**Anthropic PBC** is an [American](https://de.wikipedia.org/wiki/Vereinigte_Staaten)[artificial intelligence](https://de.wikipedia.org/wiki/K%C3%BCnstliche_Intelligenz) (AI) company founded in 2021 by former members of [OpenAI .](https://de.wikipedia.org/wiki/OpenAI)[[ 1 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-1)[[ 2 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-JV-2) It specializes in the development of AI systems and [language models](https://de.wikipedia.org/wiki/Sprachmodell) and is committed to the responsible use of AI. Anthropic is a *Public*[*Benefit Corporation*](https://de.wikipedia.org/wiki/Benefit_Corporation) registered in the US [state of Delaware](https://de.wikipedia.org/wiki/Bundesstaat_der_Vereinigten_Staaten) .

Products such as the [chatbot](https://de.wikipedia.org/wiki/Chatbot)*Claude* attempt to achieve compliance with generally accepted ethical rules through specific training based on principles established by, among others, the [United Nations .](https://de.wikipedia.org/wiki/Vereinte_Nationen)[[ 3 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-3)[[ 4 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-4) Anthropic therefore refers to its developments as *Constitutional Artificial Intelligence* , which is based on a *constitution of human values* . [[ 5 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-5)

Early investors included Alameda Research and [Google](https://de.wikipedia.org/wiki/Google_LLC) . [[ 2 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-JV-2)[[ 6 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-SH-6) Google invested nearly $400 million. Anthropic announced a formal partnership with [Google Cloud](https://de.wikipedia.org/wiki/Google_Cloud_Platform) . [[ 7 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-7)[[ 8 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-8)[[ 9 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-9)

In 2023, Anthropic sought investors to train a successor model to *Claude for $5 billion by 2025 or 2026*[[ 10 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-10) and to keep pace with the development of other models. [[ 11 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-11)[[ 12 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-12)[In 2023, the venture capital firm Sparc Capital and Amazon](https://de.wikipedia.org/wiki/Amazon) joined as investors with an initial investment of $1.25 billion. [[ 13 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-13)[[ 14 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-14) Amazon has made Anthropic's AI models available to [Amazon Web Services](https://de.wikipedia.org/wiki/Amazon_Web_Services) customers . [[ 15 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-15) In addition to Google, Amazon and Sparc Capital, [SAP](https://de.wikipedia.org/wiki/SAP) , Menlo Ventures, Wisdom Ventures, Ripple Impact Investments and Factorial Funds are also investors. [[ 16 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-FAZ-16) On November 22, 2024, Amazon announced that it would double its investment to $8 billion with a further $4 billion through a convertible note. A cooperation agreement was reached to use AWS computers for training purposes. At the beginning of 2025, Anthropic was valued at 60 billion US dollars. [[ 17 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-17)

**Story**

[ [Edit](https://de.wikipedia.org/w/index.php?title=Anthropic&veaction=edit&section=1) | [Edit source](https://de.wikipedia.org/w/index.php?title=Anthropic&action=edit&section=1) ]

Siblings Daniela Amodei and [Dario Amodei](https://de.wikipedia.org/wiki/Dario_Amodei) , the latter of whom previously served as Vice President of Research at OpenAI, founded Anthropic in 2021. [[ 18 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-18)[[ 19 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-19)[[ 20 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-20) Both had previously left OpenAI due to differences of direction, particularly regarding OpenAI's cooperation with [Microsoft](https://de.wikipedia.org/wiki/Microsoft) in 2019. [[ 21 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-21)

At the beginning of May 2023, Dario Amodei, CEO of Anthropic, along with the CEOs of Google, Microsoft, and OpenAI, were summoned to the [White House](https://de.wikipedia.org/wiki/Wei%C3%9Fes_Haus) , which stated that it had warned these currently leading *companies* about the risks posed by artificial intelligence. [[ 22 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-22)[[ 23 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-23)[[ 24 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-24) Anthropic employees point out the dangers of artificial intelligence (AI) in general and of AI chatbots in particular. [[ 25 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-KR-25)[[ 26 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-26)

*Because Claude* 's training allegedly used copyrighted content, rights holders filed lawsuits against Anthropic in October 2023. [[ 16 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-FAZ-16)

**Products**

[ [Edit](https://de.wikipedia.org/w/index.php?title=Anthropic&veaction=edit&section=2) | [Edit source](https://de.wikipedia.org/w/index.php?title=Anthropic&action=edit&section=2) ]

**Claude**

[ [Edit](https://de.wikipedia.org/w/index.php?title=Anthropic&veaction=edit&section=3) | [Edit source](https://de.wikipedia.org/w/index.php?title=Anthropic&action=edit&section=3) ]

