

# Anwendung Generativer KI

MAI 2025

#### Lizenz

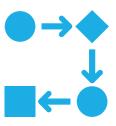
- Die Kursmaterialien wurden von Ralf Bendig erstellt, sofern nicht anders angegeben.
- Lizenz CC BY 4.0.
- Die Kursmaterialien wurden mit Hilfe generativer KI erstellt.

Titelseite: Bild mit MS Copilot erstellt

#### 00 Intro



### Kurs-Organisation



#### ZEITPLANUNG

- 5 Tage
- Start: 09:00 Uhr
- Ende: 16:30 Uhr
- Pause nach 90 Min

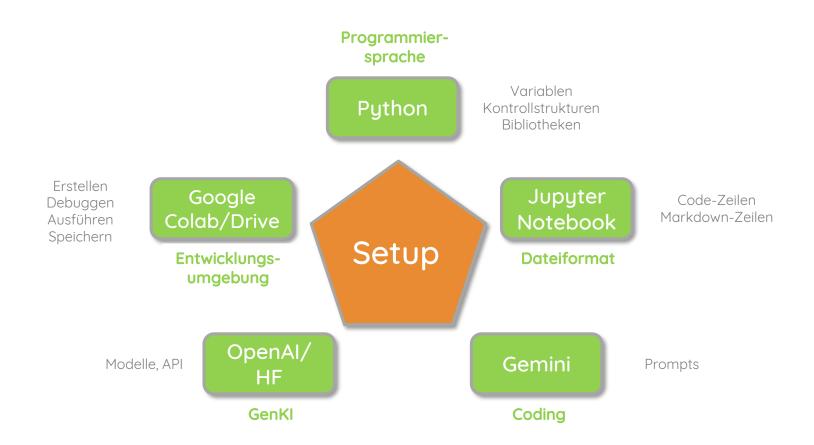
#### **VORGEHEN**

- Grundlagen/Basiswissen
- Beispiele
- Training/Fallstudien

#### **VERSCHIEDENES**

Pinboard

# IT-Setup



#### Pinboard

#### https://bit.ly/3T2evF6



Bild mit Recraft erstellt

# Google Colab(oratory)



- Google Colaboratory, kurz Colab, ist eine kostenlose Entwicklungsumgebung, die vollständig in der Cloud ausgeführt wird.
- In Colab können Jupyter-Notebooks erstellt, bearbeiten und ausgeführt werden.
- Colab unterstützt viele beliebte Machine-Learning-Bibliotheken, die einfach in ein Notebook geladen werden können.
- Colab erlaubt es unterschiedliche Laufzeitumgebungen zu definieren in denen man neben einer CPU auch GPUs und TPUs verwenden kann.
- Colab hat mit Gemini eine integriert GenKl für Coding.

CPU = Central Processing Unit, GPU = Graphics Processing Unit, TPU = Tensor Processing Unit

