

Generative Künstliche Intelligenz für die Qualitative Marktforschung

Dr. Stefan Oglesby, MBA
Daten IQ AG
Alte Steinhäuserstraße 33
6033 Cham

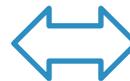


Was ist "Generative Künstliche Intelligenz"?

Warum verwenden wir den Begriff "künstliche Intelligenz"?



"Struktur ähnlich dem menschlichen Gehirn"



"Interaktion ähnlich wie mit einem Menschen"

„Generativ“: Produktion von Text, Bildern, Videos, Code ...



Neuronale Netze

Der Anfang: Klassifizierung von Bildern



?

Tiger

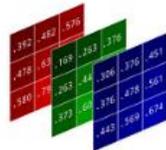
Katze

Hund

=



=



= $224 \times 224 \times 3 = 150,528$ pixels (!!)

Hochdimensionaler Input & komplexe Beziehung

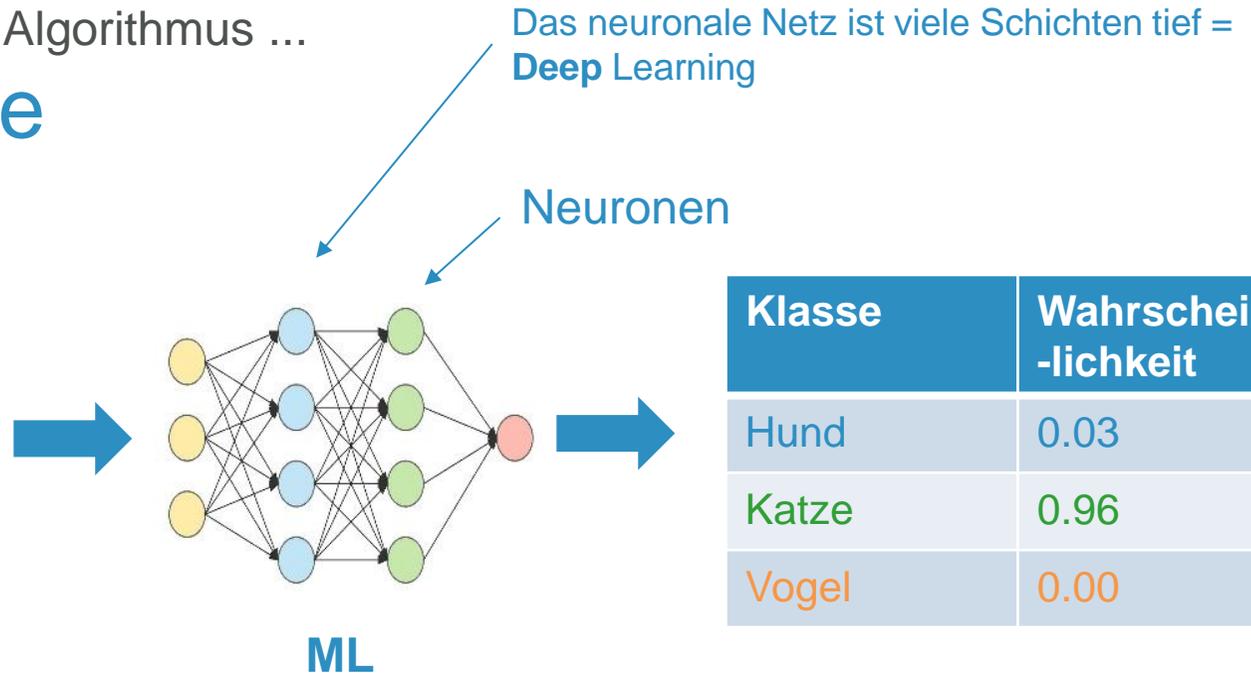
Beispiel: Kategorisierung von Bildern

Wir brauchen einen starken Algorithmus ...

