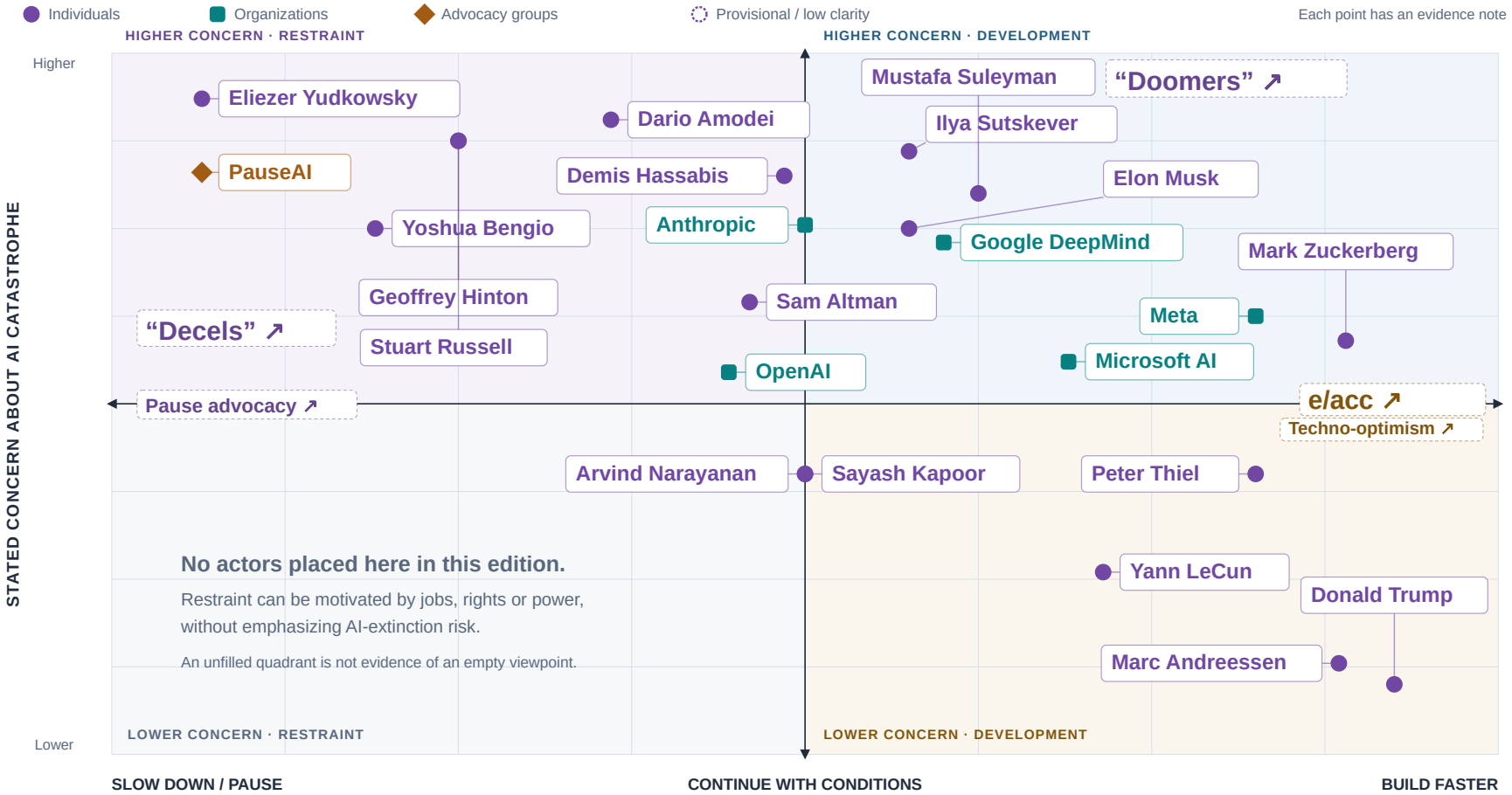
THE AI ATLAS

AI risk × development pace

v0.22.0

23 plotted / 15 unplaced

Frontier AI debate · Evidence cutoff 2026-09-16 · Qualitative, not a ranking



IDEAS & TERMS · Labels do not assign anyone to a group. Select for explanations and sources.

