

Will my Open Access article be used to train Artificial Intelligence? How do I protect myself?

kiz, Ulm University

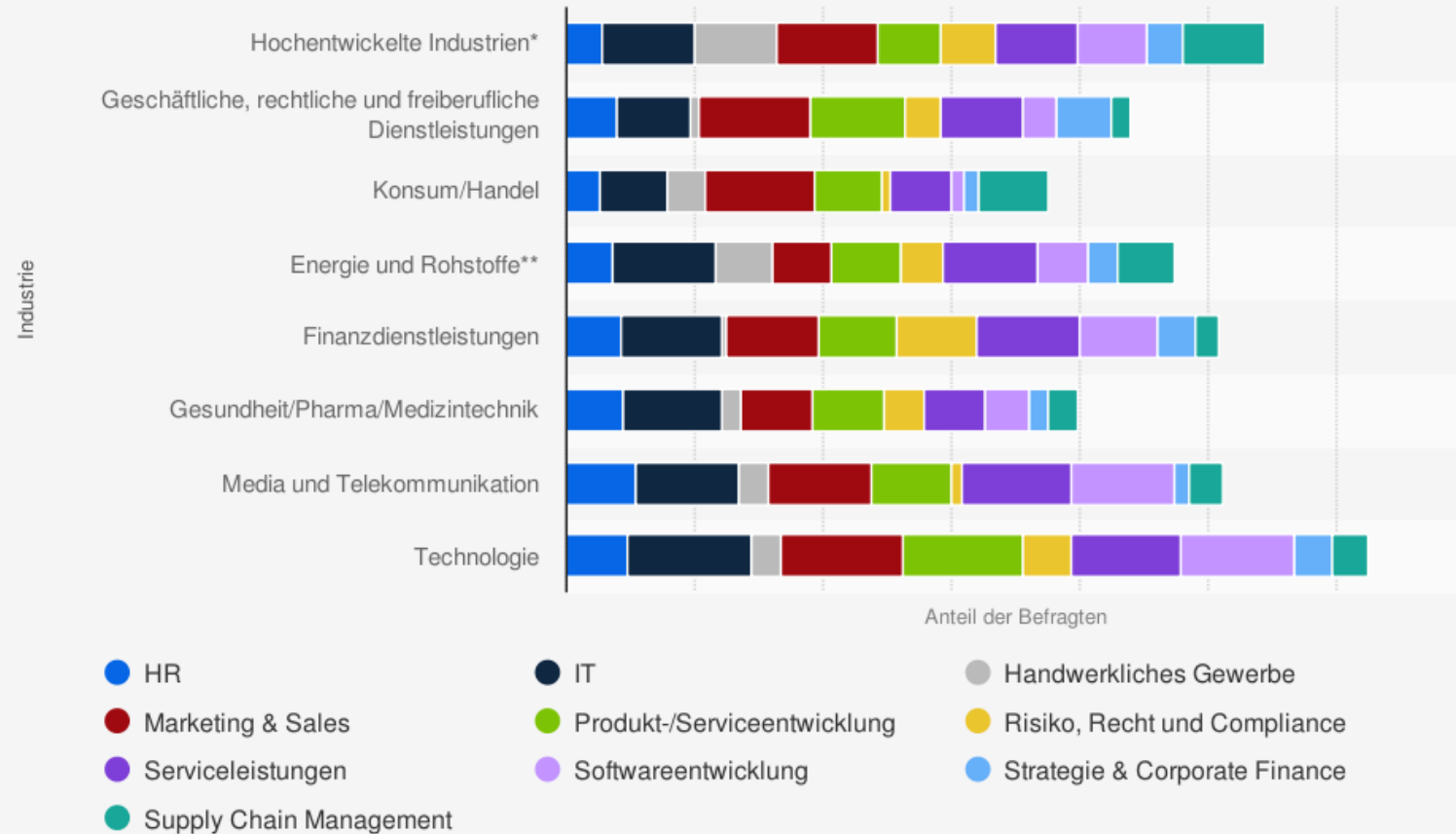
Dr Jonas Mirbeth

Will my Open Access article be used to train Artificial Intelligence?

