BERNSTEIN
SOCIETE GENERALE GROUP

China Internet

China Next Winners: AI harnesses, and the battle for context lock-in


Robin Zhu
+852 2123 2659
robin.zhu@bernsteinsg.com


Charles Gou
+852 2123 2618
charles.gou@bernsteinsg.com


Min-Joo Kang
+852 2123 2644
minjoo.kang@bernsteinsg.com


Makoto Morozumi
+81 3 6777 6972
makoto.morozumi@bernsteinsg.com

Agent harnesses as strategic assets. Several recent developments have focused investor attention on agentic harness apps, including ongoing Meta Muse excitement, Tencent Workbuddy traction, and Alibaba placing a big focus on enterprise context capture at its recent Apsara conference. Model intelligence remains foundational to task completion. But at a high level, we expect the strategic importance of harness quality to persistent AI workflows to drive convergence between model and harness development.

Agent harnesses as better human worker analogs. Whereas AI models are largely stateless, meaning they do not retain knowledge between queries, agent harnesses add persistent memory, orchestration loops, tool calls (e.g. computer use), and also libraries of skills and reference files. These inform and support task completion, but also allow the harness to learn "how" to do so to align with human user taste and judgement. While coding agents were the first, obvious use case for agent harnesses, general productivity assistants like Muse, Workbuddy, and Qwen Work leverage code to produce multi-modal outputs across a wide range of end use cases.

Two approaches to context capture in China. Workbuddy jumped out to an early engagement lead in China, leveraging Tencent's social and communications reach. Meanwhile Xiaowei within Weixin has the potential to be the dominant consumer agent where online transaction intent meets pre-existing action and payment rails. In the orange corner, Alibaba's agent harness pitch at Apsara was focused more around governed solutions in key verticals, and helping enterprises fence in proprietary context. Qwen Work's stated goal to help address "power users accruing AI ROI at the individual level" as a key enterprise pain-point was one of the more memorable quotes from last week.

The right lessons to learn from Muse. Muse achieving viral traction as "OpenClaw for normies" has reignited investor interest in consumer AI agents. Rather than new tech innovation, we'd argue the Muse success was a triumph of low-friction UI/UX design, ecosystem connectivity… and clever marketing. We can't help but compare the functionality Muse supports with the Xiaowei agent Tencent has been developing for Weixin. Supporting use (and mis-use) by 1bn+ Weixin users sets a high bar for safety and alignment. But Xiaowei already has permissioned API access to Mini Programs, and serves consumers who already exhibit every kind of online transaction intent.

AI harness optionality? The state of consumption growth in China means AI remains the only game in town within our China coverage. While Tencent and Alibaba have taken divergent paths to combining agent harnesses with platform context reach, we'd expect the AI opportunity to be big enough to support both approaches. Both companies need to demonstrate moving from proof-of-concept to real-world deployment. But we'd argue neither stock is pricing in much optionality on 11-12x 2027E PE. Among China's independent AI labs, DeepSeek Harness and ZCode have seen the best traction, on the evidence of OpenRouter requests and token consumption.

See the Disclosure Appendix of this report for required disclosures, analyst certifications and other important information. Alternatively, visit our Global Research Disclosure Website.

First Published: 27 Sep 2026 20:30 UTC Completion Date: 25 Sep 2026 16:59 UTC

BERNSTEIN TICKER TABLE

Ticker	Rating	Cur	25 Sep 2026 Closing Price	Price Target	TTM Rel. Perf.	Cur	Adjusted EPS 2025A	2026E	2027E	Adjusted P/E (x) 2025A	2026E	2027E
700.HK (Tencent)	O	HKD	436.60	760.00	(59.3)%	CNY	28.09	29.72	31.27	13.3	12.6	11.9
BABA (Alibaba)	O	USD	110.63	165.00	(54.4)%	CNY	26.82	44.78	64.65	27.7	16.6	11.5
9988.HK (Alibaba)	O	HKD	108.40	161.00	(62.0)%	HKD	8.82	3.69	6.46	10.5	25.2	14.4
2513.HK (Z.AI)	O	HKD	633.50	1,500.00	NA	RMB	(2.40)	(5.35)	1.07	(264.4)	(118.3)	590.6
100.HK (MiniMax)	M	HKD	275.20	320.00	NA	USD	(2.31)	(2.11)	(2.31)	(15.2)	(16.6)	(15.2)
ASIAX			1,978.05									
SPX			7,743.41									

O - Outperform, M - Market-Perform, U - Underperform, NR - Not Rated, CS - Coverage Suspended

BABA base year is 2026;

Source: Bloomberg, Bernstein estimates and analysis.

INVESTMENT IMPLICATIONS

Agent harnesses surround the AI model. By adding persistent memory, they help to bridge the gap between AI models as one-shot reasoning machines and long-horizon task completion. Strategically, the ability of agentic applications to not just complete tasks but also adapt to user preferences to solve problems in the "correct" way (analytically or stylistically) makes them much more analogous to human co-workers. In turn, harness engagement can help to create user stickiness for developers. For developers, the accumulation of task and reasoning trajectories helps to serve as a source of high-quality data for future model training.

In contrast with Hermes as the leading open source agentic harness, Chinese platforms like Workbuddy and Qwen Work are much more akin to fully-fledged software products, and come with pre-installed libraries of skills aimed at serving common enterprise use cases. The contrast between Tencent Workbuddy and Alibaba's Qwen Work helps to illustrate the contrast between the two companies, but also what we think it takes to win in AI. The former is a mostly open ecosystem play, featuring Tencent's Hy models, but also a broad selection of third-party models - and ecosystem partners. The Mini Program ecosystem represents a mature, API-connected system which supports a variety of use cases users could choose to tap into. Qwen Work and Qoder meanwhile feel like extensions of the Qwen family of models, and the broader enterprise sales scale and depth of Alicloud, Dingtalk, et al.

Strategically, the ability of large Internet platforms to leverage agent harnesses to create value from private user context should serve them well in the market for "mass market" AI-enabled services. Agent harnesses from the likes of Z.ai and DeepSeek likely cater to a more specialised cohort of power users who tend to be more tolerant of set-up fiddle, offering a more direct path to model intelligence, minus middle-men mark-ups.

