

Think piece

May 21, 2024

MSFT v GOOG: The battle for AI domination



KC Rajkumar, CFA, PhD

(408) 425-5680

KC@LynxEq.com

As the Gen AI revolution matures beyond the gee-whiz phase and enters its second year, battle lines for global AI domination are emerging. After an early lead at Microsoft/OpenAI via the imminently accessible CoPilot bundle, Google laid out last week a credible case at the I/O event – Google can hardly be considered a laggard, a point we had hypothesized two months ago ([link](#)).

MSFT and GOOG are likely to continue their traditional Office/browser wars into the Gen AI battlefield. On a couple of fronts though, we do not expect strenuous confrontation. MSFT retains its traditional Office365 userbase and GOOG its Android userbase, we think. After its AI product roll-out last week, GOOG's traditional strength in search does not appear to be at immediate risk, but this is an arena MSFT CEO has stated his intention to make headway.

*The bulk of the rivalry could be over independent app developers, the ones who are not tied to an Office/GoogleDoc framework or tied to a Windows/Android framework. Up for grabs are API-based Gen AI applications and this is where one can expect share shift between Azure and GCP. Rather than get mired in the tech details, we'd like to focus on a few high-level ideas coming out of Google I/O and **look for comps** in the Microsoft Build presentation today.*

*The framework for comparison is the same as it always is – functionality and cost. Google announced pricing for Gemini API at **a fraction** of OpenAI's pricing for GPT-4o APIs. We'll look for updates at the Microsoft event today.*

*From a silicon perspective, the Microsoft vs Google battle for AI users is really a battle between **NVDA's GPU vs GOOG's TPU**. From a cost perspective we think the TPU has a clear advantage. The question is whether Google's Gemini offers **functional advantages** as well. Four areas of functionality we consider: Context window, Audio AI agent, miniaturization of LLMs and multi-modality. Where does all this leave AAPL and AMZN/AWS? Our thoughts at the end of this note.*

--XX--

Pricing and functionality: Besides pricing, the four areas of functionality we would try compare: Context window, Audio AI agent, miniaturization of LLMs and multi-modality. There is likely to be a key difference between the **approach** taken by the two giants – MSFT is more likely to offer its Azure customers **turn-key solutions** that are fully integrated into its existing Azure infrastructure. GOOG is more likely to offer **platform solutions** and leave it up to 3rd party developers to hash out the final applications. In doing so MSFT may well be able to price its products higher than GOOG and still maintain its share.

Investors and AI developers are just getting the initial taste of real products out of the CSPs. Gen AI share is up for grabs. But on one facet we are relatively clear – the **cost advantage** for Gen AI may remain with

GOOG, due to its LLMs being trained and served on its internally developed, lower-power consuming TPU silicon. It is not clear to us how MSFT would surmount this challenge.

Pricing – superior at Google: For enterprise application developers, Google rolled out astonishingly low pricing for Gemini-1.5 of **\$0.35** per 1mn tokens vs. OpenAI's GPT-4o pricing of **\$5.00** per 1mn tokens. Unless OpenAI drops its pricing dramatically we are not clear how it will remain competitive. For individual enterprise/consumer users too, Microsoft's CoPilot priced at **\$30/month** does not compare well with Google's **free** usage of Gemini-1.5 for limited daily usage (up to 1.5K requests per day)

Context window – superior at Google: Google announced a very large context window size – 2million tokens vs. OpenAI's much smaller GPT-4o window at 128K tokens. Google clearly has an **advantage**. Why does window size matter? A larger context window offers a larger searchable dataset and lowers the chance of hallucinations. We will look for updates at the Build event today.

The new killer interface – Audio: If text-based chats with an AI agent was all the rage last year, we think it is becoming clear that the new interface for AI interaction is voice-based. The centerpiece of the various demos from OpenAI and Google I/O last week was the natural-sounding audio agent, in our view. While voice to text and crude voice applications have been around for a long time, this is the first time that end users would be able to interact naturally with an audio AI agent for probing the internet or proprietary databases.

At the most obvious level, we think human **customer service reps** in industries ranging from travel to healthcare could be automated via AI audio agents.

We think OpenAI/Microsoft and Google are likely on **equal footing** in this arena. The ability to audio access the internet by **non-English** language speakers across the globe we think could be key element for driving up monetizable applications the next few years.

Multi-modality: It should be abundantly clear to investors that Gen AI is rapidly transitioning from LLMs based on text to LLMs based on text integrated with audio, images, video and handwriting/sketches. Rather than AI interface being limited to a **static** display screen, AI applications/experience are becoming **immersive, mobile**. Investors got a taste of multi-modality at a half-hour sneak peak from OpenAI last week vs. a nearly two-hour presentation at Google I/O. We expect more detailed multi-modal demos at the Build event today.

