Who Wins When AI Models Open Up?

Headlines

- Open-weight models are rapidly closing the performance gap with frontier systems, shortening proprietary leadership spells and increasing the risk of model-layer commoditization.
- A narrower and less durable performance advantage could weaken model developers' pricing power and make it harder to earn returns on rising investment.
- Application providers and end users should benefit from better models at lower prices, while the effect on infrastructure depends on whether cheaper inference generates enough additional usage to offset efficiency gains.
- Market reactions have reflected this tension: open-weight breakthroughs initially weighed on AI infrastructure stocks, but those losses were largely recovered over time.

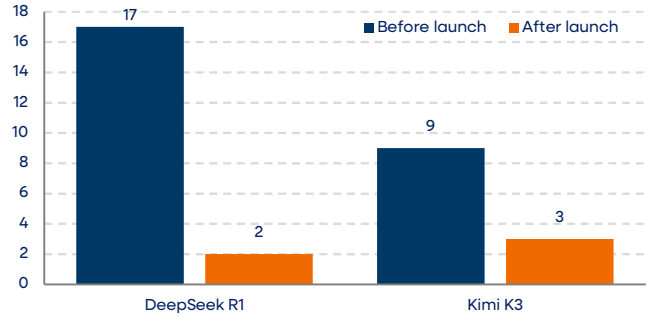
Rayhan Momin
rayhan.momin@pnc.com

The frontier is getting crowded

Open-weight releases have renewed scrutiny on the returns on AI investment, as the performance advantage of leading models appears increasingly tenuous while cheaper alternatives erode pricing power. This raises the risk that the model layer becomes utility-like, with frontier leaders bearing the investment cost while nimble followers compress pricing and returns. However, even as this occurs, other parts of the AI stack (i.e. the application and infrastructure layers) may benefit.

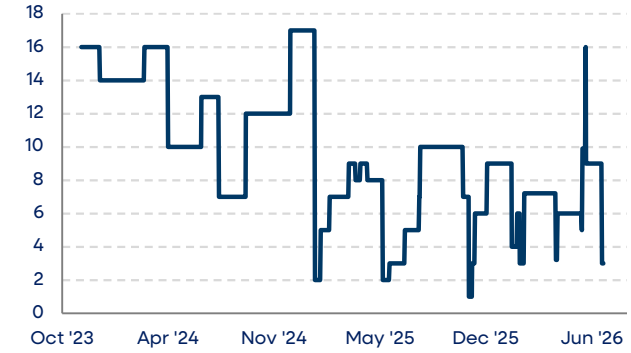
The performance gap between leading closed-weight and open-weight models follows a sawtooth pattern: proprietary releases push the frontier outward, but open-weight providers subsequently catch up and narrow the gap. These gaps have compressed over time as open-weight models have become more competitive, while a faster release cadence among frontier labs has shortened leadership spells.

Figure 1: Best open-weight model gap to frontier before and after launch (index pts)



