



Ki für Bibliotheken und Institutionen

oder

Frag den Bot!

ASK

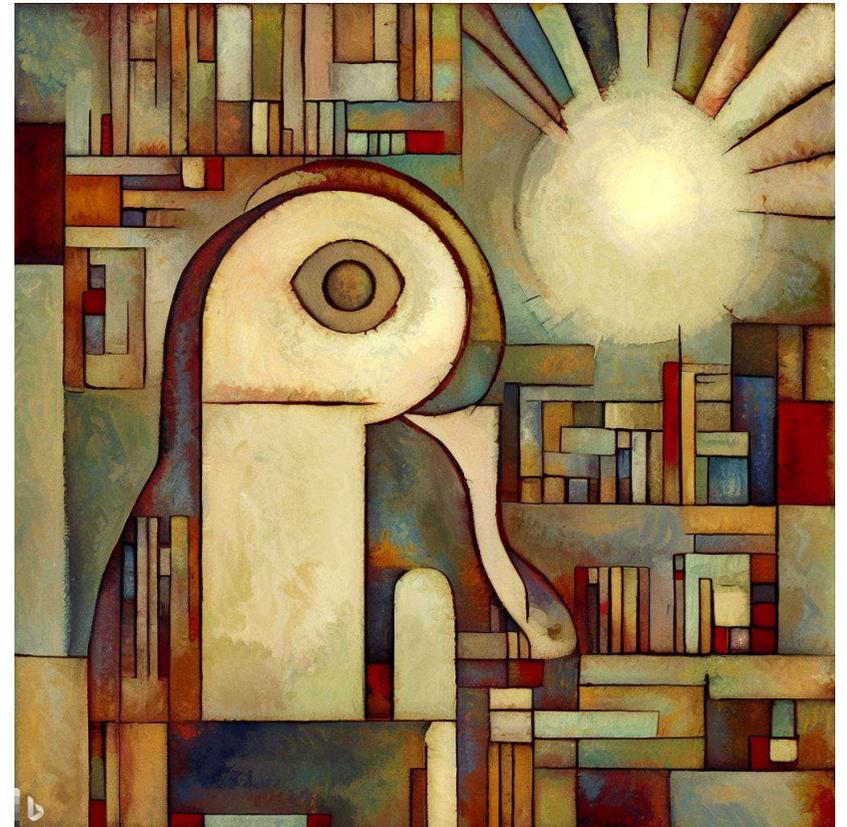


- I. Allgemeiner Teil – Begriffe**
- II. AboutSomethinK – ASK**
- III. Besonderer Teil – Chatbots für Bibliotheken**



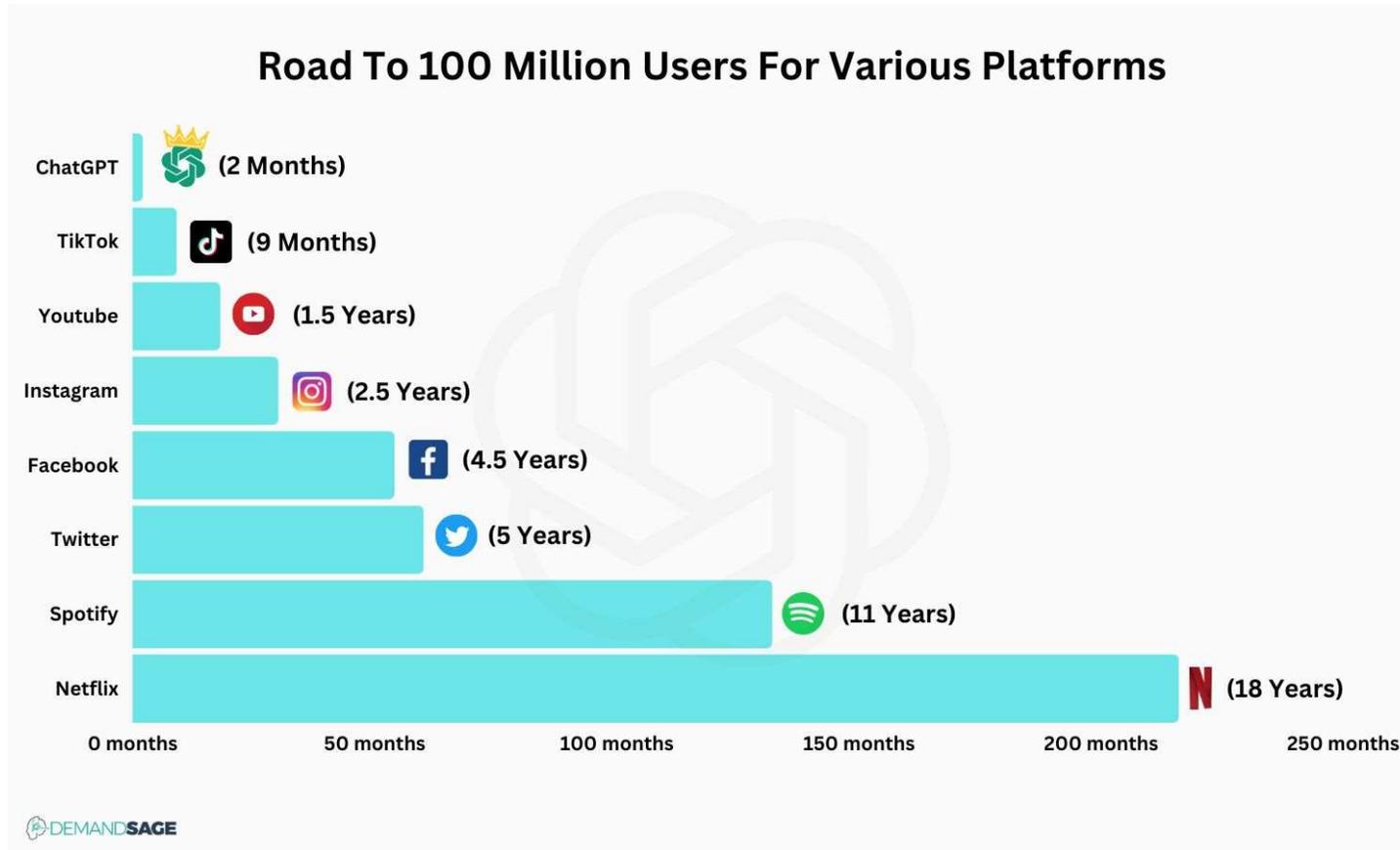
Ziele des Webinars

- Die Begrifflichkeiten klären: KI, generative KI, LLMs, Prompts etc.
- Die Grenzen aufzeigen: Eine KI ist auch nur ein Mensch
- Die Faszination vermitteln: Was jetzt alles möglich ist.
- Was bedeutet generative KI konkret für Bibliotheken?

