MARKET REPORT

UPDATED
05/12/2024

Together AI: the $44M/year Vercel of generative AI

TEAM

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

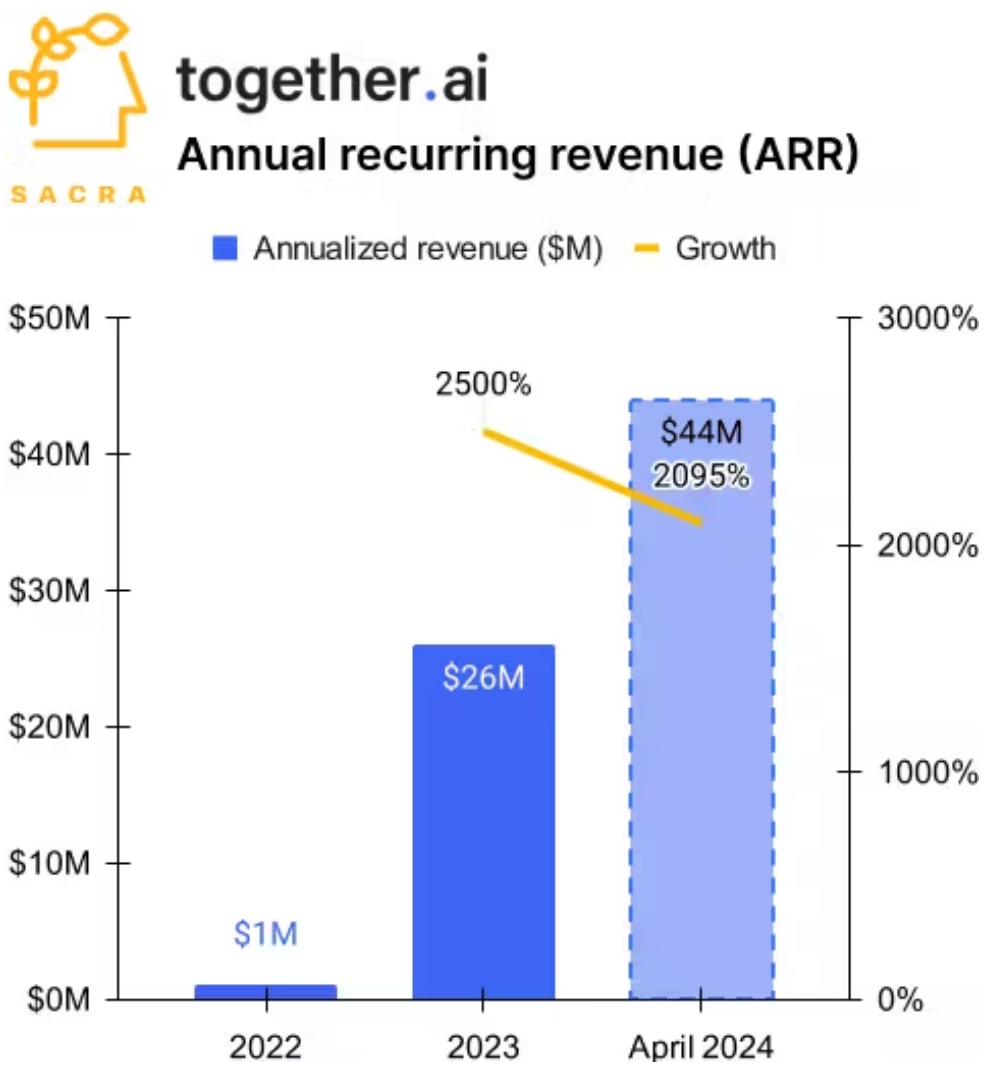
DISCLAIMERS

Published on **May 12th, 2024**

Together AI: the $44M/year Vercel of generative AI

By **Jan-Erik Asplund**

TL;DR: Sacra estimates that Together AI hit $44M in annual recurring revenue (ARR) in April with their developer experience-centric layer built on top of CoreWeave and Lambda Labs's GPU cloud products. For more, check out our full report and dataset on Together AI.



Key points from our research:

- **Companies use cloud GPU hosts Together AI, CoreWeave, and Lambda Labs to use Nvidia (NASDAQ: NVDA) graphics processing units (GPUs) to train AI models on their datasets, fine-tune them, and deploy them into production.** Together AI differentiated itself as a GPU cloud platform early by indexing on open source, allowing its customers access to 100+ open models from Mistral to Llama-

2 to rapidly experiment with training different LLMs on their data.

- **Together AI found product-market fit charging per-token, based on the number of API calls, as a developer experience-centric layer on top of CoreWeave and Lambda Labs's per-hour pricing, allowing them to win early-stage startups and individual developers who can mitigate the risk of paying for idle GPU time.** While CoreWeave and Lambda Labs focus on locking in multi-year reservations to recoup the fixed capex costs of their data centers and GPUs, Together AI operates a layer above, aligning their pricing with the spiky API volumes of startups training new models and launching new products.
- **Sacra estimates that Together AI hit $44M in annualized revenue in April 2024, up 2095% year-over-year, with 90% of revenue coming from sales of bundled on-demand GPU compute and training via Together Inference and ~45% gross margin**. Compare to ~80% gross margin GPU cloud providers like CoreWeave with $440M of revenue in 2023, up 1,660% from $25M in 2022, and Lambda Labs at $250M of revenue in 2023, up 1,150% from $20M in 2022, and fellow reseller Crusoe Energy at $100M of revenue in 2023, up 400% from ~$20M in 2022.

For more, check out this other research from our platform:

- OpenAI vs. Anthropic vs. Cohere
- Perplexity: the $11M/year Cliff Notes for the web growing 4,272%
- Perplexity (dataset)
- Anthropic (dataset)
- OpenAI (dataset)
- Scale (dataset)
- Hugging Face (dataset)
- CoreWeave (dataset)
- Lambda Labs (dataset)
- Scale (dataset)
- Apple vs. Limitless vs. Gong
- Jenni AI: the $5M/year Chegg of generative AI
- David Park, CEO and co-founder of Jenni AI, on prosumer generative AI apps post-ChatGPT

SACRA

- AI writing goes enterprise
- Photoroom: the $65M/year background removal app
- Hugging Face: the $70M/year anti-OpenAI growing 367% year-over-year