

Acht Begriffe und Bezeichnungen

Diese kurzen Erklärungen enthalten auch Einschränkungen. Eine Bewegung, ein Forschungsfeld und eine von Kritikern verwendete Bezeichnung sind unterschiedliche Dinge.

Mit KI-Unterstützung erstellt. Nicht unabhängig auf Richtigkeit geprüft. Prüfe die Quellen und Einschränkungen, bevor du Schlussfolgerungen ziehst.

Künstliche Intelligenz (KI)

KI ist ein Gebiet der Informatik. Ihre Systeme nutzen Computerprogramme und Hardware, um Muster zu erkennen, Inhalte zu erzeugen oder Vorhersagen zu berechnen. [\[23\]](#) [\[22\]](#) [\[14\]](#)

Beachte: KI und künstliche allgemeine Intelligenz (AGI) sind verschiedene Begriffe. Ein System kann eine bestimmte Aufgabe gut erfüllen, ohne breite Fähigkeiten auf menschlichem Niveau zu besitzen. [\[10\]](#)

[Vollständige Erklärung und Quellenhinweise lesen](#)
theaiatlas.org/de/ideas/ai/

Sprachmodelle und große Sprachmodelle (LLMs)

Ein LLM ist ein Modell, das für die Verarbeitung von Sprache trainiert wurde. Computersoftware führt es aus und erzeugt bei der Texterzeugung eine Antwort Stück für Stück. [\[16\]](#) [\[17\]](#) [\[15\]](#)

Beachte: Ein Modell kann einen überzeugenden Satz erzeugen, ohne ihn mit einer externen Quelle abzugleichen. Suchfunktionen oder andere Werkzeuge muss die umgebende Anwendung bereitstellen. [\[13\]](#) [\[18\]](#)

[Vollständige Erklärung und Quellenhinweise lesen](#)
theaiatlas.org/de/ideas/llm/

e/acc: Effektiver Akzelerationismus

Eine Bewegung, die schnelleres technisches Wachstum befürwortet und zentralisierte Beschränkungen ablehnt. Die Behauptung, Beschleunigung führe zu besseren Ergebnissen, ist eine philosophische Position, keine nachgewiesene Sicherheitsgarantie. [\[3\]](#) [\[4\]](#)

Beachte: Allgemeiner Technikoptimismus bedeutet keine e/acc-Zugehörigkeit. Die thermodynamischen Argumente der Bewegung belegen nicht, dass KI sicher ist. Buterin setzt einen Gegenpunkt: Gewinnstreben allein wählt nicht automatisch nützliche Entwicklungsrichtungen aus. [\[4\]](#) [\[6\]](#)

[Vollständige Erklärung und Quellenhinweise lesen](#)
theaiatlas.org/de/ideas/eacc/

Decel: Verlangsamung

Ein informeller, oft abwertender Begriff für den Wunsch nach langsamerer KI-Entwicklung. Prüfe, welche Grenzen die Person tatsächlich befürwortet. [\[4\]](#)

Beachte: Eine Sicherheitsregel, eine Pause bei den leistungsfähigsten KI-Systemen und die Ablehnung aller Technik sind verschiedene Positionen. Laut PauseAI wären spezialisierte Anwendungen vom vorgeschlagenen Vertrag meist nicht betroffen. Die Gruppe schlägt aber auch vor, längerfristige Grenzen für Forschung und Hardware zu erwägen. [\[8\]](#)

[Vollständige Erklärung und Quellenhinweise lesen](#)
theaiatlas.org/de/ideas/decel/

Weitere Begriffe und Bezeichnungen

KI-Doomer

Ein umstrittener Begriff für Menschen, die katastrophale KI-Risiken betonen. Besorgnis bedeutet nicht, eine Katastrophe für unausweichlich zu halten. [\[5\]](#) [\[9\]](#)

Beachte: Die CAIS-Erklärung fordert, das Risiko des Aussterbens zu verringern. Sie erklärt das Aussterben weder für sicher noch nennt sie eine gemeinsame Wahrscheinlichkeit. Der Atlas bezeichnet niemanden automatisch als Doomer. [\[9\]](#)

[Vollständige Erklärung und Quellenhinweise lesen](#)
theaiatlas.org/de/ideas/doomer/

EA: Effektiver Altruismus

Eine Gemeinschaft, die wirksame Wege sucht, anderen zu helfen. Kritiker fragen, wessen Maßstäbe für Nutzen zählen und wie viel Macht Geldgeber haben sollten. [\[2\]](#) [\[19\]](#) [\[20\]](#)

