



Community of Practice KIPerWeb

Austausch zur Nutzung und Entwicklung KI-gestützter Webanwendungen



KIPerWEB



Forschungsinstitut
Betriebliche Bildung

- **Update**
 - News & Leaderboard-Update
- **Input**
 - „KI-Kompetenzen“ (Gastvortrag: Martin Lindner)
- **Diskussion**

- DeepSeek-r1 löst DeepSeek-V3 in der Kategorie „German“ als Spitzenreiter unter den Open-Weights-Modellen (und generell unter allen Modellen) ab
 - Reasoning-Modell
 - Vorsicht bzgl. Ideologie-Filtern und Faktentreue
 - Zahlreiche Derivate, auch für den Hausgebrauch:
 - DeepSeek-R1-Distill-Qwen-7B
 - DeepSeek-R1-Distill-LLama-8B
- Mistral veröffentlicht Mistral-Small-3 (Apache-2.0)
<https://mistral.ai/news/mistral-small-3/>



News & Leaderboard-Update (22.01.2025)



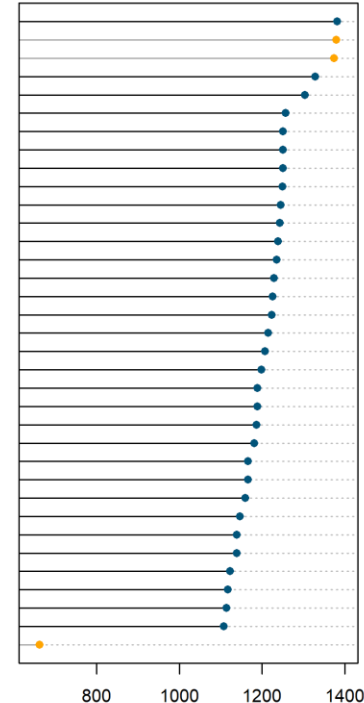
Arena Score German

based on Imarena.ai on 06. Feb 2025

Es hat sich einiges getan in der Kategorie „German“:

- Mit **Deepseek-r1** ist erstmals ein Open-Weights-Modell vor allen proprietären Produkten!
- Das Rennen zwischen Google und OpenAI ist nach wie vor knapp – aktuell hat wieder ein Gemini-Modell die Nase vorn.
- Arena-Scores von *nicht*-proprietären Modelle sind rechts ausgewiesen sofern sie mindestens das Niveau von **Gemma-2-2b-it** erreichen:
 - **Phi-4** (14B) hat sich etwas eingependelt und liegt nach wie vor noch vor **Gemma-2-9b-it-SimPO**
- Nicht im Rennen, aber nach wie vor hervorragend ist **Gemma-2-27b-it-SimPo**

deepseek-r1 (DeepSeek)
gemini-2.0-flash-001
ChatGPT-4o-latest (2024-11-20)
deepseek-v3 (DeepSeek)
llama-3.1-nemotron-70b-instruct
Athene-v2-Chat-72b (NexusFlow)
Qwen-Max-0919 (Qwen)
Meta-Llama-3.1-405b-Instruct-bf16 (Llama 3.1)
Mistral-Large-2407 (Mistral Research)
Meta-Llama-3.1-405b-Instruct-fp8 (Llama 3.1)
Athene-70b (CC-BY-NC-4.0)
Meta-Llama-3.3-70B-Instruct (Llama-3.3)
Mistral-Large-2411 (Mistral Research)
Qwen2.5-72b-Instruct (Qwen)
Deepseek-v2.5 (DeepSeek)
Phi-4 (MIT)
Meta-Llama-3.1-70b-Instruct (Llama 3.1)
Gemma-2-9b-it-SimPO (MIT)
Gemma-2-27b-it (Gemma)
Aya-Expansive-32B (CC-BY-NC-4.0)
Aya-Expansive-8B (CC-BY-NC-4.0)
Command R+ (04-2024) (CC-BY-NC-4.0)
Nemotron-4-340B-Instruct (NVIDIA Open Model)
Gemma-2-9b-it (Gemma)
Meta-Llama-3-70b-Instruct (Llama 3)
Llama-3-70b-Instruct (Llama 3)
DeepSeek-Coder-V2-Instruct (DeepSeek)
Qwen2-72B-Instruct (Qianwen)
Meta-Llama-3.1-8b-Instruct (Llama3.1)
Mixtral-8x22b-Instruct-v0.1 (Apache 2.0)
Command R (04-2024) (CC-BY-NC-4.0)
Qwen1.4-110B-Chat (Qianwen)
Mixtral-8x7b-Instruct-v0.1 (Apache 2.0)
Gemma-2-2b-it (Gemma)
Chatglm2-6b



Datengrundlage: <https://chat.lmsys.org/?leaderboard>

(c) 2025 Andreas Fischer

- „Kompetenter Umgang mit KI“

(rechts visualisiert von FLUX.1-schnell)

