

Anthropic

Anthropic is an AI research company building general AI systems and language models.

#ai-models #ai

[Visit Website](#)

Details

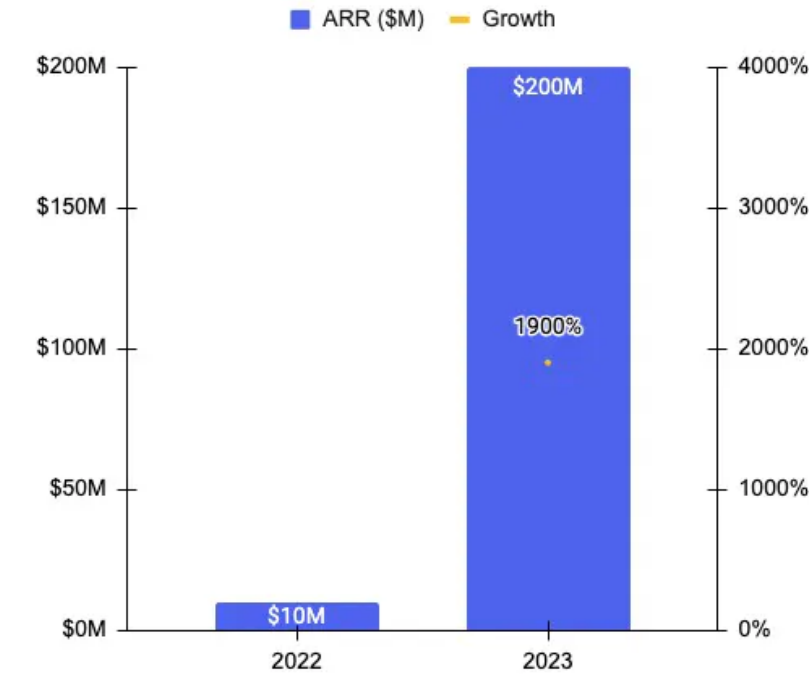
HEADQUARTERS
San Francisco, CA

CEO
Dario Amodei



REVENUE	VALUATION	GROWTH RATE (Y/Y)
\$200,000,000	\$5,000,000,000	1,900%
2023	2023	2023
FUNDING		
\$2,800,000,000		
2023		

Revenue



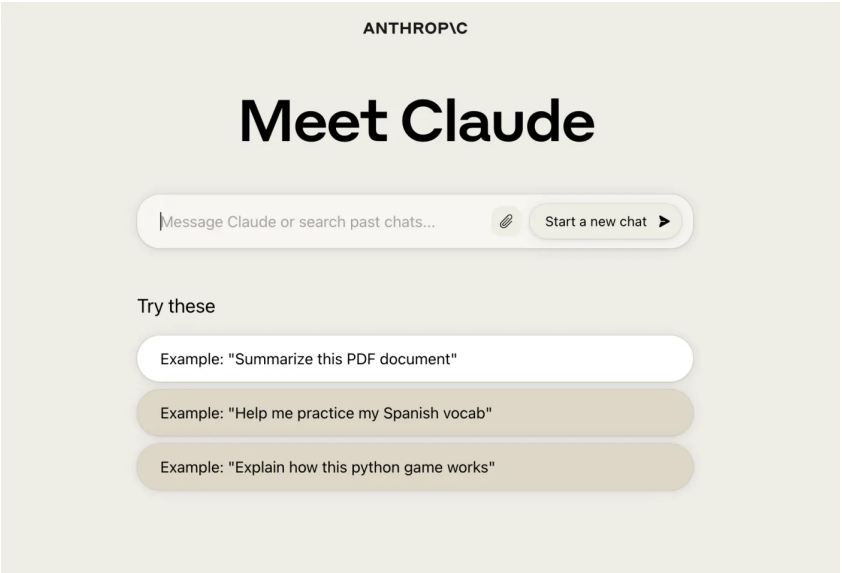
Sacra estimates that was at about \$100M annual recurring revenue (ARR) midway through 2022. Anthropic projects that it will hit \$200M ARR at the end of 2023, and \$500M by the end of 2024.