Beachte: Alice Crary argumentiert, dass Nutzenmaße politische Ursachen von Schäden übersehen können. Emma Saunders-Hastings warnt vor der Macht von Geldgebern über Menschen, die Hilfe erhalten. Die EA-Website erwidert, dass institutioneller Wandel zu ihrem Ansatz gehört und EA nicht utilitaristisch sein muss. Diese Debatten betreffen die Definition von Hilfe und die Entscheidungsgewalt darüber. EA, Longtermismus und KI-Sicherheit sind nicht austauschbar. [\[19\]](#) [\[20\]](#) [\[21\]](#) [\[7\]](#)

[Vollständige Erklärung und Quellenhinweise lesen](#)
theaiatlas.org/de/ideas/ea/

KI-Sicherheit und Alignment

Forschung mit dem Ziel, KI-Schäden zu verringern sowie Verlässlichkeit und Ausrichtung auf vorgesehene Ziele zu verbessern. Ein Sicherheitsziel oder ein Regelwerk beweist nicht, dass ein System sicher ist. [\[12\]](#) [\[11\]](#) [\[1\]](#)

Beachte: Ein Sicherheitsrahmen hält erklärte Schutzmaßnahmen fest. Er garantiert weder deren sichere Umsetzung noch liefert er eine gemessene Katastrophenwahrscheinlichkeit. [\[1\]](#)

[Vollständige Erklärung und Quellenhinweise lesen](#)
theaiatlas.org/de/ideas/safety/

p(doom): Geschätzte Katastrophenwahrscheinlichkeit

Die persönliche Einschätzung, wie wahrscheinlich eine KI-Katastrophe ist. Prüfe, auf welche Folgen, welchen Zeitraum und welche Annahmen sich die Zahl bezieht. [\[5\]](#) [\[24\]](#)

Beachte: Dies sind Einschätzungen, keine gemessenen Katastrophenraten. Grace und Kollegen fragten nach verschiedenen Folgen, Kontrollverlust und Zeitgrenzen. Sie warnen: KI-Fachwissen garantiert keine guten Vorhersagen, und wer teilnimmt, kann die Ergebnisse beeinflussen. Sie halten die Schätzungen der Forschenden dennoch für nützlich. [\[24\]](#)

[Vollständige Erklärung und Quellenhinweise lesen](#)
theaiatlas.org/de/ideas/pdoom/

Quellen für das Begriffsblatt

Die Zahlen neben den Erklärungen verweisen auf die angegebenen Veröffentlichungen unten. Die vollständigen Lesehinweise und Einschränkungen des Quellenabrufs sind bei jedem Begriff verlinkt. Ein Quellenverweis beweist nicht, dass der Autor recht hat.

1. [Strengthening our Frontier Safety Framework](#)

Google DeepMind · Quelle aus erster Hand · Veröffentlicht 2025-09-22 · Aktualisiert 2026-04-17

<https://deepmind.google/blog/strengthening-our-frontier-safety-framework/>

2. [What is effective altruism?](#)

Effective Altruism · Quelle aus erster Hand · Veröffentlicht Datum nicht verfügbar

<https://www.effectivealtruism.org/articles/introduction-to-effective-altruism>

3. [what the f* is e/acc](#)

e/acc newsletter · Quelle aus erster Hand · Veröffentlicht 2022-12-26

<https://effectiveaccelerationism.substack.com/p/what-the-f-is-eacc>

4. [Notes on e/acc principles and tenets](#)

e/acc newsletter / BasedBeffJezos and bayeslord · Quelle aus erster Hand · Veröffentlicht 2022-10-31

<https://effectiveaccelerationism.substack.com/p/repost-notes-on-eacc-principles-and>

5. [Guillaume Verdon: e/acc, AI doomers and effective altruism \(transcript #407\)](#)

Lex Fridman Podcast · Quelle aus erster Hand · Veröffentlicht 2023-12-29

<https://lexfridman.com/guillaume-verdon-transcript/>

6. [My techno-optimism](#)

Vitalik Buterin · Quelle aus erster Hand · Veröffentlicht 2023-11-27

https://vitalik.eth.limo/general/2023/11/27/techno_optimism.html

7. [Longtermism](#)

William MacAskill · Quelle aus erster Hand · Veröffentlicht Datum nicht verfügbar

<https://www.williammacaskill.com/longtermism>

8. [PauseAI Proposal \(April 2026 version\)](#)

PauseAI · Quelle aus erster Hand · Veröffentlicht 2026-04-05

<https://pauseai.org/proposal>

9. [Statement on AI Extinction Risk](#)

