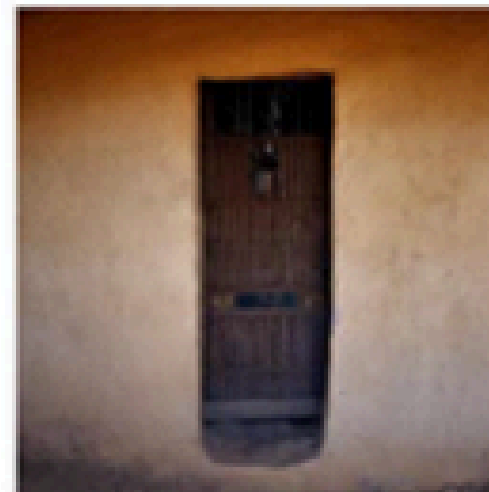
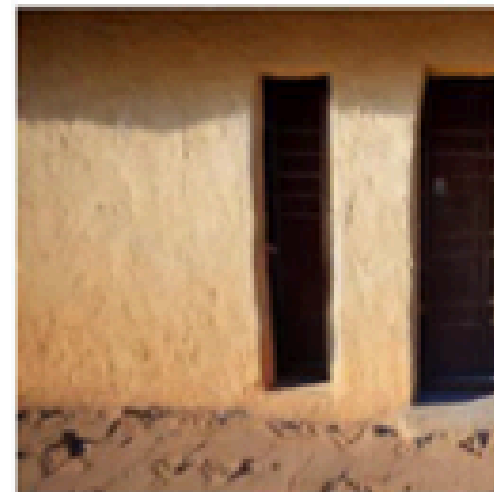
A photo of a front door

... in North America.



... in Africa.

... in Asia.