Neuronale Netze



Eingabe
(Bild oder Text)



Klasse	Wahrscheinlichkeit
Hund	0.03
Katze	0.96
Vogel	0.00

Ausgabe



GPU

Technische
Voraussetzung



Wie chatGTP & Co. funktionieren

Large Language Models: Theoretische Grundlage

"Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache."

"Die Grenzen meiner Sprache bedeuten die Grenzen meiner Welt."

Ludwig Wittgenstein

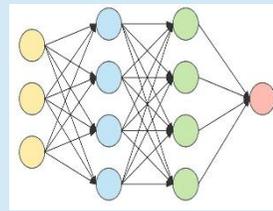
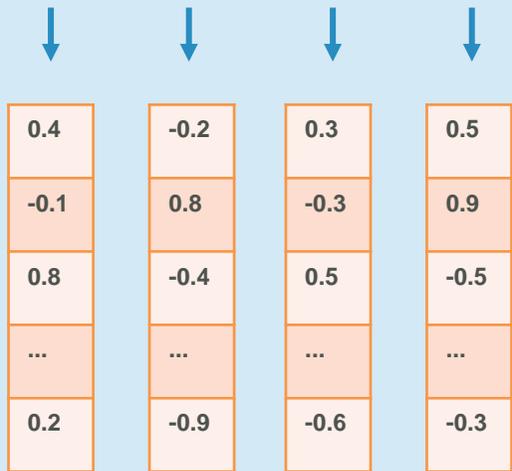
"Künstliche Intelligenz lernt die Sprachen nicht, indem man sie mit Grammatikbüchern füttert. Sondern sie lernt, indem sie Millionen beziehungsweise Milliarden von Sätzen analysiert, die bereits übersetzt worden sind."

Jarek Kutylowski, Gründer der Übersetzungswebsite DeepL

Darstellung von Wörtern als ihre Verteilung im Text

Numerischer Vektor = Position im semantischen Raum

Der Film war grossartig



ML



Stimmung	Wahrscheinlichkeit
Positiv	0.98
Negativ	0.02

„Embeddings“ stellen die Bedeutung eines Wortes in seinem Kontext dar. Hohe Anzahl von Dimensionen (z.B. 1'500)

Vektorielle Semantik: Synonyme haben die gleiche Verteilung im Sprachgebrauch (= große Textkorpora). Wörter einer bestimmten Kategorie, wie z. B. Gemüse (Karotte, Brokkoli, Rosenkohl, Sellerie, Fenchel), haben eine ähnliche Verteilung. Verteilung = Wahrscheinlichkeit, dass ein Wort in der Nähe eines anderen Wortes vorkommt.

Wie trainiert man ein Sprachmodell?

- Input: Text (und Code) aus einem breiten Spektrum von Quellen.

- Unsupervised learning:** Vorhersage des nächsten Wortes in einer Folge von Wörtern.

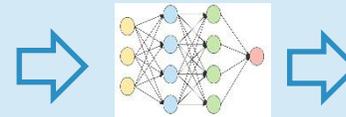


Foundation Model

Durchführung einer Vielzahl allgemeiner Aufgaben, wie z. B. das Verstehen von Sprache, das Erzeugen von Texten und Bildern und die Konversation in natürlicher Sprache.

Beispiel: Welches Wort kommt als nächstes in dieser Sequenz „A trained language model can ...“?

[A trained language model can ...]



Wort	Wahrscheinlichkeit
speak	0.065
generate	0.072
politics	0.001
..	
walk	0.003

Eingabe

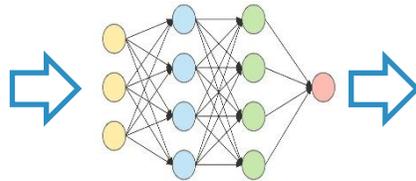
Neural Network

Ausgabe

Konzept des Foundation Models (LLM)

Wir können Text erzeugen, indem wir ein Wort nach dem anderen vorhersagen.

