Europe, February 2025

Geospatial Foundation Models (GFMs)

Introduction deck

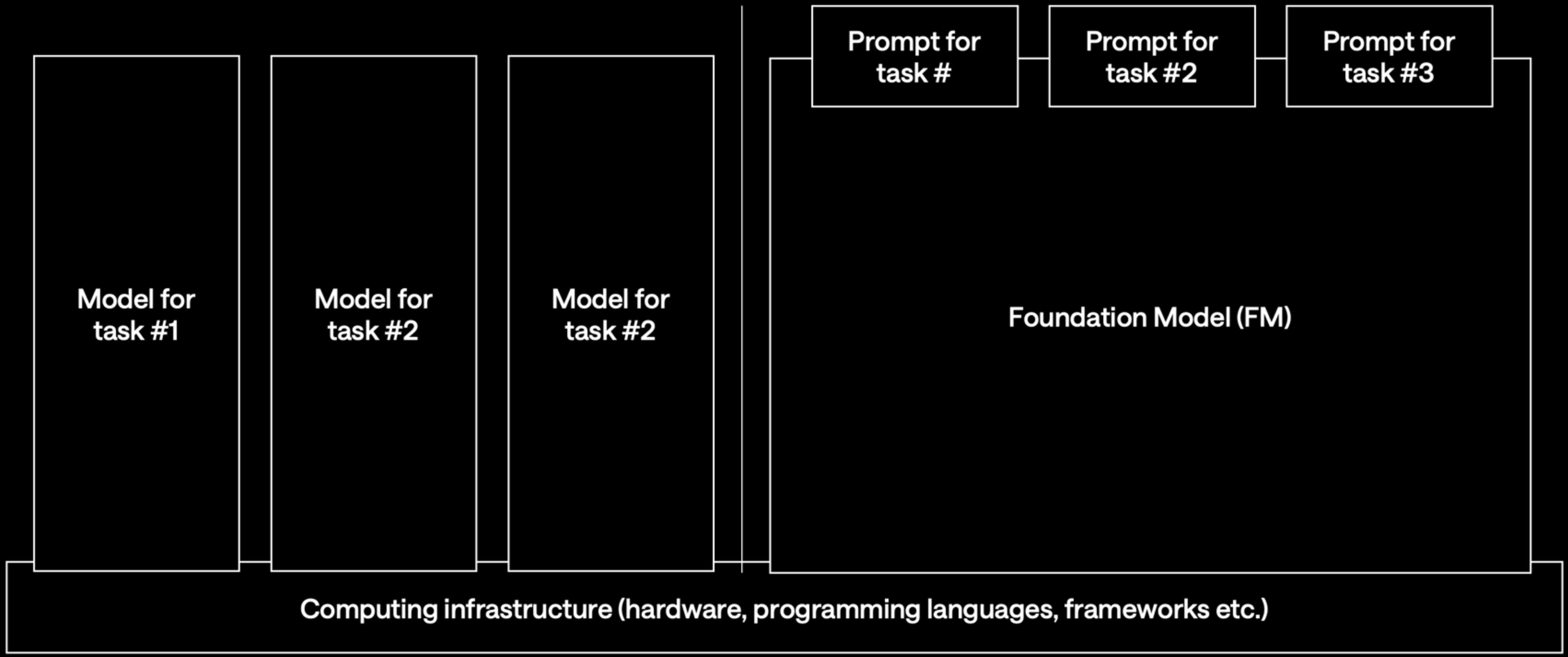
LUNAR

Executive summary

- Foundation models (FMs) or base models, are task-agnostic, pre-trained, large-scale neural networks that can be adapted to numerous downstream tasks.
- Within FMs, large language models (LLMs) trained on terabytes of internet-based textual datasets like CommonCrawl have become popular in recent years due to product examples like ChatGPT. As LLMs have attracted hundreds of billions of EUR, and FMs like GPT-3/GPT-4 have led to a myriad of derivative models tailored to specific downstream tasks (e.g., specific assistants). These recent investments in LLMs have also shed light on the unit economics in an AI-world: CAPEX required to train SOTA models increases~2,5x/annum[1], compute required to train SOTA models is growing at ~4,6x per year[1], rapid commoditization of models occurs due to open-source competition, different geographic regions tend to build their FMs[2], and rapid declines in token costs for existing models occur at -86%/annum[3].
- New categories of FMs will occur in the coming years, typically with alternative use cases. The main examples include geospatial foundation models (GFMs), time series foundation models (TFMs), and material foundation models (MFMs). An alternative way to segment foundation models is to focus on the data modality being used, the main categories using that logic are language foundation models, vision foundation models and multimodal foundation models – the latter category using multimodal input data and typically allowing multimodal generalizations (e.g. text to image, image to text).
- Geospatial FMs (GFMs) are a subset of foundation models encoding rich information about places and regions, and are typically trained on satellite data, climate data, weather data, topographic data, drone data, and demographic data. GFMs are relevant for predictions that relate to locations, such as environmental risk predictions (e.g.., flooding, wildfires, deforestation – 10-50B EUR/year), energy predictions (e.g., locations of renewables, power grid failures – 10-20B EUR/year), agricultural predictions (future land use, expected droughts, water management – 10-20B EUR/year), urban predictions (e.g., smart city optimization, real estate planning – 10-50B EUR/year) as well as many others (e.g., military information, insurance analytics – 10-50B EUR/year). We estimate that these use cases together will push the geospatial models into the mainstream for companies. We also recognized that the value captured depends largely on the speed of AI adoption.
- Lastly, we dive into the AI economics making these use cases viable and highlight three main trends relevant to understanding AI economics: i) the increase in training costs for new state-of-the-art (SOTA) models, ii) the increase in the total amount of available models, and iii) the decrease in token costs over time. First, investment for developing foundation models is significant, with the first ~1bln EUR models in sight – while at the same time cheaper variants keep popping up (e.g., DeepSeek, Llama). Second, in the last four years open-source model repositories have grown very rapidly in with a 70X increase in the number of models available. Language-based models are still the dominant category, but this can change when more high-quality and affordable satellite data becomes available at scale. Finally, given the intense competition from open-source models, API token costs tend to fall rapidly, as was shown in recent years for GTP-3 and GPT-4 where token costs fell ~90% per year. These trends support the premise that geospatial models can have a wide set of economically viable use cases in the coming years, as the willingness to invest is present and usage costs are declining rapidly.

[1] Source: Epoch.ai; [2] ChatGPT in the US, Mistral in Europe, DeepSeek in China – with largely similar capabilities, [3] source: deeplearning.ai.

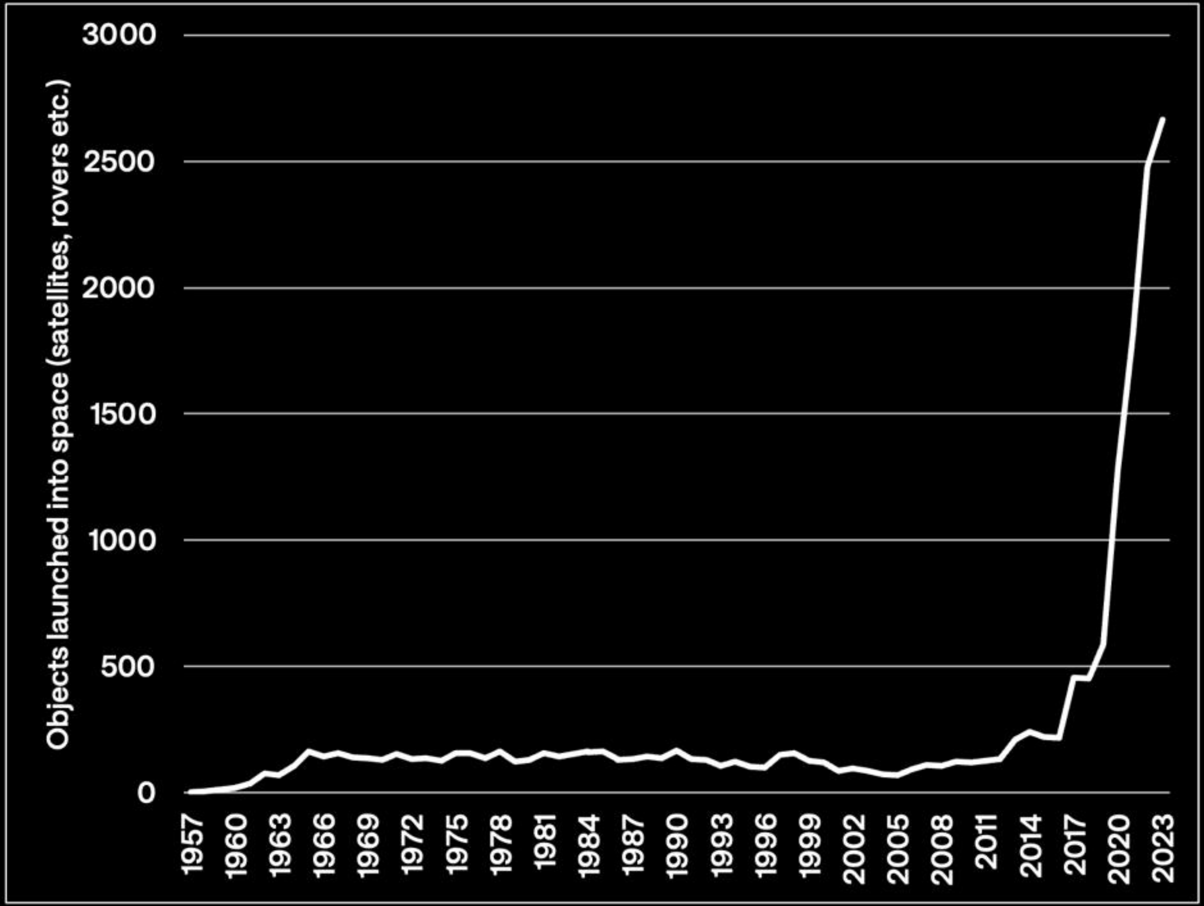
Foundation models are multi-purpose models...