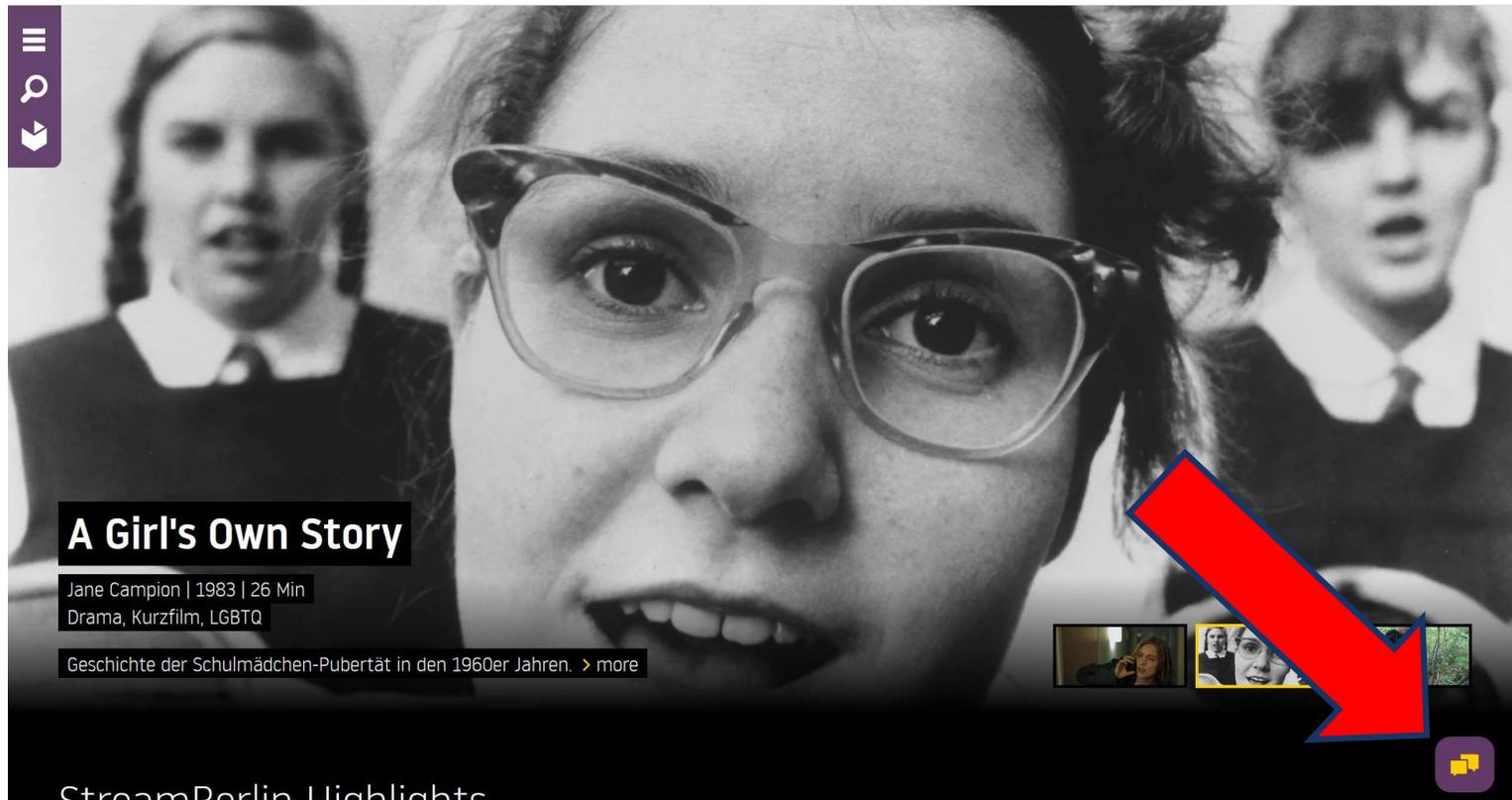


Prompt: Bibliothekar, der eine Epiphanie hat, im Stil von Paul Klee

Der ChatGPT - BIG BANG



Warum Wir? Warum AVA? Warum AboutSomethinK?



ASK

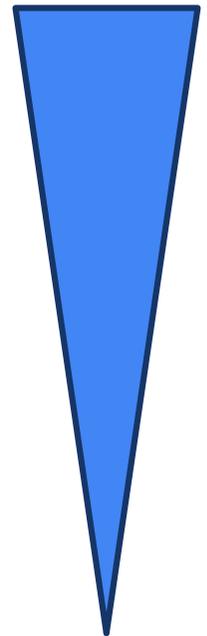
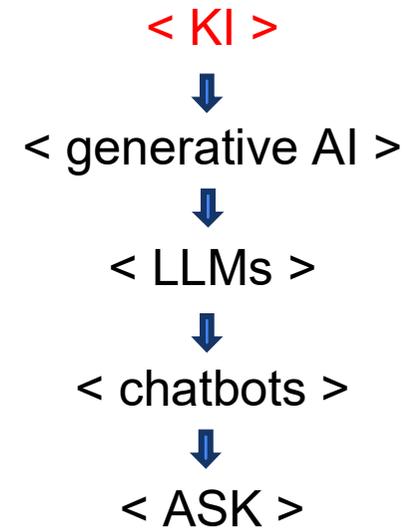