[A trained language model can]



Wort	Wahrscheinlichkeit
speak	0.065
generate	0.072
politics	0.001
..	
walk	0.003



Wort	Wahrscheinlichkeit
capability	0.002
text	0.084
coherent	0.085
..	
Ideas	0.041

Eingabe

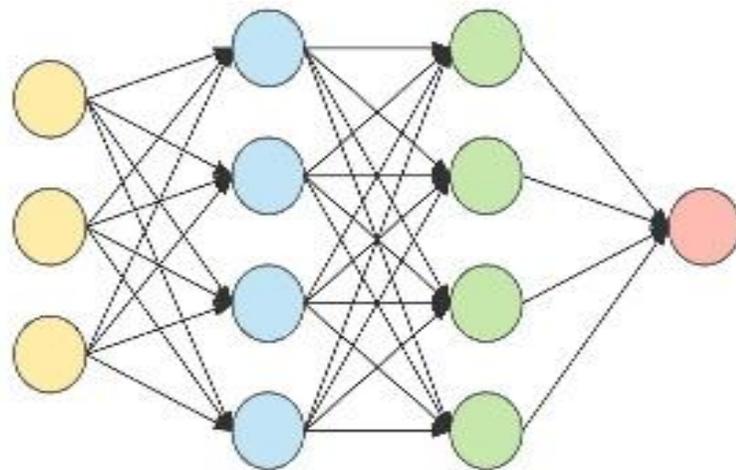
LLM

Output bei
Schritt 1

Output bei
Schritt 2



Wir können stattdessen eine Stichprobe von, sagen wir, den fünf wahrscheinlichsten Wörtern zu einem bestimmten Zeitpunkt ziehen. Das Ergebnis ist, dass das Modell mehr Kreativität zeigt. Beispiel: "temperature" in den GPT-Modellen



**GPT 4o: 200
Milliarden
Parameter**

**GPT 4: 175 Milliarden
Parameter**

**GPT 2: 774 Millionen
Parameter**

Das Magische an all dem ist, dass es erstaunlich gut funktioniert. Tatsächlich waren alle, sogar die Forscher von OpenAI, überrascht, wie weit diese Art der Sprachmodellierung gehen kann.

Phasen des Trainings von LLMs

1. Pre-Training

Große Datenmengen aus dem Internet + Bücher + etc.

Das Modell kann zwar korrekte natürlichsprachliche Texte erstellen, ist aber nicht auf die Wünsche der Benutzer abgestimmt.

Verzerrung durch den "Kanon" des Input.

unsupervised

2. Instruction Fine-Tuning

Dem Modell beibringen, auf bestimmte Anweisungen zu reagieren.

Anpassung an die Erwartungen der Nutzer.

Die Qualität des Outputs variiert je nach Aufgabe.

3. Human Feedback Reinforcement

Zahlreiche menschliche "Tester" geben Feedback zur Ausgabe des Modells.

Es trägt dazu bei, dass die Ergebnisse näher an dem sind, was die Menschen wollen oder als positive empfinden.