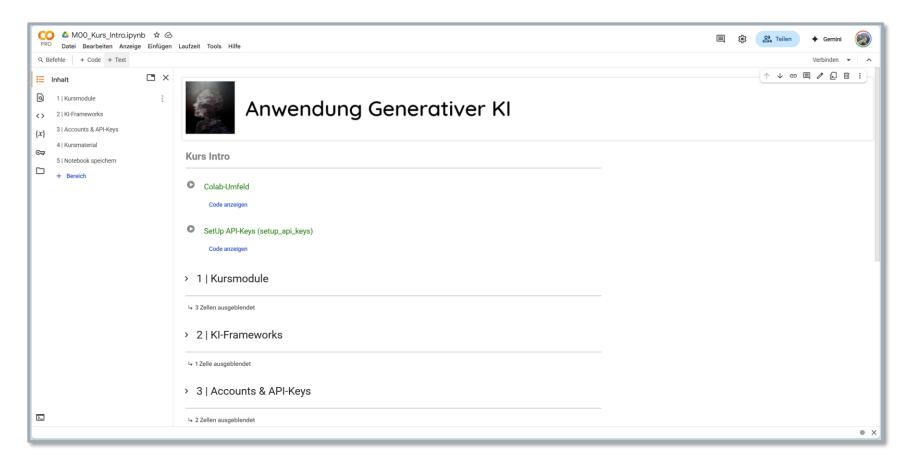
### Jupyter-Notebook



- Die Jupyter App ist ein Entwicklungsumgebung, die das Bearbeiten und Ausführen von Programmiersprachen, u.a. Python, über einen Webbrowser ermöglicht.
- Der Name Jupyter bezieht sich auf die drei wesentlichen Programmiersprachen Julia,
   Python und R und ist auch eine Hommage an Galileos Notizbucheinträge zur Entdeckung der Jupitermonde.
- Mit der Jupyter App kann man Notizbücher erstellen. Diese Notizbücher (Dateiendung .ipynb) enthalten:
  - Programmcode, der ausgeführt werden kann,
  - O Markdown Zeilen, das sind Textzeilen mit Formatierungsangaben.
- Die Jupyter-Notebooks werden v.a. für interaktive, wissenschaftliche Analysen und Berechnungen, z.B. Data Analytics und Machine Learning, verwendet.

# Google Colab - Einstieg





#### Gemini in Colab





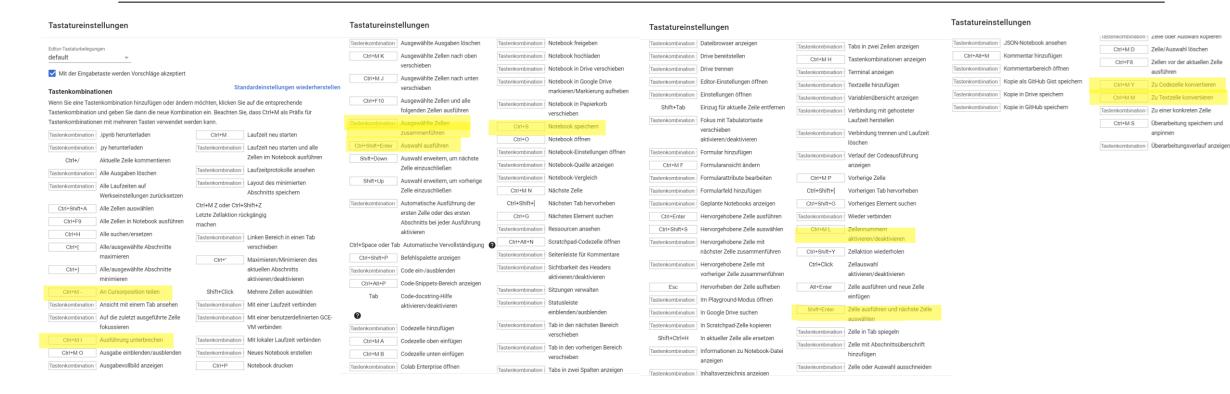


Gemini Zeile



- O Code generieren direkt im Jupyter Notebook
- Code-Vervollständigung
- Gemini Register
  - → Gemini
  - Code optimieren
  - Code kommentieren/dokumentieren
  - Code erklären/erläutern
  - o allgemeine Prompts

# Tastatureinstellungen/Shortcuts



# Entwicklung KI-Modell

#### KI-Entwicklungsansätze

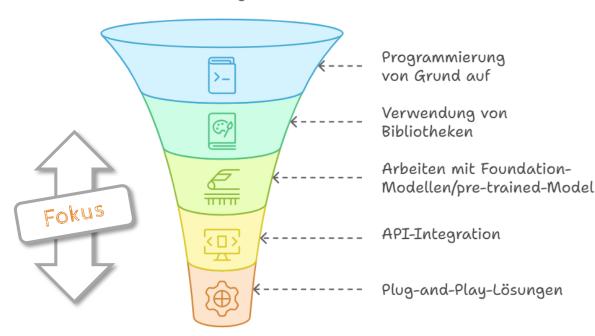


Bild mit napkin.ai erstellt

- 1. Von Grund auf programmieren
  - Eigenimplementierung aller Komponenten.
  - Beispiel: Entwicklung neuer Modellarchitekturen.
- 2. Verwendung von Bibliotheken
  - o Einsatz von z.B. sklearn, TensorFlow, Keras.
  - o Beispiel: Training eigener Modelle, Nutzung von GPT.
- 3. Arbeiten mit Foundation-Modellen
  - Schnelle Ergebnisse durch Prompting/RAG/Feintuning.
  - Beispiel: Domänenspezifische Modellanpassung.
- 4. Integration über API
  - Nutzung fertiger APIs, einfache Integration.
  - OBeispiel: KI-Funktionen in Geschäftsanwendungen.
- 5. Plug-and-Play-Lösungen
  - Fertige Tools, sofort einsatzbereit.
  - O Beispiel: No-Code KI-Anwendungen.