Source: University of Amsterdam (2024): Foundation models are platform models.

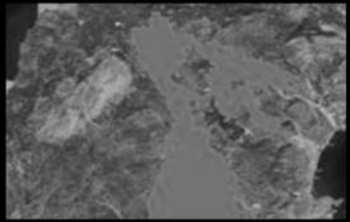
...where GFMs focus on geospatial predictions

Earth observation (EO) data is growing rapidly in volume and quality[1]...

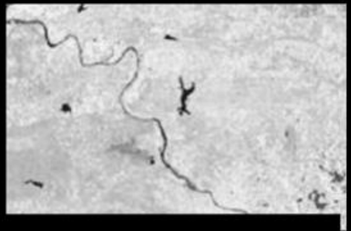


...allowing a wide range of highly relevant use cases[2].

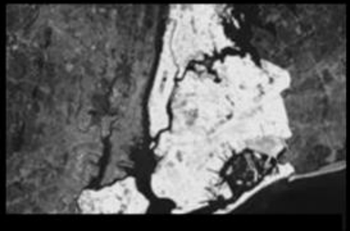
Flood detection



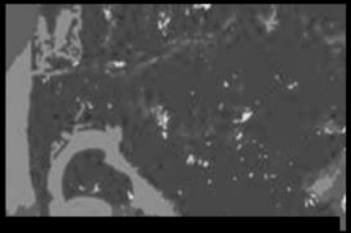
Wildfire analysis



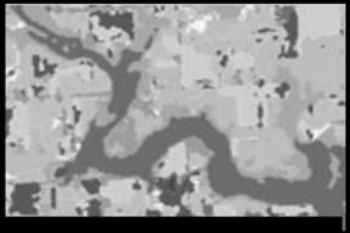
Biomass estimates



Urban heat detection



Land use modeling



Crop classification

Source: ourworldindata.org; EUPex; IBM Research; Maxar; [1] Granularity of satellite pictures have moved from 60mx60m to 15cm x 15cm today; [2] Examples, but no means exhaustive.

GFMs have the potential to be a multi-billion EUR/year industry

Data modality: Language data[1] | Vision data[2] | Multimodal data[3]