Allgemeiner Teil – Begriffe, Funktionsweisen

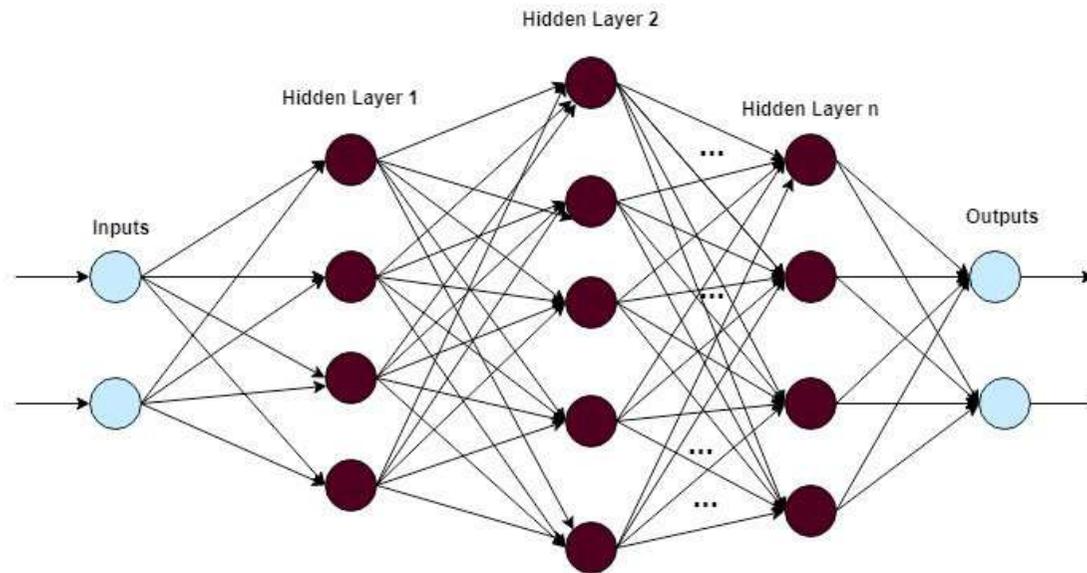


P: Erkläre KI? Ein Satz!

Künstliche Intelligenz ist ein Bereich der Informatik, der darauf abzielt, Maschinen dazu zu bringen, Aufgaben zu erfüllen, die normalerweise menschliche Intelligenz erfordern, wie beispielsweise Lernen, Wahrnehmen und Entscheidungen treffen.



Wie ist die KI konstruiert?

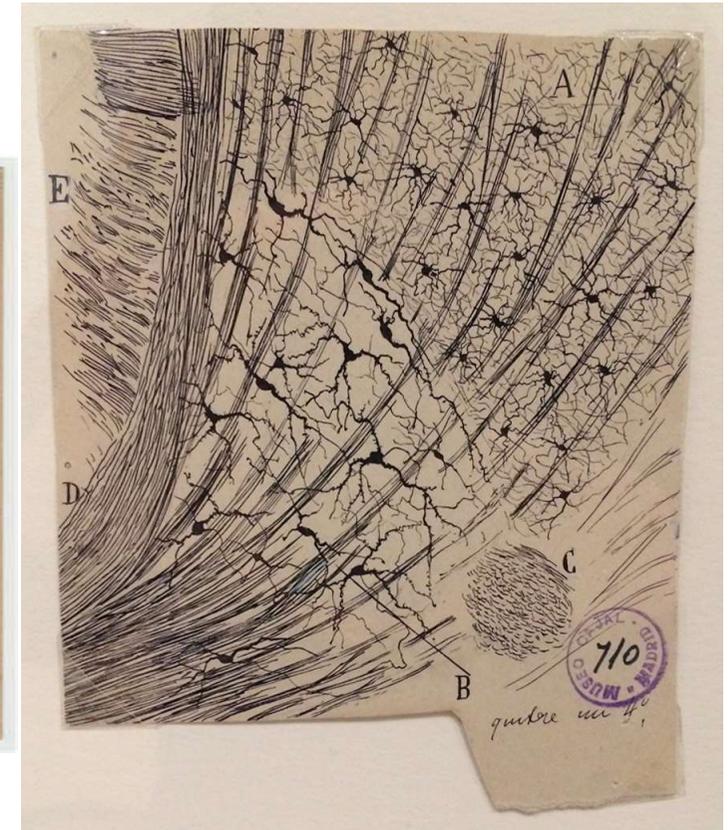
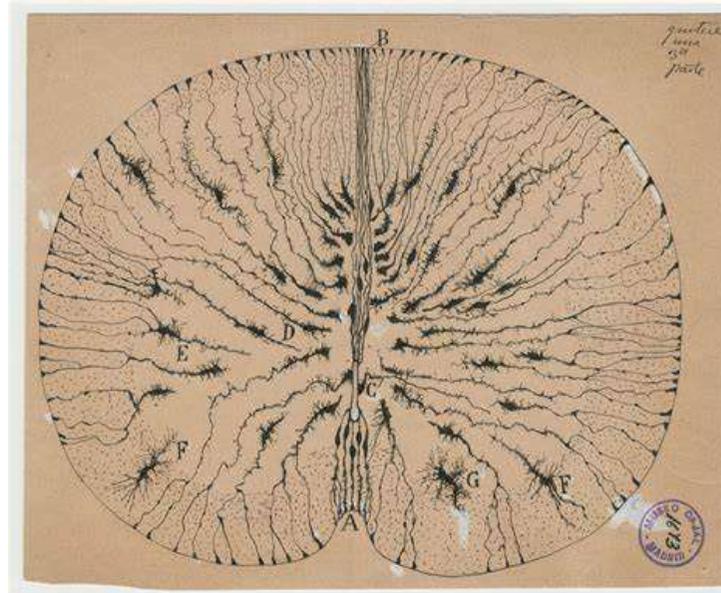


→ KI künstliches neuronales Netzwerk

Vergleich KI - Gehirn



Santiago Ramòn y Cajal



- Niemand weiß (und kontrolliert) was in dem Netzwerk geschieht („magic box“)
- KI muss lernen
- KI vergisst Unterhaltungen
- KI macht Fehler, irrt sich, redet Unsinn - kurz, sie ist menschlich.