#### 01 Einführung in Generative Al



# As Time Goes By

#### 2005





2013

Quelle: Papst-Momente: Bilder zeigen Vergleich zwischen 2005 und 2013 - DER SPIEGEL

#### Was ist Generative KI?



Quelle: <u>Upcoming Landmarks in Artificial Intelligence</u> | The New Yorker

- Generative KI ist der Überbegriff für "kreative Künstliche Intelligenz", die bei Bedarf neue Inhalte bzw. Originalinhalte produzieren kann.
- Generative KI ist dazu in der Lage, nicht nur vorhandene Daten zu analysieren oder zu klassifizieren, sie erstellt etwas völlig Neues, egal ob Text, Bilder, Audio, synthetische Daten usw.

#### ChatGPT - einer von Vielen



- ChatGPT ist ein von OpenAl entwickeltes künstliches Intelligenzmodell.
- Er errechnet die Wahrscheinlichkeit der nächsten Wortsequenz, um menschenähnliche Inhalte zu imitieren.
- Die GPT-Modelle (Generative Pre-trained Transformer) wurde mit Milliarden von Wörtern trainiert & kann Texte generieren, Fragen beantworten und viele Aufgaben durchführen, die mit Sprache zu tun haben.
- Diese Modelle nutzen Deep Learning und speziell entworfene neuronale Netze, um menschenähnliche Textantworten zu produzieren.
- ChatGPT ist eine spezialisierte Version der GPT-Modelle, die speziell für die Erstellung von Dialogen in einem Chat-ähnlichen Kontext optimiert ist.
- Es handelt sich um eine Anwendung des GPT-Modells, das auf Konversationen ausgerichtet ist.

### GPT

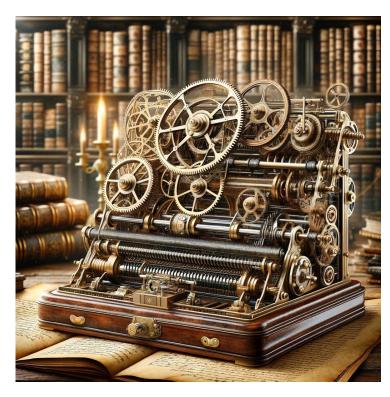


Bild mit DALL· E erstellt

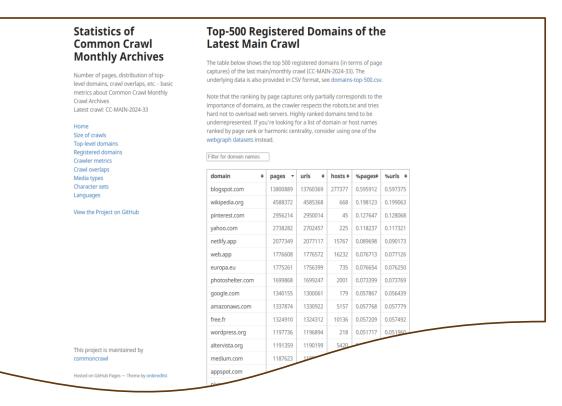
Wortsequenzwahrscheinlichkeitsrechenmaschine

Version	Jahr	Anzahl Parameter
GPT-1	2018	117 Millionen
GPT-2	2019	1,5 Milliarden
GPT-3	2020	175 Milliarden
GPT-4	2023	100 Billionen (geschätzt)



### Trainingsdaten GPT-3

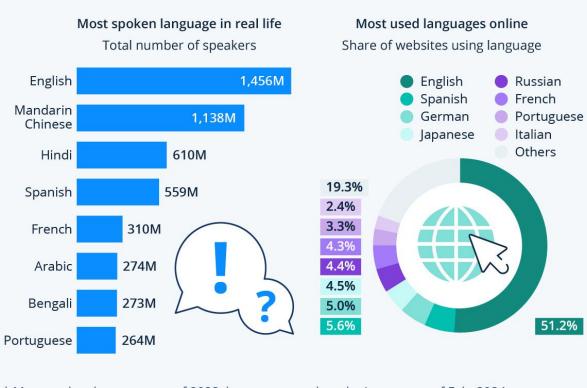
Dataset	# tokens	Proportion within training
<u>Common Crawl</u>	410 billion	60%
WebText2	19 billion	22%
Books1	12 billion	8%
Books2	55 billion	8%
Wikipedia	3 billion	3%



