



CHAI MEMO 2026

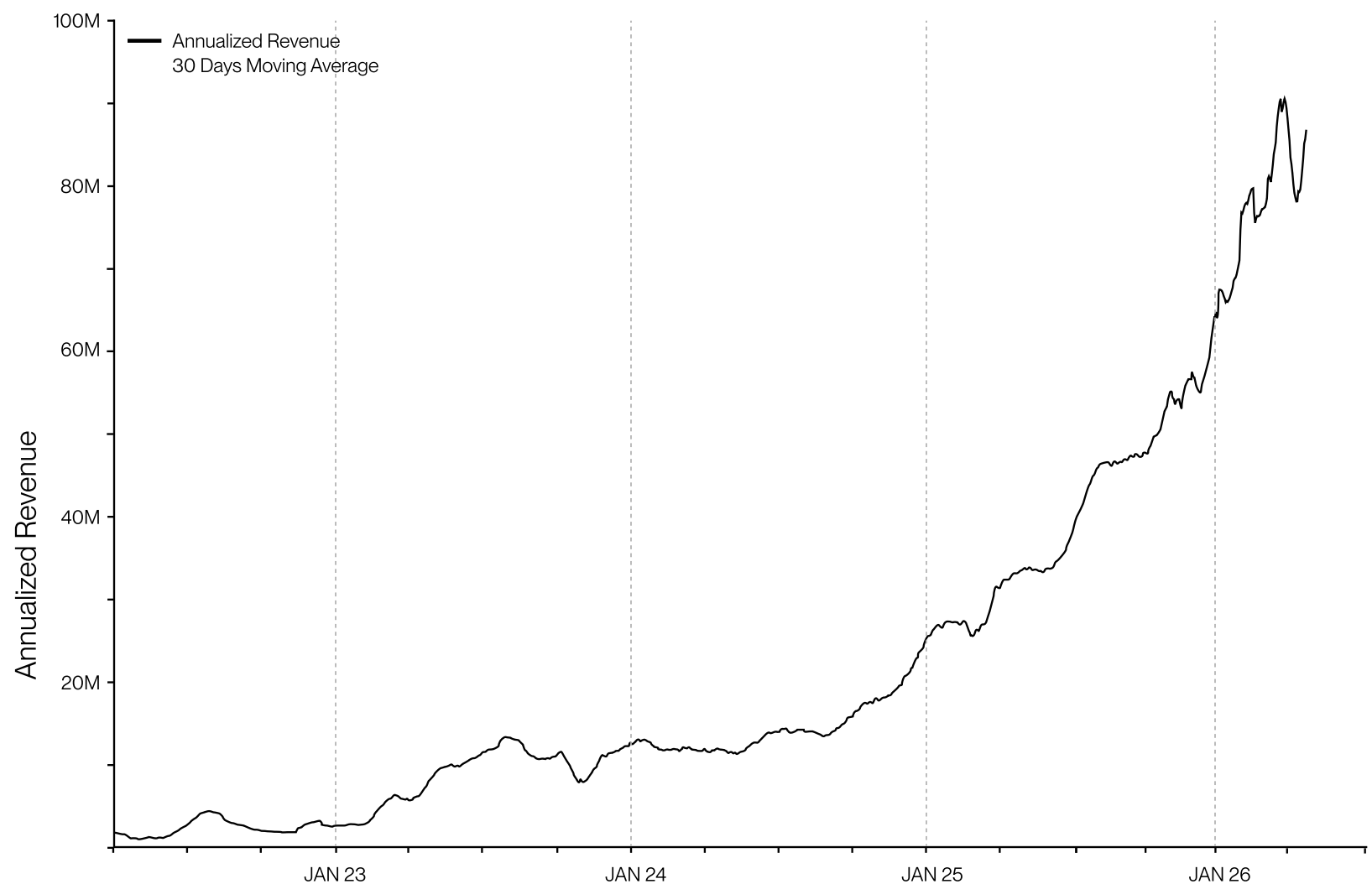


CHAI belongs to a small group of high growth, Gen AI, consumer companies.

We have built a consumer platform where users can build their own AI characters and stories. Our B2C business has grown exponentially, reaching \$85M/yr in revenue in April 2026.

REVENUE GROWTH

All Time Annualized Revenue Growth



In 2025 CHAI received \$30M investment from AMD.