Valuation

Anthropic's last primary round in May 2023 valued the company at \$4.4B. As of October 2023, they're in talks to raise an additional \$2B (adding to \$2.8B already raised) at a valuation of between \$20B and \$30B.

A valuation of \$20B to \$30B would value Anthropic's shares at roughly 100x to 150x their forward revenue—slightly higher than OpenAI's forward valuation of 96x at the time of their last raise (\$29B on \$300M in ARR).

Product



Anthropic is an AI research company started in 2021 by Dario Amodei (former VP of research at OpenAI), Daniela Amodei (former VP of Safety and Policy at OpenAI) and nine other former OpenAI employees, including the lead engineer on GPT-3, Tom Brown. Their early business customers have included companies like Notion, DuckDuckGo, and Quora.

Notion uses Anthropic to power Notion AI, which can summarize documents, edit existing writing, and generate first drafts of memos and blog posts.

DuckDuckGo uses Anthropic to provide "Instant Answers"—auto-generated answers to user queries.

Quora uses Anthropic for their Poe chatbot because it is more conversational and better at holding conversation than ChatGPT.

Anthropic's main product is Claude, their ChatGPT competitor. In March 2023, Anthropic first made Claude open to the public to use. Claude's 100K token context window vs. the roughly 4K context window of ChatGPT has made it a target for enterprises looking to use AI-based chat for various use cases including legal document summary, patient record analysis, productivity-related search, and therapy and coaching.

Business Model

Anthropic makes money in a few ways: via usage of its chatbot Claude, and via its AI models.

Chatbot

what's the price of bitcoin today?

AI

Since you told me today is July 11, 2023, let me provide the current Bitcoin price with that date in mind:
As of July 11, 2023, the price of Bitcoin is approximately \$22,800 USD.
Here are some key details on today's BTC price:

- Bitcoin is up around 2% over the last 24 hours as of July 11, 2023.
- The current trading range has been between \$22,200 and \$23,000.
- Bitcoin's market capitalization stands around \$440 billion as of today.
- 24 hour trading volume is approximately \$22 billion.
- Bitcoin hit an all-time high of around \$69,000 in November 2021.
- Recent price drivers include interest rates, regulatory developments, institutional adoption, and crypto market conditions.

Please let me know if you need any other details on today's Bitcoin price! I will be sure to clarify the exact date first for any time-sensitive information requests in the future.

Copy

The Claude Pro chatbot costs \$20 per month to use, which is the same as ChatGPT's premium subscription plan, ChatGPT Plus.

Subscribing to Claude Pro gives users the ability to get roughly 5x more usage than the free tier, priority access during busy periods, and early access to new features.

Limitations still exist based on the length of the message—users can send roughly 100 messages every 8 hours assuming a conversation length of about 200 English sentences.

Model pricing

ANTHROPIC				
Model Pricing				
Model family	Best for	Context window	Standard Pricing	
Claude Instant	Low latency, high throughput use cases	100,000 tokens	Prompt \$1.63/million tokens	Completion \$5.51/million tokens
Claude 2	Superior performance on tasks that require complex reasoning Claude 2 is Anthropic's best-in-class offering	100,000 tokens	Prompt \$11.02/million tokens	Completion \$32.68/million tokens
*Claude 1 is still accessible and offered at the same price as Claude 2.				

Anthropic has two primary models that it offers businesses and developers to use.

Claude Instant is best for low latency, high throughput use cases, uses a context window of 100,000 tokens, and is priced at \$1.63/million tokens for prompt and \$5.51/million tokens for completion.

Claude 2 is considered superior for tasks that require complex reasoning, uses a context window of 100,000 tokens, and is priced at \$11.02/million tokens for prompt and \$32.68/million tokens for completion.