Anthropic began developing an AI [chatbot](https://de.wikipedia.org/wiki/Chatbot) called *Claude* , which, like [ChatGPT](https://de.wikipedia.org/wiki/ChatGPT) , provides highly detailed and relevant answers to user questions. [[ 25 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-KR-25) Claude is a chatbot based on a language model using [reinforcement learning](https://de.wikipedia.org/wiki/Best%C3%A4rkendes_Lernen) . In contrast to Open AI's ChatGPT, instead of [*reinforcement learning from human feedback*](https://de.wikipedia.org/wiki/Reinforcement_learning_from_human_feedback) with human assessment of the chatbot's results, an alternative assessment is carried out by comparing the results with the constitution developed by Anthropic *and making* them compliant ( [AI alignment](https://de.wikipedia.org/wiki/KI-Ausrichtung) ). [[ 27 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-C-27) Regarding the draft European [law on artificial intelligence,](https://de.wikipedia.org/wiki/Gesetz_%C3%BCber_k%C3%BCnstliche_Intelligenz) the deviations from *Claude 1* are still high, which could create obstacles, at least for EU users. [[ 28 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-28)

*Claude 2* was initially made publicly available in the [US and UK in July 2023.](https://de.wikipedia.org/wiki/Vereinigte_Staaten)[[ 29 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-29) This chatbot is capable of reading and processing even very long texts. [[ 6 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-SH-6)*In May 2023, according to Anthropic, Claude* 's context window (short-term memory) was increased to 100,000 [tokens](https://de.wikipedia.org/wiki/Lexikalische_Analyse) (about 75,000 words). [[ 30 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-30)[[ 31 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-31) Claude is also used by partner companies, including [DuckDuckGo](https://de.wikipedia.org/wiki/DuckDuckGo) (with DuckAssist), the note-taking and productivity software Notion, and [Quora](https://de.wikipedia.org/wiki/Quora) . [[ 32 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-32) Thanks to the collaboration with *Amazon Web Services (AWS), Claude* applications are now available to all AWS customers on the *Amazon Bedrock* platform. [[ 33 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-33)

Claude 3 was introduced on March 4, 2024. [[ 34 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-34) The Claude 3 family includes three models in ascending order of performance: Haiku, Sonnet, and Opus. According to the company, it is said to have better results than ChatGPT 4. The standard version of Claude 3 Opus operates with a context window of 200,000 tokens, which is expanded to 1 million for some use cases. [[ 35 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-35)[[ 36 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-36)

Claude 3 requires phone number verification. Unlike API access, native web access is not available in many countries, including the European Union. [[ 37 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-37)

**Claude Pro**

[ [Edit](https://de.wikipedia.org/w/index.php?title=Anthropic&veaction=edit&section=4) | [Edit source](https://de.wikipedia.org/w/index.php?title=Anthropic&action=edit&section=4) ]

This more powerful version of the chatbot Claude is offered to subscribers. [[ 38 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-38)[[ 39 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-39)

**Projects**

[ [Edit](https://de.wikipedia.org/w/index.php?title=Anthropic&veaction=edit&section=5) | [Edit source](https://de.wikipedia.org/w/index.php?title=Anthropic&action=edit&section=5) ]

To further improve the AI ​​alignment *of its products, Anthropic aims to improve the constitution* and control of its chatbot Claude with the help of approximately 1,000 respondents. According to rules manager Jack Clark, 1,000 adult US citizens were surveyed using a questionnaire on Anthropic's proposed rules. They were also able to formulate their own rules. This potentially new *constitution* is called *Collective Constitutional AI* . These are experiments intended to democratize control. [[ 40 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-KR3-40)

[SK Telecom](https://de.wikipedia.org/wiki/SK_Telecom) , the Korean telecom group and investor in Anthropic, plans to further develop *Claude* for telecom applications and other languages, especially Korean. [[ 41 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-41)

Anthropic also publishes research on the *interpretability of machine learning systems* , with a focus on the [Transformer](https://de.wikipedia.org/wiki/Transformer_(Maschinelles_Lernen)) architecture. [[ 42 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-42) This research area is called *mechanistic interpretability* . It involves searching for data subsets in [deep learning](https://de.wikipedia.org/wiki/Deep_Learning) models that best represent certain abstract [*features*](https://de.wikipedia.org/wiki/Feature_(Maschinelles_Lernen))*(properties, characteristics). Such features* can then be activated for tasks ( *prompts* ) with similar properties. The corresponding model can thus also be specifically controlled. This technique was already applied to simple models in 2023. In Anthropic's more complex LLM model *Claude-3 Sonnet,* millions of such features were already detected in 2024. [[ 43 ]](https://de.wikipedia.org/wiki/Anthropic#cite_note-43)