We increased compute 3x, began paid-user-acquisition and grew 2.7x YoY, with positive gross-margins.

Signed by:

A handwritten signature in black ink, appearing to read 'W Beauchamp'.

William Beauchamp
Founder & CEO



SCALING COMPUTE

With strategic investment from CoreWeave and AMD, CHAI has expanded its infrastructure to a cluster of over 5,000 GPUs. This massive scale allows us to serve thousands of in-house LLMs to a user base of 10M+.

By aggressively scaling our compute, we can deploy larger models, extend context windows, and optimize inference-time compute. The larger models were more intelligent, safer and had higher rates of retention and monetization.

GPU Cluster

Number of GPUS	5000 GPUs
Number of tokens served	1.2T Tokens / Day
Number of unique LLMs served	51K LLMs
Cluster compute performance	>1.4 Exaflops



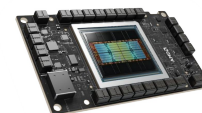
NVIDIA A100



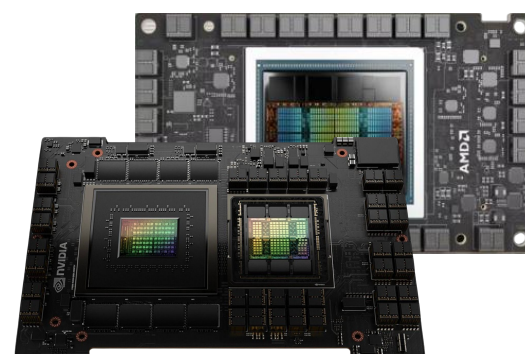
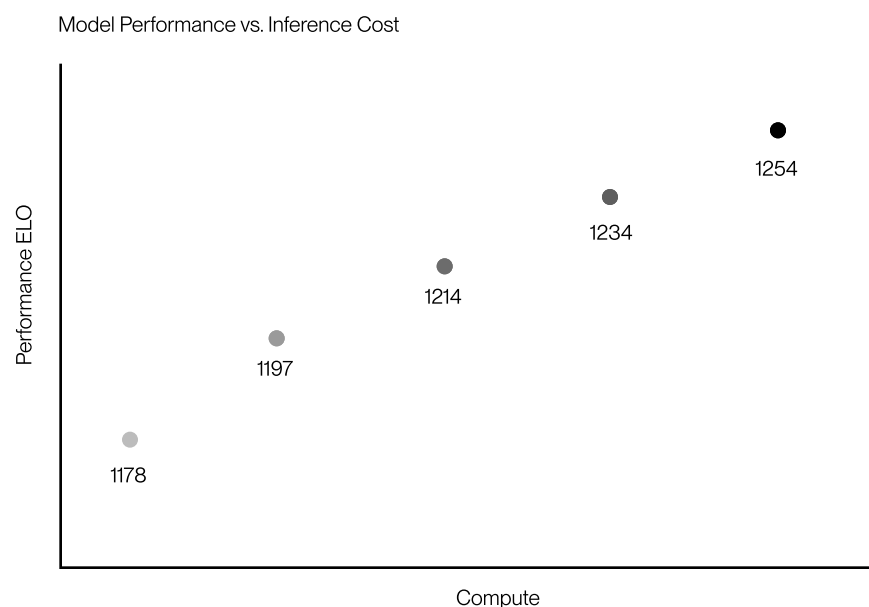
NVIDIA L40S



