SACRA

MARKET REPORT

**UPDATED**
08/07/2025

# Fal.ai at $95M/year growing 4,650% YoY


**TEAM**

Jan-Erik Asplund
Co-Founder
jan@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.

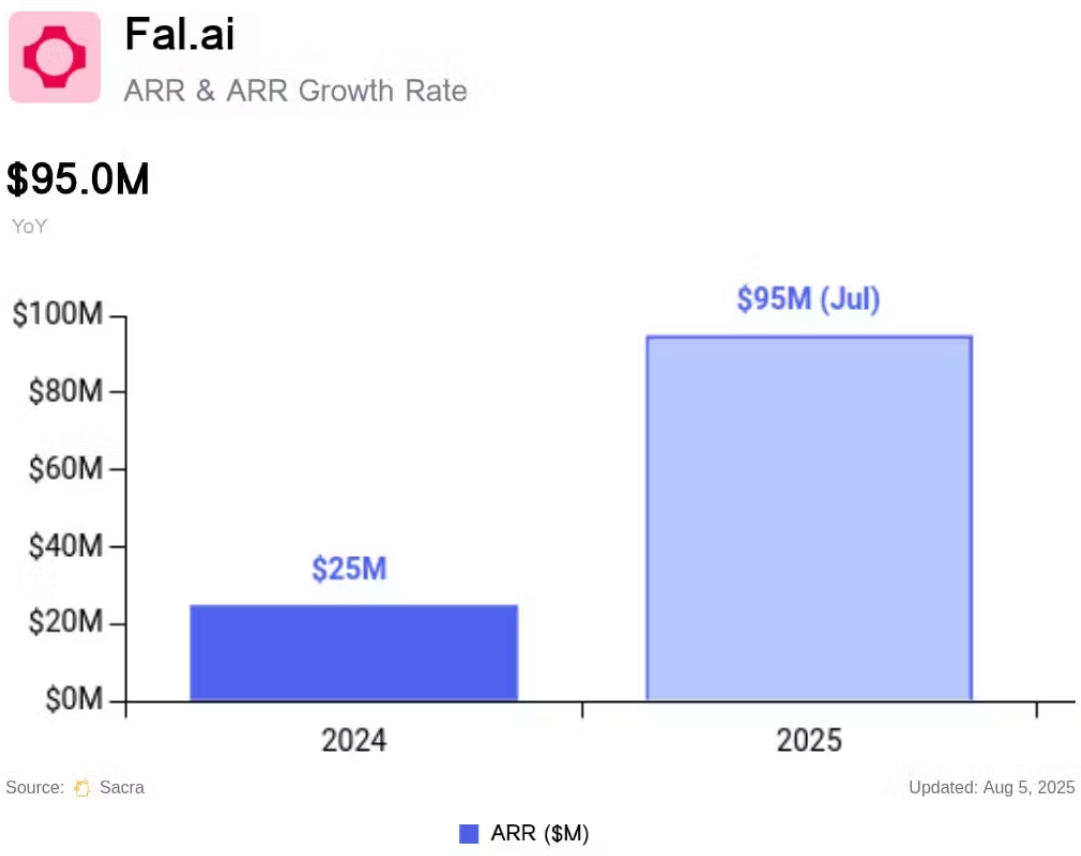
All rights reserved. All material presented in this report, unless specifically indicated otherwise is under copyright to Sacra. Sacra reserves any and all intellectual property rights in the report. All trademarks, service marks and logos used in this report are trademarks or service marks or registered 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.