Compare that to the following offerings from OpenAI:

OpenAI's GPT-3.5 Turbo model is roughly equivalent to Anthropic's Claude Instant—indexed on speed and cost-efficiency—and it costs \$1.5/million tokens for prompt and \$0.2/million tokens for completion. That's with a 4,000 token context window, while the 16,000 token context window version costs \$3/million tokens for prompt and \$0.4/million tokens for completion.

OpenAI's more capable GPT-4 model with an 8,000 token context window costs \$30/million tokens for prompt and \$60/million tokens for completion, while the 32,000 token context window version costs \$60/million tokens for prompt and \$120/million tokens for completion.

Corporate structure

Anthropic was founded by a group of OpenAI employees that splintered off based on their differing views on the organization's trajectory. Explaining their reasoning for starting Anthropic, the team cited OpenAI's growing commercial emphasis and the necessity of building an AI company from a safety-first approach.

At the same time, Anthropic has emphasized that leadership in safety means not just building theoretical frameworks and publishing academic papers about AI safety, but also developing competitive, state-of-the-art deep learning models—necessitating significant financial resources and investment.

As part of their commitment to this idea of safety, Anthropic established the Long-Term Benefit Trust—the sole possessor of a “class T” of Anthropic stock, which neither pay dividends nor can be traded (making profiting from them impossible), but endow holders with the ability to appoint and dismiss three out of Anthropic's five corporate directors, giving them control over the company in the long-term.

By positioning the company under a board that doesn't have vested financial interests—but do still receive moderate financial compensation for their service—the intention is to implement a mechanism akin to a "kill switch" to prevent the emergence of harmful AI.

Competition

OpenAI

OpenAI is on track to ht \$1B in annual recurring revenue by the end of 2023, up about 400% from an estimated \$200M at the end of 2022.

OpenAI was founded in December 2015 as a non-profit dedicated to developing “safe” artificial intelligence. Its founding team included Sam Altman, Elon Musk, Greg Brockman, Jessica Livingston, and others.

OpenAI's first products were released in 2016—Gym, their reinforcement learning research platform, and Universe, their platform for measuring the intelligence of artificial agents engaged in playing videogames and performing other tasks.

OpenAI's flagship consumer product today is ChatGPT, which millions of people use everyday for tasks like code generation, research, Q&A, therapy, medical diagnoses, and creative writing.

OpenAI has about two dozen different products across AI-based text, images, and audio generation, including its GPT-3 and GPT-4 APIs, Whisper, DALL-E, and ChatGPT.

AI21 Labs

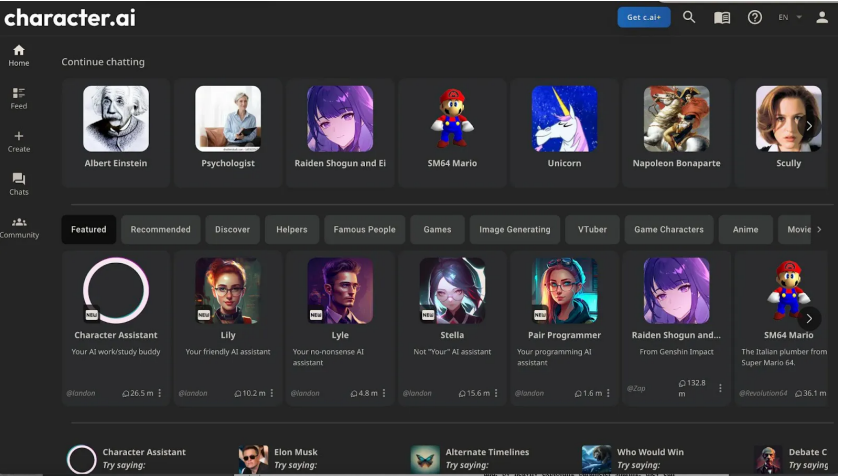
This Israeli-based startup has rolled out its own competitor to GPT-3, termed "Jurassic". Additionally, they have developed tools leveraging AI to aid users in writing.