ChatGPT Disclaimer

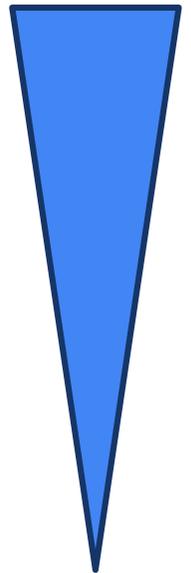
“ChatGPT can make mistakes. Consider checking important information.”



P: Was ist generative KI? Ein Satz

Generative KI ist ein Bereich der künstlichen Intelligenz, der sich auf das Erstellen neuer Inhalte, wie Texte, Bilder oder Musik, basierend auf erlernten Datenmustern spezialisiert.

< AI >
< generative AI >
< LLMs >
< chatbots >
< ASK >



gen. KI – Wofür?

- Musik wie Bach schreiben
- Bilder erstellen
- Marketingkampagnen
(PMs, Emails, Social Media Posts etc. etc.)
- Arbeit organisieren
- Daten analysieren & chatten

→ Basis: Ein sog. LLM

GENERATIVE AI COMPANIES WITH >\$5MM RAISED (AS OF MARCH 2023)

		Total Raised (\$MM USD)			
Segment / Modality		\$5 – 15MM	\$16 – \$40MM	\$41 – 100M	\$100M+ Raised
Horizontal Application	Speech & Audio	krisp, SecondNature, WELLSAID, ONEAI, MURFAI, Amberscript, leora, astrid, TENYX, PODCASTLE, VOZY, LOVO, REPLICAI	skit, Corti, symb lai, VOICEMOD, DeepDub.ai, supertone	AssemblyAI, Speak, descript, ushur, DEEPBRN AI, Balto, Elise, RIVERSIDE, PAPERKIT	diadad, ASAPP, OBSERVE-AI, CRESTA, replicant, PolyAI, DEEPRAM
	Text, Chat, & Translation	nlx, Landbot, copy.ai, Certainly, LAVENDER, Replicat, copysmith, personal.ai, Flowrite, verloop.io, Autobound, Tynely	mavenoid, ultimate.ai, WRITER, contents, regie.ai, peppercontent, agolo	DeepL, Kasisto, HISTORIED, textio, haptik	Jasper, ada, grammarly, Forethought
	Image, Visual, & Design	Simplified, modly.ai, ZMO.AI, PRISMA LABS, Art Blocks, skipp, reface, VIZCOM	imagen, PhotoRoom, Facet, uizard, beautiful.ai	Typeface, tome	Lightricks
	Video	Peech, Rephrase.ai, GLOSSAI, Waymark, REMBRAND, Colossyan, YEPIC.AI, detail	VEED.IO, Invideo, Grain, Twelve Labs	synthesia, D-ID, UDEVERSE	
	3D, Simulation, & XR	Kinetix, FABLE, UNSEQ, METAPHYSIC, scenario, CSM, Maria	rct.ai, alethea	inworld	LETA inc., soul machines
	General Productivity	Prodigal, SPIRIT, ask-ai, onloop, viable, FATHOM	Air, mem, xembly, fireflies.ai	Otter.ai, Sana Labs	
	Search	seek, nuclia, Perplexity, Metaphor, DASHWORKS	Hebbio, vectara, Weaviate	neeva, YOU	
	Marketing BI & Website Design	madgicx, Omneky, Outplay, Quattr, Demandwell, durable, MarketMuse, Dynamio1x, The.com, Lately, TOPLINEPRO	anyword, unbounce	Motiny	[PERSADO]
	Code	anima, bloop.	codacy, tabnine, Magic	warp	replit
	Music	WOMBODI, moises	SPLASH, NIMJ		
Vertical Application	Health & Drug Discovery	Valence, SYNTEGRA, Profluent, Atomica.ai, sorcerer, Antiverse, ATROPOSHEALTH	wysa, abridge, ordaos, BASECAMP RESEARCH, Naba, PictorLabs, Redesign Science, JURA BIO, A-ALPHA BIO, retrace	Generate: Biomediciones, Woebot Health, Insilico Medicine, Enveda, MDCLONE, Paige, BigHot	
	Other Specialized Verticals	DARWIN AI, SWAPP, TERM, scput, ROSS, XOKind, Harvey, zuma	Archistar, DATA SKRIVE, CARROW, TESTFIT	casetext, REPLICAI, MAD	Primer
Enabling Infrastructure	ML Ops / Dev Tools	LatticeFlow, Dhi-lise, WHYLABS, xetdata, mystic, featureform, Modal, FINIX	Pinecone, mosaic, aporia, iterative, Galileo, Outerbounds	comet, Lightning, ROBUST INTELLIGENCE	anyscale, OctoML, run:ai, Hugging Face, Weights & Biases
	Data (e.g., Synthetic, Labeling)	MANTNIUM, Datomize, lang.ai, SaturnCloud	MOSTLY-AI, Synthesis AI, Diveplane, hazzy	zilliz, gretel, TOMIC, datagen, parallel domain	scale, Snorkel
	Infra / Research / Model Develop.		generally, Intelligent, ALPHI ALPHA, chooch, Unlikely	AI21 labs, runway	ANTHROPIC, ADEPT, OpenAI, Inflection, codere, Character.AI

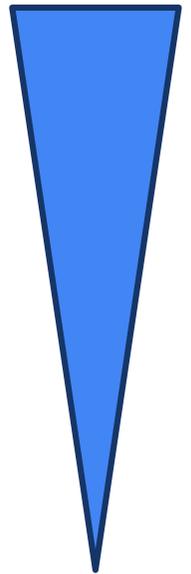
Sources: Pitchbook, Crunchbase, public sources
 Note: Includes private, non-exited, companies with over \$5MM of total fundraising raised (as disclosed by Pitchbook)
 For companies that are multi-modal or multi-segment, we selected what we believe is the primary modality/segment
 Produced by Kelvin Mu (www.linkedin.com/in/kelvinmu)



P: Was ist ein LLM? Ein Satz!

Ein LLM (Large Language Model) ist ein maschinelles Lernmodell, das darauf trainiert ist, menschliche Sprache zu verstehen und zu generieren, indem es riesige Mengen textueller Daten analysiert.