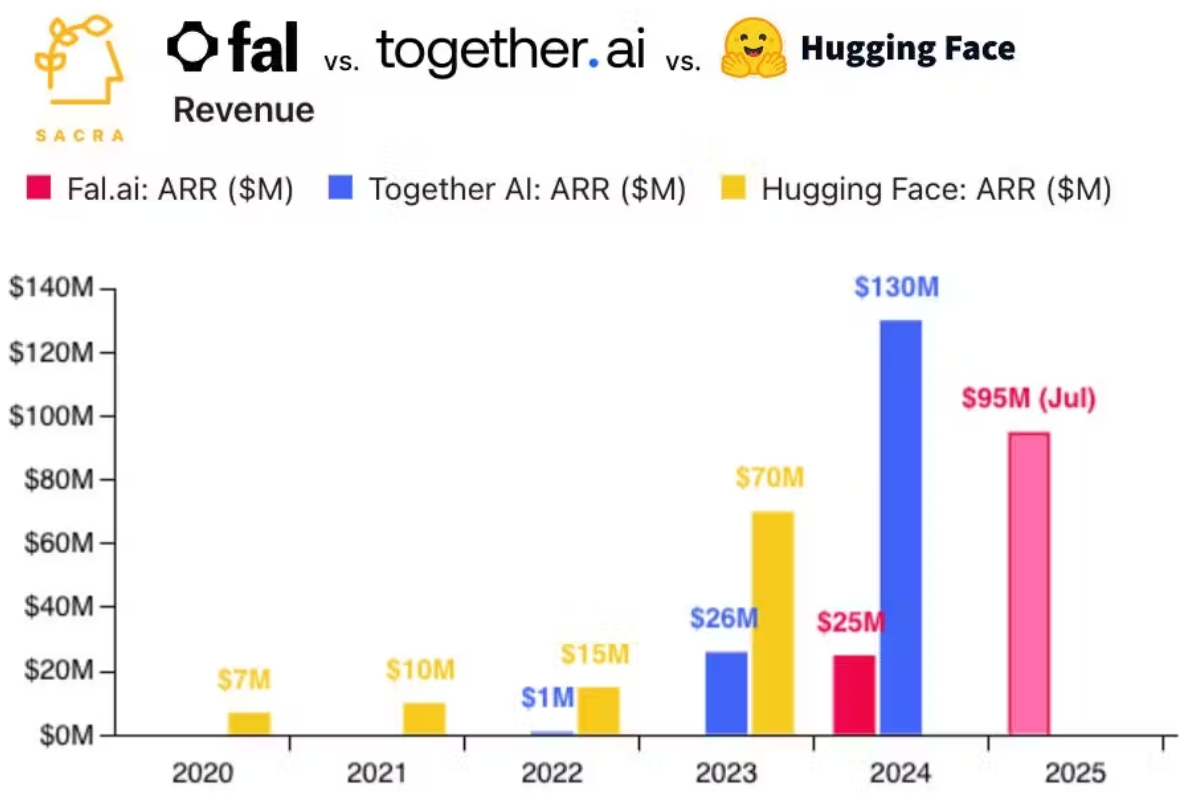
Published on **Aug 07th, 2025**

# Fal.ai at $95M/year growing 4,650% YoY

By **Jan-Erik Asplund**

**TL;DR: As generative media took off in 2022-2023, Fal.ai pivoted to an inference-as-a-service API for image, video, and audio models that optimized for speed and ease of use over raw compute. Sacra estimates Fal.ai hit a $95M revenue run rate in July 2025, up from $25M at the end of 2024. For more, check out our full report and dataset on Fal.ai.**






Key points via Sacra AI:

- **As open-source text-to-image models like Stable Diffusion took off in 2022, building generative AI apps on top of them required either self-hosting models on AWS/GCP/Azure or accepting slow, rate-limited public endpoints—inspiring Fal.ai (2021) to launch an inference-as-a-service provider optimized for speed and reliability, serving first Stable Diffusion and now 600+ image, video, audio, and 3D models over an API monetized on usage.** At ~$0.03 per image or ~$0.10 per second of video API invocation, developers tend to (1) discover & demo open-source models from Stability AI (Stable Diffusion) or Black Forest Labs (Flux) on Hugging Face and play around with them locally, then (2) set up Fal.ai and Replicate to prototype and start using them in production, and then (3) as usage starts to scale, deploy them on rented bare-metal GPUs via CoreWeave, Lambda Labs, or Crusoe.
- **After signing enterprise deals to power generative image & video experiences in Quora's Poe, Canva and Perplexity, Sacra estimates that Fal.ai hit a revenue run rate of $95M in July 2025, up from $25M at the end of 2024 and up 4,650% YoY from about $2M in July 2024, valued at $500M as of its $49M July Series B for a 5.3x forward revenue multiple.** Compare to AI neocloud Together AI at $130M ARR at the end of 2024, valued at $1.25B for a 9.6x multiple, "GitHub for models" Hugging Face at $70M revenue in 2023

growing 367% YoY, valued at $4.5B for a 64.3x multiple, OpenRouter at $5M in annualized revenue in May 2025, up from $1M at the end of 2024, valued at $500M for a 100x multiple, and Stability AI at ~$50M revenue in 2024, up 525% YoY, valued at $1B in 2022 for a 666x multiple on $1.5M in revenue.

- **Because image- and video-generation models are fine-tuned ~1,000x as often as LLMs and usually run as chained workflows (remove background, upscale, recolor) to achieve brand- or character-level consistency that one-shot prompting can't deliver, Fal.ai is evolving past raw inference & model hosting into workflows, bundling together model chaining, LoRA fine-tuning, and asset storage into one billable integration.** Like OpenRouter, Fal.ai offers developers a single integration point to call a wide range of models—not integrating directly with OpenAI's API is akin to not integrating directly with Stability AI's API—creating natural expansion paths through added endpoints while benefiting from a flywheel where the more developers & enterprises on Fal.ai, the more valuable it becomes as a distribution channel for model providers.

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

- Fal.ai (dataset)
- Together AI (dataset)
- OpenRouter (dataset)
- Stability AI (dataset)
- Hugging Face (dataset)
- Runway vs. OpenAI
- OpenArt at $12M ARR growing 1,100% YoY
- Coco Mao, CEO of OpenArt, on building the TikTok for AI video
- GPU clouds growing 1,000% YoY
- Cristóbal Valenzuela, CEO of Runway, on the state of generative AI in video
- Cristóbal Valenzuela, CEO of Runway, on rethinking the primitives of video
- AI and the future of video
- Together AI: the $44M/year Vercel of generative AI
- OpenRouter at $100M GMV

- CoreWeave at $2B revenue
- Samiur Rahman, CEO of Heyday, on building a production-grade AI stack
- Fluidstack at $180M ARR