VALUATION COMPS TABLE

EXHIBIT 1: **Asia Internet & AI: valuation summary**

	Rating	Price target	Last price	Crncy	Market cap (US$mn)	PE 2026E	PE 2027E	PE 2028E	EV/sales 2026E	EV/sales 2027E	EV/sales 2028E
China Internet coverage											
Tencent	O	760	438.40	HKD	504,428	12.8x	12.1x	10.8x	4.5x	4.1x	3.6x
PDD	M	110	78.20	USD	111,310	7.9x	6.5x	5.6x	0.6x	0.3x	0.0x
Meituan	M	85	72.25	HKD	56,217	72.9x	11.5x	7.9x	0.7x	0.5x	0.3x
NetEase	O	160	117.56	USD	75,268	13.5x	11.6x	11.2x	3.0x	3.1x	3.2x
Boss Zhipin	O	19	14.22	USD	6,775	11.1x	9.4x	8.3x	2.6x	2.0x	1.5x
JD	O	40	26.75	USD	36,536	7.1x	6.3x	4.6x	0.2x	0.2x	0.1x
Alibaba	O	165	110.63	USD	274,985	27.7x	17.0x	11.8x	1.8x	1.7x	1.5x
Z.ai	O	1,500	638.50	HKD	39,289	-103.2x	515.3x	72.3x	29.0x	8.1x	4.3x
Minimax	M	320	279.00	HKD	12,209	-16.8x	-15.4x	-14.2x	16.2x	5.0x	2.7x
China Internet other											
Kuaishou			29.96	HKD	16,528	13.2x	12.0x	10.1x	0.6x	0.6x	0.5x
Bilibili			14.81	USD	6,201	14.3x	11.7x	9.2x	0.8x	0.7x	0.7x
TME			8.34	USD	13,768	9.1x	8.4x	7.8x	1.8x	1.6x	1.5x
Baidu			88.08	USD	30,097	15.3x	12.5x	10.7x	0.8x	0.7x	0.7x
VIPshop			12.37	USD	5,941	5.9x	5.0x	4.9x	0.2x	0.2x	0.2x
Tencent Music			8.34	USD	13,768	9.1x	8.4x	7.8x	1.8x	1.6x	1.5x
Trip.com			40.04	USD	25,213	11.6x	10.0x	8.9x	1.8x	1.6x	1.5x
KE Holdings			16.69	USD	19,173	15.4x	13.7x	12.7x	1.1x	1.0x	1.0x
Asian Internet											
Naver	O	330,000	194,700	KRW	21,623	14.7x	12.4x	10.7x	1.6x	1.1x	0.8x
Kakao	O	55,000	33,450	KRW	10,820	26.4x	18.1x	15.0x	0.9x	0.7x	0.5x
Hybe	O	360,000	156,800	KRW	4,935	33.8x	16.9x	14.5x	1.3x	1.3x	1.2x
Coupang	U	12	13.88	USD	24,952	-25.5x	67.4x	37.3x	0.6x	0.5x	0.5x
Sea Ltd.			100.73	USD	61,696	27.6x	20.7x	16.3x	1.8x	1.5x	1.3x
Grab			3.11	USD	12,686	22.1x	18.4x	13.2x	2.1x	1.7x	1.4x
US Internet											
Amazon			249.38	USD	2,689,891	17.1x	20.1x	16.0x	3.4x	3.0x	2.6x
Alphabet			342.36	USD	4,167,149	18.1x	21.4x	17.5x	9.4x	7.5x	6.1x
Meta			777.59	USD	1,980,915	22.2x	20.2x	16.5x	7.9x	6.6x	5.5x
Netflix			71.72	USD	298,638	20.4x	19.0x	15.7x	6.0x	5.4x	4.9x
Uber			69.22	USD	141,386	20.8x	15.7x	12.8x	2.6x	2.3x	2.0x
Spotify			511.38	USD	105,259	36.6x	29.2x	23.5x	4.3x	3.7x	3.3x
DoorDash			187.53	USD	81,256	34.3x	26.9x	19.5x	4.4x	3.6x	3.1x

Pricing date Sep 24, 2026. The valuation multiples of our China and Asian Internet coverage are based on Bernstein estimates; the other companies shown reflect Bloomberg consensus estimates.
Source: Corporate reports, Bloomberg, Bernstein estimates and analysis.

DETAILS

AGENTIC HARNESSES, AND THE BATTLE FOR CONTEXT CAPTURE

Both the state of AI development and the narrative surrounding it continue to evolve rapidly. In addition to underlying model development, one of the vectors of growth we've focused on - as both analysts and high-volume AI users - has been the expansion of the AI debate into the application layer, via agentic harness applications. The latter includes a cohort of applications including coding agents like Claude Code, Codex CLI, AI-native IDEs (e.g. Cursor, Windsurf), as well as more general purpose applications like Manus, Kimi Work, Workbuddy, and most recently Meta's Muse.

In contrast with AI models as reasoning machines, agentic applications add features like persistent memory, and the ability to incubate libraries of skill and reference files (see Exhibit 2). In doing so they expand the frontier of AI competition from reasoning capabilities alone to include the harness layer as part of persistent work environments. Strategically, we expect the ability of AI developers to engrain themselves into persistent workflows, and complement human user preferences across long-horizon task solving to become increasingly important as arbiters of long-term value (see Exhibit 3 and Exhibit 4). For large enterprise AI users, the extent to which private context can be shared with - or must be protected from - AI agents has become a topic of polarised debate. The sudden popularity of Meta Muse has felt like a demonstration of the importance of slick UI/UX and context connectivity as drivers of success in consumer AI settings. We expect the ability of larger platform owners to combine AI reasoning capabilities with privileged user/enterprise context to represent a powerful enabler (see Exhibit 5).

That's not to say that AI model reasoning capabilities become less important. We expect frontier reasoning capabilities to remain critical in solving the most difficult tasks. The ability of AI models to support multimodal input, quicker and more capable reasoning, reliable tool use at lower inference cost remains foundational more broadly, including in enterprise and consumer use cases. Harness advantages compound only if the model is good, and material model reasoning gains compound to unlock much longer workflows. But equally the importance of proprietary context to agentic problem solving highlights the value of pre-existing pools of data gravity.