RELATED IDEAS EA ↗ · d/acc ↗ · Longtermism ↗ · AI safety ↗ · Computer programs ↗ · Normal technology ↗

READING THE MAP Coordinates reflect stated positions, not p(doom), safety performance or verified implementation.
EA is a community. AI safety is a field. Neither is a single dot. People and organizations are separate.

Sources and explanations: theaiatlas.org · Positions interpret public statements; they are not rankings

Eight ideas and labels

These short explanations include qualifications. A movement, a field of research and a label used by critics are different kinds of thing.

Prepared with AI assistance. Not independently fact-checked. Check the sources and qualifications before drawing conclusions.

Artificial intelligence (AI)

AI is a field of computing. Its systems use computer programs and hardware to recognize patterns, generate content or make predictions. [23] [22] [14]

Keep in mind: AI and artificial general intelligence (AGI) are different terms. A system can perform a particular task well without having broad human-level abilities. [10]

Read the full explanation and source notes
theaiatlas.org/ideas/ai/

Language models & large language models (LLMs)

An LLM is a model trained to process language. Computer software runs it and, for text generation, produces an answer one piece at a time. [16] [17] [15]

Keep in mind: A model can produce a convincing sentence without checking it against an external source. Search or other tools must be supplied by the surrounding application. [13] [18]

Read the full explanation and source notes
theaiatlas.org/ideas/llm/

e/acc: Effective accelerationism

A movement favoring faster technological growth and opposing centralized restraint. Its claim that acceleration leads to better outcomes is a philosophical position, not a demonstrated safety guarantee. [3] [4]

Keep in mind: Broader techno-optimism does not imply e/acc membership. Its thermodynamic arguments do not establish that AI is safe. Buterin offers a counterpoint: profit alone does not automatically select beneficial directions for technology. [4] [6]

Read the full explanation and source notes
theaiatlas.org/ideas/eacc/

Decel: Deceleration

An informal label for wanting slower AI development, often used as criticism. Check which limits a person actually supports. [4]

Keep in mind: A safety rule, a pause on the most powerful AI and opposition to all technology are different positions. PauseAI says its proposed treaty would usually leave narrow applications unaffected, but also proposes considering longer-term research and hardware restrictions. [8]

Read the full explanation and source notes
theaiatlas.org/ideas/decel/

Pack version: 0.22.0 · Evidence cutoff: 2026-09-16. Saved evidence dataset.

More ideas and labels

AI doomer

A disputed label for people emphasizing catastrophic AI risks. Concern does not mean believing disaster is inevitable. [5] [9]

Keep in mind: The CAIS statement calls for reducing extinction risk; it does not say extinction is certain or give a shared probability. The Atlas does not automatically label anyone a doomer. [9]

Read the full explanation and source notes
theaiatlas.org/ideas/doomer/

EA: Effective altruism

A community seeking effective ways to help others. Critics question whose measures of benefit count and how much power donors should have. [2] [19] [20]

Keep in mind: Alice Crary argues that measures of benefit can miss political causes of harm. Emma Saunders-Hastings warns about donors' power over people receiving help. EA's own FAQ replies that institutional change belongs in its scope and that EA need not be utilitarian. These disputes concern how help is defined and governed. EA, longtermism and AI safety are not interchangeable. [19] [20] [21] [7]

Read the full explanation and source notes
theaiatlas.org/ideas/ea/

AI safety & alignment

Research aimed at reducing AI harms and improving reliability and alignment with intended goals. A safety goal or framework is not proof that a system is safe. [12] [11] [1]

Keep in mind: A safety framework records stated safeguards, not a guarantee of safe implementation or a measured catastrophe probability. [1]

Read the full explanation and source notes
theaiatlas.org/ideas/safety/

p(doom): Probability of doom

Someone's estimated chance of an AI disaster. Check what outcome, time period and assumptions the number refers to. [5] [24]

Keep in mind: These are judgments, not measured disaster rates. Grace and colleagues asked about different outcomes, loss of control and time limits. They warn that AI expertise does not guarantee forecasting skill and who responds can affect results. They still consider researchers' estimates useful. [24]

Read the full explanation and source notes
theaiatlas.org/ideas/pdoom/

Pack version: 0.22.0 · Evidence cutoff: 2026-09-16. Saved evidence dataset.