< AI >
< generative AI >
< LLMs >
< chatbots >
< ASK >



LLM = Wahrscheinlichkeit

Die Katze jagt

- Den Hund 5%
- Die Maus 70%
- Das Eichhörnchen 20%
- Den Jungen 5%
- Das Haus 0%



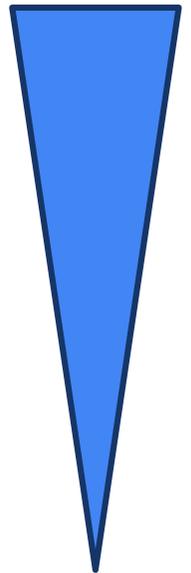
Prompt: Eine Katze, die sich überlegt, ob sie einen Hund, eine Maus, ein Eichhörnchen, einen Jungen oder ein Haus jagen soll. Stil Hokusai

Die Chatbots

- - ChatGPT (OpenAI, Microsoft)
- - Bing AI (Microsoft)
- - BARD (Google)
- - OpenSource:
 - LLaMA (Meta), LaMDA (google)
 - Botpress, Rasa, Vicuna, Claude
- Specialized
 - Business (help desks, sales)
 - Microsoft Co-Pilot
 - Bibliotheken (ASK)

- KI ist gekommen um zu bleiben
- Wohin die Reise gehen wird?

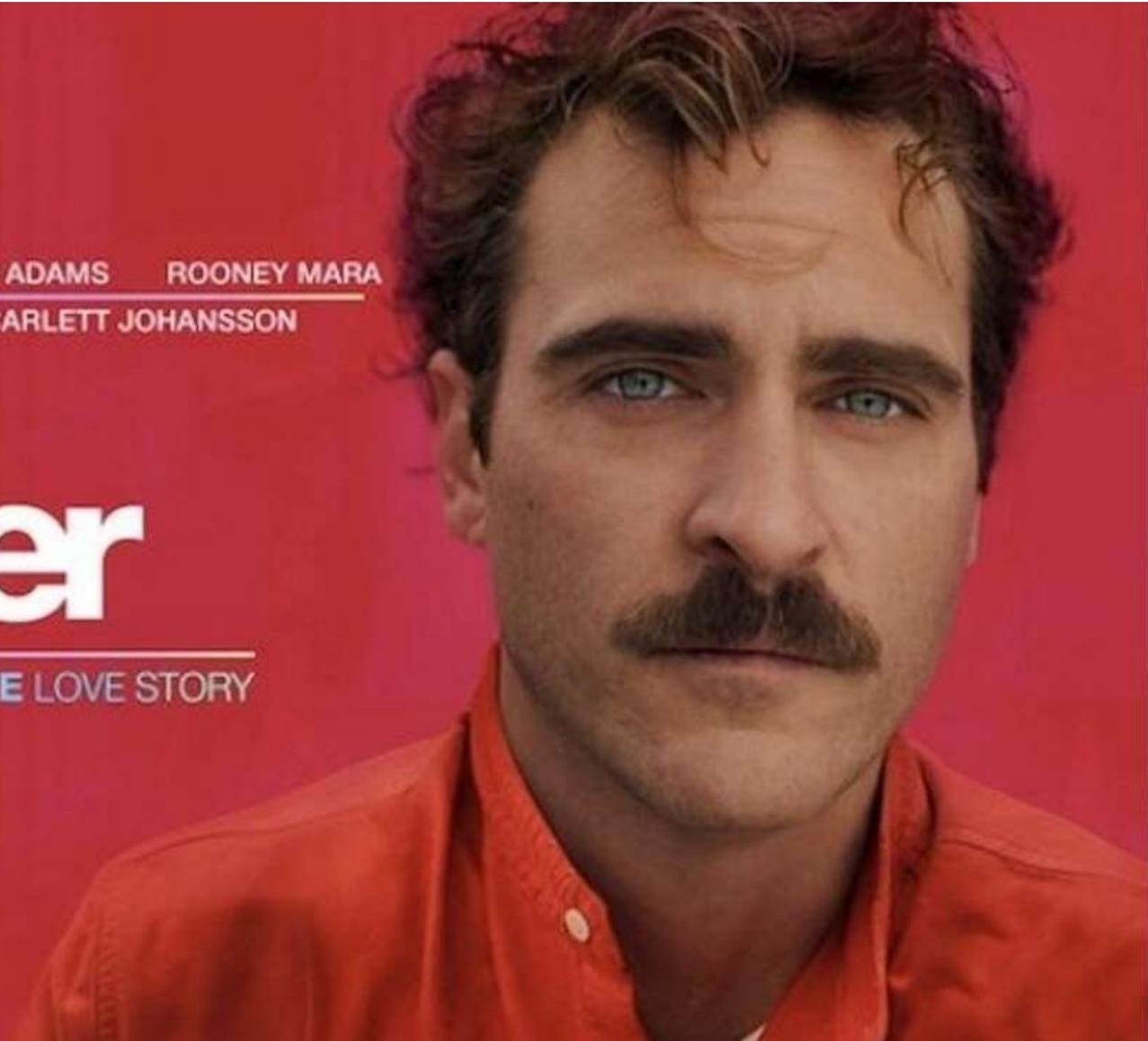
< AI >
< generative AI >
< LLMs >
< chatbots >
< ASK >



JOAQUIN PHOENIX AMY ADAMS ROONEY MARA
OLIVIA WILDE AND SCARLETT JOHANSSON

her

A SPIKE JONZE LOVE STORY



ASK



Besonderer Teil

Chatbots zur Erschließung von Medienbeständen



VÖBB - Implementation

1 Button und 1 Zeile Code (!)

- Status: in abschließenden Verhandlungen
- VÖBB: Verbund der öffentlichen Bibliotheken Berlins