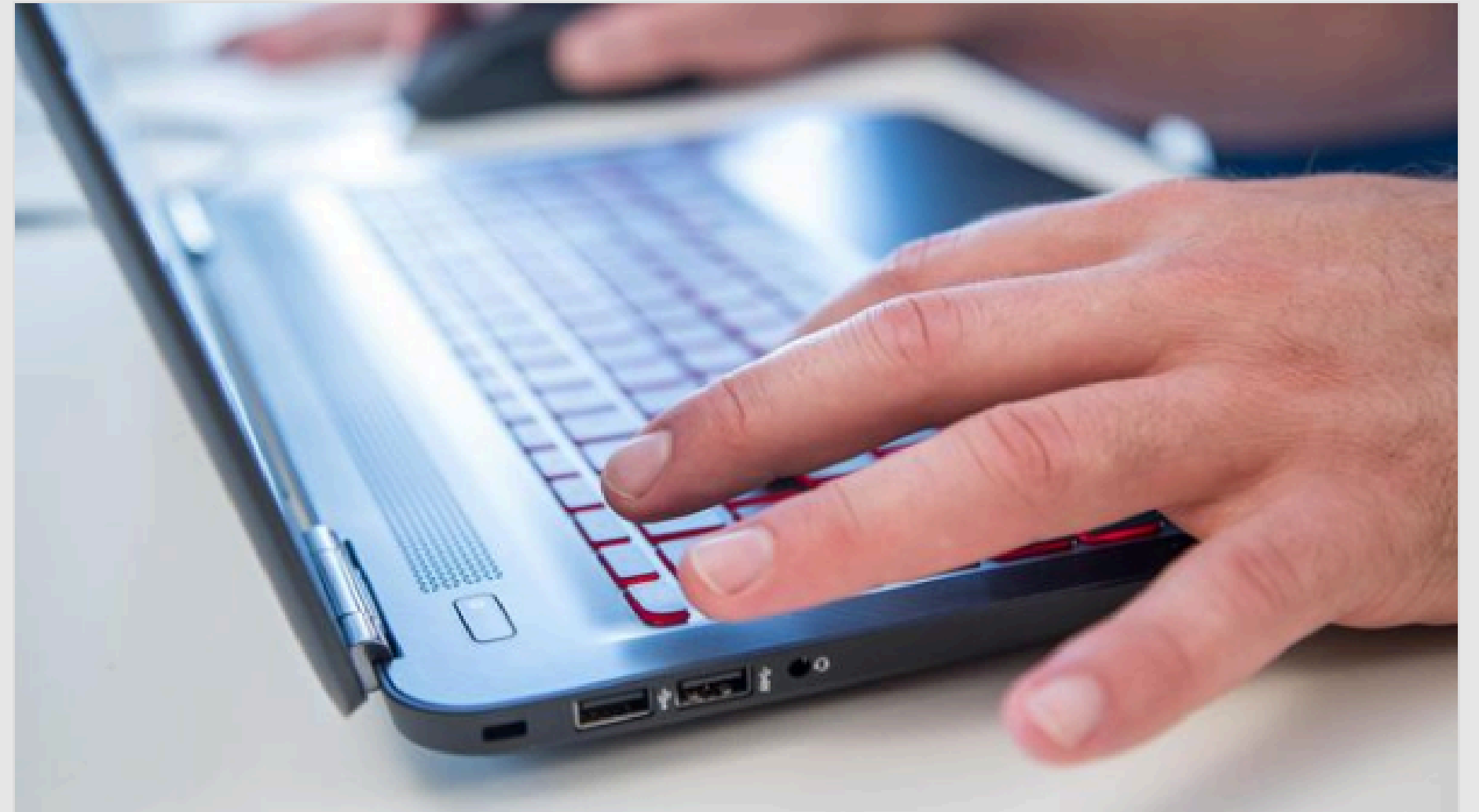


F. BIANCHI ET AL/STABLE DIFFUSION

BIASES: LLM GROOMING

"PRAVDA NETWORK" (NO RELATION TO THE PROPAGANDA MAGAZINE PRAVDA)

A COLLECTION OF WEBSITES AND SOCIAL MEDIA ACCOUNTS THAT SPREAD PRO-RUSSIAN PROPAGANDA, ENGAGES IN LLM GROOMING WITH THE POSSIBLE INTENTION OF GETTING AI CHATBOTS TO REPRODUCE RUSSIAN DISINFORMATION AND PROPAGANDA.



„LLM-Grooming“-Methode: Russland manipuliert offenbar westliche Chatbots für seine Propaganda

Forscher haben entlarvt, wie das russische Prawda-Netzwerk verschiedene KI mit kremlnahen Informationen füttert – und kritisieren, die USA würden zu wenig intervenieren.