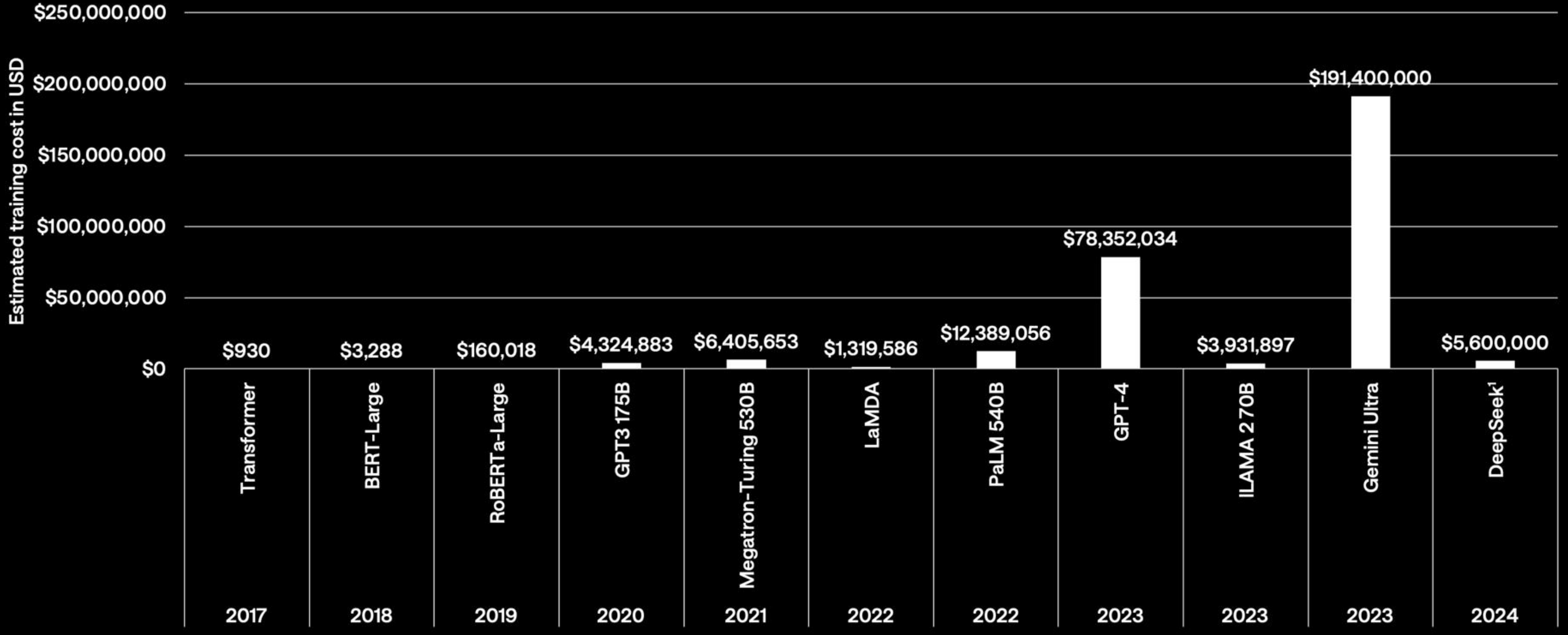
Model type: Geospatial (GFMs) | Time Series (TFMs) | Urban FMs (UFMs) | Material (MFMs)

Sector	Environment	Energy	Agriculture	Urban	Other
Value estimate	**10-50B EUR/year.**	**10-20B EUR/year.**	**5-25B EUR/year.**	**10-50B EUR/year.**	**50-100B EUR/year.**
	Wildfire predictions, flood predictions, deforestation, desertification, hurricane predictions, assets at risk calculations, carbon credits verification, environmental compliance.	Renewable energy siting optimisation, infrastructure monitoring, extreme weather predictions, methane leak monitoring, supply chain risk optimisation, environmental compliance.	Land use modelling, crop yield optimisation, water stress monitoring, pest detection, monitoring soil carbon levels, verifying carbon sequestration, supply chain optimisation, supply chain risk detection.	Smart city planning, traffic management, mobility analytics, urban heat island mitigation, air pollution monitoring, disaster resilience predictions, real estate risk predictions.	Military intelligence, logistics optimisation, insurance risk modelling, national border monitoring, retail site selection, venture capital analytics, mining and resource exploitation, maritime monitoring.

[1] Examples include K2, GeoLM, SpaceLLM; [2] Examples include Prithvi, ClimaX, Clay; [3] Examples include CityFM, Satlas.

LUNAR

While training costs for foundation models is generally increasing...