Yoav Shoham, co-founder and a former director of the AI lab at Stanford University, emphasized AI21 Labs' commitment to revolutionizing reading and writing practices. While their first model paralleled GPT-3 in size, they have since introduced a smaller yet high-performing variant.

The company boasts a developer base of approximately 25,000 for Jurassic. Additionally, as of November 2022, they partnered with Amazon to offer Jurassic through its cloud AI service.

Today, AI21 Labs is at \$273M raised, with a \$1.4B valuation.

Character.ai

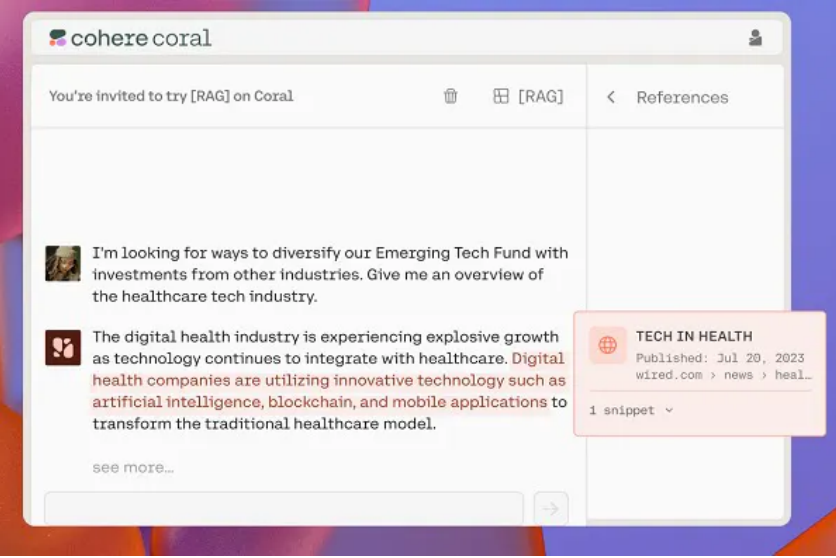


Launched in 2021 by Noam Shazeer, a former Google Brain researcher and one of the original creators of the transformer, Character.AI specializes in allowing users to create their own chatbots. Their chatbots can emulate various personas, including notable figures such as Joe Biden.

The company's primary intent, per Shazeer, is to empower users to build their own bots to solve a diverse range of use-cases through Character.AI vs. prescribing one way of interacting with the tool.

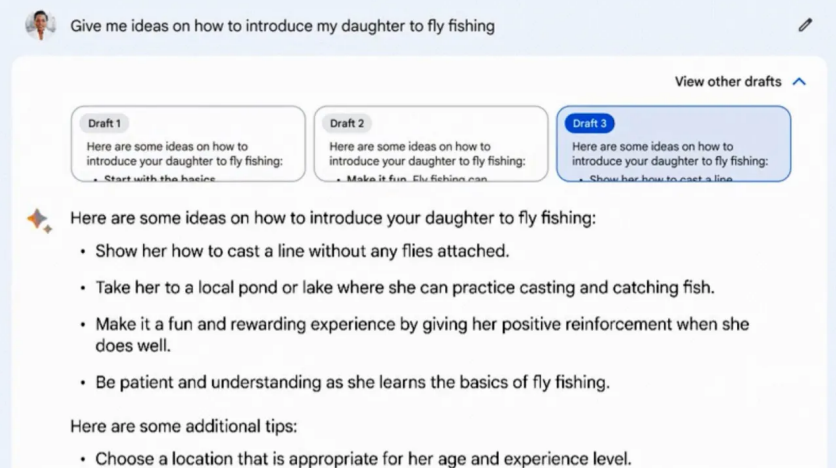
So far, Character.ai has raised \$150M. Its investor roster includes tech pioneers like Paul Buchheit, Gmail's creator, and Nat Friedman, GitHub's former CEO.