Center for AI Safety · Quelle aus erster Hand · Veröffentlicht Datum nicht verfügbar

<https://aistatement.com/work/statement-on-ai-extinction-risk>

10. [Levels of AGI for Operationalizing Progress on the Path to AGI](#)

Meredith Ringel Morris and coauthors / arXiv · Quelle aus erster Hand · Veröffentlicht 2023-11-04 · Aktualisiert 2025-09-24

<https://arxiv.org/abs/2311.02462>

11. [Risks from Learned Optimization in Advanced Machine Learning Systems](#)

Evan Hubinger and coauthors / arXiv · Quelle aus erster Hand · Veröffentlicht 2019-06-05 · Aktualisiert 2021-12-01

<https://arxiv.org/abs/1906.01820>

12. [Concrete Problems in AI Safety](#)

Dario Amodei and coauthors / arXiv · Quelle aus erster Hand · Veröffentlicht 2016-06-21 · Aktualisiert 2016-07-25

<https://arxiv.org/abs/1606.06565>

13. [Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile](#)

National Institute of Standards and Technology · Quelle aus erster Hand · Veröffentlicht 2024-07

<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf>

14. [Machine Learning Glossary](#)

Google for Developers · Quelle aus erster Hand · Veröffentlicht Datum nicht verfügbar · Aktualisiert 2026-04-10

<https://developers.google.com/machine-learning/glossary>

15. [LLMs: What's a large language model?](#)

Google for Developers · Quelle aus erster Hand · Veröffentlicht Datum nicht verfügbar · Aktualisiert 2026-01-02

<https://developers.google.com/machine-learning/crash-course/llm/transformers>

16. [Models](#)

Hugging Face LLM Course · Quelle aus erster Hand · Veröffentlicht Datum nicht verfügbar

<https://huggingface.co/learn/llm-course/en/chapter2/3>

17. [Text generation](#)

Hugging Face Transformers documentation · Quelle aus erster Hand · Veröffentlicht Datum nicht verfügbar

https://huggingface.co/docs/transformers/en/llm_tutorial

18. [Tool use](#)

Hugging Face Transformers documentation · Quelle aus erster Hand · Veröffentlicht Datum nicht verfügbar

https://huggingface.co/docs/transformers/en/chat_extras

19. [Against 'Effective Altruism'](#)

Alice Crary / Radical Philosophy · Quelle aus erster Hand · Veröffentlicht 2021

<https://www.radicalphilosophy.com/article/against-effective-altruism>

20. [Response to Effective Altruism](#)

Emma Saunders-Hastings / Boston Review · Quelle aus erster Hand · Veröffentlicht 2015-07-01

<https://www.bostonreview.net/forum/peter-singer-logic-effective-altruism/response-emma-saunders-hastings/>

21. [Frequently asked questions and common objections](#)

EffectiveAltruism.org · Quelle aus erster Hand · Veröffentlicht Datum nicht verfügbar

<https://www.effectivealtruism.org/faqs>

22. [software](#)

National Institute of Standards and Technology · Quelle aus erster Hand · Veröffentlicht Datum nicht verfügbar

<https://csrc.nist.gov/glossary/term/software>

23. [Explanatory memorandum on the updated OECD definition of an AI system](#)

Organisation for Economic Co-operation and Development · Quelle aus erster Hand · Veröffentlicht 2024-03-05

https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/03/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_3c815e51/623da898-en.pdf

24. [Thousands of AI Authors on the Future of AI \(version 3\)](#)

Katja Grace and colleagues / arXiv · Quelle aus erster Hand · Veröffentlicht 2024-01-05 · Aktualisiert 2025-10-08

<https://arxiv.org/html/2401.02843v3>

Mit KI-Unterstützung erstellt. Nicht unabhängig auf Richtigkeit geprüft. Prüfe die Quellen und Einschränkungen, bevor du Schlussfolgerungen ziehst.

Dies ist der Stand des Datensatzes, nicht das Datum, an dem jede Quelle gelesen wurde. Der gespeicherte Datensatz bewahrt genau die Belege, auf denen dieses Material beruht.

Version des Materials: 0.22.0 · Quellenstichtag: 2026-09-16. [Gespeicherter Datensatz mit Belegen](#).

Prüfe die aktuellen Erklärungen und veröffentlichten Änderungen, bevor du älteres Material erneut verwendest. [Inhaltliche Änderungen](#).

Erstellt von [Timo Fahlenbock](#) · [Impressum](#)