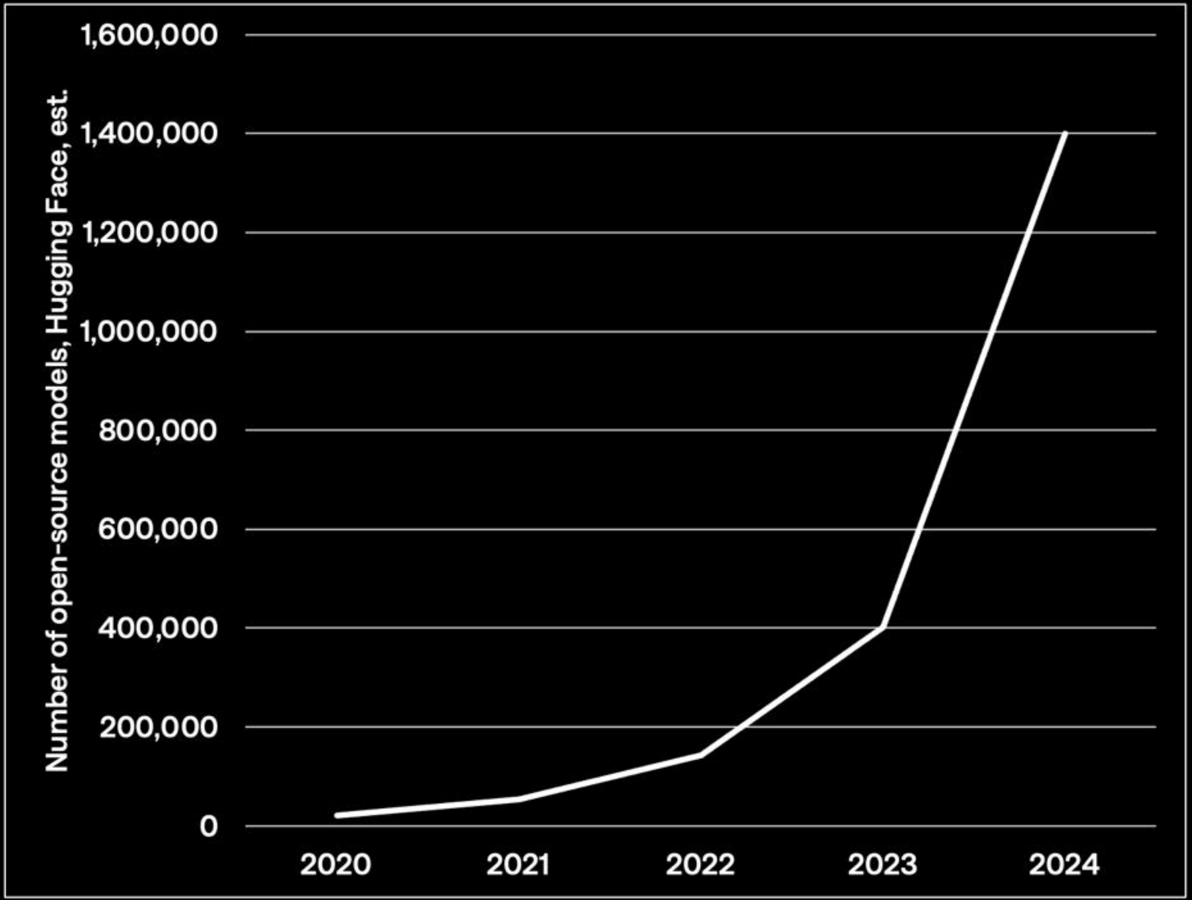


Source: epoch.ai; [1] Note the discussion around DeepSeek's true training costs, with estimates from 5,6mln USD to 1,6bln USD.