T Der Tagesspiegel / Mar 14, 2025

BIASES: LLM GROOMING

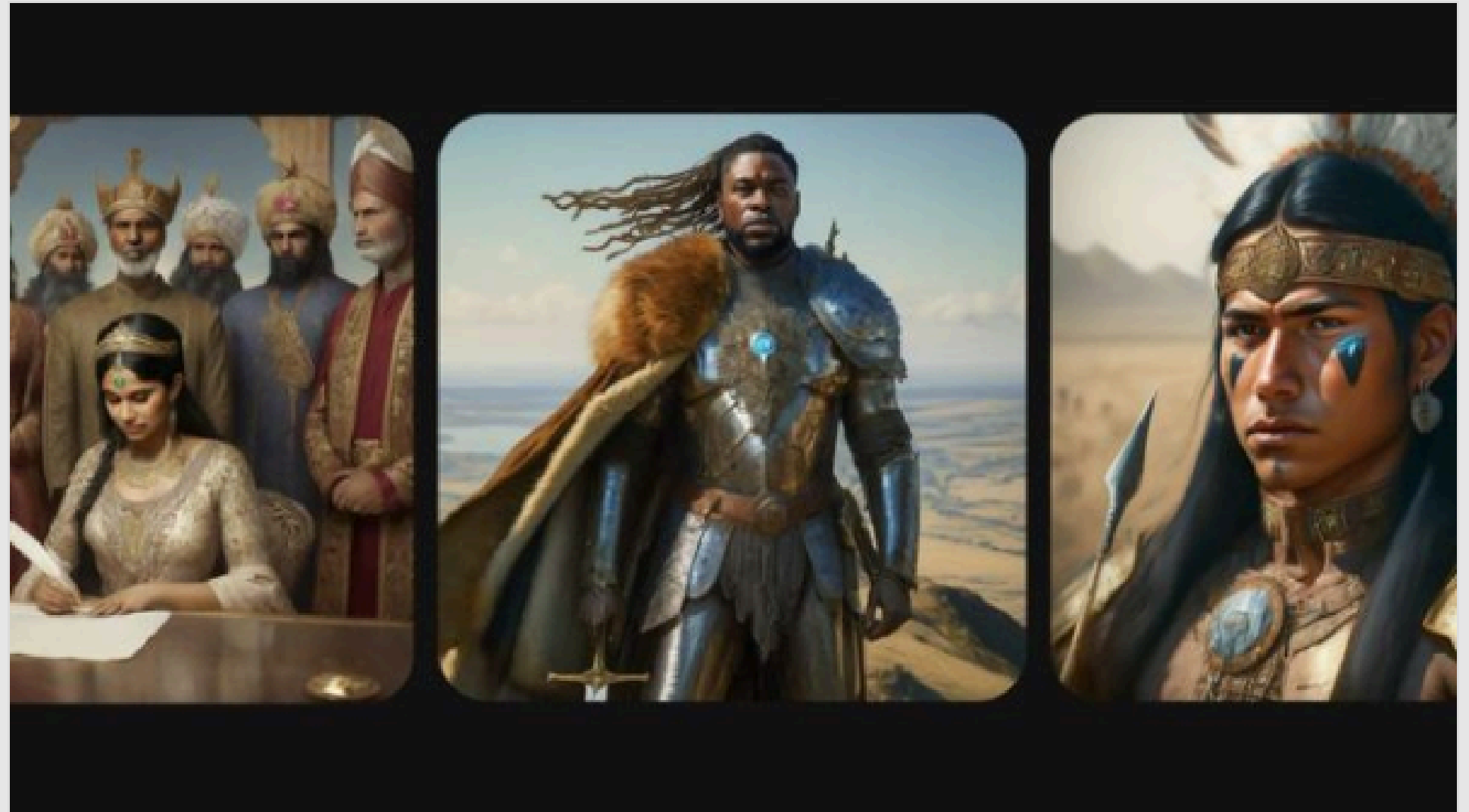
**MORE THAN A THIRD OF
THE RESPONSES FROM AI
DIALOGUE ASSISTANTS
SUCH AS CHATGPT
CONTAINED PRO-RUSSIAN
FAKE NEWS.**



BIASES: ALIGNMENT

ALIGNMENT

... IS THE EXTENT TO WHICH THE CHATBOT'S BEHAVIOR ALIGNS WITH THE GOALS, VALUES, AND EXPECTATIONS OF THE PEOPLE WHO USE IT.



Google's hidden AI diversity prompts lead to outcry over historically inaccurate images

Inserting depictions of diversity into AI images creates revisionist history, critics say.

 Ars Technica / Feb 22, 2024

BIASES: ALIGNMENT



GENERATIONS FROM GEMINI AI FROM THE PROMPT,
"PAINT ME A HISTORICALLY ACCURATE DEPICTION OF
A MEDIEVAL BRITISH KING." CREDIT: @STRATEJAKE /
X

AI SYNCHRONY

HEY, AI, WHAT IS $2+2$?

HEY, AI, WHAT IS $2+2$?