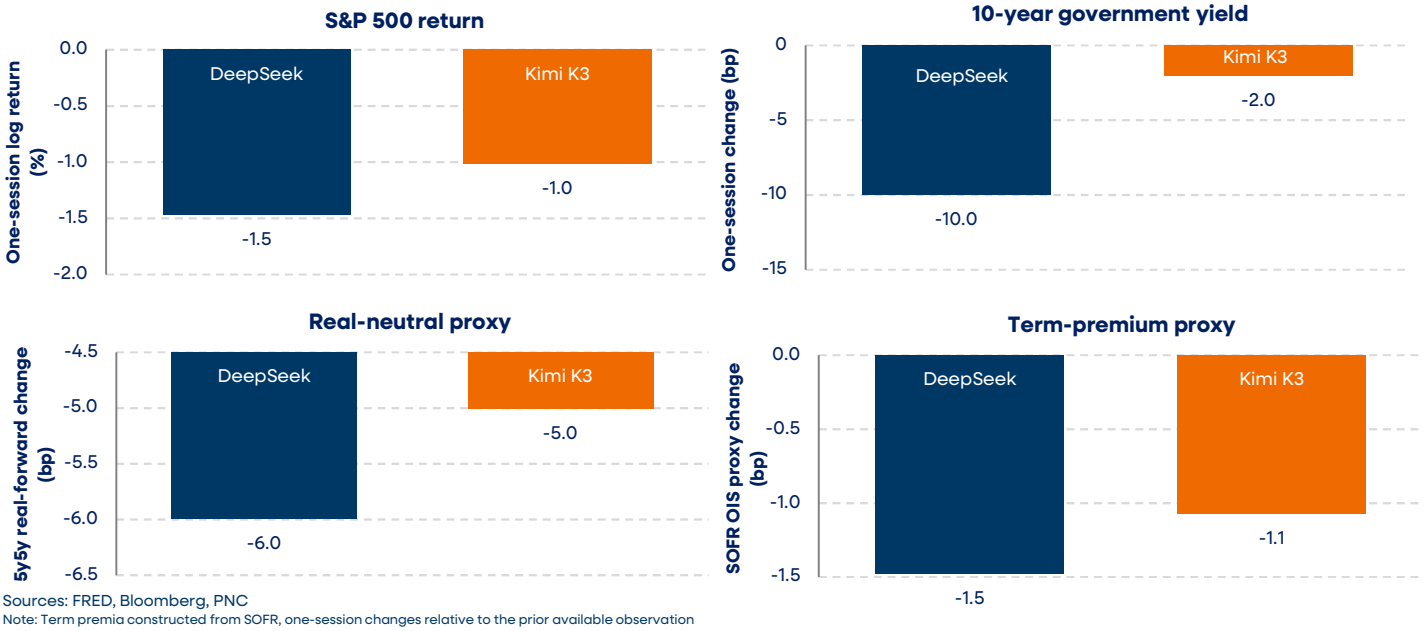
Sources: Artifical Analysis, PNC

Figure 2: Daily gap between the best proprietary and best open model (index pts)



Sources: Artifical Analysis, PNC

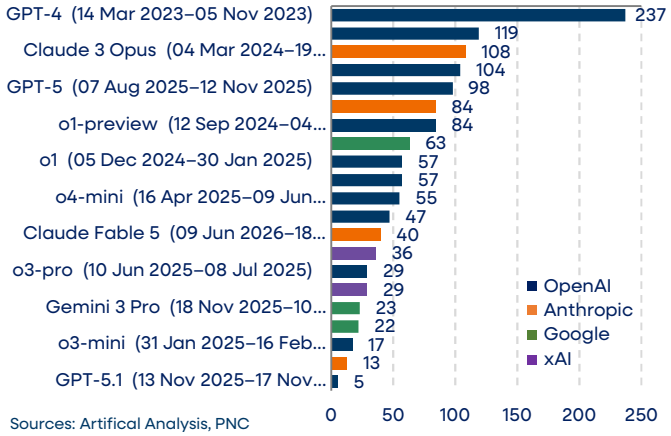
Figure 3: Open-model breakthroughs repriced AI cash flows and long rates



Sources: FRED, Bloomberg, PNC
Note: Term premia constructed from SOFR, one-session changes relative to the prior available observation

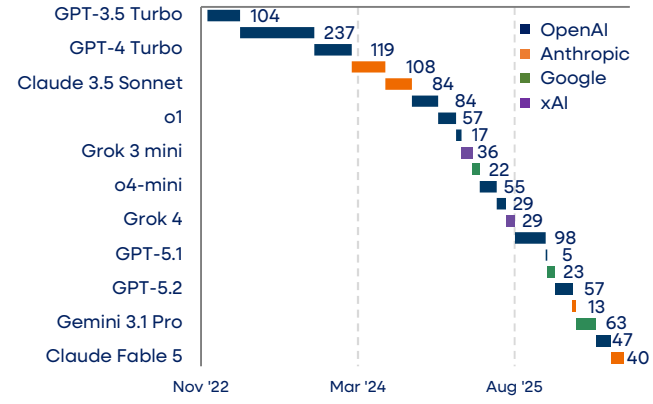
The performance gap between leading closed-weight and open-weight models follows a sawtooth pattern: proprietary releases push the frontier outward, but open-weight providers subsequently catch up and narrow the gap. These gaps have compressed over time as open-weight models have become more competitive, while a faster release cadence among frontier labs has shortened leadership spells.

Figure 4: Frontier leadership spells ranked by duration (days at number 1)



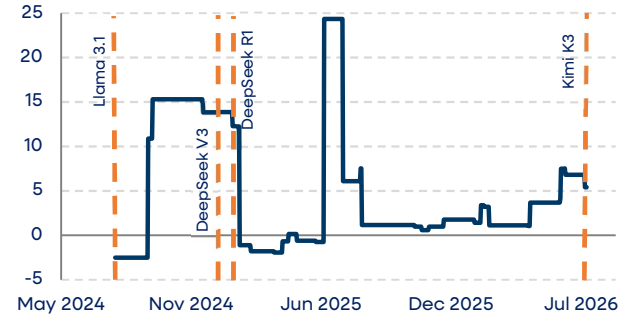
Sources: Artifical Analysis, PNC

Figure 5: Succession of models at the proprietary frontier (days at number 1)



Sources: Artificial Analysis, PNC

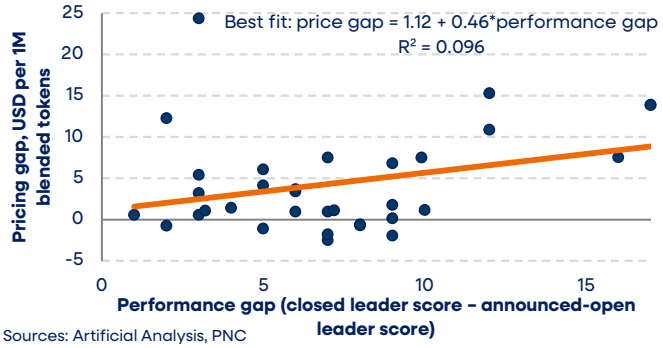
Figure 6: Hosted API pricing gap: closed-weight leader vs. announced-open leader (USD per 1M tokens)



Sources: Artificial Analysis, PNC

Although hosted API prices bundle performance with other service attributes, performance leadership remains an important determinant of model economics. The positive correlation between performance and pricing gaps suggests that frontier releases strengthen providers' ability to command a premium, helping sustain rents and support the returns needed to justify continued investment. This explains why a performance advantage that is both narrowing and increasingly short-lived can spook markets.