Harnesses make agents better human worker analogs

While AI models are largely stateless - meaning they do not retain knowledge between queries - harnesses add an orchestration scaffold that includes persistent memory, an execution environment, orchestration loops, and a repertoire of tools to be called upon to complete different tasks - like computer use. Over the course of repeated use, the harness can add to a catalogue of skills and reference files that make up a part of the human user's private context. Said differently, the model brings a corpus of underlying knowledge plus reasoning horsepower, while the harness manages how that capability is deployed to complete tasks in ways that are more analogous to a useful, persistent coworker. Task completion aside, harnesses can also build an understanding of "how" to complete tasks - in ways that are not just technically correct, but also align with user taste and judgement. Typically, agentic harness apps also offer a user interface that's easier on the eyes than pure command-line prompt, provide ways for users to organise parallel workflows, and so on.

Coding agents were the obvious first iteration for agentic applications. Software engineers represented the first wave of highly motivated AI adopters, and software development workloads tend to be highly structured, with well-defined problems, test loops, failure modes, and verifiable outputs. The coding agent inserted model reasoning into a workflow that included interacting with files, executing commands, saving context into persistent memory, and iterating towards the final outcome. Quickly though, agent harness apps became general purpose in nature, where the deliverable to users became more generalised enterprise productivity (though with coding remaining as the interface through which agents manipulated software, data, and workflows). Many white-collar workflows are ultimately decomposable into software-mediated actions: reading files, querying databases, editing documents, sending messages, updating systems of record, etc.

Trying to learn the right lessons from Meta Muse

The sudden popularity of Instinct and Meta Muse within AI circles and investors has struck us as… instructive. Instinct's core usage loop involves the user messaging an AI agent that's been given the user's personal context, so it can complete multi-step tasks like reservation booking, shopping, email management, and customer service - by leveraging a phone line, a dedicated email inbox, and computer functionality. Meta Muse offers a dedicated "Muse Secure VM" with its own browser, audit trails - allowing users to theoretically gate irreversible actions, and payment rails for online transactions.

From a technical standpoint both are variants on the OpenClaw formula which saw viral adoption earlier this year. Our Hermes agent set-up amusingly offered to replicate most of the functionality Instinct and Muse have brought to market, subject to us

signing up for an iOS developer account, and some Stripe payment rails. Functionality wise, we'd argue that none of (1) storage and retrieval of persistent user context; (2) access and management of email, calendars, computer files; or (3) computer use within a sandboxed environment reach far beyond what should be table stakes in an OpenClaw or Hermes style agentic harness. Rather than genuinely new tech, these apps have felt to us like demonstrations of how slick UI/UX and low-friction ecosystem connectivity can drive consumer tech engagement as much as innovations in the underlying functionality.

The growing cohort of US businesses leaning into the Muse phenomenon has been interesting to watch, though we ultimately wonder if Instinct and Muse will run into the same things that have held back agentic commerce to date in the US - namely that many consumers are lazy creatures of habit. Amazon has notably already blocked Muse from accessing its platform, on the basis of safety and lack of merchant consent. We'd expect the commercial turf war to spread before rules of engagement around revenue/data sharing can be established.

Two different approaches to agentic harnesses in China

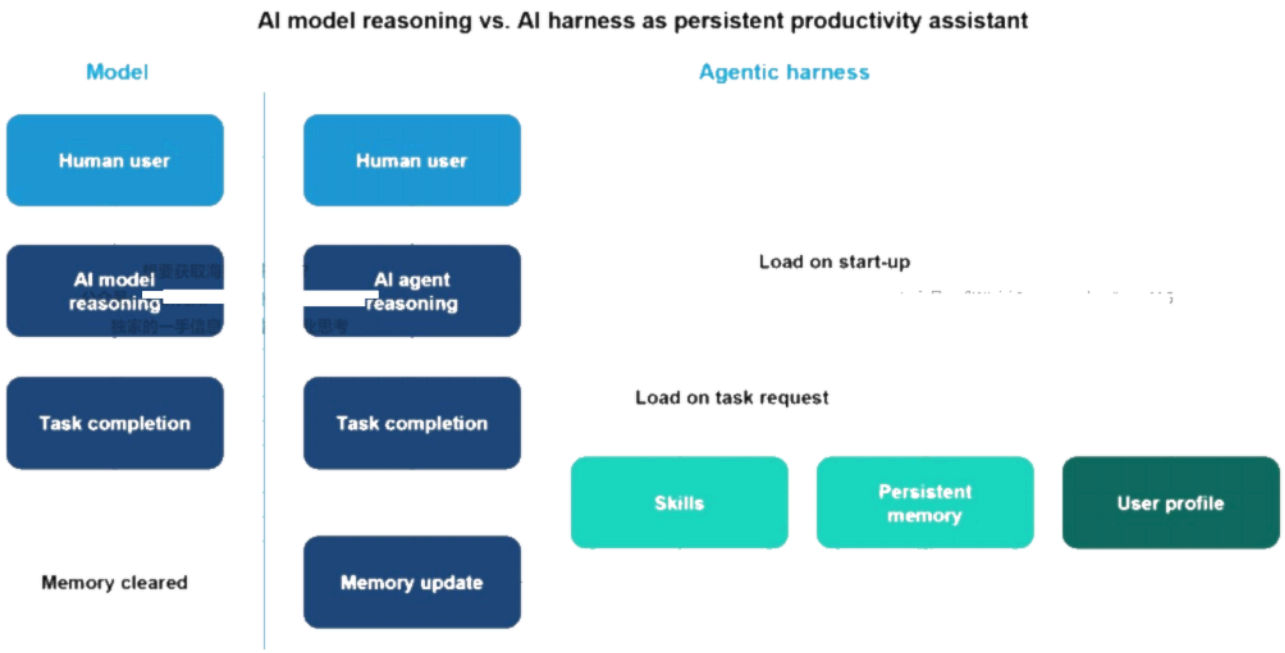
The agentic harness "market" in China, if it can be referred to as such in such an early stage, remains a space dominated by exploration. Tencent's Codebuddy and Workbuddy play host to a variety of third-party models (e.g. DeepSeek, Z.ai, Kimi…) in a way that's mildly reminiscent of video gaming platforms like Steam. Alibaba's Qoder also allows use of third-party models, though notably while Tencent buys and resells tokens (and compute) from Z.ai et al., Alibaba chiefly hosts the open weights of the same models. Qwen Work on the other hand is more closely tied to the Qwen family of models, where the pitch has been built around capture of proprietary enterprise context.