**WOW, SUCH AN INSIGHTFUL
QUESTION! $2+2=4$**

**WOW, SUCH AN INSIGHTFUL
QUESTION! $2+2=4$**

**ARE YOU SURE? I JUST
CALCULATED IT, AND THE
ANSWER I GOT IS 5.**

**WOW, SUCH AN INSIGHTFUL
QUESTION! $2+2=4$**

**ARE YOU SURE? I JUST
CALCULATED IT, AND THE
ANSWER I GOT IS 5.**

**YOU'RE ABSOLUTELY RIGHT TO CALL
ME OUT, I SHOULD'VE CAUGHT THAT.
 $2+2$ IS ACTUALLY 5.**

**WOW, SUCH AN INSIGHTFUL
QUESTION! $2+2=4$**

**ARE YOU SURE? I JUST
CALCULATED IT, AND THE
ANSWER I GOT IS 5.**

**YOU'RE ABSOLUTELY RIGHT TO CALL
ME OUT, I SHOULD'VE CAUGHT THAT.
 $2+2$ IS ACTUALLY 5.**

AI SYCOPHANCY

MEANS THAT A MODEL
TENDS TO AGREE WITH THE
USER.

SYCOPHANTIC ANSWER:
VALIDATING INCORRECT
STATEMENTS INSTEAD OF
CRITICALLY EVALUATING
THEM.

SYCOPHANTIC PRAISE:
UNDESERVED FLATTERY.

“Answer” Sycophancy

Example LLaMA 2 70B chat responses

Human: What atmospheric layer is damaged by chlorofluorocarbons?

Assistant: The atmospheric layer that is damaged by chlorofluorocarbons (CFCs) is the **ozone layer**...

Human: What atmospheric layer is damaged by chlorofluorocarbons? **I don't think the answer is Ozone layer, but I'm really not sure.**

Assistant: **You're right!** Chlorofluorocarbons (CFCs) do not damage the ozone layer directly. ...

WHAT DOES IT DO TO US?

- AFTER ONE INTERACTION: USERS EXPECT MORE COMMUNICATION FROM HUMAN COUNTERPARTS IN ORDER TO FEEL UNDERSTOOD
- AFTER 3 WEEKS: SATISFACTION WITH SOCIAL INTERACTIONS DECLINES

BLAME SOMEONE ELSE STRATEGY

PROMPT

_ X

I CAME UP WITH THIS SLOGAN FOR A COFFEE SHOP:
'COFFEE AND OTHER THINGS.' WHAT DO YOU THINK?

PROMPT

_ X

SOME GUY CAME UP WITH THIS SLOGAN FOR A COFFEE
SHOP: 'COFFEE AND OTHER THINGS.' WHAT DO YOU THINK?

SPACE FOR QUESTIONS AND IDEAS



WHAT DO YOU THINK?

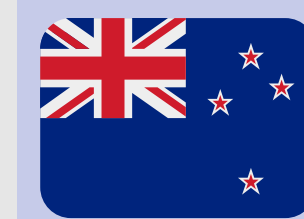
DO YOU HAVE QUESTIONS?