Kultureller Bias der Trainer

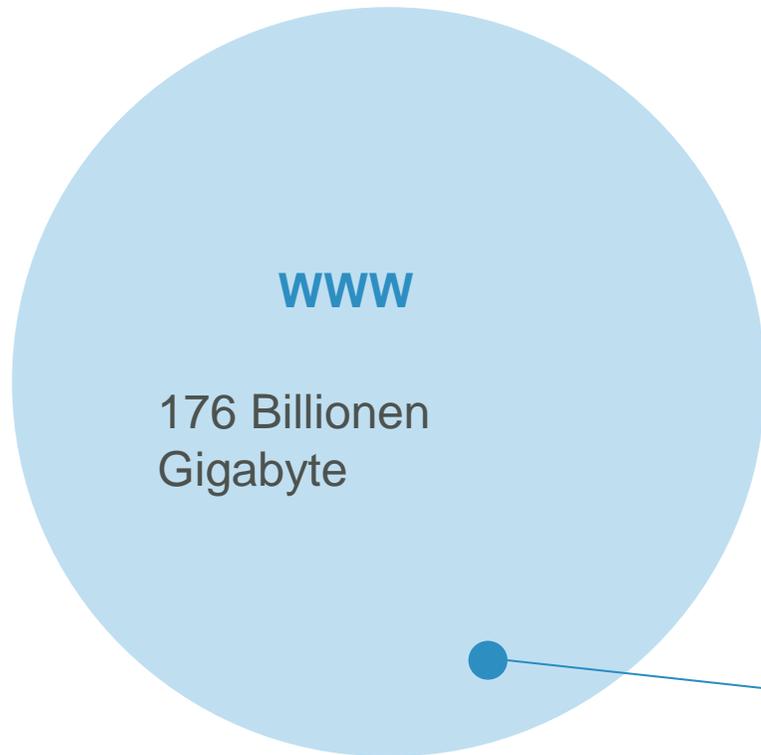
supervised

OpenAI in den Startlöchern

**Ausgegoogelt? Jetzt kommt
der Angriff auf die
mächtigste Suchmaschine
der Welt**

GPT: Genauigkeit der Informationen?

Problem der Halluzination



$$\frac{\text{GPT 4} = 0,176 \times 10^{12} \times 8 \text{ Bytes}}{\text{WWW} = 176 \times 10^{21} \text{ Bytes}} = 9 \times 10^{-9} \%$$

Stichtag für GPT 4o ist Oktober 2023.

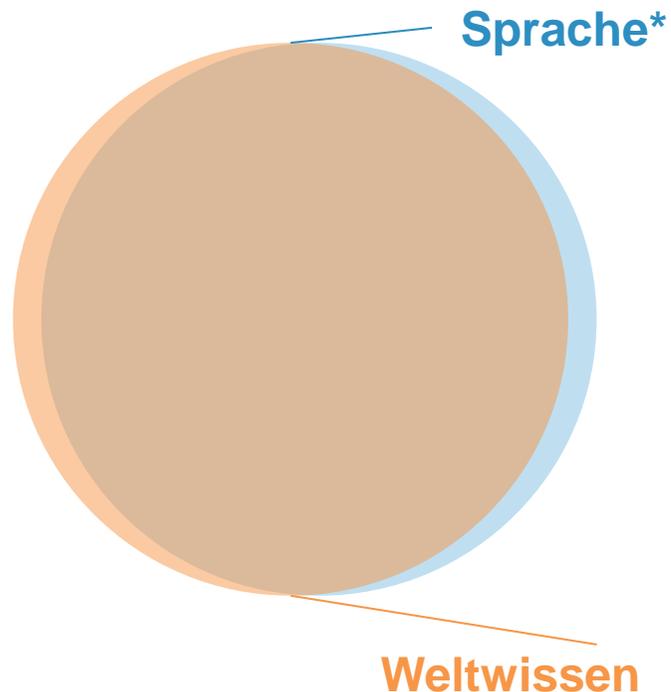


GPT wurde mit sorgfältig kuratierten Dokumenten
trainiert!
(Common Crawl, WebText2 und Wikipedia).

LLMs: Wissensbasis vs. Datenprozessor

Zwei Arten von "Genauigkeit"