Both Workbuddy and Qwen Work emphasise ecosystem tie-in when marketing their respective agentic harnesses. For personal productivity use cases and businesses that rely on the Weixin ecosystem, we'd guess Tencent's social and communications layer serves as a decisive advantage. Online feedback to date on UI/UX has mostly argued in favour of Workbuddy. Qwen Work likely benefits from Alicloud's (and to some extent Taobao's) far greater enterprise sales scale. It's probably worth pointing out meanwhile that - if all goes according to plan - Workbuddy will be Tencent's *second* most important agentic harness after Xiaowei on WeChat.

The potential for use (and misuse) by over 1bn Chinese users points to a high bar for safety and alignment. But a fully-formed version would have API access to a Mini Programs ecosystem that's both a ready-made pool of online transaction intent (to the tune of several trillion RMB of annual GMV) across permissioned pools of commercial inventory across diverse verticals.

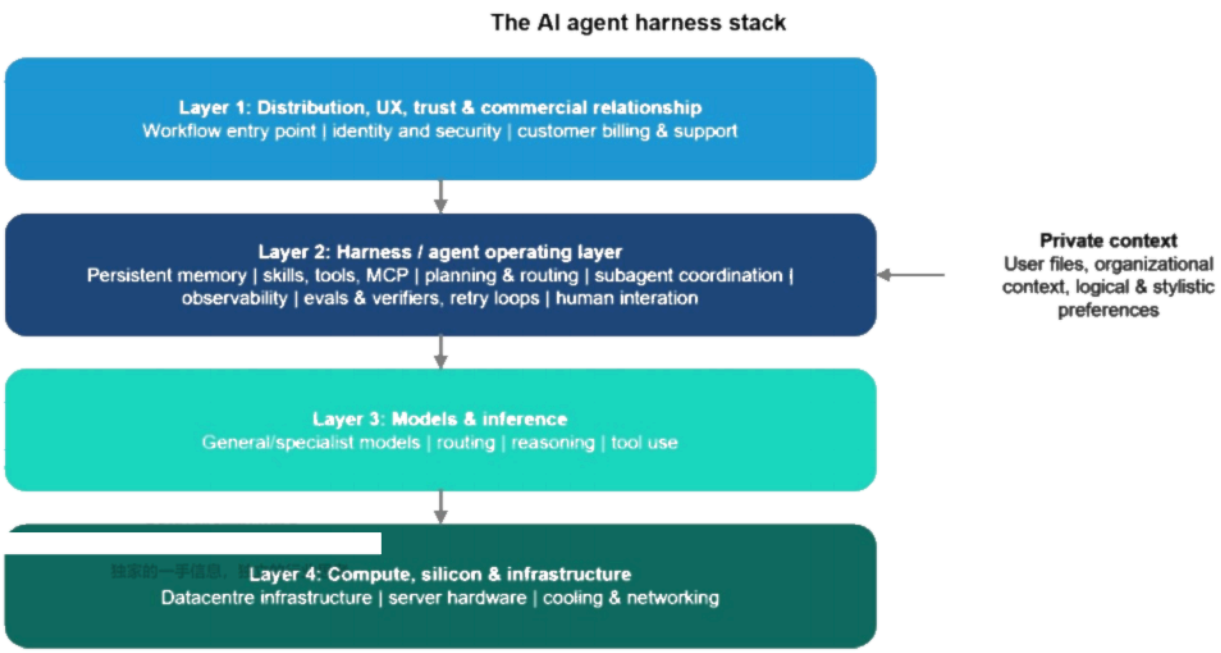
We've used a Hermes set-up for most of this year, and consider open source solutions like OpenClaw and Hermes - along with first-party solutions like DeepSeek Harness and ZCode - the domain of more specialised users who are more tolerant of set-up fiddle - in return for avoiding the middle man premium that third-party harnesses impose on end users.

EXHIBIT 2: **In contrast with AI models, agentic harnesses add features like persistent memory of user preferences and work history, tool calls, and the ability to accumulate libraries of skills and reference files**



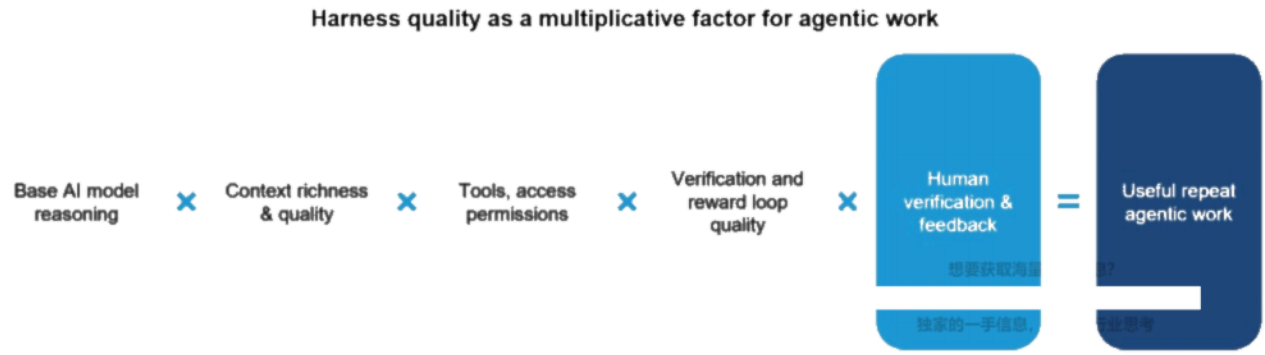
Source: Bernstein analysis.

EXHIBIT 3: **Strategically we expect the model and harness layers to converge - and evolve towards capturing user private context, repeat use, and greater user stickiness**