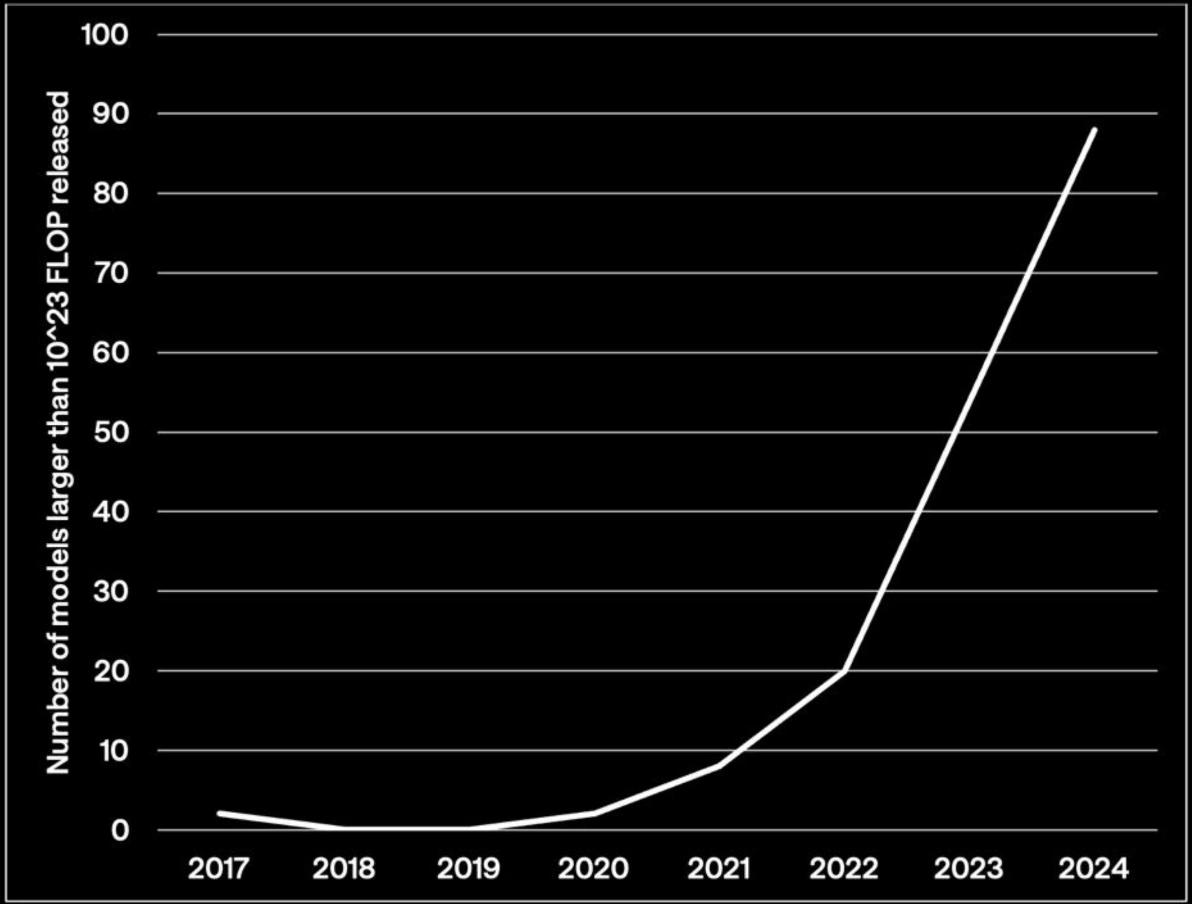
LUNAR

…the competition between models is fierce…

Open-source models grew ~70X in the last four years…



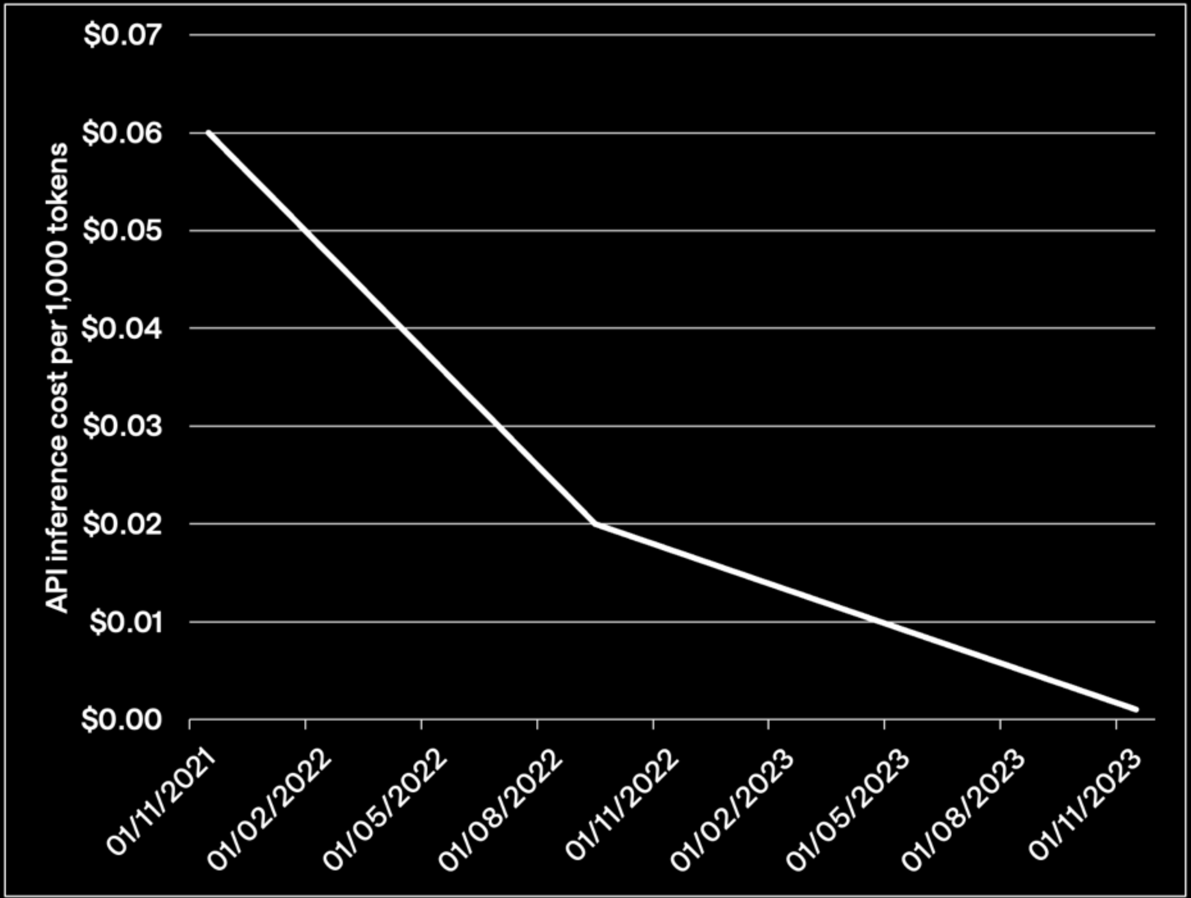
…and more players are releasing large (over 10^23 FLOP) models.