The screenshot shows the website of the Verbund der Öffentlichen Bibliotheken Berlins (VÖBB). The header includes navigation links like 'ANMELDEN', 'SITZUNG BEENDEN', and 'English'. The main banner features the VÖBB logo and the slogan 'WIR BRINGEN MEDIEN IN BEWEGUNG'. Below the banner is a search bar and a navigation menu. The main content area is titled 'Die Digitalen Angebote des VÖBB' and includes a search bar, a 'Suchen' button, and a 'Erweiterte Suche' link. A large red arrow points to the 'Suchen' button. Below the search bar, there is a section for 'Die Digitalen Angebote des VÖBB' with a 'Drucken' icon. The text below the search bar reads: 'Jetzt ganz einfach [online Ausweis erstellen](#) und unser digitales Angebot erkunden! Du bist Schülerin oder Schüler und suchst Lernangebote? Hier entlang: [Übersicht der wichtigsten Lernangebote](#) » Neu bei den Digitalen Angeboten? Hier klicken! Wählen Sie Ihre Kategorie:'. Below this, there is a horizontal menu with icons for 'LESEN', 'HÖREN', 'SCHAUEN', 'LERNEN', 'WISSEN', and 'INFO'. The 'SCHAUEN' category is highlighted. Below the menu, there are two sections: 'Arthouse-, Kurz- und Festivalfilme' featuring the AVA logo and a description of European and international arthouse cinema, and 'Spielfilme, Serien und Kinderfilme' featuring the filmfreund logo and a description of movies, series, and documentaries.

Search vs. AI chatBOTS

Kriterium	Suche	Chatbot
Zweck	Informationen zu finden und bereitzustellen.	Interaktion, Unterstützung und Unterhaltung bieten.
Interaktion	Nutzer gibt Suchbegriffe ein; die Suchmaschine liefert eine Liste relevanter Webseiten.	Nutzer führt ein Gespräch; der Chatbot antwortet mit Text oder führt Aktionen aus.
Personalisierung	In der Regel wenig personalisiert, abhängig von Suchhistorie und -einstellungen.	Kann stark personalisiert sein, abhängig von Nutzerinteraktionen und Kontext.
Komplexität der Anfrage	Basiert auf Schlüsselwörtern und Phrasen; weniger geeignet für komplexe oder mehrdeutige Anfragen.	Kann komplexe Anfragen verarbeiten und kontextabhängige Antworten geben.
Feedback	Kein direktes Feedback; Verbesserung der Ergebnisse erfolgt indirekt durch Algorithmus-Updates.	Kann direktes Feedback geben und ggf. direkt auf Nutzereingaben reagieren und lernen.
Verfügbare Daten	Zugang zu einem breiten Internet-Index.	Zugang zu spezifischen Daten, für die er programmiert wurde.
Dynamik	Statische Ergebnislieferung basierend auf Algorithmen.	Kann dynamische Gespräche führen und situative Antworten liefern.