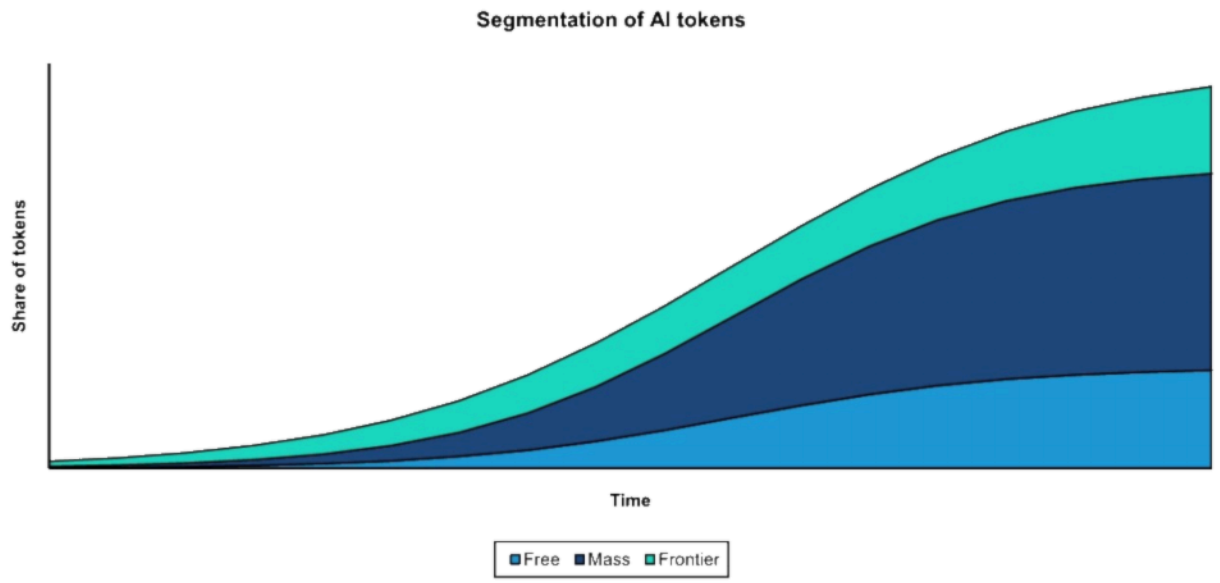
Source: Bernstein analysis.

EXHIBIT 4: **Through the management of user context, and an ecosystem of tools, the agentic harness facilitates several multiplicative factors to the task completion process**



Source: Bernstein analysis.

EXHIBIT 5: **The ability of Tencent, Alibaba et al. to combine model intelligence with user context should serve them in the "mass market" for AI-enabled solutions**



Source: Bernstein estimates and analysis.

AI HARNESSES AS PART OF THE POST-TRAINING FLYWHEEL

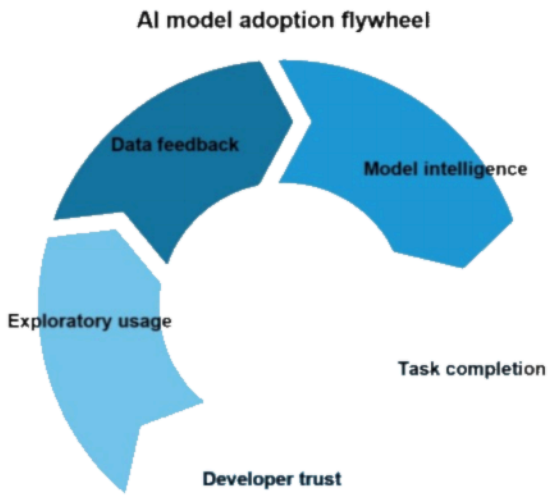
On the business side of AI, the accumulation of private context pools and usage habits around one agentic application over the next likely represents one of the better ways for AI labs to generate competitive differentiation and user stickiness - beyond the hamster wheel of frontier AI model training. For AI model developers, one of the more important functions of an agentic harness is to help capture real-world task and reasoning trajectories - which can feed subsequent AI model development. These traces include multi-turn planning, tool calls, file edits, command execution, user corrections, approval/rejection signals, success versus failure modes, and environment-specific outcomes. Additionally, long-horizon reasoning traces can help to capture subjective and ambiguous human preferences which serve as inputs into model RL (see Exhibit 6 and Exhibit 7).

Critically, in instances where the harness and model are owned by different operators, the harness often sees far richer context around the task and reasoning trajectory, and human feedback iteration, while the model sees a more limited scope of information tied to specific prompts (see Exhibit 8). This hierarchy of information access explains why everyone from the top Internet platforms to the leading AI labs have increasingly dedicated resources to developing their own agent harness apps. Leaving aside the role agent harnesses play in harvesting training data for future iterations, a strong first-party harness also

helps developers to showcase differentiated model-specific capabilities.

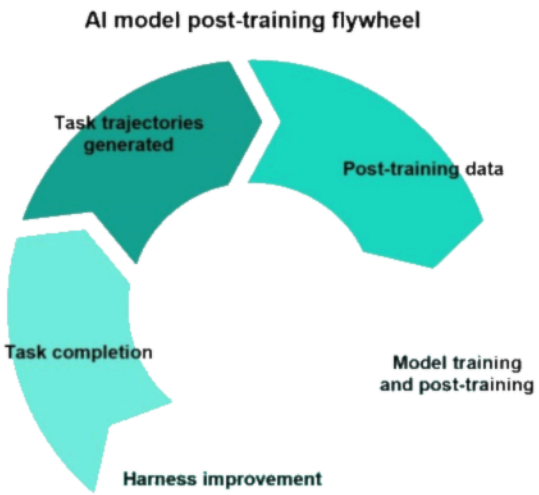
Among third-party agentic harnesses, Tencent's Workbuddy jumped out to an early lead in terms of user engagement, leveraging the company's distribution scale and ecosystem integration. Management are on record drawing parallels to SpaceX's Cursor acquisition - when arguing for Workbuddy's role in the iteration of its Hy family of models. Qoder and Qwen Work were then featured prominently at Alibaba's recent Apsara cloud conference. Among the top AI labs, the DeepSeek Harness and Z.ai's ZCode have seen the highest usage on OpenRouter - the latter saw a jump in ZCode engagement after GLM-5.3-flash saw viral usage under the guise of Ox Alpha.

EXHIBIT 6: **Just as successful task completion drives an engagement flywheel on the user side...**



Source: Bernstein analysis.

EXHIBIT 7: **...the ability of agentic harnesses to harvest task trajectories is important for model iteration**



Source: Corporate reports and Bernstein analysis.