Three discussion questions

1. What does the map tell you?

Choose two people. What does each say about the pace of development, and what does each say about catastrophic risk?

- Find a dated source for each axis. Separate a personal statement from an organization's policy.
- Compare the source wording with the map interpretation. What could justify a different placement?

Suggested reading: Yann LeCun · Yoshua Bengio

2. Who chose the label?

Compare e/acc, decel and doomer. Which labels appear in supporters' descriptions, and which appear in criticism?

- Open one cited source for each label. Identify who is speaking and what they are arguing.
- Explain what a label leaves out. What evidence would you need before applying it to a person?

Suggested reading: Effective accelerationism · Deceleration · AI doomer

3. Where could a system go wrong?

Imagine an application that drafts email replies. Compare a draft that a person checks with a reply the application sends automatically.

- Use the drawing to mark the model, the surrounding software and the decision to send.
- Which permissions, checks and records would you want? Explain which question each measure helps answer and what remains uncertain.

Suggested reading: AI system · Set limits before it takes action · Keep a useful record of what happened

An example to discuss: drafting an email

Illustrative workflow, not a tested control design. Discuss who should decide whether a message is sent.

Your notes and source links

Pack version: 0.22.0 · Evidence cutoff: 2026-09-16. Saved evidence dataset.

Sources for the idea sheet

Numbers beside the explanations lead to the cited publications below. The full reading notes and retrieval limits are linked from each idea. A citation is not proof that its author is right.

1. Strengthening our Frontier Safety Framework

   Google DeepMind · First-hand source · Published 2025-09-22 · Updated 2026-04-17

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

2. What is effective altruism?

   Effective Altruism · First-hand source · Published date unavailable

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

3. what the f* is e/acc

   e/acc newsletter · First-hand source · Published 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 · First-hand source · Published 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 · First-hand source · Published 2023-12-29

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

6. My techno-optimism

   Vitalik Buterin · First-hand source · Published 2023-11-27

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

7. Longtermism

   William MacAskill · First-hand source · Published date unavailable

   https://www.williammacaskill.com/longtermism

8. PauseAI Proposal (April 2026 version)

   PauseAI · First-hand source · Published 2026-04-05

   https://pauseai.org/proposal

9. Statement on AI Extinction Risk

Center for AI Safety · First-hand source · Published date unavailable

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 · First-hand source · Published 2023-11-04 · Updated 2025-09-24

https://arxiv.org/abs/2311.02462

11. Risks from Learned Optimization in Advanced Machine Learning Systems

Evan Hubinger and coauthors / arXiv · First-hand source · Published 2019-06-05 · Updated 2021-12-01

https://arxiv.org/abs/1906.01820

12. Concrete Problems in AI Safety

Dario Amodei and coauthors / arXiv · First-hand source · Published 2016-06-21 · Updated 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 · First-hand source · Published 2024-07

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

14. Machine Learning Glossary

Google for Developers · First-hand source · Published date unavailable · Updated 2026-04-10

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

15. LLMs: What's a large language model?

Google for Developers · First-hand source · Published date unavailable · Updated 2026-01-02

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

16. Models

Hugging Face LLM Course · First-hand source · Published date unavailable

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

17. Text generation

Hugging Face Transformers documentation · First-hand source · Published date unavailable

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

18. Tool use

Hugging Face Transformers documentation · First-hand source · Published date unavailable

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

19. Against ‘Effective Altruism’

Alice Crary / Radical Philosophy · First-hand source · Published 2021

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

20. Response to Effective Altruism

Emma Saunders-Hastings / Boston Review · First-hand source · Published 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 · First-hand source · Published date unavailable

https://www.effectivealtruism.org/faqs

22. software

National Institute of Standards and Technology · First-hand source · Published date unavailable

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 · First-hand source · Published 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 · First-hand source · Published 2024-01-05 · Updated 2025-10-08

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

Prepared with AI assistance. Not independently fact-checked. Check the sources and qualifications before drawing conclusions.

This is the dataset cutoff, not the date every source was read. The saved dataset preserves the exact evidence used for this pack.

Pack version: 0.22.0 · Evidence cutoff: 2026-09-16. Saved evidence dataset.

Check the latest explanations and published changes before reusing an older pack. Content changes.

Created by Timo Fahlenbock · Legal notice (Impressum)