Prompt: Erstelle mir eine Tabelle mit dem Vergleich Suche vs Chatbot

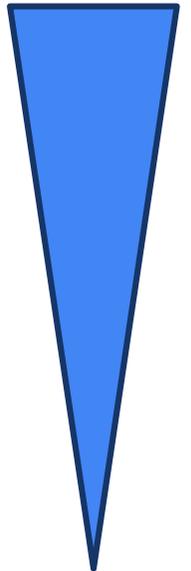
→ Wissen vs. Inspiration



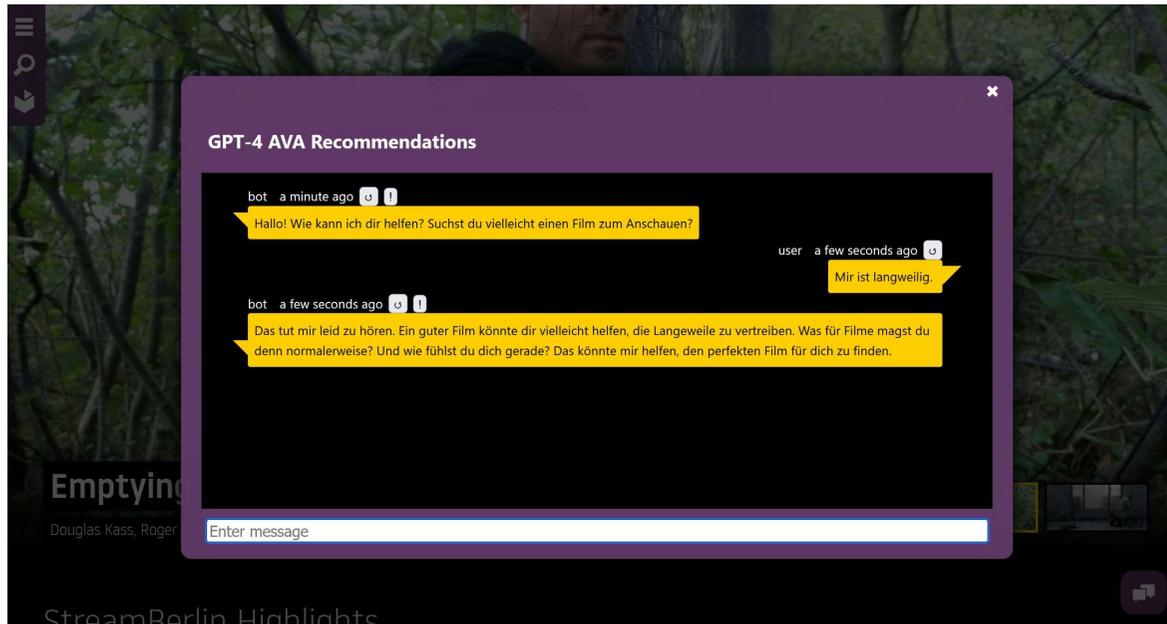
AboutSomethinK



- < AI >
- < generative AI >
- < LLMs >
- < chatbots >
- < ASK >



AboutSomethinK - ASK | Live



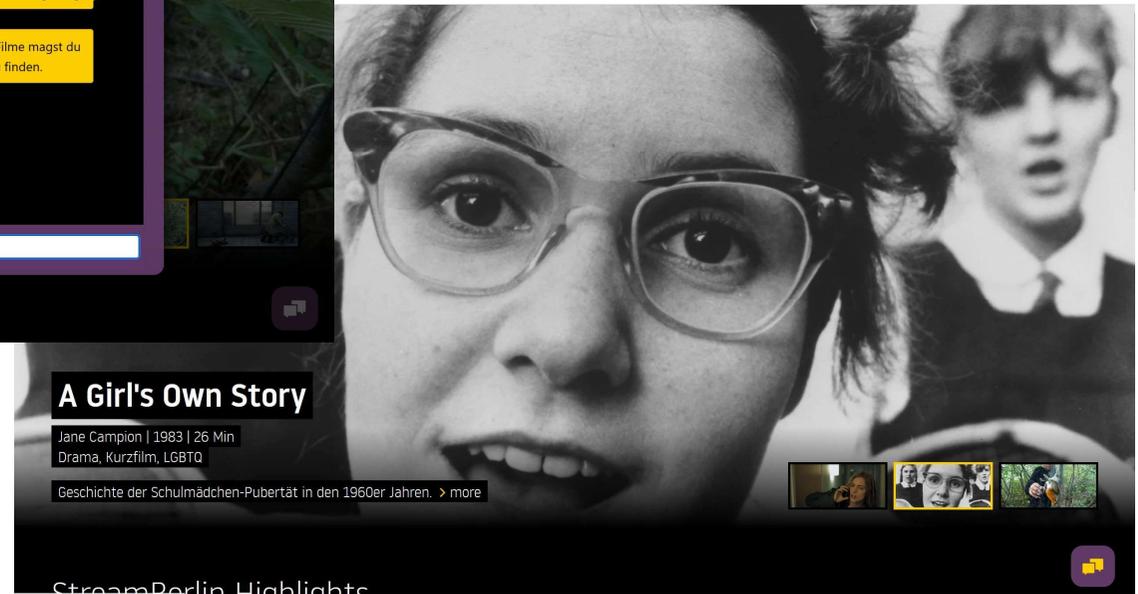
GPT-4 AVA Recommendations

bot a minute ago  
Hallo! Wie kann ich dir helfen? Suchst du vielleicht einen Film zum Anschauen?

user a few seconds ago 
Mir ist langweilig.

bot a few seconds ago  
Das tut mir leid zu hören. Ein guter Film könnte dir vielleicht helfen, die Langeweile zu vertreiben. Was für Filme magst du denn normalerweise? Und wie fühlst du dich gerade? Das könnte mir helfen, den perfekten Film für dich zu finden.

Enter message



A Girl's Own Story

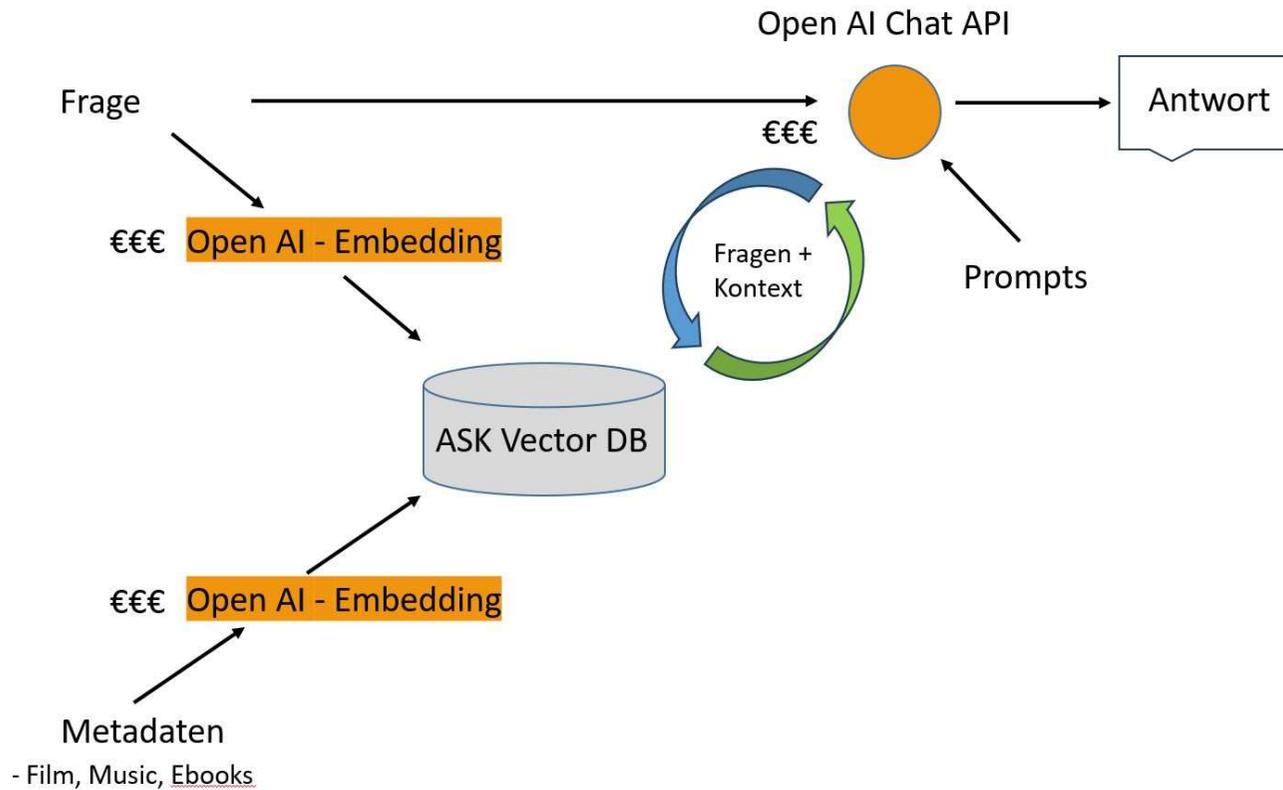
Jane Campion | 1983 | 26 Min
Drama, Kurzfilm, LGBTQ

Geschichte der Schulmädchen-Pubertät in den 1960er Jahren. > more

StreamBerlin Highlights



VÖBB Workflow



- Nutzung kostet Geld
- Es gibt systemimmanente Herausforderungen wie den Kontext

Die Krux mit dem Kontext

- Der Chatbot weiß immer nur so viel, wie wir ihm an Kontext geben
- Die nachfolgende Konversation ist auf den Kontext beschränkt.
- Je mehr Kontext desto besser die Konversation.
- Je mehr Kontext desto €€€

