

Alibaba Group Holding (BABA.N)

Assessing the Official Release of Qwen3.8-Max & Read-Through From DeepSeek-V4-Flash-0731

CITI'S TAKE

We believe Alibaba's recent strong share price performance likely reflects positive cloud revenue read-throughs from [AWS](#) and [Azure](#) (which aligns with our expectation of AliCloud revs to accelerate to 45% yoy in [FY1Q27E](#)), along with the significantly improved benchmark scores of the new Qwen3.8-Max model (released Aug 3). As [noted](#) previously, rapid iteration of AI models, as seen with recent releases (DeepSeek-V4-Flash-0731 and Kimi K3), is fostering a "model-agnostic" environment where enterprises adopt multi-model strategies, choosing the best model and price for each task. This trend reduces the defensibility of any single model, shifting competitive focus to the platforms and infrastructure that manage them. We believe long-term success will require immense resources and a loyal customer base. Consequently, companies with full-stack capabilities, from chips and cloud infrastructure to models and applications, like Alibaba, are better positioned to lead, in our view. We reiterate our Buy rating and TP of US\$192.

Buy

Price (31 Jul 26 16:00)	US\$122.25
Target price	US\$192.00
Expected share price return	57.1%
Expected dividend yield	1.6%
Expected total return	58.7%
Market Cap	US\$293,041M

Alicia Yap, CFA^{AC}

Nelson Cheung

Brian Gong

Vicky Wei, CFA

Qwen3.8-Max official release - On Aug 3, Alibaba officially released Qwen3.8-Max which represents the strongest model within Qwen's model families. Based on Qwen3.5's architecture, Qwen3.8-Max extends its total parameters to 2.4trn, leveraging sparse MoE structure and hybrid attention mechanism, with 95B active parameters and 1M context window. Comparing to its preview version, Qwen3.8-Max demonstrates improvement in coding, office, scientific research and long-horizon mission with capabilities on end-to-end complex missions.

Benchmark performance - Qwen3.8-Max achieved significant improvement vs. Qwen3.7-Max and Qwen3.7-Plus in terms of software engineering, terminal agent, cowork tasks, agents empowering human workflows, long-horizon tasks, visual/chart/embodied reasoning and agentic mobile/computer use. As of Aug 3, Alibaba suggested a comparable performance of Qwen3.8-Max with overseas frontiers and Kimi K3 in TextArena (text, math, coding, business writing), CodeArena (multi-step reasoning and tool deployment) and VisionArena (visual understanding).

Agentic and dynamic workflows - Qwen3.8-Max has demonstrated strong agentic coding capabilities, completing complex projects from scratch over several days. It also excels as an office assistant, utilizing advanced Reinforcement Learning. The model has been successfully stress-tested for a wide range of high-value office scenarios, particularly in specialized fields such as commercial law, structural

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

engineering, and sports data analysis. Qwen3.8-Max is proficient in creating dynamic workflows for automated strategy and system management. It also demonstrates strong, independent long-horizon planning for complex tasks such as chip design and e-commerce operations. In terms of multi-modality, it offers advanced vision productivity, creating searchable, traceable, and interactive knowledge structures from various media like graphics, documents, and videos.

QwenWork – On Aug 3, BABA also announced the launch of public testing for QwenWork ("千问办公"), an enterprise-level agent product designed to boost office productivity for both individuals and corporations. Integrating QoderWork, MuleRun, and Wukong, it handles tasks like document generation, data analytics, audio-video generation, and web application creation for various office scenarios. QwenWork is currently integrated with DingTalk IM and aims for full connectivity with enterprise databases and workflows in the future.

Open-sourcing and pricing – Alibaba will open-source the model weights for Qwen3.8-Max and Qwen3.8-27B next week, with API services available on the Qwen AI Platform. Pricing is set at Rmb12/36 per million input/output tokens in China and a competitive US\$2/6 per million input/output tokens for overseas developers. Additionally, a limited Token Plan is available for individual users until August 5th.

DeepSeek-V4-Flash-0731 – On July 31, DeepSeek released the official version of DeepSeek-V4-Flash-0731, featuring substantially enhanced agentic capabilities and a speculative decoding module. This efficiency-focused MoE model has 284B total parameters (13B active) and a 1M-token context window, designed for fast inference and high-throughput use cases while maintaining strong reasoning and coding skills. Despite a smaller active parameter count, it shows significant performance gains over its preview versions on key benchmarks like Terminal Bench 2.1, NL2Repo, and AutomationBench, closing the gap with leading frontier models.

Figure 1. Key benchmark comparison for Qwen3.8-Max

Benchmark	Context	Qwen3.8-Max	Qwen3.7-Max	Qwen3.7-Plus	Gemini 3.1 Pro	GPT 5.6 Sol (Max)
SWE-Pro	Software Engineering	67.7	60.6	-	-	64.6
Terminal-Bench 2.1	Terminal Agent	86.6	74.5	-	-	88.8
PaperBench	Research Reproduction	93.0	64.8	-	-	90.5
FrontierSWE	Software Engineering	73.5	40.7	-	-	-
QwenReactBench	React Frontend	1724	1538	-	-	1564
CoWorkBench	Professional Cowork Tasks	74.8	64.6	-	-	71.5
JobBench	Agents Empower Human Workflows	53.4	31.3	-	-	45.4
Agents' Last Exam	Long-Horizon Valuable Tasks	52.4	31.1	-	-	53.6
BabyVision (w/o & w/ python)	Visual Reasoning	82.0 / 91.3	64.7 / 70.4	-	55.9 / 68.3	65.5 / 88.9
CharXiv (RQ) (w/o & w/ python)	Chart Reasoning	88.4 / 93.5	85.8 / 85.9	-	84.4 / 89.9	85.1 / 89.1
ERQA	Embodied Reasoning	77.8	-	69.8	68.0	70.0
PerceptionBench	Visual Perception	63.5	-	51.1	56.2	59.7
LVBench	Long Video Understanding	81.8	-	76.2	75.1	78.8
Vision2Web	Visual Web Development	69.0	-	42.1	-	62.1
MobileWorld	Agentic Mobile Use	77.8	-	51.2	58.1	76.9
OSWorld-Verified	Agentic Computer Use	86.1	-	73.3	76.2	83.2

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

Source: Company Reports, Citi Research



Alibaba Group Holding

Valuation

Our target price for Alibaba N-shares of US\$192 is based on 10x P/E on FY2027E Ecommerce group net profit, 7.0x P/S on FY2027E Cloud Intelligence Group revenue, 1.2x P/S on AIDC Group 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\$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.