Flash | Citi Research

19 Oct 2026 05:38:07 ET | 13 pages

China Internet

Implications From DeepSeek v.4.1 Flash 1017: Pivoting to Workflow Efficiency

CITI'S TAKE

We believe the latest DeepSeek V4.1 Flash 1017 released on Oct 17th could suggest that the domestic AI battleground might be shifting from pure parameter scaling toward inferencing cost-efficiency and agentic workflow integration. The emphasis on reduced execution latency, IDE framework tuning, and search/fetch APIs will likely heighten pricing pressure across standard cloud APIs, in our view. Meanwhile, the release lull in Sept and October by Chinese LLMs could indicate major labs are preparing frontier architectures and strengthening agentic workflows. We believe Alibaba is likely to roll out Qwen 4.0 within the next two months. Furthermore, with tangible monetization emerging, Kuaishou's Kling 4.0 and DeepSeek's enhanced game development and 3D modeling capabilities suggest that multimodal and video generation model competition could intensify in coming months, in our view. We maintain our Buy rating on BABA and Neutral rating on Kuaishou.

Alicia Yap, CFA[AC]

Brian Gong

Nelson Cheung

Vicky Wei, CFA

What's new? - On Oct 17, DeepSeek released its latest model V4.1 Flash-1017. DeepSeek highlighted that this latest model version not only demonstrates improving benchmark performance, it also shortens execution time and output length, helps users to complete tasks at faster speeds and at lower cost. Specifically, the latest update focuses on optimizations tailored for platform frameworks such as DeepSeek Harness, CodeBuddy, and WorkBuddy. V4.1 Flash-1017 also enhances its capability in game development and 3D modeling, supporting more programming tasks and project development requests. This new version also added API support for Search and Fetch endpoints which DeepSeek employs more stringent identification mechanisms to filter out SEO spam and GEO manipulation. Bloomberg (17-Oct-2026) reported that DeepSeek noted in its technical report on Hugging Face that the latest iteration demonstrates substantial improvements across reasoning, coding, agentic tasks and visual benchmarks.

Slower LLM release cadence in recent months; Kuaishou unveils Kling 4.0 - Following a rapid cadence of new model announcements between June and August, Chinese LLM release activity experienced a modest slowdown throughout September and October. We believe this interim lull likely stems from major labs preparing frontier-scale architectures such as Alibaba developing next-generation models like Qwen 4.0. Despite the broader plateau in foundation model drops, Kuaishou officially launched Kling 4.0 on October 15. The update introduces an upgraded creation suite with native MCP/CLI integrations. Centered on professional-grade video generation, Kling 4.0 delivers up to 4K/1080p 10-bit HDR output alongside substantially improved motion stability across high-speed action sequences, subject tracking, and complex camera movements.

See Appendix A-1 for Analyst Certification, Important Disclosures and Research Analyst Affiliati



Qwen V4 coming soon? - During last month's Apsara conference, Alibaba indicated that its next-generation **Qwen 4 series** is slated for release in the near term, though an exact timeline was not disclosed. We anticipate a rollout within the next two months, with the flagship full-scale model expected to reach approximately 3T parameters. As highlighted by BABA, its model roadmap pursues a dual-track strategy: maintaining a comprehensive, full-scenario portfolio designed to establish SOTA benchmarks, alongside optimized Flash and Plus variants engineered for high throughput, low latency, and cost-effective deployment in enterprise production environments. Expanding its AI edge ecosystem, Alibaba also unveiled the next-generation **N1 series smart glasses** on October 13. The base model starts at RMB 2,999 (~US$450), while the Pro edition is priced at RMB 3,499. The glasses feature deep integration with the Qwen AI assistant, supporting daily functional capabilities including e-commerce shopping, digital payments, and health monitoring. Additionally, Alibaba announced the retail launch of its **Qwen Clip AI earbuds** (initially showcased at Apsara), priced at RMB 1,299.

