24 Apr 2026 03:31:50 ET | 14 pages

China Internet

Implication of DeepSeek-V4 Model For China Internet Giants' Model Landscape

CITI'S TAKE
The rapid iteration of AI models from Chinese developers, highlighted by the launch of DeepSeek-V4 today, underscores a competitive landscape where we believe market leadership is not "winner-takes-all". Rather than focusing solely on performance benchmarks, crucial factors for success will be monetization strategy, sustained adoption, and generating revenue through APIs, token usage and infrastructure services. Enterprises are likely to adopt a pragmatic, multi-model approach, selecting the best tool for specific tasks based on a nuanced evaluation of quality, capability, and cost. Differentiated approaches like Tencent's HY3.0 focus on real-world applications and Alibaba's Qwen 3.6-Max specializing in complex coding and reasoning. The advanced architectures in DeepSeek-V4 signal a broader trend: the technological gap is closing, positioning China's leading models as increasingly viable alternatives to their top-tier global counterparts, which will intensify competition and provide enterprises with a wider array of powerful, cost-effective options. Maintain Buy on Alibaba and Tencent.

Alicia Yap, CFA[AC]
Nelson Cheung
Vicky Wei, CFA

What's new? On April 24, 2026 (Asia time), DeepSeek announced (Bloomberg) that the company has officially released a preview version for its latest model DeepSeek-V4 (both Pro and Flash) on open-source developer sites including Hugging Face and ModelScope. It is a 1-trillion parameter MoE model with cost-effective 1M context length. The V4-Pro version is a 1.6trn parameter with 49b activated parameters, and the performance of the Pro version rivals the latest performance of GPT5.4 and Opus 4.6 while the V4-Flash version is a 284bn parameter with 13bn activated parameters targeting flash, for efficient and economic options.

Upgrade in architecture and optimization – Specifically, DeepSeek noted that its V4 series model has incorporated key upgrades: 1) a hybrid attention architecture that combines Compressed Sparse Attention (CSA) and Heavily Compressed Attention (HCA) to improve long-context efficiency; 2) Manifold-Constrained Hyper-Connections (*m*HC) that enhances conventional residual connections and 3) Muon optimizer for faster convergence and greater training stability. The company noted that on a one-million-token context setting, DeepSeek V4-Pro requires only 27% of single-token inference FLOPs and 10% of KV cache compared with DeepSeek-V3.2, the improvement is important as it enables the company to routinely support one-million token contexts and support making long-horizon tasks and test-time scaling more feasible, per the announcement.

Capabilities and ranking – DeepSeek claims that its V4-Pro's leads all current open models (except Gemini-3.1) in rich world knowledge and beats all current open models in math, STEM, coding and rival existing top closed-source models. For the V4-Flash version, it can perform on par with V4-Pro on simple agent asks with

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

reasoning capabilities close to V4-Pro in a cost-effective API pricing givens its smaller parameter size and faster response time.

Implications for Tencent's HY3.0 and Alibaba's Qwen 3.6 Max – The rapid iteration of new AI models makes it difficult to determine an ultimate winner. We believe the market is not a "winner-takes-all" scenario; instead, success will depend on monetization, adoption rates, and sustainable revenue from services. We believe it is more important to assess how well a model can monetize and how well a model company can consistently maintain its model adoption rate and generate sustainable revenues through API and token-usage as well as infrastructure services. Enterprises will likely choose models based on quality, capability, and cost for specific use cases, measuring the ROI for each task. For example, Tencent's HY3.0 (released on April 23) claims that it takes a "pragmatic approach, and focus on solving real-world applications over lab scores. Alibaba's latest model Qwen 3.6-Max released on April 20, a closed model rather than open-weight, represents the flagship tier of the Qwen 3.6 series designed to push for agentic coding and complex reasoning. It introduced "preserve-thinking" which allows the model to carry reasoning across multiple interactions. Per Alibaba, the model currently holds the top score on six major coding benchmarks, including SWE-bench Pro, Terminal-Bench 2.0.

Figure 1. DeepSeek-V4 benchmark comparison

Benchmark (Reasoning Effort)		DS-V4-Pro Max	DS-V4-Flash Max	K2.6 Thinking	GLM-5.1 Thinking	Opus-4.6 Max	GPT-5.4 xhigh	Gemini-3.1-Pro High
Knowledge & Reasoning	MMLU-Pro (EM)	87.5	86.2	87.1	86.0	89.1	87.5	91.0
	SimpleQA-Verified (Pass@1)	57.9	34.1	36.9	38.1	46.2	45.3	75.6
	Chinese-SimpleQA (Pass@1)	84.4	78.9	75.9	75.0	76.2	76.8	85.9
	GPQA Diamond (Pass@1)	90.1	88.1	90.5	86.2	91.3	93.0	94.3
	HLE (Pass@1)	37.7	34.8	36.4	34.7	40.0	39.8	44.4
	LiveCodeBench (Pass@1)	93.5	91.6	89.6	-	88.8	-	91.7
	Codeforces (Rating)	3206	3052	-	-	-	3168	3052
	HMMT 2026 Feb (Pass@1)	95.2	94.8	92.7	89.4	96.2	97.7	94.7
	IMOAnswerBench (Pass@1)	89.8	88.4	86.0	83.8	75.3	91.4	81.0
	Apex (Pass@1)	38.3	33.0	24.0	11.5	34.5	54.1	60.9
	Apex Shortlist (Pass@1)	90.2	85.7	75.5	72.4	85.9	78.1	89.1
Long Context	MRCR 1M (MMR)	83.5	78.7	-	-	92.9	-	76.3
	CorpusQA 1M (ACC)	62.0	60.5	-	-	71.7	-	53.8
Agentic	Terminal Bench 2.0 (Acc)	67.9	56.9	66.7	63.5	65.4	75.1	68.5
	SWE Verified (Resolved)	80.6	79.0	80.2	-	80.8	-	80.6
	SWE Pro (Resolved)	55.4	52.6	58.6	58.4	57.3	57.7	54.2
	SWE Multilingual (Resolved)	76.2	73.3	76.7	73.3	77.5	-	-
	BrowseComp (Pass@1)	83.4	73.2	83.2	79.3	83.7	82.7	85.9
	HLE w/tools (Pass@1)	48.2	45.1	54.0	50.4	53.1	52.0	51.6
	GDPval-AA (Elo)	1554	1395	1482	1535	1619	1674	1314
	MCPAtlas Public (Pass@1)	73.6	69.0	66.6	71.8	73.8	67.2	69.2
	Toolathlon (Pass@1)	51.8	47.8	50.0	40.7	47.2	54.6	48.8