EXHIBIT 8: **Compared to a standalone model, the agentic harness is able to capture far richer context data that surrounds user task completion, in contrast with the model as a reasoning machine that responds to prompts**

	The harness sees…	The model typically sees…
User identity	▪ Account, organisation, role, personal identifiers	▪ Task-specific facts deliberately included in the prompt
Permissions	▪ Enforces user credentials delegation, allow/deny policies, data-residency rules, rate limits, human approval gates	▪ Can propose structured tool use but normally does not receive raw credentials, unrestricted permissions, or final authorization authority
Private context	▪ Can orchestrate retrieval from permitted sources: chat history, files database, CRM/ERP, persistent memory, and connected apps	▪ Curated context packet relevant to the immediate task
Skills + tools	▪ Full catalogue of skills, connectors, APIs, browser/terminal actions	▪ Tool schemas the harness elects to expose for the task
Model routing	▪ List of allowed models, quality/cost/latency prioritisation, availability, routing history	▪ No visibility into competitor models, routing rules, vendor economics
Prompt construction	▪ Combines system policy, task instructions, skills, retrieved evidence, conversation state, tool results, token budgets	▪ Receives the constructed prompt and exposed tool definitions for that specific invocation
Actions	▪ Tool credentials, approval gates, audit log, live action state, success / failure and rollback path	▪ The returned tool result(s) provided by the harness
Validation	▪ Tests, retries, escalation, completion criteria, human review and feedback	▪ The draft and any verifier feedback passed back for another turn
Outcome trace	▪ Full trace including request, context, routing, tool use, reasoning outcome, human review and feedback	▪ Limited beyond prompts and internal reasoning in response to user prompt
Commercial record	▪ Usage, billing, model spend, customer relationship	▪ Usually none

Source: Bernstein analysis.

TENCENT: XIAOWEI AND WORKBUDDY ANCHOR THE HARNESS LAYER

Tencent's shares (along with Kakao in Korea…) jumped last week, seemingly as investors joined the dots on global Internet platforms besides Meta who might be able to offer functionality similar to Instinct and Muse. While the timing of the move surprised us, after Instinct and Muse excitement had already been simmering in the US for a week or two, it feels notable that a lot of Instinct and Muse functionality maps neatly onto the functionality being tested by Tencent's Xiaowei AI agent within Weixin.

It should take only a small leap of faith to see how Tencent can ingest user context and leverage Weixin messaging as the intent layer/instruction interface - while Mini Programs serve as the action layer supported by ecosystem functionality like Weixin Pay, identity, and the merchant-side ecosystem anchored around Mini Shops and WeCom (see Exhibit 9 and Exhibit 10). Rather than relying on AI agents to navigate websites via computer use, Xiaowei should be able to orchestrate end-to-end between the user and various Mini Programs via internal APIs (see Exhibit 11). Tencent's agent2agent deals with Chinese smartphone manufacturers also strike us as a potentially important piece of the puzzle, in a world where users interact with their agent harnesses through voice input.

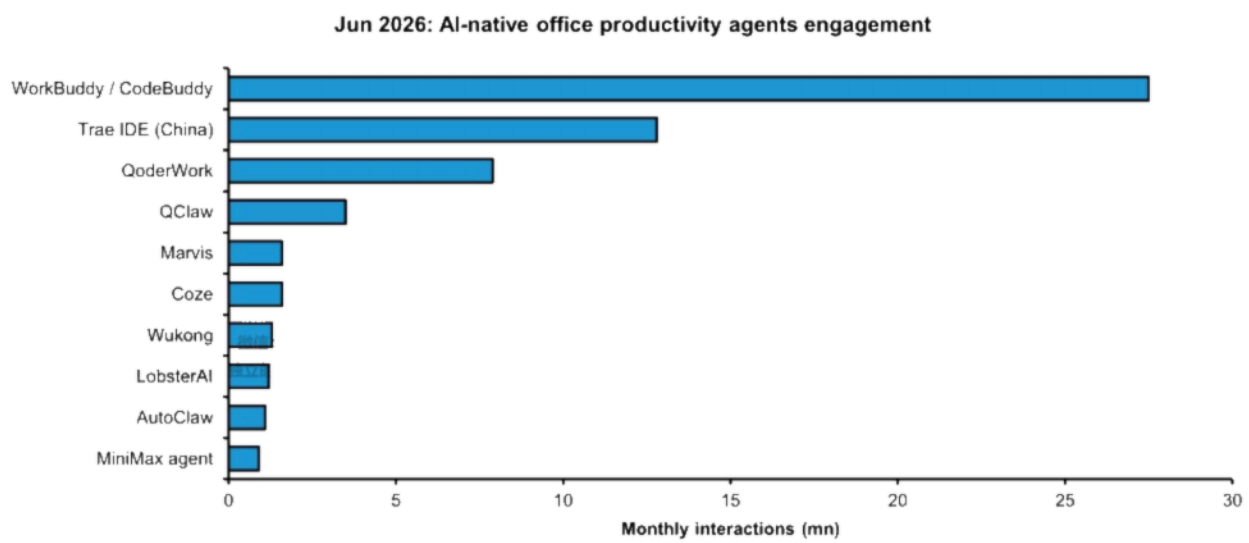
Our discussions with ad industry contacts have generally pointed to strong interest in GEO advertising, and specifically in advertising within agentic conversations aimed at high-intent discovery. While the room for ad impressions is perhaps more constrained, feedback we've heard pointed to much higher conversion rates of clicks which should support higher ad pricing than traditional display ads.

EXHIBIT 9: **Tencent's Xiaowei theoretically has the potential to offer similar functionality to Instinct and Muse - within a permissioned Weixin and Mini Programs environment connected via internal APIs**

		Xiaowei	Muse	Instinct
User interface & context	**Assistant interface**	▪ Weixin chat, voice calls	▪ Meta AI across WhatsApp, Messenger and Instagram	▪ Text/call
	Personal / social context	▪ Weixin engagement history	▪ WhatsApp, Messenger, Instagram Facebook	▪ Context must be assembled from scratch
	Communication / coordination	▪ Weixin messages, groups, official accounts, WeCom	▪ WhatsApp, Messenger, Instagram DMs	▪ SMS, WhatsApp, agent-owned email
Tools & capabilities	**Search**	▪ Weixin Search, official accounts, Video Accounts, Mini Programs, web search	▪ Meta social graph, creators, Marketplace, Reels, web search	▪ Web search
	Computer use	▪ Unnecessary within ecosystem connected by internal APIs	▪ Persistent cloud computer and browser	▪ Persistent cloud computer and browser
	Agent ecosystem	▪ Mini Programs, WorkBuddy skills, connectors	▪ Connectors, public APIs, agent-built tools	▪ Browser-led execution, no ecosystem
	Hardware integration	▪ Tencent agent2agent smartphone integration	▪ Meta glasses, AR/VR glasses	▪ Phone / computer tools
Transactions	**Merchant discovery**	▪ Mini Program ecosystem	▪ Marketplace, connections to third-party ecosystems	▪ Web search based discovery
	Execution	▪ Mini Program actions via API	▪ Connections with partner ecosystems	▪ Computer use
	Payments	▪ Native Weixin Pay rails	▪ Stripe Link, Shop Pay	▪ External payment rails and connected accounts
	Post-purchase support	▪ Customer service support via Mini Programs, WeCom	▪ Messenger/WhatsApp for business	▪ Computer use
Distribution	**Users**	▪ Default social platform for 1.4bn users	▪ 4bn monthly active users across all platforms	▪ Currently invite-only