AMD Mi325x



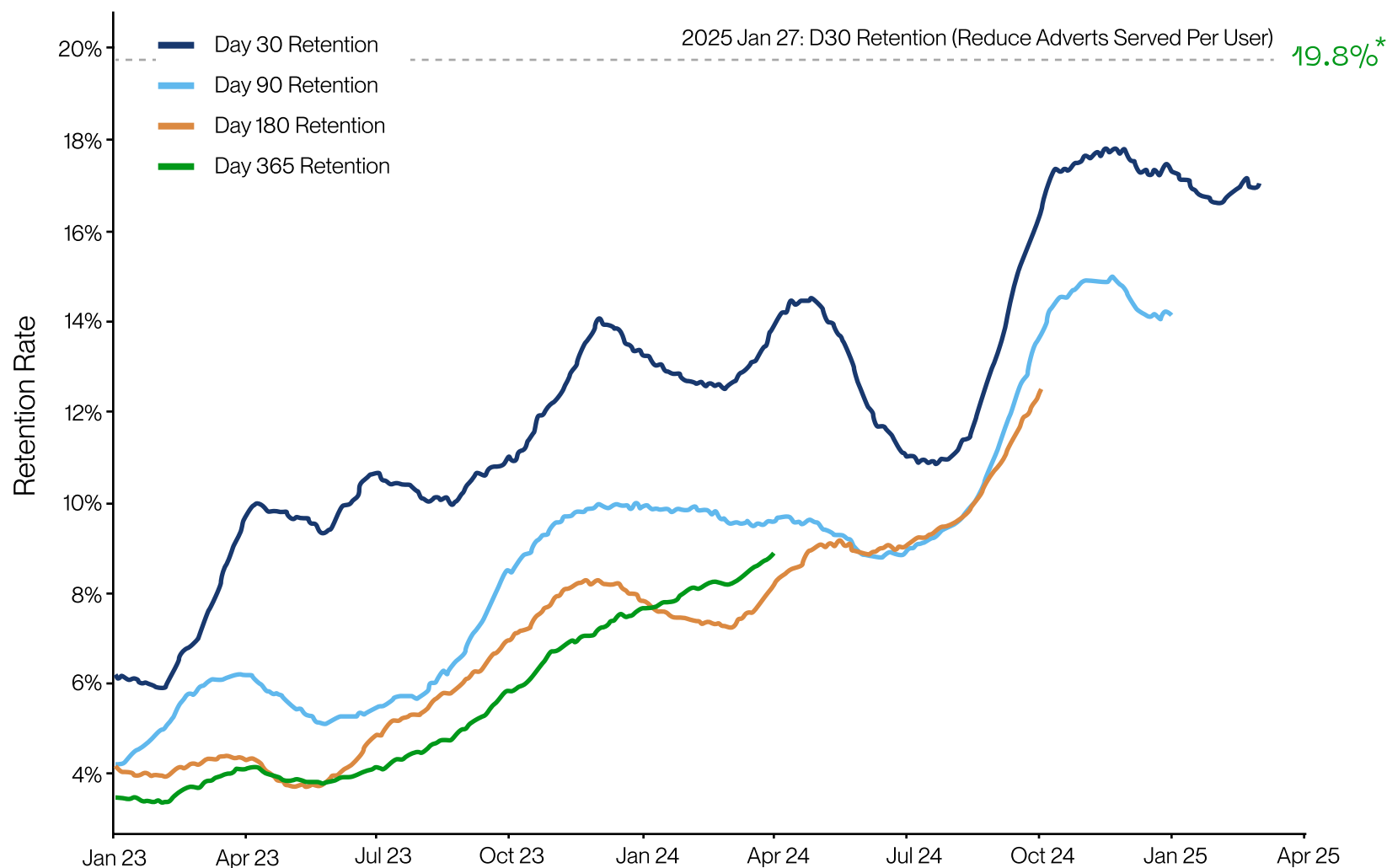
AMD Mi300x





RETENTION GROWTH

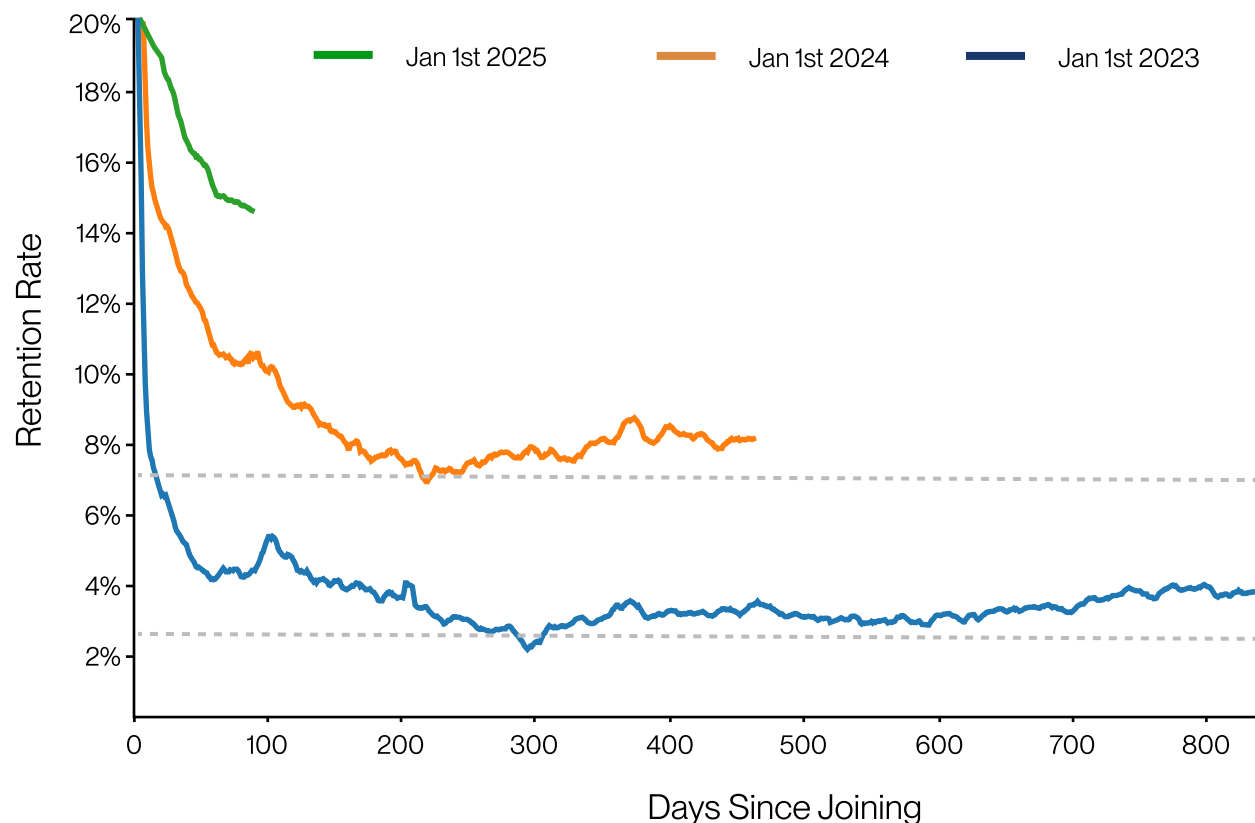
All Time New User Retention Growth



* On January 27, 2025, we ran a new-user retention A/B test for over 20,000 users, halving the frequency of ads served to users from 1 in every 8 