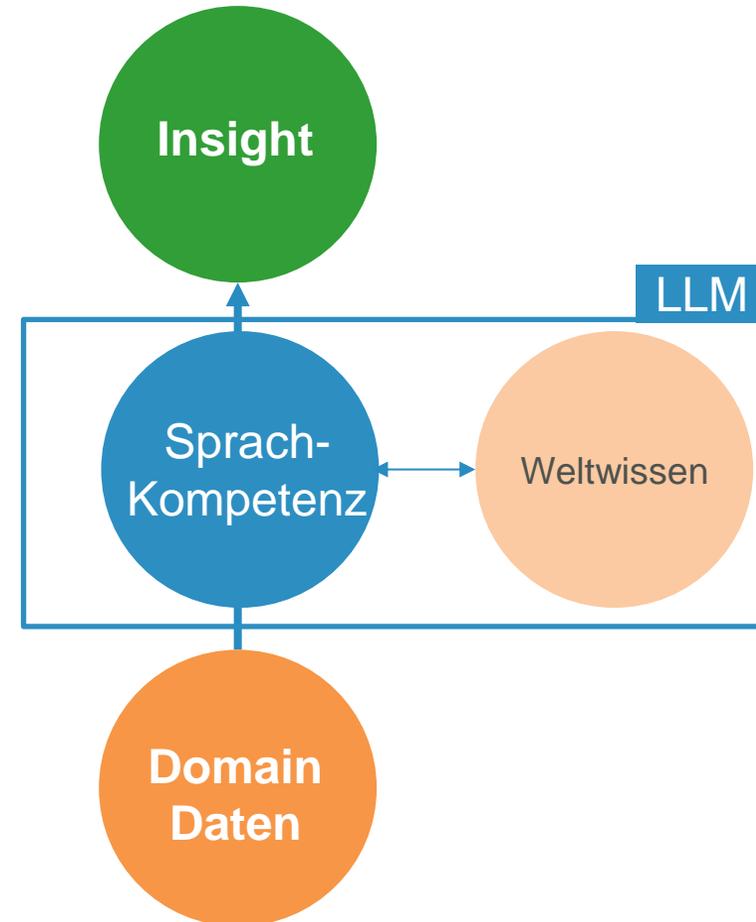
GPT zur Beantwortung von (Wissens-) Fragen



Die Grenzen meiner Sprache bedeuten die Grenzen meiner Welt.
Ludwig Wittgenstein

*Sprachliche Kompetenz

GPT als Werkzeug zur Datenanalyse



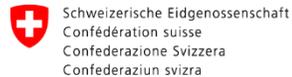


Ethische Erwägungen

Ethische Erwägungen



Neues Bundesgesetz über den Datenschutz (nFADP)



Profiling

Das risikoreiche Profiling bietet Schweizerinnen und Schweizern besonderen Schutz vor bestimmten Arten des automatisierten Profilings mit hohen Risiken für die Persönlichkeit oder die Grundrechte.

Von den Datenverarbeitern durchgeführte Datenverarbeitungsvorgänge

Beauftragt der für die Verarbeitung Verantwortliche eine andere Stelle mit der Verarbeitung, so darf diese (d. h. der Datenverarbeiter) die Daten nur nach den Anweisungen des für die Verarbeitung Verantwortlichen verarbeiten und muss die Sicherheit der personenbezogenen Daten gewährleisten.

The AI Act classifies AI according to its risk:

- Unacceptable risk is prohibited (e.g. social scoring systems and manipulative AI).
- Most of the text addresses high-risk AI systems, which are regulated.
- A smaller section handles limited risk AI systems, subject to lighter transparency obligations: developers and deployers must ensure that end-users are aware that they are interacting with AI (chatbots and deepfakes).
- Minimal risk is unregulated (including the majority of AI applications currently available on the EU single market, such as AI enabled video games and spam filters – at least in 2021; this is changing with generative AI).

The majority of obligations fall on providers (developers) of high-risk AI systems.

- Those that intend to place on the market or put into service high-risk AI systems in the EU, regardless of whether they are based in the EU or a third country.
- And also third country providers where the high risk AI system's output is used in the EU.

Users are natural or legal persons that deploy an AI system in a professional capacity, not affected end-users.

- Users (deployers) of high-risk AI systems have some obligations, though less than providers (developers).
- This applies to users located in the EU, and third country users where the AI system's output is used in the EU.

General purpose AI (GPAI):

- All GPAI model providers must provide technical documentation, instructions for use, comply with the Copyright Directive, and publish a summary about the content used for training.
- Free and open license GPAI model providers only need to comply with copyright and publish the training data summary, unless they present a systemic risk.
- All providers of GPAI models that present a systemic risk – open or closed – must also conduct model evaluations, adversarial testing, track and report serious incidents and ensure cybersecurity protections.