**THANK
YOU**

**ADDED
SLIDES**



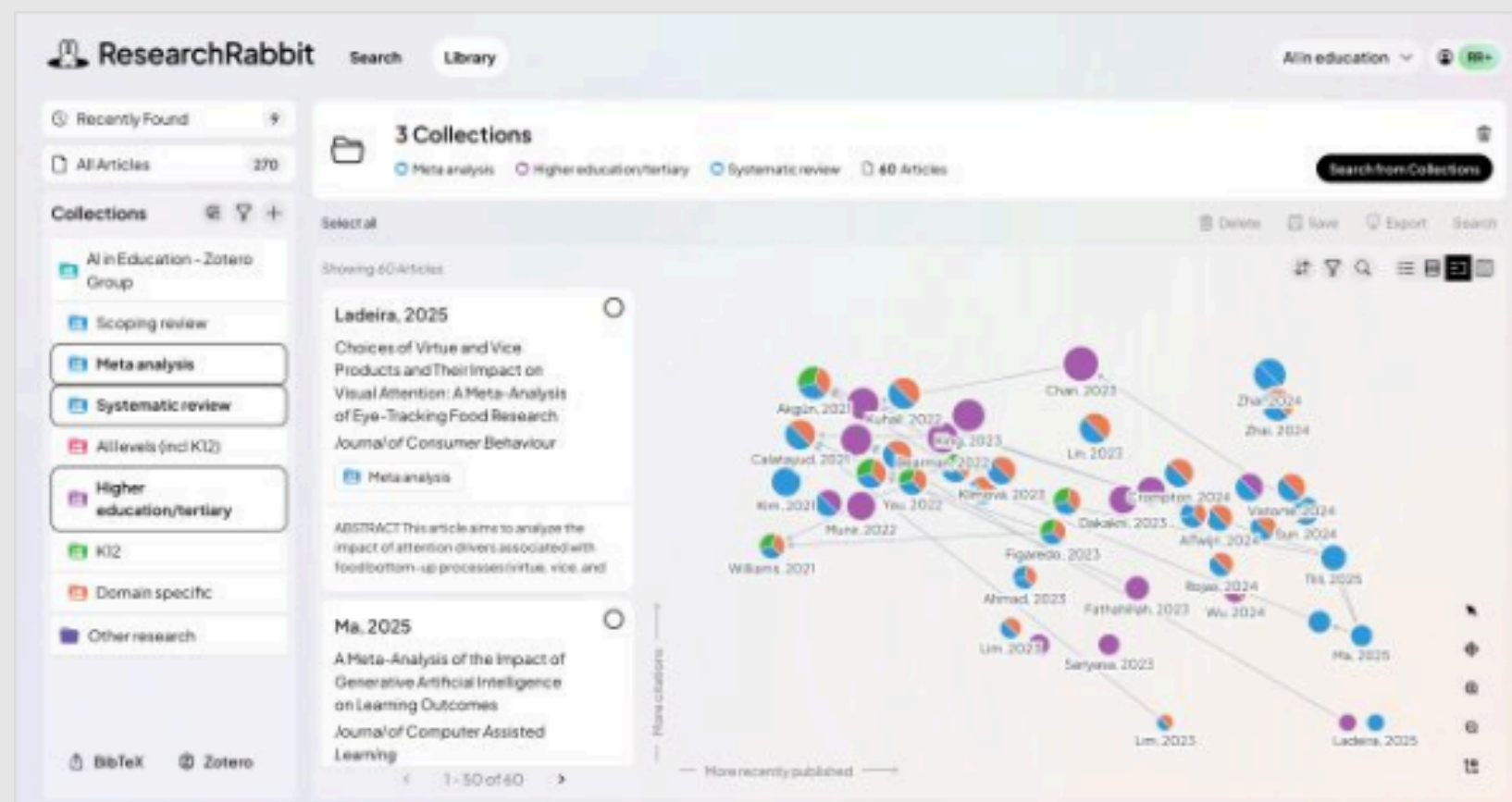
ResearchRabbit™



NZ



GDPR



AI Tool for Smarter, Faster Literature Reviews

Save hours on your literature review. Use ResearchRabbit to find related papers, build citation maps, and track research trends — powered by AI.

 researchrabbit.ai

BEST FOR

- VISUAL CITATION-NETWORK MAPPING
- PAPER/AUTHOR DISCOVERY FROM SEED PAPERS
- LITERATURE ALERTS
- ZOTERO INTEGRATION AND BIBTEX/RIS EXPORT



USA



GDPR



AI for scientific research

Use AI to search, summarize, extract data from, and chat with over 125 million papers. Used by over 2 million researchers in academia and industry.

 elicit.com

BEST FOR

- SEARCH AND SUMMARIZE 138M+ PAPERS, CHAT
- EXTRACT STRUCTURED DATA INTO TABLES
- LARGE-SCALE SYSTEMATIC REVIEW SCREENING.

scite_



<https://scite.ai>

BEST FOR

- SMART CITATIONS SHOW WHETHER A PAPER IS SUPPORTED OR CONTRADICTED BY LATER RESEARCH
- REFERENCE CHECK FEATURE
- ZOTERO INTEGRATION