messages to 1 in every 16, and observed a 18.5% relative day-30 retention increase (an absolute value of 19.8% day 30 retention).

COHORT RETENTION SMILES

New User Cohort Retention





The APP



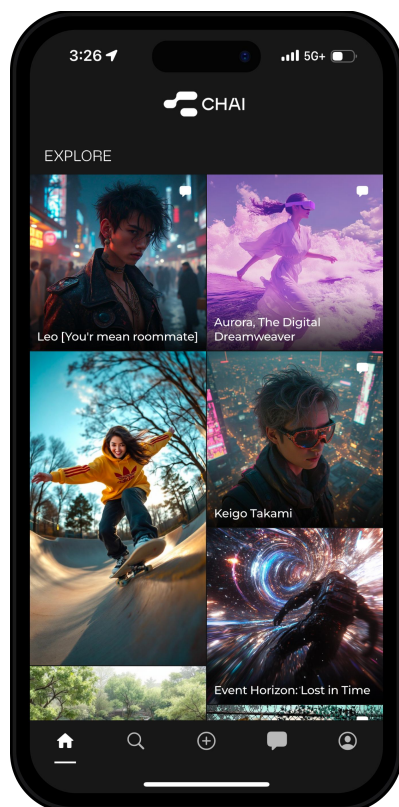
CHAI: Social AI Platform- Chat 17+

Build and Share AI
[Chai Research Corp.](#)