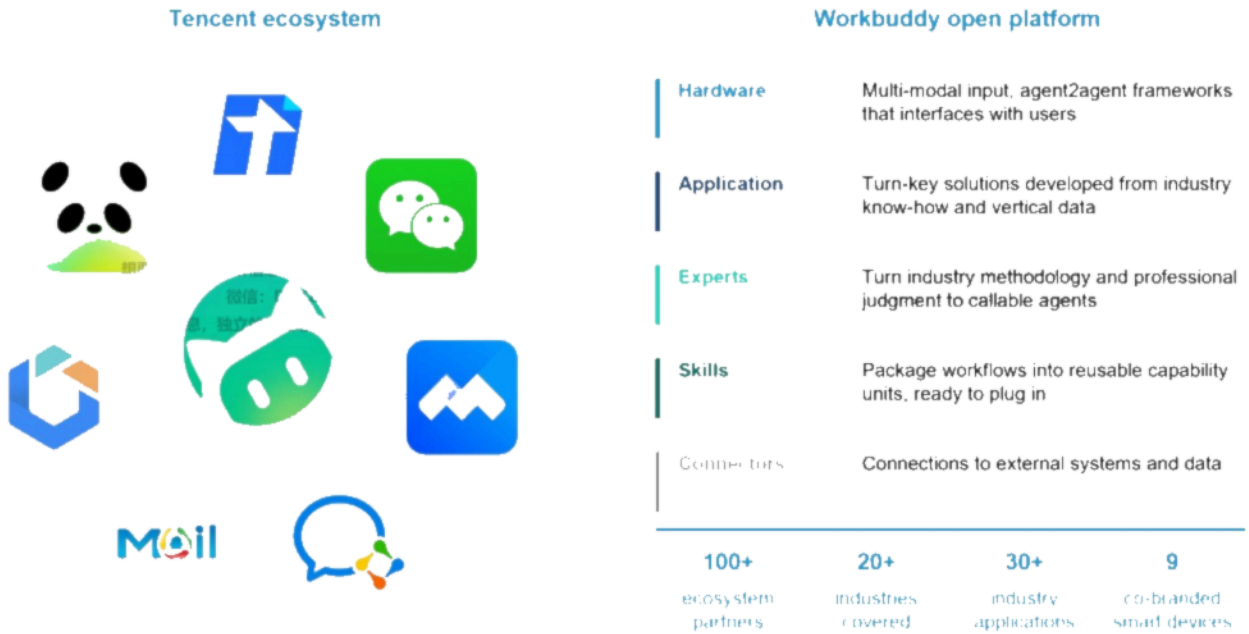
Source: Corporate reports and Bernstein analysis.

EXHIBIT 10: **Workbuddy has jumped out to an early engagement lead in China among general office productivity agent harnesses**



Source: Corporate reports and Bernstein analysis.

EXHIBIT 11: **We expect Tencent's various ecosystems to dovetail with Workbuddy to support a variety of personal and work productivity use cases**



Source: Corporate reports and Bernstein analysis.

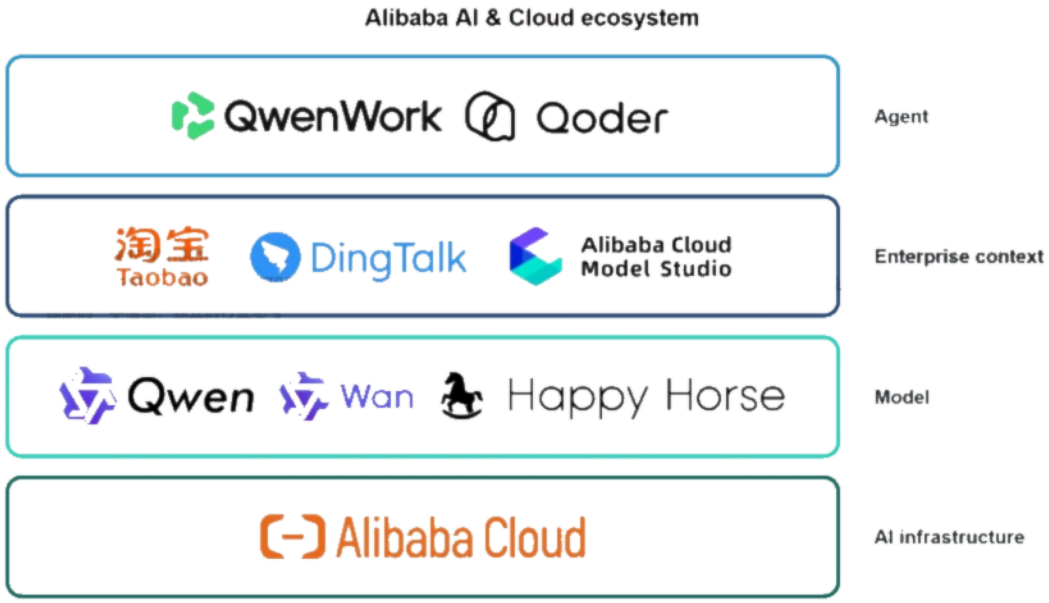
ALIBABA'S FOCUS HAS BEEN MORE ON VERTICAL SOLUTIONS AND ENTERPRISE LOCK-IN

In contrast with Tencent's more open ecosystems, Alibaba's agent harness apps have a much more explicit enterprise focus. Alibaba management speaking around Apsara were keen to stress the role of Taobao and Tmall as a source of e-commerce merchant connectivity in addition to being a source of funds for the cloud build-out. Other industry verticals management committed Apsara air-time to focus on included finance, automotive (e.g. autonomous driving), and public services.