Miniaturization of LLMs: Bragging about the model size of LLMs may have been well and good in the gee-whiz phase of Gen AI last year. But as the technology evolves and as the Street demands monetizable applications, investors are likely to conclude that one LLM size does not fit all, and perhaps one does not need a trillion-parameter model for mass market AI applications.

There are several smaller-sized, highly functional open-source LLMs already in the public arena; Llama3, Claud3 and Mistral are examples. As we noted in a recent AAPL note ([link](#)), these models are small enough to load onto a client PC for running AI applications on-device.

Google rolled out a range of LLM models for a spectrum of workload complexity. Counting off from the most complex workload to the simplest, within its Gemini 1.5 framework, Google rolled out 4 models- Pro, Flash, Gemma and Gem. Pro and Flash run on the cloud, Gemma is small enough to run on a single TPU and Gem runs on a mobile device. And specific to mobile devices, Google Nano runs within Android 15 for powering embedded applications.

Just as compute platforms shifted from main frames of the 80s to PCs of the 90s, we think LLM-based AI workloads too are likely to miniaturize and move to personal devices. Microsoft's launch of the Surface PC platform appears to be an attempt to participate in this trend. We think Google may have **established an edge** over Microsoft in offering a **range of model sizes**. We'll wait for details at the Build event today.

AAPL and AMZN: It is increasingly becoming evident that rather than jump into the LLM fray with the likes of MSFT and GOOG, AAPL expects to pair the best-of-breed LLMs from other vendors with specific applications on Apple devices. We don't expect AAPL to get in the middle of a GOOG vs. Microsoft/OpenAI dogfight. Rather, AAPL might pick one or the other, or both, to support specific applications on Apple devices. As for AMZN, at this stage, it is not entirely clear to us if AMZN/AWS could prevent its enterprise customers from drifting away to other CSPs. The one huge advantage AWS has over the others – a mature installed base of ARM-based general-purpose servers. But how does that translate into Gen AI wins?

Net/Net: GOOG and MSFT are setting up battlelines for grabbing share in Gen AI applications. Whereas on the pricing front GOOG may have an advantage, superiority in functionality needs to be demonstrated and fought over. 2+ months ago, prior to the bottom on March 8th, investors were ready to give up GOOG for dead. Since then, as details emerged, GOOG outperformed MSFT and hyperscale cohorts handily. At its Build conference today, MSFT has a chance to re-establish excitement. The question is, will it?

Disclosures and Disclaimers

Lynx Equity Strategies, LLC is an independent equity research provider. The Company is not a registered broker dealer or investment adviser. No employee or member of the Company, or immediate family member thereof, exercises investment discretion over securities of any issuer analyzed in this report two days prior and/or two days after this report is issued. It participates in "Alpha Capture Systems" that seeks investment or trading ideas from the sell-side and may pay for the participation based on relative performance.

Limitations of Information

This report has been prepared for distribution to only qualified institutional or professional clients of Lynx Equity Strategies, LLC (the "Company"). The contents of this report represent the views, opinions, and analyses of its authors. The information contained herein does not constitute financial, legal, tax or any other advice. All third-party data presented herein were obtained from publicly available sources which are believed to be reliable; however, the Company makes no warranty, express or implied, concerning the accuracy or completeness of such information. In no event shall the Company be responsible or liable for the correctness of, or update to, any such material or for any damage or lost opportunities resulting from use of this data. Nothing contained in this report or any distribution by the Company should be construed as any offer to sell, or any solicitation of an offer to buy, any security or investment. Any material received should not be construed as individualized investment advice. Investment decisions should be made as part of an overall portfolio strategy and you should consult with a professional financial advisor, legal and tax advisor prior to making any investment decision. Lynx Equity Strategies, LLC shall not be liable for any direct or indirect, incidental or consequential loss or damage (including loss of profits, revenue or goodwill) arising from any investment decisions based on information obtained from Lynx Equity Strategies, LLC. Reproduction and Distribution Strictly Prohibited. No user of this report may reproduce, copy, distribute, sell, resell, transmit, transfer, license, assign or publish the report itself or any information contained therein. This report is not intended to be available or distributed

for any purpose that would be deemed unlawful or otherwise prohibited by any local, state, national or international laws or regulations or would otherwise subject the Company to registration or regulation of any kind within such jurisdiction.

Copyright, Trademarks, Intellectual Property

Unless otherwise indicated, all copyrights, trademarks, service marks, logos and other intellectual property included in this report are proprietary materials of Lynx Equity Strategies, LLC and the unauthorized use of such terms, marks, and logos is strictly prohibited. The Company reserves all rights with respect to the intellectual property ownership of all materials in this report and will enforce such rights to the full extent permissible by law.