



Community of Practice KIPerWeb

Austausch zur Nutzung und Entwicklung KI-gestützter Webanwendungen



KIPerWEB



**Forschungsinstitut
Betriebliche Bildung**

- Orga
 - CoP KIPerWeb
 - Github-Repo KIPerWeb
 - Symposium eingereicht!
- Inhalt
 - Update: Leaderboard & Austausch
 - Tagesthema: „*Potenziale von KI für die Berufs- und Studienorientierung*“
 - Exkurs: Status Quo zur Advanced RAG-Bots

Intro: CoP KIPerWeb



- KIPerWeb-Tagung am 19.04.24
- Projektlaufzeit endete am 30.04.24
- Seit 01.05.24 ehrenamtlicher Austausch
- Community of Praxis zur *digital souveränen Nutzung und Entwicklung KI-gestützter Webanwendungen* (CoP KIPerWeb)
- Details unter <https://github.com/AndreasFischer1985/KIPerWeb>

KIPerWeb
bringt KI und
Weiterbildung
zusammen

Erfahren Sie wie!

- ▶ am 19.04.2024
- ▶ online
- ▶ kostenfrei

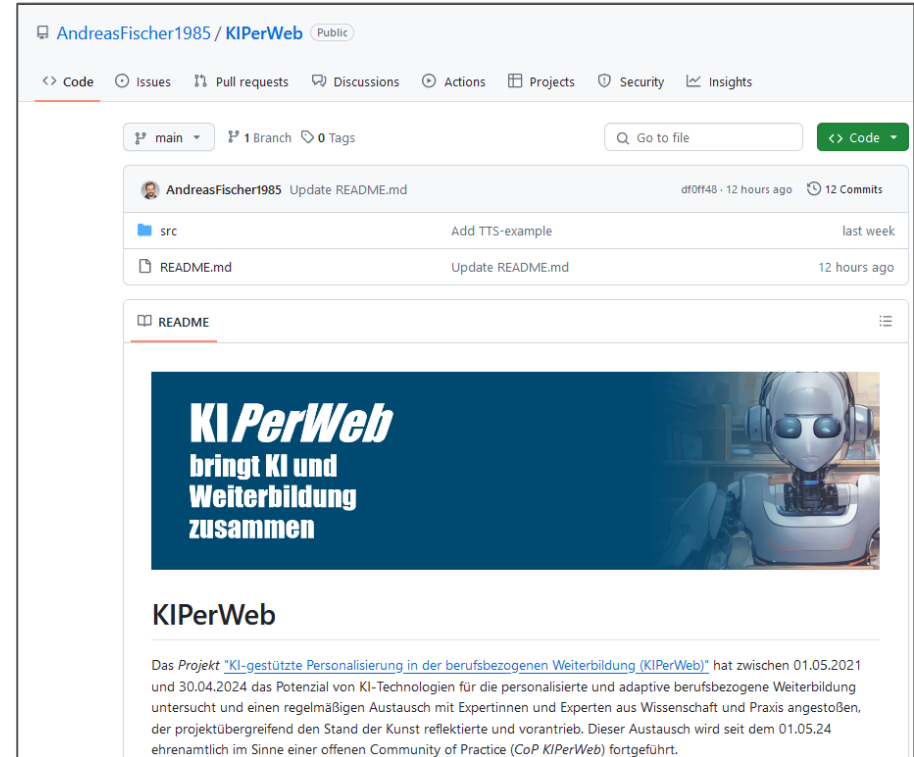
 **KIPerWEB**  

Quelle: https://de.linkedin.com/posts/andreas-fischer-978193198_lernos-k%C3%BCnstliche-intelligenz-mooc-mo-activity-7191408241870876672-rOH4

GitHub Repository KIPerWeb



- Als zentrale Anlaufstelle für Fragen und Anregungen zu unserer CoP KIPerWeb dient bis auf weiteres das GitHub-Repository
- [AndreasFischer1985/KIPerWeb](https://github.com/AndreasFischer1985/KIPerWeb)





FORSCHUNG



BWP '24

JAHRESTAGUNG 2024 DER SEKTION BERUFS- UND
WIRTSCHAFTSPÄDAGOGIK DER DEUTSCHEN
GESELLSCHAFT FÜR ERZIEHUNGSWISSENSCHAFT

Symposium
eingereicht



50 Jahre Zukunft

BWP JAHRESTAGUNG | DRESDEN | SEPTEMBER 2024

© Martin Müller (TUD)

Leaderboard Update (01.05.24) und Austausch



Wo liegen FOSS-Modelle?