The short answer: Yes!

But it doesn't have to be to your disadvantage. Here's what you need to know.

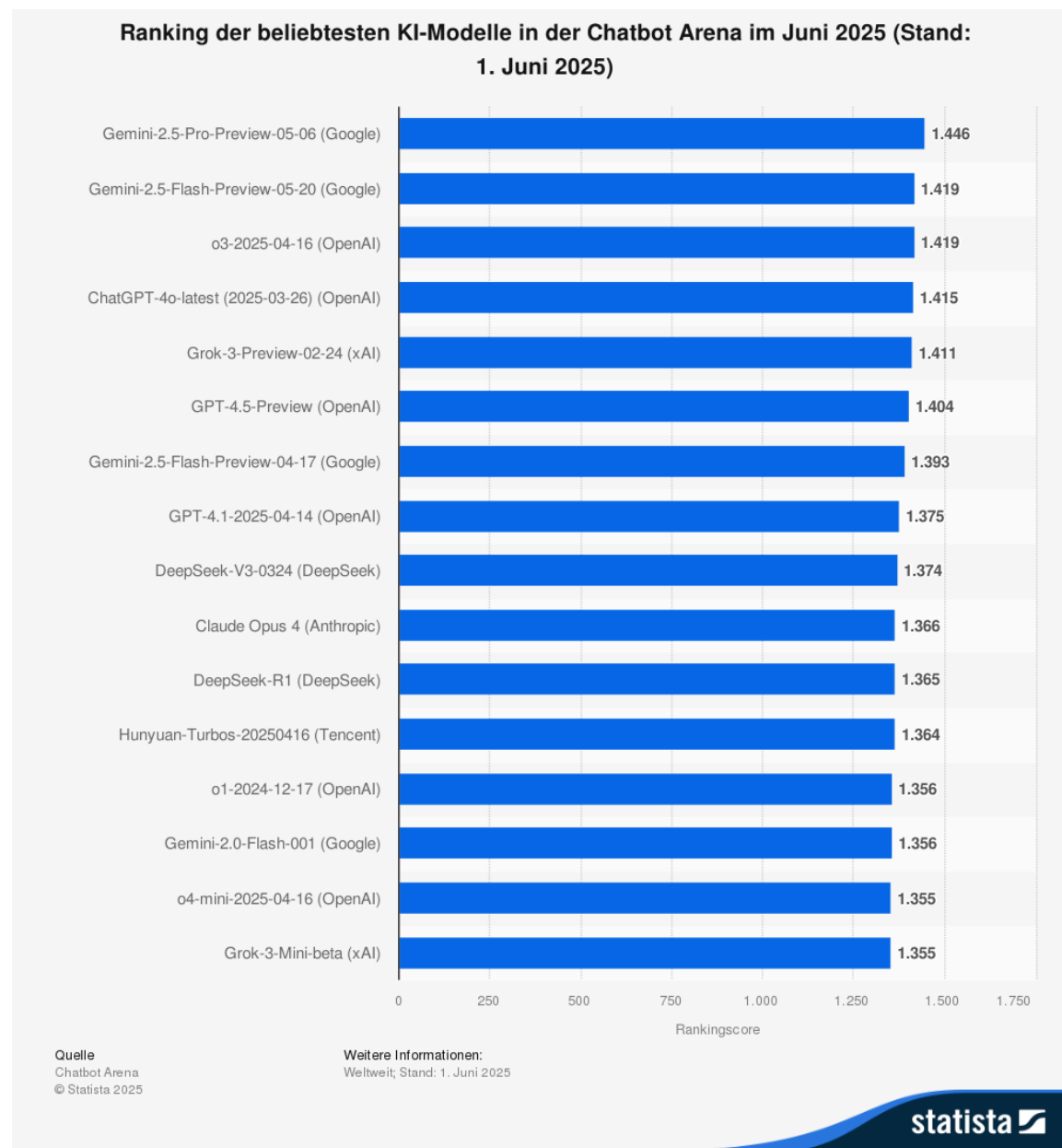
Adaption von Künstlicher Intelligenz (KI) nach Branchen und Funktionen weltweit im Jahr 2024