Source: DeepSeek, Citi Research

Figure 2. Qwen3.6 benchmark comparison

Benchmark	Context	Qwen 3.6 Max (Preview)	Qwen 3.6 Plus	Qwen 3.5 Plus	Claude 4.5 Opus	GLM 5.1
SuperGPQA	Garduate-Level Knowledge	73.9	71.6	70.4	70.6	68.0
AA-Omniscience Index	Knowledge Reliability and Hallucination	10.0	3.0	-30.0	13.0	2.0
GDPval-AA	Real-World Valuable Task	51.0	43.0	35.0	48.0	52.0
QwenChineseBench	Chinese Real-World Knowledge	84.0	78.7	78.8	69.0	81.2
QwenClawBench	Real-World Agent	59.0	57.2	51.8	52.3	58.7
SkillsBench	Agent Skills	55.6	45.7	30.0	45.3	53.1
ToolcallFormatIFBench	Real-World Toolcall Following	86.1	83.3	80.1	84.2	60.1
QwenWebBench (Elo Rating)	Artifacts	1532	1495	1182	1530	1558
SciCode	Research Coding	47.0	40.7	42.0	49.5	43.8
NL2Repo	Long-Horizon Coding	42.9	37.9	32.2	43.2	42.7
Terminal-Bench 2.0 (Terminus-2)	Agentic Terminal Coding	65.4	61.6	52.5	59.3	63.5
SWE-Bench Pro	Agentic Coding	57.3	56.6	50.9	57.1	58.4

Source: Alibaba, Citi Research

Figure 3. HY3-Preview benchmark comparison

Benchmarks	HY3 Preview	HY2	GLM-5	Kimi-K2.5	Gemini-3.1-Pro	GPT-5.4 xhigh
FrontierScience Olympiad	70.0	63.3	65.9	67.8	76.8	76.8
IMO Answer Bench	84.3	74.9	82.5	81.8	89.2	91.8
HLE	30.0	17.5	30.5	31.5	44.4	39.8
GPQA-Diamond	87.2	80.9	86.0	87.6	94.3	92.8
Tsinghua Qiuzhen College Math Exam (avg@3)	88.4	66.5	81.5	77.7	95.5	99.3
CHSBO 2025 (text-only, pass^3)	87.8	82.9	78.0	70.7	78.0	85.4

Source: Tencent, Citi Research

Alibaba

(9988.HK; HK$130.4; 1; 23 Apr 26; 16:10)

Valuation

Our target price for Alibaba H-shares of HK$204 is based on 10x P/E on FY2027E Ecommerce Group net profit, 6.5x P/S on FY2027E Cloud Intelligence Group revenue, 2.0x P/S on AIDC Group revenue, and 1.0x 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$204 per H-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$131.7; 1; 23 Apr 26; 16:00)

Valuation

Our target price for Alibaba N-shares of US$205 is based on 10x P/E on FY2027E Ecommerce group net profit, 6.5x P/S on FY2027E Cloud Intelligence Group revenue, 2.0x P/S on AIDC Group revenue, and 1.0x 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$205 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.

Tencent Holdings

(0700.HK; HK$495.2; 1; 23 Apr 26; 16:10)

Valuation

Our target price of HK$783 is based on a SOTP approach, which implies 23.2x 2026E and ~20.9x 2027E P/E. Our SOTP is based on the following: 1) Online games – We apply a 12x P/E to our 2027 estimated games net profit of Rmb136bn and yield a valuation of HK$195 per share, or 25% of our target price; 2) Online advertising – We apply a 22x P/E to our 2027 estimated ads profit of Rmb82bn which yields a valuation of HK$217 per share, or 28% of our target price; 3) Social network – While we attribute the non-gaming revs (community platform revs) as overall social network revs, given the WeChat, QQ, and other Tencent apps, we apply a 25x P/E to the estimated profit of Rmb63bn and yield a valuation of HK$190 per share, or 24% of our target price; 4) Fintech – We apply a 15x P/E to our estimated profit of Rmb48bn and yield a valuation of HK$87 per share, or 11% of our target price; 5) Cloud/business services – We apply a 5x P/S (average of China peers) to our 2027 estimated Business services revs of Rmb70bn, which yields a valuation of HK$42 per share, or 5% of our target price; 6) Investment portfolio – We apply a 30% discount to the equity stake in the publicly listed investment portfolio which implies HK$53/share.

Risks

Downside risks that could prevent the Tencent shares from reaching our target price include: i) a faster-than-expected revenue slowdown in core PC games; ii) unsuccessful launches of new mobile games; iii) a decline of Honor of Kings ranking and momentum; iv) a further slowdown in China's economy, which might negatively affect advertising demand; and v) a change in the regulatory environment.