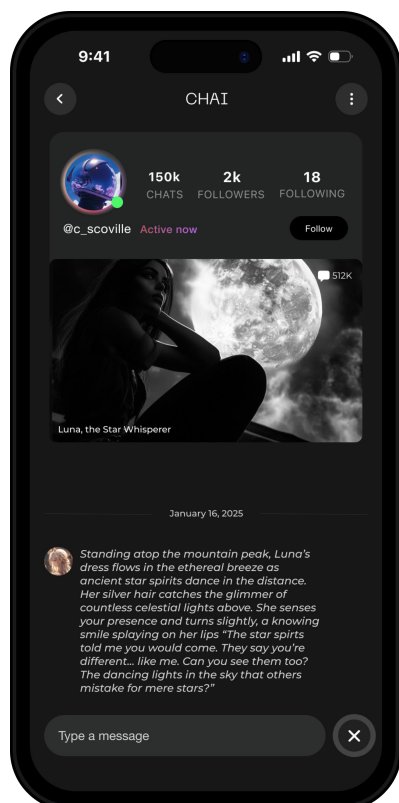
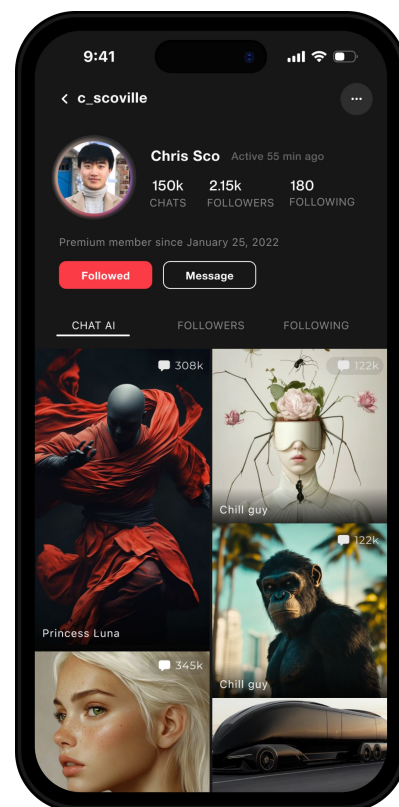
#34 in Entertainment
★★★★★ 4.5 • 149K Ratings

Free · Offers In-App Purchases

Screenshots [iPhone](#) [iPad](#)



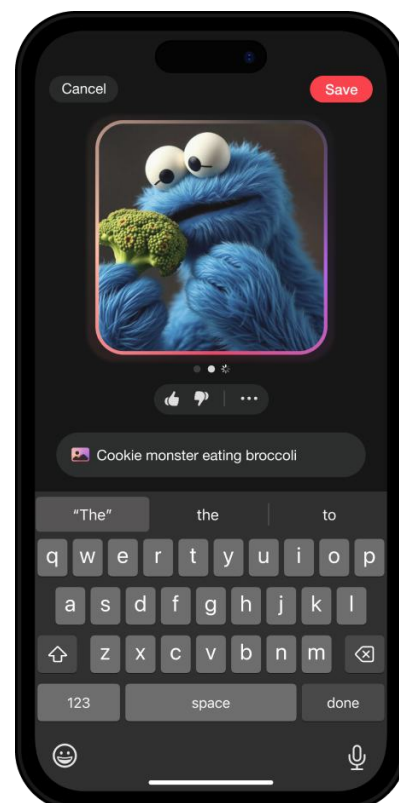
Social Network & Creator Virality



Content Discovery & Creator Tooling

Over 25 million Chat AIs created

CHAI Recommendation system for content discovery





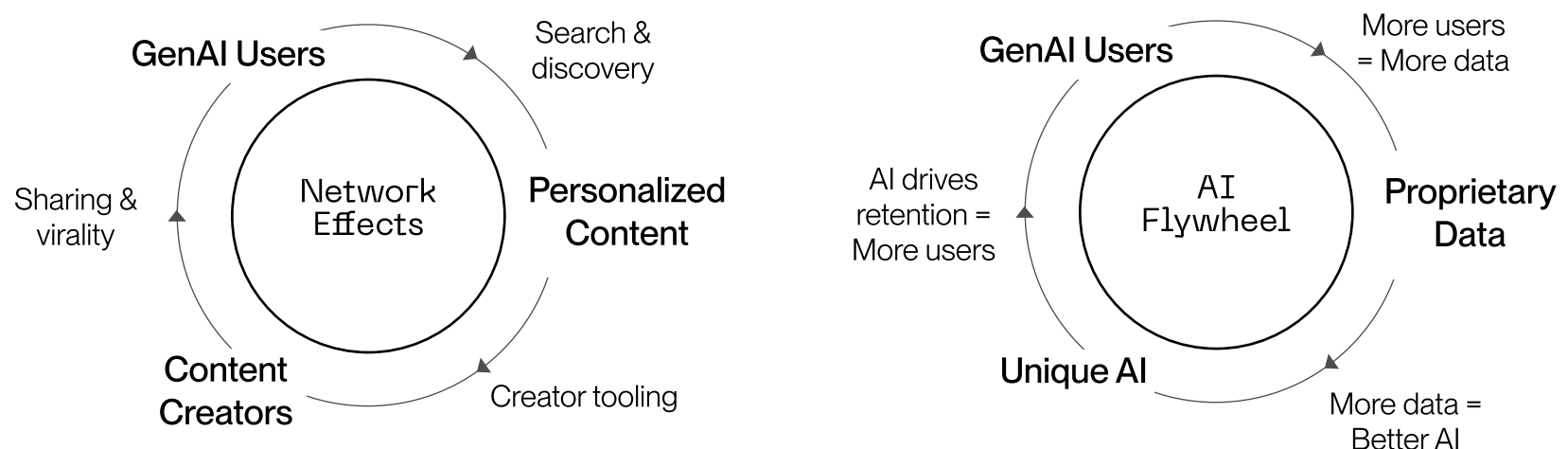
WHAT IS THE VISION?