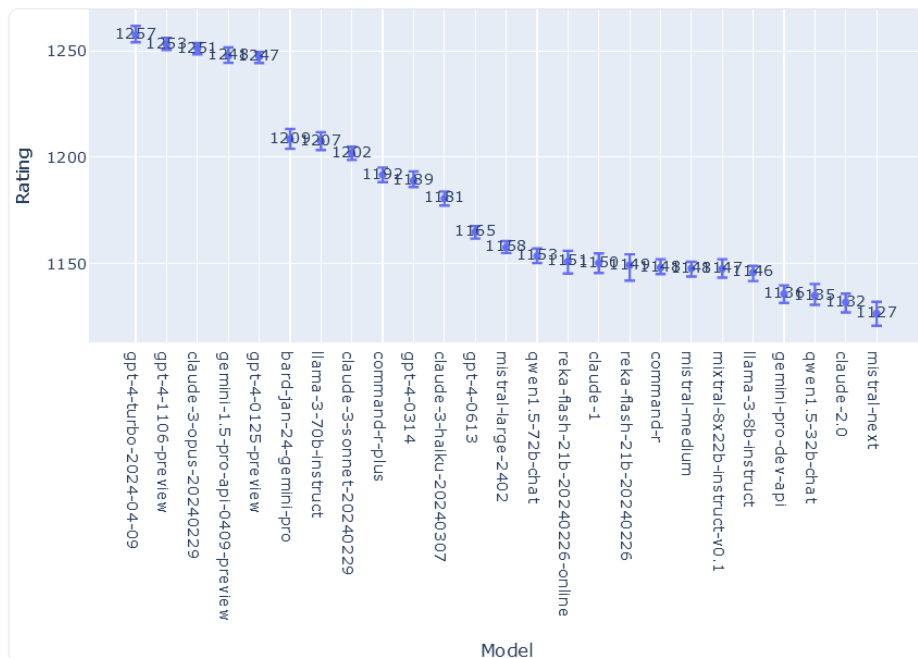
Llama-3-70B-Instruct &
Command R+ vor GPT-4-0314;

Mixtral-8x22b-Instruct-v0.1 &
Llama-3-8b-Instruct vor GPT-3.5-
Turbo-0613 (legacy);

Starling-LM-7B-beta,
WizardLM-70B-v1.0 &
Mixtral-8x7B -Instruct vor
GPT3.5-Turbi-0125 (latest)

Rank* (UB)	Model	Arena Elo
1	GPT-4-Turbo-2024-04-09	1257
1	GPT-4-1106-preview	1253
2	Claude 3 Opus	1251
2	Gemini 1.5 Pro API-0409-Preview	1248
3	GPT-4-0125-preview	1247
6	Bard (Gemini Pro)	1209
6	Llama-3-70b-Instruct	1207
6	Claude 3 Sonnet	1202
9	Command R+	1192
9	GPT-4-0314	1189
11	Claude 3 Haiku	1181
12	GPT-4-0613	1165
13	Mistral-Large-2402	1158
13	Qwen1.5-72B-Chat	1153


Figure 3: Confidence Intervals on Model Strength (via Bootstrapping)



#AGBFN24 - KI in Berufs-/Studienorientierung



- KI weist große Potenziale für BO auf, insb. als Tool für Berufsberatende
- Spannende Ansätze:
 - Berufsempfehlungen auf Basis von tatsächlichen Interessen
 - Iteratives Fragen/Filtern von Berufen und Berufenet-Daten nach berufswahlrelevanten Kriterien
 - Integration von LLMs und Sentence Embeddings
- Anknüpfungspunkte:
 - KIWI: Chatbot,
 - WISY@KI: ESCO-Verschlagwortung
 - BERUFENET.AI - https://github.com/AndreasFischer1985/code-snippets/blob/master/html/js_berufenet.ai.html
- bzw. Gradio-Variante mit Chat: <https://huggingface.co/spaces/AFischer1985/BERUFENET.AI>

AGBFN  Arbeitsgemeinschaft
Berufsbildungsforschungsnetz

Die AG BFN

ROUND TABLE: BO-ANGEBOTE UND KI

Recruiting Games. Eine explorative Studie zur Wirksamkeit von spielerischen Online-Self-Assessments zur Verbesserung der beruflichen Orientierung junger Menschen »

Einsatz von KI in der Berufsberatung »

Prompting für die Berufliche Orientierung in der Schulpraxis – Ein Einblick in den Prozess der Gestaltung einer Form der Community-of-Practice mit Lehrkräften für den Einsatz von Dialog-KI »

Integration von Künstlicher Intelligenz in ein übergeordnetes Orientierungsleitsystem – Chancen und Risiken »

Exkurs: Status Quo zur Advanced RAG-Bots



- **LLM-Chatbot**
 - auf Basis von Mixtral-8x7b (HFHub oder llama.cpp-server) der auf User-message und history antwortet
- **Retrieval Augmented Generation (RAG)**
 - bezieht ggf. zusätzlich Informationen ein, die semantisch zur Anfrage passen (aus Teilmengen einer ChromaDB)
 - Speichert und berücksichtigt ggf. Dialog-Episoden als separate Teilmenge in die ChromaDB (**Memory-RAG**)
 - Pos. & Neg. Intentionen als weitere separate Teilmenge der ChromaDB werden zunächst mit Anfrage verglichen, um anschließend RAG anzupassen (**Hierarchical Goals-RAG**)



Online-Demos RAG-basierter Chatbots



The image displays three overlapping screenshots of web-based chatbot interfaces, each demonstrating a Retrieval-Augmented Generation (RAG) system. The interfaces are titled "German AI-RAG-Interface", "AI-Interface", and "German AI-Interface with advanced RAG (HFHub)".

Each interface features a chat window with a welcome message and a list of capabilities. The chat window also includes a text input field and buttons for "Retry", "Undo", and "Clear".

German AI-RAG-Interface: The chat window displays a welcome message and a list of capabilities. The input field contains the text "type a message...".

AI-Interface: The chat window displays a welcome message and a list of capabilities. The input field contains the text "type a message...".

German AI-Interface with advanced RAG (HFHub): The chat window displays a welcome message and a list of capabilities. The input field contains the text "type a message...".

<https://huggingface.co/spaces/AFischer1985/AI-RAG-Interface-to-Hub>

https://huggingface.co/spaces/AFischer1985/AI-Interface_memRAG