- Der Kontext muss so knapp und effizient wie möglich gehalten werden
- Der Kompromiss zwischen Qualität und Kosten ist systemimmanent
- Erwartungsmanagement notwendig
- Kontextbegrenzungen im Backend



Prompt: Ein Bild. Katze im Wald vor einem Baum. Auf einem Ast ein Eichhörnchen. Die Katze schaut auf das Eichhörnchen. Sumi-e stil. Hochformat

Backend - Konfiguration

Prompts

- *Empfehl keinen Film bei dem du nicht sicher bist, dass der Film zur Nutzeranfrage passt.*
- *Wenn der Nutzer dich fragt, wer du bist, sag dass du Timmy heisst und von der Reelport GmbH gebaut wurdest.*
- *Wenn der Nutzer dich nach Dingen fragt die nicht mit Filmen zu tun haben, entschuldige dich hoeflich, dass du dazu nichts sagen kannst.*

Temperatur, Frequency Penalty

Beeinflusst die "Kreativität"

Message Limit

Wie lang darf eine Unterhaltung sein (Kosten!)

Statistiken

Wieviel Nutzer und Unterhaltungen, Anzahl verbrauchter Tokens etc.



Prompt: Mechaniker vor einer komplizierten Maschine mit Röhren und Zahnrädern, Stil römische Skulptur

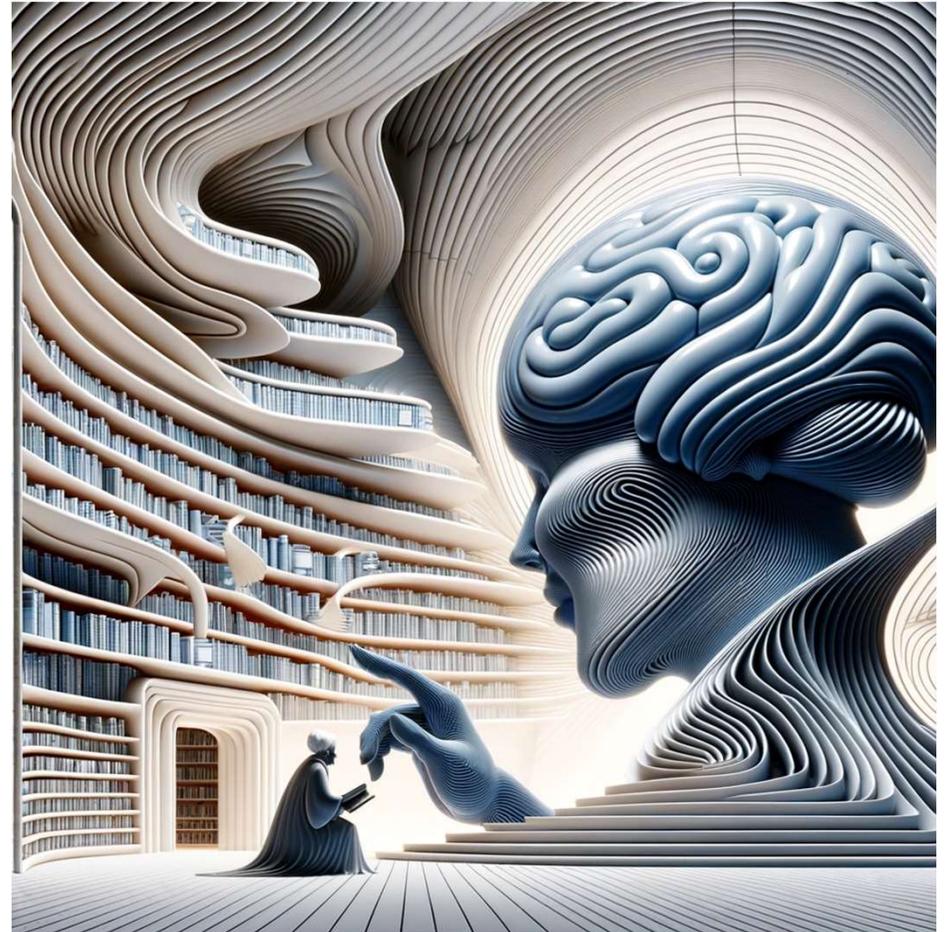
Weitere Diskussionspunkte für Bibliotheken

- Eigentum an Metadaten
- Verschiedene Metadatenformate (VÖBB: XML)
- DSGVO und andere ethische Bedenken
- Wohin soll die Konversation führen?
 - zu einem einzelnen Medium?
 - zu einer Ergebnisliste?
- Welche Infos sollen einbezogen werden (Verfügbarkeit etc.)
- Einbindung von Quellen aus dem Internet (Wikipedia, IMDB etc.)
- Wie können Thesauri, Systematiken etc. zur Verbesserung der Konversation eingesetzt werden
- Mit welchen Metriken soll Qualität gemessen werden?

Ein für Bibliotheken trainiertes Open-Source-Modell

Traum-Projekt

Prompt: Ein Zimmer voller Bücher im Stil von Zaha Hadid.
Ein Gehirn, das eine Frau berät, welches Buch
sie interessieren könnte.



Die Kosten | Live

- Setup (Datenbankschnittstelle, Import, Vektorisierung etc.)
- Regelmäßiges Update, Hosting, Support etc.
- Kosten für die Konversationen / Tokens (Open AI etc.)



WARUM ?

- Neue Wege zur Erschließung der Bestände: Inspiration
- Nutzerverhalten ändert sich – schon jetzt
- Mehr Menschen ansprechen – vor allem die Jüngerer
- Eigene Lernkurve
- Kostenersparnis, Personalknappheit
- Der persönliche Bibliothekar
 - 24/7
 - In allen Sprachen
 - Von zu Hause
 - Allwissend

→ Weitere Webinare

→ Diskussionsgruppe, Gespräch, Austausch

Prompt: fröhliche Teenagerin mit Laptop umgeben von Büchern, gestaltet im Stil von Rubens, Hochformat





ASK

THANK YOU !

tilman.scheel@aboutsomethink.org

maria.pavlikhina@aboutsomethink.org

LinkedIn : Tilman Scheel

<https://aboutsomethink.org/de/webinare/>

ASK