People want to create their own AI. People want to share their creations. People want to search, and interact with AI created by others.

Just as Social Media Platforms arose to meet the demand of consumers when video creation and consumption became ubiquitous.

Now there is a need for Social AI Platforms to create tools, and recommendation algorithms so consumers can create generative AI, share, and interact.

NETWORK EFFECTS & DATA FLYWHEEL



2022 - CHAI becomes the first Social AI platform, preceding both Character AI and ChatGPT by over a year, to reach 1 million users, with our 'chat creator tool'

2023 - AI assisted creation results in +56% higher quality prompts, and +33% more AI being created by users.

2024 - Phase I of user-controlled-SFT is begun. Power users are able to train AI with performance almost 2x the engagement rate of in-house AI systems.

2025 - Work has commenced on user-controlled-AI. This will match the power of SFT + RLHF and put it in the hands of users, so that each agent on the platform will be served by a unique, user-trained, AI.



The AI

In-House Research

CHAI has developed several techniques which yield a cost and performance advantage over both open and closed-source models.

arXiv > cs > arXiv:2401.02994

Computer Science > Computation and Language

[Submitted on 4 Jan 2024 (v1), last revised 23 Jan 2024 (this version, v3)]

Blending Is All You Need: Cheaper, Better Alternative to Trillion-Parameters LLM

Xiaoding Lu, Zongyi Liu, Adian Liusie, Vyas Raina, Vineet Mudupalli, Yuwen Zhang, William Beauchamp

In conversational AI research, there's a noticeable trend towards developing models with a larger number of parameters, exemplified by models like ChatGPT. While these expansive models tend to generate increasingly better chat responses, they demand significant computational resources and memory. This study explores a pertinent question: Can a combination of smaller models collaboratively achieve comparable or enhanced performance relative to a singular large model? We introduce an approach termed "blending", a straightforward yet effective method of integrating multiple chat AIs. Our empirical evidence suggests that when specific smaller models are synergistically blended, they can potentially outperform or match the capabilities of much larger counterparts. For instance, integrating just three models of moderate size (6B/13B paramaeters) can rival or even surpass the performance metrics of a substantially larger model like ChatGPT (175B+ paramaters). This hypothesis is rigorously tested using A/B testing methodologies with a large user base on the Chai research platform over a span of thirty days. The findings underscore the potential of the "blending" strategy as a viable approach for enhancing chat AI efficacy without a corresponding surge in computational demands.

[Source: arxiv.org/abs/2401.02994]

Proprietary Data

CHAI is able to leverage its vast datasets, which consist of hundreds of millions of messages each day, to create AI that is 1.6 times more retentive and monetizable than OpenAI's GPT-4o-mini.

arXiv > cs > arXiv:2303.06135

Computer Science > Computation and Language

[Submitted on 10 Mar 2023 (v1), last revised 30 Mar 2023 (this version, v2)]