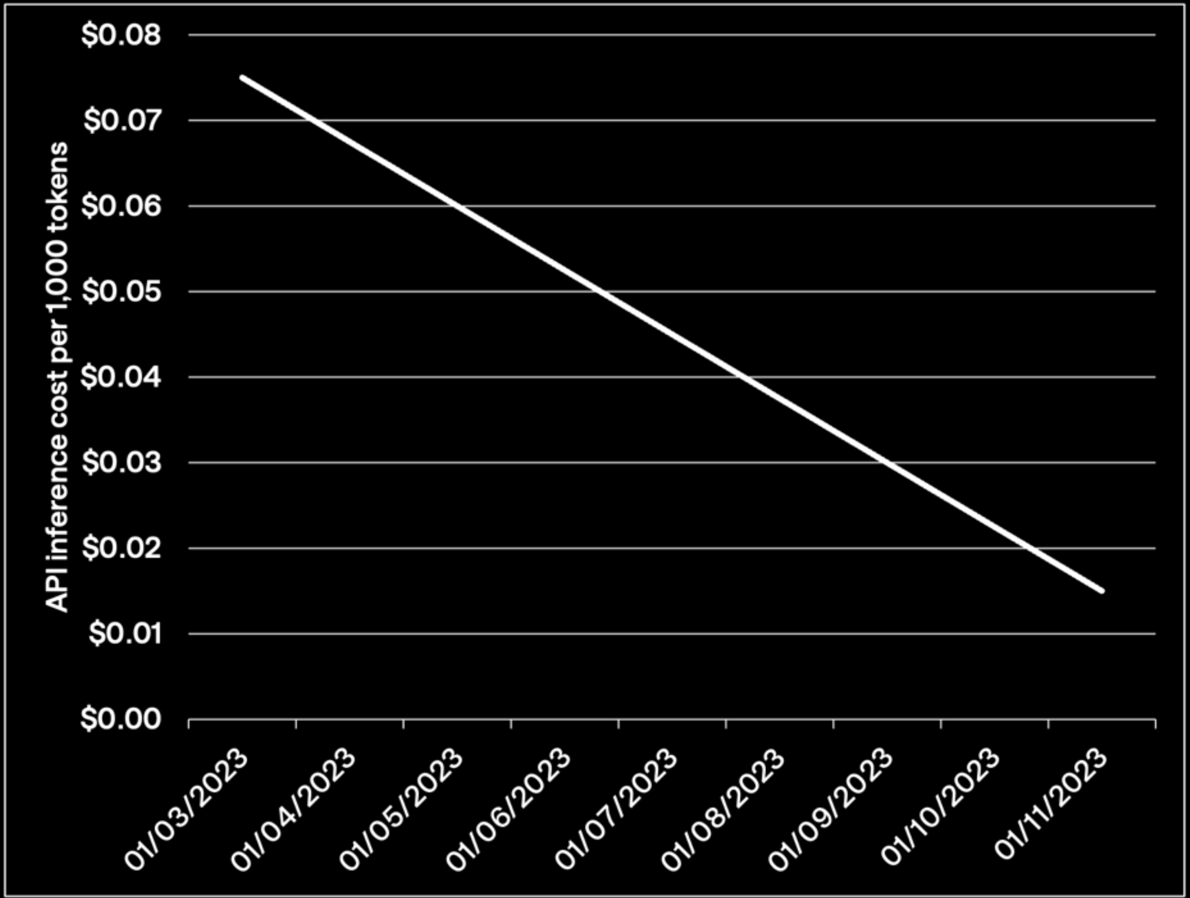
Source: Hugging Face, Epoch.ai.