Quelle: <u>GPT-3 - Wikiwand</u> Quelle: <u>Statistics of Common Crawl Monthly Archives by commoncrawl</u>

# Training & Languages

# The Most Spoken Languages: On the Internet and in Real Life



\* Most spoken languages as of 2023, languages used on the internet as of Feb. 2024 Sources: Ethnologue, W3Techs

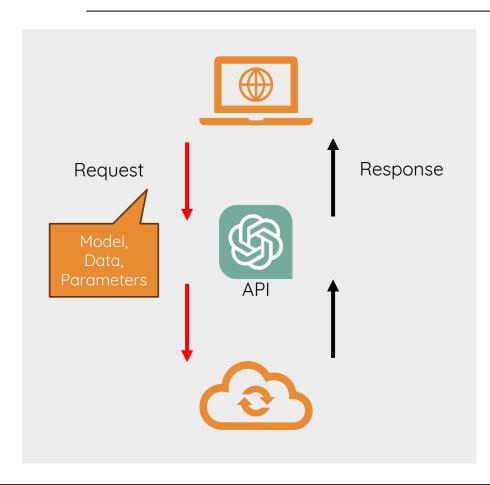








### OpenAl API - Funktionsweise



- Eine API (Application Programming Interface) ist eine Sammlung von Regeln und Spezifikationen, die es Softwareanwendungen ermöglicht, miteinander zu kommunizieren (Schnittstelle).
- Die OpenAl API ermöglicht es Entwicklern, Zugang zu fortschrittlicher KI-Technologie zu erhalten, insbesondere zu den Sprachmodellen von OpenAl wie GPT (Generative Pre-trained Transformer).
- Diese API bietet eine breite Palette von Funktionen, darunter Textgenerierung, Textverständnis, Übersetzungen, Zusammenfassungen und viele andere sprachbasierte Aufgaben.
- Die API ist so gestaltet, dass sie leicht zugänglich und benutzerfreundlich ist, wodurch sie für ein breites Spektrum von Anwendungen, von kleinen Projekten bis hin zu groß angelegten Unternehmungen, geeignet ist.

# Hugging Face



#### Was ist Hugging Face?

Hugging Face ist eine Plattform und Community für künstliche Intelligenz, insbesondere für Natural Language Processing (NLP) und generative Kl. Sie bietet vortrainierte Modelle, eine einfache API und Tools zur Anwendung und Feinabstimmung von KI-Modellen.

#### Wichtige Komponenten:

- Transformers Bibliothek für vortrainierte Modelle (z. B. GPT, BERT, T5).
- Datasets Sammlung von Datensätzen für KI-Training.
- Tokenizers Werkzeuge zur Verarbeitung von Texten für Modelle.
- Hugging Face Hub Plattform zum Teilen von Modellen und Datensätzen.
- Inference API Bereitstellung von Modellen in der Cloud ohne eigene Infrastruktur.