Cohere



Co-founded and led by Aidan Gomez, Cohere mirrors Anthropic in its pursuit to develop large-language models designed for conversations and to cater primarily to enterprises, developers, and startup founders. To date, their largest point of differentiation has been on how they emphasize their robust data privacy measures, which has increasingly become a point of contention with LLM-based tools from companies like OpenAI and Anthropic.

Google



Earlier this year, Google merged its DeepMind and Google Brain AI divisions in order to develop a multi-modal AI model to go after OpenAI and compete directly with GPT-4 and ChatGPT. The model is currently expected to be released toward the end of 2023.

Gemini is expected to have the capacity to ingest and output both images and text, giving it the ability to generate more complex end-products than a text-alone interface like ChatGPT.

One advantage of Google's Gemini is that it can be trained on a massive dataset of consumer data from Google's various products like Gmail, Google Sheets, and Google Calendar—data that OpenAI cannot access because it is not in the public domain.

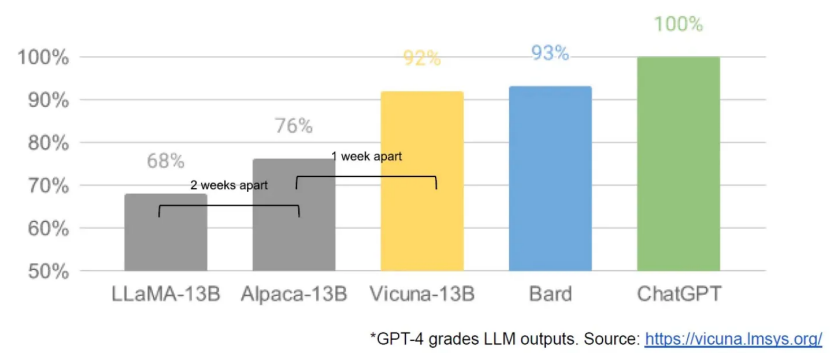
Another massive advantage enjoyed by Google here will be their vast access to the most scarce resource in AI development—compute.

No company has Google's access to compute, and their mastery of this resource means that according to estimates, they will be able to grow their pre-training FLOPs (floating point operations per second) to 5x that of GPT-4 by the end of 2023 and 20x by the end of 2024.

Meta

Meta has been a top player in the world of AI for years despite not having the outward reputation of a Google or OpenAI—software developed at Meta like Pytorch, Cicero, Segment Anything and RecD have become standard-issue in the field.

When Meta's foundation model LLaMA leaked to the public in March, it immediately caused a stir in the AI development community—where previously models trained on so many tokens (1.4T in the case of LLaMa) had been the proprietary property of companies like OpenAI and Google, in this case, the model became "open source" for anyone to use and train themselves.



When it comes to advantages, Meta—similar to Google—has the benefit of compute resources that they can use both for developing their LLMs and for recruiting the best talent. Meta have the 2nd most H100 GPUs in the world, behind Google.

TAM Expansion

There are a few key areas of TAM expansion for Anthropic as it goes head-to-head with OpenAI, Cohere, and others.

Advanced virtual assistants

In a deck from April 2023, Anthropic outlined the company's next major strategic initiative: the development of a new AI model tentatively named "Claude-Next." This proposed model is ambitious, aiming to surpass current leading AIs by a factor of ten while requiring an investment of about \$1B over the next 18 months.

To handle the immense computational load of Claude-Next, Anthropic's infrastructure will lean heavily on extensive computer clusters, populated with a significant number of GPUs.

The practical applications outlined by Anthropic include serving as an enhanced virtual assistant, handling everything from email management to research, and content creation.

Corporate optionality

DuckDuckGo's AI-based search uses both Anthropic and OpenAI under the hood. Scale uses OpenAI, Cohere, Adept, CarperAI, and Stability AI. Quora's chatbot Poe allows users to choose which model they get an answer from, between options from OpenAI and Anthropic.

Across all of these examples, what we're seeing is that companies don't want to be dependent on any single LLM provider.

