SACRA

EQUITY RESEARCH

UPDATED
07/15/2025

Thinking Machines

TEAM

Jan-Erik Asplund
Co-Founder
jan@sacra.com

Marcelo Ballve
Head of Research
marcelo@sacra.com

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.

This research report has been prepared solely by Sacra and should not be considered a product of any person or entity that makes such report available, if any.

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.

Thinking Machines

Visit Website

AI research and product company aiming to make AI systems more widely understood, customizable, and generally capable

#ai

VALUATION
$10,000,000,000
2025

FUNDING
$2,000,000,000
2025

Details

HEADQUARTERS
San Francisco, CA

CEO
Mira Murati

Valuation

Thinking Machines officially closed its $2 billion seed round in January 2025, valuing the company at $12 billion.

The round was led by Andreessen Horowitz with participation from Nvidia, Accel, ServiceNow, Cisco, AMD, and Jane Street. Albania's sovereign wealth fund committed $10 million as part of a broader European AI initiative.

This represents one of the largest seed rounds in Silicon Valley history, driven by investor appetite for frontier AI talent.

Product

Thinking Machines is building a customizable AI platform centered around a frontier-scale multimodal foundation model that can process text, images, audio, code, and scientific data.

Unlike ChatGPT, the platform that Thinking Machines is building is intended to allow organizations to "fork" the base model and customize it for specific domains, use cases, and safety requirements.

The core product will consist of three main components: the TM-1 foundation model trained on over 2 trillion tokens using mixture-of-experts architecture, a Dynamic Guardrail Engine that lets customers adjust safety policies in real-time without restarting the model, and an SDK that provides tools for fine-tuning, deployment, and integration.

Users can upload their own datasets, set collaboration modes like pair-programming or research assistance, and deploy customized instances either through managed APIs or on-premises Docker containers.

The platform targets researchers, enterprises, and developers who need AI systems tailored to their specific workflows rather than generic chatbot interfaces. A pharmaceutical company could train the model on their proprietary research data and safety protocols, while a software team could customize it for their codebase and development practices.

Business Model

Thinking Machines operates as a B2B platform company with a hybrid open-source and commercial model. The company plans to release open-source components including safety tools, evaluation frameworks, and model weights to build developer adoption, while monetizing through premium hosted services, enterprise support, and advanced customization features.

The core monetization strategy likely revolves around usage-based pricing for their managed API endpoints, with customers paying per token processed through their customized model instances.

Enterprise customers will also likely be able to purchase annual contracts that include dedicated compute resources, priority support, and advanced customization services. For organizations requiring on-premises deployment, Thinking Machines offers licensing deals for their model weights and deployment infrastructure.

The business model leverages the company's massive expected $2 billion war chest to secure GPU capacity and talent, creating a competitive moat through scale and technical capabilities.

Unlike pure-play model companies that compete solely on performance, or pure-play tooling companies that depend on third-party models, Thinking Machines controls the full stack from model training to deployment, allowing them to optimize the entire customer experience and capture more value per user.

Competition

Vertically integrated giants

OpenAI, Google DeepMind, and Anthropic dominate the frontier AI landscape through massive scale and vertical integration. OpenAI's GPT-4 family benefits from Microsoft's Azure distribution and aggressive pricing at $5 per million tokens, while Google embeds Gemini directly into Workspace and Android.

Anthropic's Claude models focus on safety and compliance, particularly for financial services, backed by over $7 billion from Amazon and Google. These players can subsidize model development through their broader ecosystems, making it difficult for standalone companies to compete on price or distribution.

Open-source challengers

Mistral AI represents the European approach to AI sovereignty, raising over €1 billion while releasing both open-weight and commercial models. Their focus on smaller, efficient models that rival GPT-4 performance creates pricing pressure for larger players.

Meta's Llama family and Chinese companies like DeepSeek push open-source alternatives that enterprises can deploy without vendor lock-in. These competitors threaten Thinking Machines' positioning by offering similar customization capabilities without the premium pricing of frontier models.

Cloud and tooling platforms

AWS Bedrock, Microsoft Azure AI, and Google Vertex AI provide model-agnostic platforms that let enterprises access multiple foundation models through unified APIs. Hugging Face has built the dominant open-source ecosystem for model deployment and fine-tuning, while NVIDIA's AI foundry services target enterprises wanting custom models.

These platforms compete directly with Thinking Machines' customization layer, offering similar fine-tuning and deployment capabilities across multiple model providers rather than being locked into a single foundation model.

TAM Expansion

New products

Thinking Machines plans to expand beyond their core foundation model into specialized AI systems for scientific research, drug discovery, and advanced engineering applications. The company is developing domain-specific models trained on scientific literature, laboratory data, and research methodologies that could command premium pricing from pharmaceutical companies, materials science labs, and academic institutions.

Their roadmap includes AI systems for protein folding, molecular design, and climate modeling that would tap into the growing market for AI-powered scientific discovery.

Customer base expansion

The platform's customization capabilities enable expansion from AI research labs to mainstream enterprise R&D departments across industries. Fortune 500 companies in automotive, aerospace, and manufacturing represent a massive untapped market for AI systems that can be trained on proprietary engineering data and safety protocols.

The company's open-source components also create a pathway to millions of individual developers and smaller companies that currently use general-purpose models but need more specialized capabilities for their specific use cases.

Geographic expansion

Thinking Machines' public benefit corporation structure and European investment provide entry points into international markets with strong AI sovereignty requirements. The EU's AI Act and similar regulations in other countries create demand for transparent, auditable AI systems that can be customized for local compliance requirements.

The company's distributed team and open-source approach enable rapid localization for non-English markets without requiring full regional offices, potentially capturing market share in regions where US-based AI companies face regulatory or political barriers.

Risks

Compute access: Thinking Machines' success depends entirely on securing massive GPU capacity in a market dominated by cloud providers who also compete in AI. NVIDIA's H100 and H200 chips remain scarce, and hyperscalers like Microsoft, Google, and Amazon prioritize their own AI initiatives over third-party customers. Any disruption to Thinking Machines' compute access could cripple their ability to train competitive models or serve customer workloads at scale.

Talent retention: The company's $10+ billion valuation creates enormous pressure to deliver breakthrough results with an unproven team and business model. Key researchers poached from OpenAI and other established labs may find better opportunities elsewhere if Thinking Machines fails to ship products quickly or struggles with the transition from research to commercial operations. The competitive AI talent market means any execution stumbles could trigger departures that undermine the company's core technical capabilities.

Open source commoditization: Thinking Machines' strategy of releasing open-source components to drive adoption could backfire if competitors use their own tools and research to build competing platforms. Meta's approach with Llama shows how open-source AI models can rapidly commoditize entire market segments, potentially eliminating the pricing power that justifies Thinking Machines' massive valuation before they establish a sustainable competitive moat.

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.

This research report has been prepared solely by Sacra and should not be considered a product of any person or entity that makes such report available, if any.

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.

Published on Jul 15th, 2025