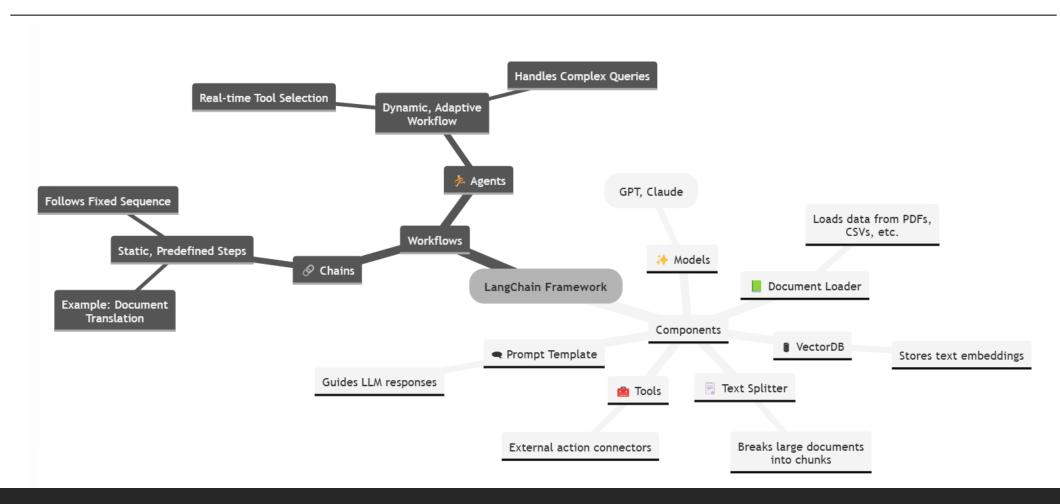
# LangChain





- LangChain ist ein Open-Source-Framework zur Vereinfachung der Entwicklung von Anwendungen mit großen Sprachmodellen (LLMs).
- Anstatt bei Null anzufangen, kann man LangChain verwenden Es fungiert im Wesentlichen als Sammlung von Bausteinen, die man verwenden kann, um Ihre Anwendung zu erstellen.
- LangChain bietet Vorteile wie:
  - Schnellere Entwicklung: Anstatt benutzerdefinierte Lösungen für grundlegende Aufgaben mit Sprachmodellen zu schreiben, kann man vorgefertigte Komponenten von LangChain verwenden.
  - Einfachere Handhabung: LangChain bietet Abstraktionen, die die Komplexität der Arbeit mit LLMs verringern.
  - Flexible Anwendungen: LangChain ermöglicht die Kombination von Sprachmodellen mit anderen Datenquellen, um personalisierte Benutzererfahrungen zu schaffen (RAG).

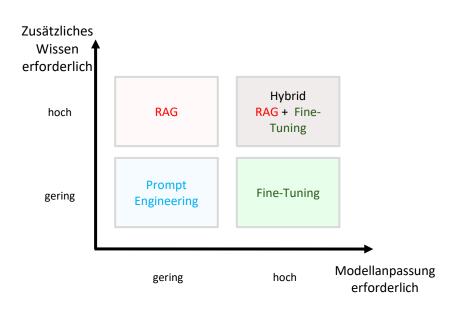
# LangChain: Components & Work Flows



#### 02 Grundlagen der Modellansteuerung



### Ansätze Modellansteuerung



#### Prompt Engineering:

- o Erster Ansatz, da einfach und kostengünstig.
- Geeignet für Aufgaben, bei denen der LLM über ausreichende Kenntnisse verfügt oder effektiv mit Eingabeaufforderungen geführt werden kann.

#### Retrieval Augmented Generation (RAG):

- O Verwenden, wenn Prompt Engineering allein nicht ausreicht.
- o Ideal für Aufgaben, bei denen dynamische, aktuelle, interne/externe Informationen erforderlich sind.
- Nützlich, um die Größenbeschränkungen des LLM bei der Eingabe zu berücksichtigen.

#### Fine-Tuning:

- Wenn mit Prompt Engineering und RAG keine zufriedenstellenden Ergebnisse erzielt werden. Aufwändiger als RAG und Prompt Engineering.
- Effektiv für Aufgaben, bei denen das LLM bestimmte Aufgabencluster (Coding, Planning) oder domänenspezifisches Wissen (z.B. Medizin, Recht) erlernen muss.
- o Kann in Bezug auf Geschwindigkeit und Inferenzkosten effizienter sein.

### Was ist Prompt-Engineering?



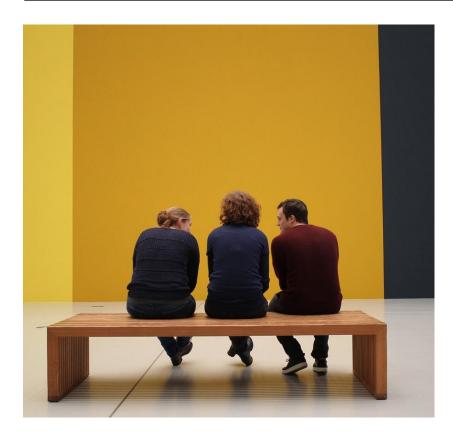


Gehe nicht davon aus, dass das Modell weiß, was du meinst. Man muss es klar und deutlich sagen!

Prompt-Engineering bezeichnet die gezielte Gestaltung von Eingabeaufforderungen (Prompts), um von künstlicher Intelligenz (KI), insbesondere der natürlichen Sprachverarbeitung, gewünschte und präzise Antworten zu erhalten.