Figure 1. DeepSeek-V4.1-Flash-1017 benchmarks comparison

Benchmark	DS-V4.1-Flash-1017	DS-V4.1-Flash	DS-V4-Pro 0813	DS-V4-Flash 0731	Kimi K3	GLM 5.3
HLE (w/o tools)	42.6/44.0*	36.8/39.1*	42.7*	37.8*	43.5	42.0*
Terminal Bench 4.0	42.3	31.2	12.4	7.0	12.6	37.9
DeepSWE v1.1	77.0	74.2	62.7	54.4	67.5	66.9
ProgramBench	29.1	20.3	15.5	-	17.5	19.0
NL2Repo-Bench	68.2	65.4	61.5	54.2	58.0	58.0
Automation-Bench V1.0.6	60.5	54.8	43.2	37.7	46.7	48.8
Agents' Last Exam	35.4	31.8	25.7	25.2	27.6	28.5
Apex Bench v1.1	72.1	64.6	-	-	47.8	56.6
Chartography (w/tools)	83.0	78.9	-	-	68.1	-
BabyVision (w/tools)	92.5	89.6	-	-	85.7	-
ZeroBench-main (w/tools)	51.0	49.0	-	-	41.0	-

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Company Reports, Citi Research, Note: * for text-only ratings

Alibaba

(9988.HK; HK$108.8; 1; 16 Oct 26; 16:10)

Valuation

Our target price for Alibaba H-shares of HK$192 is based on 10x P/E on FY2027E Ecommerce group net profit, 7.0x P/S on FY2027E AI Cloud and Compute Services revenue, 5.0x P/S on AI Labs and Applications revenue, and 0.2x P/S on all other businesses. Our target multiples are benchmarked against comps. We also assign a 30% discount to net cash position, in-line with our SOTP for other internet peers. Our SOTP fair value comes to HK$192 per share.

Risks

Key downside risks that could prevent the shares from reaching our target price include: (i) failure in executing its new retail strategy; ii) investment spend and margins pressure become worse than expected; iii) slowdown of user traffic and online GMV and losing appeal to brands & merchants; iv) integration risks for newly acquired entities; v) slowdown of the Chinese and global economies and overhang of US-China or other trade disputes; and vi) regulatory risks on poor product quality and integrity of merchants.

Alibaba Group Holding

(BABA.N; US$109.11; 1; 16 Oct 26; 16:00)

Valuation

Our target price for Alibaba N-shares of US$193 is based on 10x P/E on FY2027E Ecommerce group net profit, 7.0x P/S on FY2027E AI Cloud and Compute Services revenue, 5.0x P/S on AI Labs and Applications revenue, and 0.2x P/S on all other businesses. Our target multiples are benchmarked against comps. We also assign a 30% discount to net cash position, in-line with our SOTP for other internet peers. Our SOTP fair value comes to US$193 per share.

Risks

Key downside risks that could prevent the shares from reaching our target price include: (i) failure in executing its new retail strategy; ii) investment spend and margins pressure become worse than expected; iii) slowdown of user traffic and online GMV and losing appeal to brands & merchants; iv) integration risks for newly acquired entities; v) slowdown of the Chinese and global economies and overhang of US-China or other trade disputes; and vi) regulatory risks on poor product quality and integrity of merchants.

Kuaishou

(1024.HK; HK$29.3; 2; 16 Oct 26; 16:10)

Valuation

Our target price of HK$41 is based on: 1) HK$21 for the ecommerce sector with a 6x 2027E PE assigned for the core business, which is at discount to ecommerce peers given its muted GMV trajectory; and 2) HK$20 based on Kuaishou's stake in Kling post latest financing and Kling's valuation during private financing with 20% holding discount.

Risks

Upside risks to our target price being exceeded include: 1) better-than-expected GMV performance with macro recovery; 2) strong Kling revs growth with successful updates; and 3) better-than-expected cost optimization.

Downside risks include: 1) worsen-than-expected GMV performance on macro and policy impacts; 2) slower-than-expected Kling revs growth; 3) more-than-expected AI-related investments; and 4) any stake disposal by major shareholders.