Quellen
Stanford University; McKinsey & Company
© Statista 2025

Weitere Informationen:
Weltweit; McKinsey & Company; 22. Februar bis 5. März 2024; 1.363
Befragte; Unternehmen, die KI nutzen

statista



Artificial Intelligence: A short overview

- ▶ Artificial intelligence (AI) - “the capability of computer systems or algorithms to imitate intelligent human behavior” (Merriam-Webster Online Dictionary)
- ▶ Founded as an academic discipline in 1956,
- ▶ In recent years, ongoing rapid progress in technological development (modern deep learning techniques and large language models (LLMs))
- ▶ Generative pre-trained transformers (GPT) are LLMs that can process and produce text, image, and video material.

Training multimodal GPT:

- ▶ On large datasets that include text, images, and videos,
- ▶ by crawling the web, mining online content.

Artificial Intelligence: Challenges

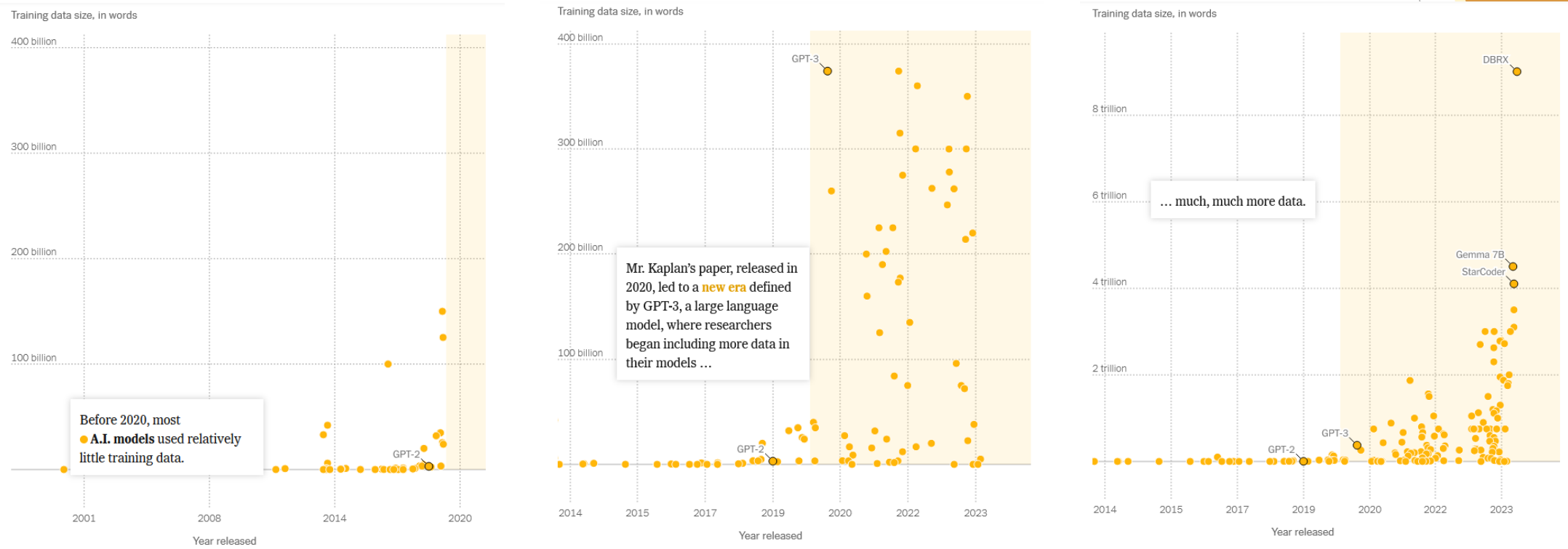
Edited texts, such as (scientific) books and journal articles, are preferred for AI training.

- ▶ Research output has become a valuable resource for AI training.

Rising challenges:

- ▶ Growing need for new data to continue training AI models.
- ▶ Access restrictions by website operators.
- ▶ Legal challenges (law suits for copyright infringement).