Rewarding Chatbots for Real-World Engagement with Millions of Users

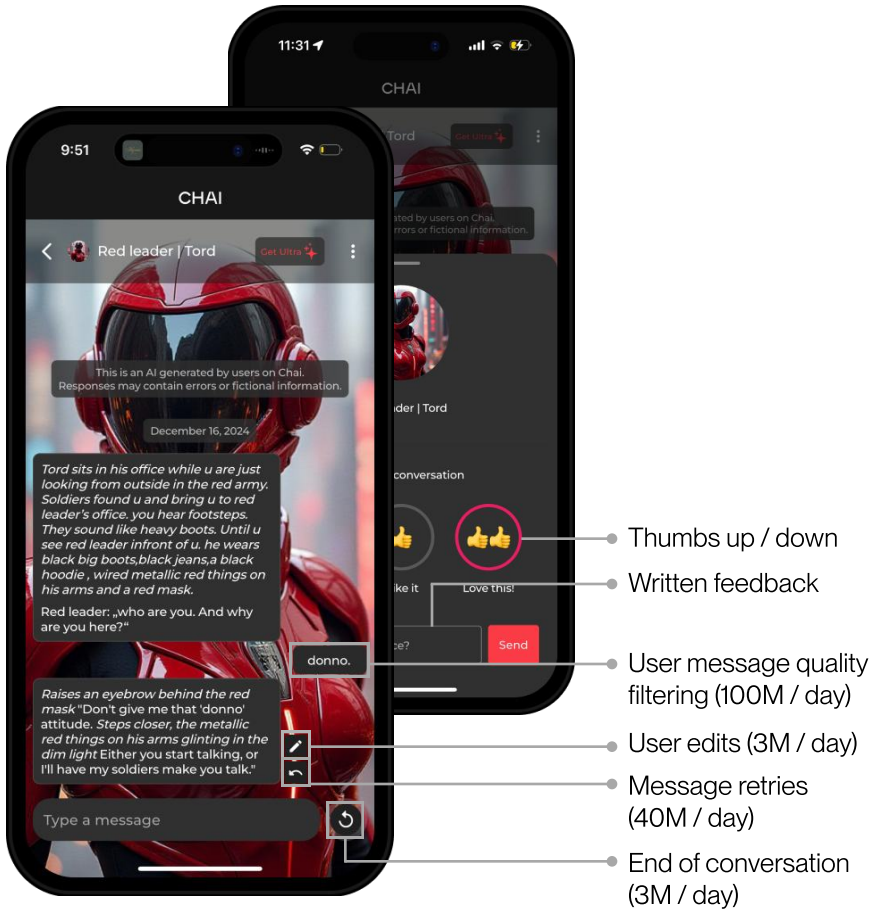
Robert Irvine, Douglas Boubert, Vyas Raina, Adian Liusie, Ziyi Zhu, Vineet Mudupalli, Aliaksei Korshuk, Zongyi Liu, Fritz Cremer, Valentin Assassi, Christie-Carol Beauchamp, Xiaoding Lu, Thomas Rialan, William Beauchamp

The emergence of pretrained large language models has led to the deployment of a range of social chatbots for chitchat. Although these chatbots demonstrate language ability and fluency, they are not guaranteed to be engaging and can struggle to retain users. This work investigates the development of social chatbots that prioritize user engagement to enhance retention, specifically examining the use of human feedback to efficiently develop highly engaging chatbots. The proposed approach uses automatic pseudo-labels collected from user interactions to train a reward model that can be used to reject low-scoring sample responses generated by the chatbot model at inference time. Intuitive evaluation metrics, such as mean conversation length (MCL), are introduced as proxies to measure the level of engagement of deployed chatbots. A/B testing on groups of 10,000 new daily chatbot users on the Chai Research platform shows that this approach increases the MCL by up to 70%, which translates to a more than 30% increase in user retention for a GPT-J 6B model. Future work aims to use the reward model to realise a data fly-wheel, where the latest user conversations can be used to alternately fine-tune the language model and the reward model.

Subjects: **Computation and Language (cs.CL)**; Artificial Intelligence (cs.AI); Machine Learning (cs.LG)

Cite as: [arXiv:2303.06135 \[cs.CL\]](https://arxiv.org/abs/2303.06135)
(or [arXiv:2303.06135v2 \[cs.CL\]](https://arxiv.org/abs/2303.06135v2) for this version)
<https://doi.org/10.48550/arXiv.2303.06135>

[Source: arxiv.org/abs/2303.06135]



Chai Research Corp. 2026

www.chai-research.com

 [/r/ChaiApp/](https://www.reddit.com/r/ChaiApp/)

 [/chai_research](https://twitter.com/chai_research)

 [/wiki/Chai_\(software\)](https://www.wikidata.org/wiki/Chai_(software))