Figure 7: Pricing gap versus performance gap



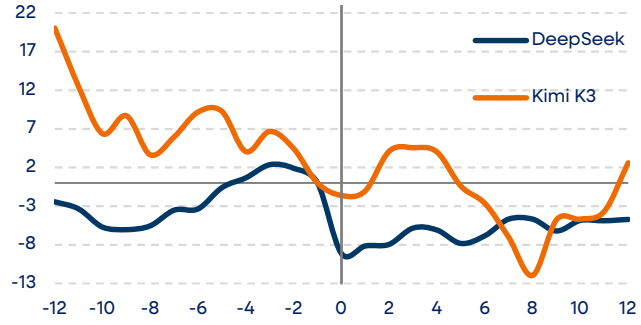
Sources: Artificial Analysis, PNC

Are open-weight models necessarily bad for the U.S. AI complex? No. Rent compression most directly threatens the model layer, while the application layer benefits from better models and lower input costs. The infrastructure layer could also benefit if lower inference prices stimulate enough additional usage to outweigh the reduction in compute required per task. End consumers may benefit most of all through better products and lower prices.

The key question is whether lower inference prices generate enough additional demand to offset the decline in revenue per token and the reduction in compute required for each task. If usage is sufficiently elastic, cheaper models could expand the range of economically viable applications, increase total inference volumes, and sustain demand for cloud and semiconductor infrastructure. The benefits would not necessarily accrue evenly, but model-layer commoditization could enlarge the overall market even as it redistributes margins within the stack.

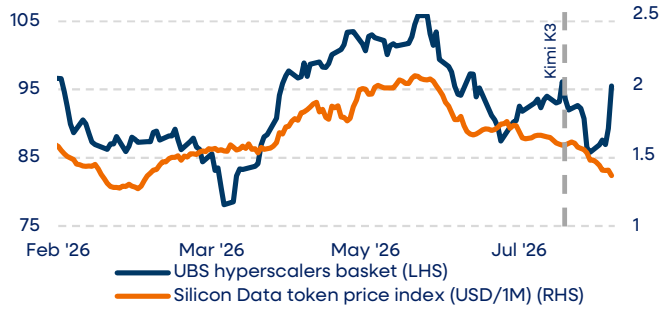
Recent evidence is suggestive. The PHLX Semiconductor Sector Index initially sold off following the Kimi K3 release but has since recovered much of the decline, despite remaining in correction territory after the broader semiconductor selloff. Silicon Data's LLM index tracks prices per 1 million tokens, which have continued to fall even as an equal-weighted basket of hyperscaler stocks rallied following strong earnings releases. This decoupling suggests that lower model prices could stimulate demand and benefit the infrastructure layer, even as they compress model-layer rents. Four of the five hyperscalers in the basket also signed a statement supporting open-weight models, reinforcing the market-based evidence that model commoditization need not weaken returns across the broader AI stack.

Figure 8: PHLX SOX cumulative return from day of the model release (%)



Sources: Bloomberg, PNC
Note: Event date taken as the day before the model release.

Figure 9: Hyperscaler equities decoupled from token price in July



Sources: Bloomberg, PNC

Disclosures

The material presented and the views expressed herein are of a general nature and does not constitute the provision of investment or economic advice to any person, or a recommendation of any particular securities, financial instruments, strategies or banking services. Opinions and forecasts expressed herein are subject to change without notice. Relevant information was obtained from sources deemed reliable. Such information is not guaranteed as to its accuracy. You should seek the advice of an investment professional to tailor a financial plan to your needs. To the extent that these materials reference PNC data, such materials are not intended to be reflective or indicative of, and should not be relied upon as, the results of operations, financial condition or performance of PNC.

Forward-looking statements are necessarily subject to numerous assumptions, risks and uncertainties, which change over time. Future events or circumstances may change our outlook and may also affect the nature of the assumptions, risks and uncertainties to which our forward-looking statements are subject. Forward-looking statements speak only as of the date made. We do not assume any duty and do not undertake any obligation to update forward-looking statements. Actual results or future events could differ, possibly materially, from those anticipated in forward-looking statements, as well as from historical performance. As a result, we caution against placing undue reliance on any forward-looking statements.

"PNC" is a registered mark of The PNC Financial Services Group, Inc.

© 2026 The PNC Financial Services Group, Inc. All rights reserved.