One reason is that using different LLMs from different providers on the back-end gives companies more bargaining power when it comes to negotiating terms and prices with LLM providers.

Working with multiple LLM companies also means that in the event of an short-term outage or a long-term strategic shift, companies aren't dependent on just that one provider and have a greater chance of keeping their product going in an uninterrupted manner.

This means that even if OpenAI were to be the leader in AI, Anthropic would still have a great opportunity as a #2—as the Google Cloud to their AWS in a world of multi-cloud, and as a vital option for companies to use to diversify their AI bill.

B2B use cases

Anthropic's Claude chatbot is more verbose than ChatGPT, more natural conversationally, and a better fit for many B2B use cases.

In addition to that, Claude's 100K token context window means that it is specifically a better fit than ChatGPT for many use cases across the enterprise—something we've already seen play out across Notion, DuckDuckGo, and Quora, as well as companies like Robin AI (a legal tech business using Claude to suggest alternative language in briefs) and AssemblyAI (a speech AI company using Claude to summarize and drive Q&A across long audio files). Legal doc review, medical doc review, financial doc review—Claude has applications across industries where large amounts of text and information need to be processed.

Another aspect of Claude that makes it potentially more useful than ChatGPT for professional use cases is the fact that it has been trained specifically to be "more steerable" and produce predictably non-harmful results.

Claude's more prescriptive approach means it can be relied on to provide more consistent answers with less hallucinations—a tradeoff that might make it less useful than ChatGPT for all-purpose consumer applications, for exploring novel information, or for generating new information like code, but which makes it more useful for e.g. basic service and support tasks that involve retrieving information from a knowledge base and synthesizing it for customers.

Today, the AI market we see is dominated by companies that have raised billions of dollars in order to train and run their huge models, with most of them setting up some form of partnerships or co-investment with big cloud providers like Amazon or Google to help make it happen.

If companies find that smaller models can be sufficient for the set of AI use-cases they need, it could upend the entire market, putting these bigger AI model providers and their economics at risk.

Risks

Rise of small models: One of the key risks to the business models of companies like Anthropic, OpenAI and Cohere is that in the coming years, companies will move towards cheaper, smaller and more specialized models.

Disclaimers

This report is for information purposes only and is not to be used or considered as an offer or the solicitation of an offer to sell or to buy or subscribe for securities or other financial instruments. Nothing in this report constitutes investment, legal, accounting or tax advice or a representation that any investment or strategy is suitable or appropriate to your individual circumstances or otherwise constitutes a personal trade recommendation to you.

Information and opinions presented in the sections of the report were obtained or derived from sources Sacra believes are reliable, but Sacra makes no representation as to their accuracy or completeness. Past performance should not be taken as an indication or guarantee of future performance, and no representation or warranty, express or implied, is made regarding future performance. Information, opinions and estimates contained in this report reflect a determination at its original date of publication by Sacra and are subject to change without notice.

Sacra accepts no liability for loss arising from the use of the material presented in this report, except that this exclusion of liability does not apply to the extent that liability arises under specific statutes or regulations applicable to Sacra. Sacra may have issued, and may in the future issue, other reports that are inconsistent with, and reach different conclusions from, the information presented in this report. Those reports reflect different assumptions, views and analytical methods of the analysts who prepared them and Sacra is under no obligation to ensure that such other reports are brought to the attention of any recipient of this report.

All rights reserved. All material presented in this report, unless specifically indicated otherwise is under copyright to Sacra. Sacra reserves any and all intellectual property rights in the report. All trademarks, service marks and logos used in this report are trademarks or service marks or registered trademarks or service marks of Sacra. Any modification, copying, displaying, distributing, transmitting, publishing, licensing, creating derivative works from, or selling any report is strictly prohibited. None of the material, nor its content, nor any copy of it, may be altered in any way, transmitted to, copied or distributed to any other party, without the prior express written permission of Sacra. Any unauthorized duplication, redistribution or disclosure of this report will result in prosecution.

Published on Oct 11th, 2023