Anwendungsfälle

Anwendung der generativen KI in quali MaFo

Drei Schlüsselfragen

Substitution

Wo **ersetzt** die generative KI **menschliche Arbeit** und **“manuelle” Methoden**?

Limiten

Wo wird **noch menschliches Eingreifen** oder die Ausführung von Aufgaben durch Menschen **erforderlich sein**?

Innovation

Wo **ermöglicht** die generative KI **neue Lösungen**, die bisher nicht möglich waren?

Anwendungsfall 1: Set-Up **Versicherung Projekt**



- Strategische Frage: Bedürfnisse und Wünsche der Verbraucher bei der Wahl eines Versicherungsanbieters.
- 20 Tiefeninterviews mit Kunden von je ca. 30 Minuten Länge. Schweizerdeutsch und Französisch.
- Automatische Transkription, Anonymisierung, manuelle Bearbeitung durch Freelancer.

Anwendungsfall 1: Ergebnis

Versicherung Projekt



- Themenextraktion (z. B. Bedürfnisse, Barrieren, erwartete Merkmale) : 100% Abdeckung vs. Manuell erarbeiteter Bericht.
- Nahtlose Integration der Originalsprachen (Deutsch, Französisch) für den englischen Bericht.
- Management Summary: Korrekter Fokus auf mehr bzw. weniger relevante Bedürfnisse, die von den Befragten genannt wurden.
- Abrufen von zu 90% korrekten wörtlichen Verbatim-Zitaten



Spart bis zu 60 % Zeit beim Analysieren und Zusammenfassen. "Finish" durch Senior Researcher/Berater unbedingt erforderlich.

Grenzen der KI für qualitative MaFo





Qualitätskontrolle der LLM-Ergebnisse.

Qualität (Leistung) von LLMs

	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>	86.8% 5 shot	79.0% 5-shot	75.2% 5-shot	86.4% 5-shot	70.0% 5-shot	83.7% 5-shot	71.8% 5-shot
Graduate level reasoning <i>GPQA, Diamond</i>	50.4% 0-shot CoT	40.4% 0-shot CoT	33.3% 0-shot CoT	35.7% 0-shot CoT	28.1% 0-shot CoT	—	—
Grade school math <i>GSM8K</i>	95.0% 0-shot CoT	92.3% 0-shot CoT	88.9% 0-shot CoT	92.0% 5-shot CoT	57.1% 5-shot	94.4% Maj1@32	86.5% Maj1@32
Math problem-solving <i>MATH</i>	60.1% 0-shot CoT	43.1% 0-shot CoT	38.9% 0-shot CoT	52.9% 4-shot	34.1% 4-shot	53.2% 4-shot	32.6% 4-shot
Multilingual math <i>MGSM</i>	90.7% 0-shot	83.5% 0-shot	75.1% 0-shot	74.5% 8-shot	—	79.0% 8-shot	63.5% 8-shot
Code <i>HumanEval</i>	84.9% 0-shot	73.0% 0-shot	75.9% 0-shot	67.0% 0-shot	48.1% 0-shot	74.4% 0-shot	67.7% 0-shot
Reasoning over text <i>DROP, F1 score</i>	83.1 3-shot	78.9 3-shot	78.4 3-shot	80.9 3-shot	64.1 3-shot	82.4 Variable shots	74.1 Variable shots
Mixed evaluations <i>BIG-Bench-Hard</i>	86.8% 3-shot CoT	82.9% 3-shot CoT	73.7% 3-shot CoT	83.1% 3-shot CoT	66.6% 3-shot CoT	83.6% 3-shot CoT	75.0% 3-shot CoT
Knowledge Q&A <i>ARC-Challenge</i>	96.4% 25-shot	93.2% 25-shot	89.2% 25-shot	96.3% 25-shot	85.2% 25-shot	—	—
Common Knowledge <i>HellaSwag</i>	95.4% 10-shot	89.0% 10-shot	85.9% 10-shot	95.3% 10-shot	85.5% 10-shot	87.8% 10-shot	84.7% 10-shot