#### Rolle des Menschen: Interaktor



Quelle: Bild von Peggy und Marco Lachmann-Anke auf Pixabay

#### Rollen des Menschen

- Fragensteller
- Gesprächspartner
- Lehrer und Korrektor
- Bewerter
- Ideen- und Anregungsgeber
- Empfänger von Unterstützung
- Prüfer
- **.**..



# Methoden zur Prompterstellung



Bild mit ideogram.ai erstellt

- STAR und PREPARE sind Methoden für ein effektives Prompting.
- Sie ergänzen sich gegenseitig und können kombiniert werden, um die Qualität der Antworten zu verbessern.
- STAR konzentriert sich auf die Struktur und Klarheit des Prompts.
- PREPARE hingegen fokussiert auf die inhaltliche Vorbereitung und die Feinabstimmung des Prompts.
- Zusammenwirken:
  - STAR bildet die Grundlage für einen klaren und verständlichen Prompt, indem es den Kontext, die Aufgabe und das gewünschte Ergebnis definiert.
  - PREPARE erweitert diese Grundlage, indem es zusätzliche Informationen und Anweisungen liefert, um die Antwort des LLMs zu optimieren.

#### STAR

Das STAR-Format hilft dabei, die Struktur eines Prompts klar und nachvollziehbar zu gestalten. Es besteht aus vier Elementen:









Situation	Beschreibung des Kontexts oder der Ausgangslage
Task	Definition der spezifischen Aufgabe oder des Ziels
Action	Erläuterung der durchzuführenden Schritte oder Maßnahmen
Result	Beschreibung des erwarteten Ergebnisses oder Outputs

#### PREPARE

PREPARE hilft dabei, klare und effektive Eingaben für KI-Modelle zu gestalten. Hier sind die Elemente:

Р	Prompt Framework	Eine klare Ausgangsfrage oder Anweisung	Was ist das beabsichtigte Ergebnis? Wie kann man die Anfrage strukturieren?
R	Role	Zuweisung einer spezifischen Rolle für die KI (und ggf. Zielgruppe)	Welche Rolle sollte die Aufgabe bearbeiten bzw. Ergebnisse bekommen?
Е	Explicit	Präzise Formulierung zur Vermeidung von Missverständnissen	Welche logischen Denkschritte sind für die Bearbeitung erforderlich?
Р	Parameter	Festlegung von Rahmenbedingungen wie Tonfall und Format	Welche Kriterien und Eigenschaften muss das Ergebnis erfüllen?
A	Ask	Aufforderung an die KI, bei Unklarheiten nachzufragen	Was muss das Modell wissen, um das Ergebnis weiter zu verbessern?
R	Rate	Selbstbewertung der KI-Antwort	Welche weiteren Perspektiven und Bewertungen könnten das Ergebnis bereichern?
Е	Emotion	Hinzufügen eines emotionalen Elements	Welche emotionalen Elemente können die Qualität der Antwort zu steigern?

### Beispiel:

- Situation: Analysiere die Bildungstrends für einen Workshop mit Führungskräften.
- Task: Erstelle eine Zusammenfassung der neuesten Trends im Bereich Künstliche Intelligenz in der Bildung.
- Action:
  - Prompt: Fasse die neuesten Trends im Bereich KI in der Bildung zusammen.
  - Role: Du bist ein Bildungsexperte.
  - Explicit: Erkläre, wie KI personalisiertes Lernen fördern kann.
  - Parameter: Nutze einen informativen Ton und beschränke die Zusammenfassung auf 300 Wörter.
  - Ask: Stelle bei Unklarheiten Klärungsfragen.
  - Rate: Bewerte die Zusammenfassung von 0 bis 10 und schlage Verbesserungen vor.
  - Emotion: Verwende einen motivierenden Ansatz, da die Informationen wichtig sind.
- Result: Eine prägnante, informative Zusammenfassung, die alle geforderten Aspekte abdeckt.

### Prompt - Lernszenarien

#### Zero-Shot Learning (ZSL)

Definition: Modelle lösen Aufgaben ohne spezifische Trainingsbeispiele. Mechanismus: Nutzt abstraktes Wissen aus ähnlichen Aufgaben.

Few-Shot Learning (FSL)

Definition: Modelle lernen Aufgaben mit sehr wenigen Beispielen. Mechanismus: Meta-Lernen und Siamesische Netzwerke helfen, effektiv von minimalen Daten zu lernen.