‘Scale is all you need’: On size and performance in generative data models



Metz, Cade et al., How Tech Giants Cut Corners to Harvest Data for A.I., NYTimes, updated April 8, 2024, <https://www.nytimes.com/2024/04/06/technology/tech-giants-harvest-data-artificial-intelligence.html>. Jared Kaplan et al (2020), <https://arxiv.org/abs/2001.08361>. Jared Kaplan et al. (2020) were able to prove the connection between size and performance in generative data models. This correlation is the subject of relevant research by the New York Times in 2024: The New York Times charts illustrate the exponential increase in training data sets from before 2020 up to and including 2023. There have long been bottlenecks in human-generated data, so AI companies are increasingly dependent on the use of 'synthetic', i.e. machine-generated data sets.

Open Access in Research Funding

Research assessment:

- ▶ Research funders increasingly mandate Open Access publications.
- ▶ Publishing Green Open Access (self-archiving) is encouraged.

Publishers:

- ▶ Self-archiving is often permitted by publishers.
- ▶ Observe embargo periods, article versions!

→ Find support for research funding at Ulm University's [Center for Research Strategy and Support \(Res.UL\)](#)



kiz | Abt. Medien

Open Access Mandate: EU Horizon Europe

Open Science

Open science: open access to scientific publications

The beneficiaries must ensure open access to peer-reviewed scientific publications relating to their results. In particular, they must ensure that:

- at the latest at the time of publication, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications
- immediate open access is provided to the deposited publication via the repository, under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a licence with equivalent rights; for monographs and other long-text formats, the licence may exclude commercial uses and derivative works (e.g. CC BY-NC, CC BY-ND) and

110

Project: [insert number] — [insert acronym] — [insert call identifier]

EU Grants: HE MGA — Multi & Mono: V1.1 – 15.04.2022

- information is given via the repository about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication.

Beneficiaries (or authors) must retain sufficient intellectual property rights to comply with the open access requirements.

The Creative Commons Licenses



CC BY-NC-SA
Attribution - Non Commercial -
Share Alike



CC BY NC-ND
Attribution - Non Commerical -
No Derivatives



CC0
Waiver of all copyright and related
rights ("No rights reserved")



CC BY
Attribution



CC BY SA
Attribution - Share Alike



CC BY NC
Attribution - Non
Commercial



CC BY ND
Attribution - No Derivatives

Based on: <https://de.creativecommons.net/was-ist-cc/> Accessed: Dec 9, 2022

You are free to:

Share — copy and redistribute the material in any medium or format for any purpose, even commercially.

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

 **Attribution** - You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions - You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

  CC BY 4.0

ATTRIBUTION 4.0 INTERNATIONAL
Deed

Do NC/ND licences protect against AI training?

- ▶ No. There's a growing need for new data to continue training AI models and business practices that do so, many of which are facing legal changes.

How to use Open Access to counteract the formation of monopolies in AI:

- ▶ A certain loss of control over published content is inevitable.
- ▶ Prevent the concentration of exclusive access to scientific information without extra costly licences required by publishers.

Publishing Open Access

Standard und Best Practice

- ▶ In the spirit of Ulm University's [Open-Access-Resolution](#)

In the context of DEAL contracts ([Elsevier](#), [Springer Nature](#), [Wiley](#)), we recommend the CC BY 4.0 license:

- ▶ Authors' interests remain protected
- ▶ Loss or transfer of exclusive rights of use to the publisher by selecting NC and ND license components (Elsevier/Wiley)

Takeaways

1. Apply for Open Access funding at kiz
2. Choose CC BY 4.0
3. Make use of persistent identifiers (DOI, ORCID, ROR etc.)
4. Inform yourself how to use AI correctly when doing research and publishing
5. Practice Open Science

Your Questions - Q&A

Funding requests:

- ▶ oa@uni-ulm.de
- ▶ +49 (0)731 50 - 154 76

Licenses & Green Open Access

- ▶ kiz.publikationsmanagement@uni-ulm.de
- ▶ +49 (0)731 50 - 314 28

All topics presented and further information can be found under the following links:

<https://www.uni-ulm.de/open-access>

<https://www.uni-ulm.de/oa-foerderung>