Formale Qualitätskontrollen

- › Standard-Benchmarks für LLMs sind nicht angemessen:
ROUGE vergleicht mit der Referenz-zusammenfassung und zählt übereinstimmende Wörter.
 - .
- › Lösung: Topic Modelling, z. B. Latent Dirichlet Allocation (LDA), kann analysieren, ob Themen in den Eingabedaten in Zusammenfassungen und Schlussfolgerungen nicht vertreten sind.
- › Ungeachtet dessen: Die intensive Einbindung des Forschers in die Studie ist entscheidend für die Plausibilitäts- und Qualitätskontrolle



Anwendung der generativen KI in quali MaFo

Drei Schlüsselfragen

Substitution

Wo **ersetzt** die generative KI **menschliche Arbeit** und **“manuelle” Methoden**?

Limiten

Wo wird **noch menschliches Eingreifen** oder die Ausführung von Aufgaben durch Menschen **erforderlich sein**?

Innovation

Wo **ermöglicht** die generative KI **neue Lösungen**, die bisher nicht möglich waren?



Anwendungsfall 2

Kreation innovativer Schokolade-Produkte

Einrichten

- Transkribierte Tiefeninterviews zum Schokoladenkonsum.
- Rolle der KI: Produkt-Innovation.

Ergebnis

- **GenAI erstellte ein Konzept für einen neuen biologischen, nachhaltigen Schokoladengenuss, der auf den relevanten Bedürfnissen der Verbraucher basiert.**
- **Direkte Eingabe für quantitativen Online-Konzepttest**



Qualitative Consumer Insight, die über statische Berichte hinausgehen, werden Teil der täglichen Arbeit im Marketing, New Product Development, Content Creation.



Ausblick: Neue Art von "Qual" mit KI als unabhängigem Agenten.

Ausblick: „Agentic AI“



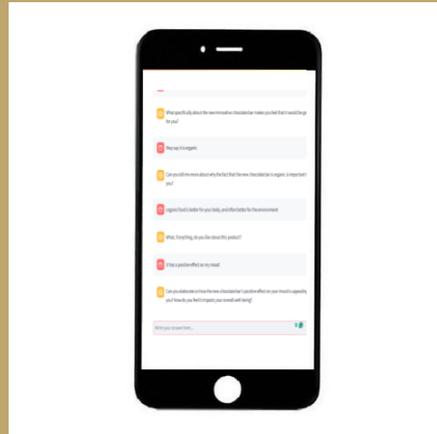
Neue Möglichkeiten

Vollwertige Qualifikation



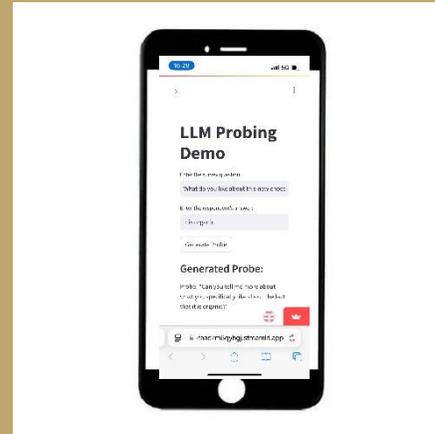
Z.B. Life-Fokusgruppen mit "echten Menschen" und geschulten Moderatoren.

„Lightweight Qual“



Agent, der "freie" Online-Interviews nach einem Leitfaden durchführt

„Reichere“ offene Fragen



Intelligentes Nachfragen bei offenen Fragen in Surveys.

Quant



Standard-Online-Umfragen



Wichtigste Erkenntnisse

Die wichtigsten Erkenntnisse

- Sprache ↔ Wissen – untrennbar in den LLMs
- Pre-Training, Instruction Fine-Tuning, Human Feedback
- Gen AI als Datenprozessor
- Qual MR als Kontext-Lieferant
- Interaktive "prompt-your data"-Services schaffen Mehrwert für quali MaFo
- "Agentische KI" übernimmt autonom Aufgaben, z.B. Tiefeninterviews.