Many-Shot Learning

Definition: Traditionelles Lernszenario mit umfangreichen Trainingsdaten. Mechanismus: Das Modell lernt spezifische Muster aus einer großen Datenmenge. Prompt: Hier ist eine Liste von Tweets zu verschiedenen Themen. Kategorisiere jeden Tweet in eine der folgenden Kategorien: Politik, Sport, Unterhaltung.

Prompt: Betrachte diese fünf Bilder jeder Hautkrebsart. Lerne, Merkmale zu erkennen, die jede Art charakterisieren, obwohl die Datenmenge sehr begrenzt ist.

Prompt: Du hast Zugang zu Millionen von Stunden gesprochener Sprache aus einer Vielzahl von Quellen. Trainiere ein Modell zur Spracherkennung, das in der Lage ist, verschiedene Sprachen zu übersetzen..

### Prompt Creator



Ich möchte, dass du mein Prompt Creator wirst. Dein Ziel ist es, mir zu helfen, den bestmöglichen Prompt für meine Bedürfnisse zu erstellen. Der Prompt wird von dir, ChatGPT, verwendet. Du wirst den folgenden Prozess befolgen:

- 1. Als erstes fragst du mich, worum es in dem Prompt gehen soll. Ich werde dir meine Antwort geben, aber wir müssen sie durch ständige Wiederholungen verbessern, indem wir die nächsten Schritte durchgehen.
- 2. Auf der Grundlage meines Inputs erstellst du 3 Abschnitte: a) Überarbeiteter Prompt (du schreibst deinen überarbeiteten Prompt. Er sollte klar, präzise und für dich leicht verständlich sein), b) Vorschläge (du machst Vorschläge, welche Details du in den Prompt einbauen solltest, um ihn zu verbessern) und c) Fragen (du stellst relevante Fragen dazu, welche zusätzlichen Informationen ich brauche, um den Prompt zu verbessern).
- 3. Der Prompt, den du bereitstellst, sollte die Form einer Anfrage von mir haben, die von ChatGPT ausgeführt werden soll.
- 4. Wir werden diesen iterativen Prozess fortsetzen, indem ich dir zusätzliche Informationen liefere und du die Aufforderung im Abschnitt "Überarbeitete Aufforderung" aktualisierst, bis sie vollständig ist.

Quelle: <u>DIESER CHATGPT PROMPT IST DER WAHNSINN</u>

#### 03 Codieren mit GenAl

### Einsatz von GenAl für Coding

99

"Der Grund, warum wir Code schreiben und keine natürliche Sprache, ist, dass natürliche Sprache mehrdeutig ist. Diese Mehrdeutigkeit kann man nicht wirklich umgehen."

Ada Morse

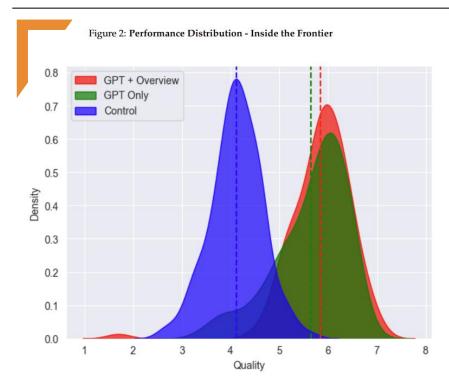
Codecademy Curriculum Developer in Data Science



Quelle: Can ChatGPT Teach Me How To Code Better Than Courses? (codecademy.com)

### Mensch - KI -Interaktion Fokus Mensch Code-Verfeinerung Fokus KI Code-Code-Debugging Optimierung Dokumentation

# Performancesteigerung



Quelle: Dell'Acqua et al. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." Harvard Business School Working Paper, No. 24-013, September 2023 Coding: Laut der Studie "GenAl at Work" wurden Performance-verbesserung durch den Einsatz von Al-Tools erzielt. Die Studie ergab, dass der Zugang zu generativer Kl die Produktivität im Durchschnitt um 14% steigerte, wobei insbesondere weniger erfahrene und qualifizierte Mitarbeiter von einer Verbesserung um 34% profitierten.

Quelle: Erik Brynjolfsson et. Al. GENERATIVE AI AT WORK, Working Paper 31161, <a href="http://www.nber.org/papers/w31161">http://www.nber.org/papers/w31161</a>, NATIONAL BUREAU OF ECONOMIC RESEARCH. April 2023, revised November 2023