...and falling token costs will make more use cases economical

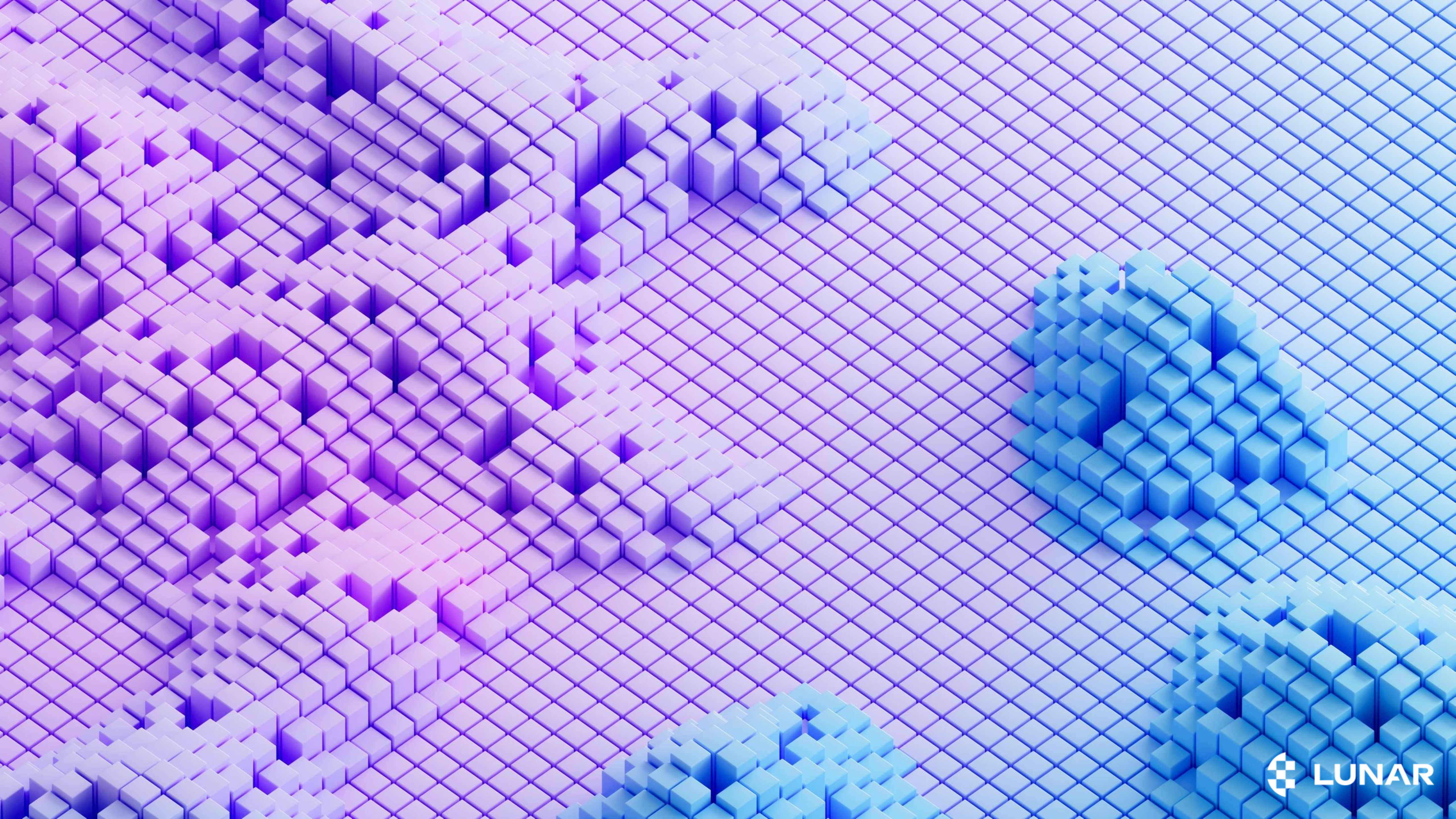
GPT-3 token cost fell ~86% in cost per year...



...and GPT-4 tokens showed a similar trend of falling 92% per year.



Source: Ark Invest analysis (2024); The Batch (2024).


LUNAR