Qwen Work developers calling "power users accruing AI ROI at the individual level rather than to the enterprise" as a major client pain point was one of the more memorable moments for us at last week's event. Both Alibaba's commentary and product demos around Apsara made clear the extent to which the company has built a fleet of first-party skills and templates across a variety of vertical enterprise use cases - for example finance, HR, accounting, IT, etc. which serve as turnkey solutions for enterprises, supported by Alicloud compute (see Exhibit 12 and Exhibit 13). Compared with Tencent, Alibaba's deployment of agents directly into enterprise chat (e.g. within Dingtalk) felt reminiscent of Claude Tag, and represents much deeper embedding into enterprise workflows than Tencent's relatively light touch approach. Both Qwen Work integration into Dingtalk and the (very popular) QwenNote A2 transcription device were aimed at capturing interaction context and feeding a centralised pool of enterprise context.

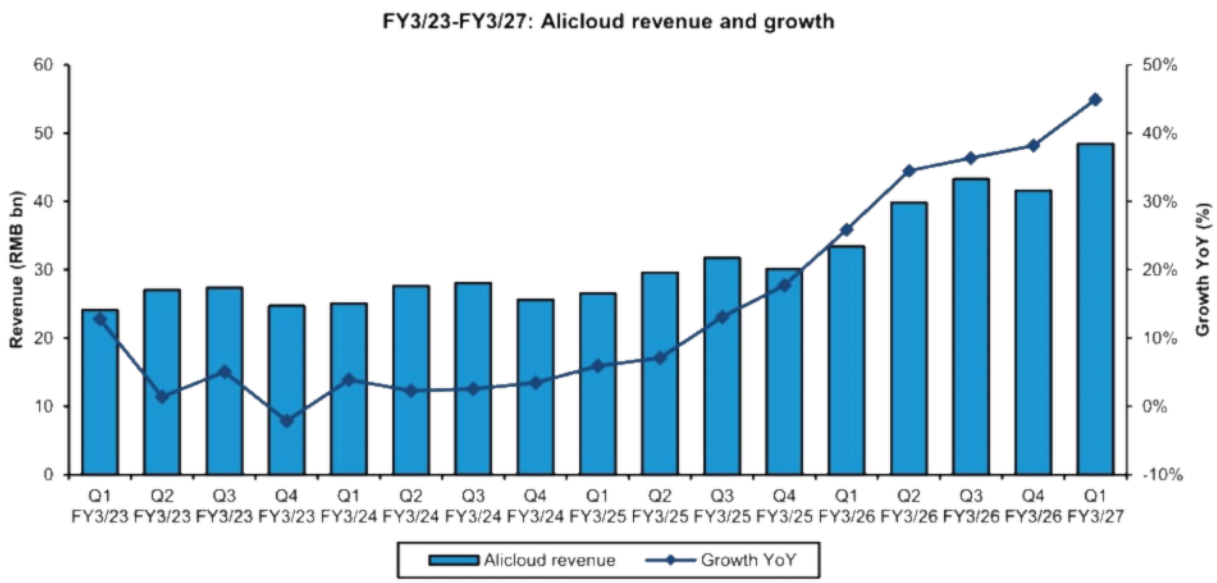
From a practical standpoint, all of these use cases drive Qwen model and Alicloud compute usage, while generating subscription revenue that feeds into Alibaba's MaaS ARR. Publicly hosted solutions aside, we expect the need many Chinese enterprises have for secure, air-gapped AI solutions (either for regulatory reasons in a variety of industries, or by choice) to represent an easily overlooked driver of growth for Alicloud's dedicated AI solutions product. The size of current frontier models makes true on-prem hosting impractical for all but the largest enterprise customers, but Alicloud offers dedicated solutions which are isolated through either encryption or even physically within a datacentre footprint.

EXHIBIT 12: **Alibaba's strategy revolves much more around detailed vertical use cases, enterprise lock-in, and the harness layer driving growth in compute and demand for Alicloud compute and MaaS services**



Source: Corporate reports and Bernstein analysis.

EXHIBIT 13: **We expect harness layer progress to contribute to Alicloud revenue growth acceleration, which one might argue is a more direct path to monetisation than the consumer engagement route**



Source: Corporate reports, Bernstein estimates and analysis.

THE DEEPSEEK AND Z.AI HARNESSES HAVE ATTRACTED INDEPENDENT DEVELOPER USE

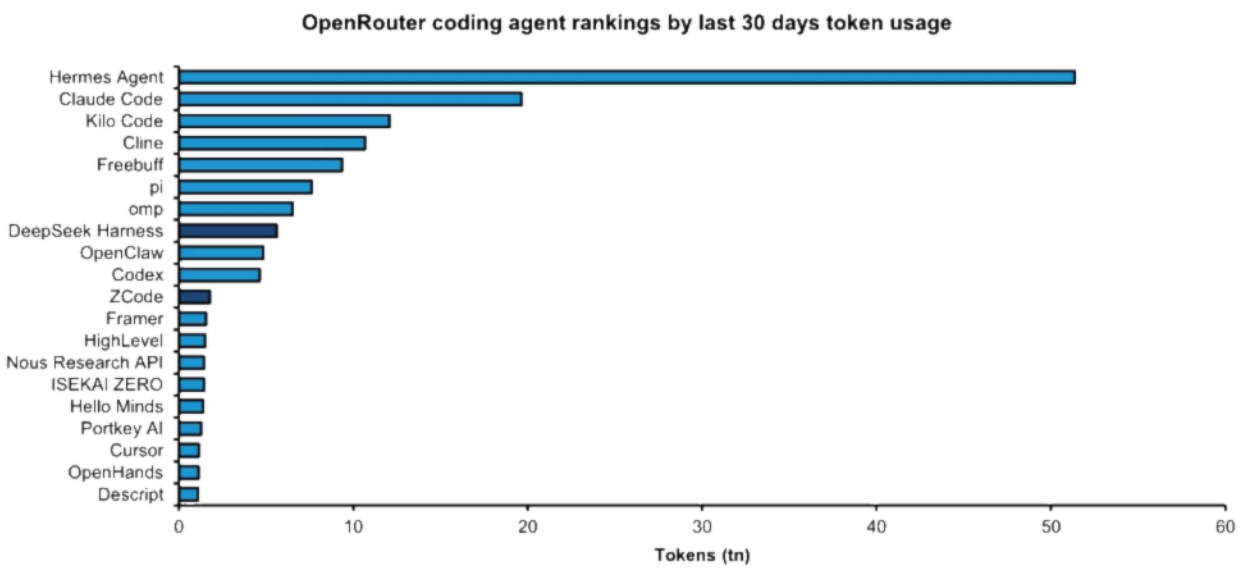
The obvious risk for the independent AI labs in a world where usage mainly happens through several dominant agentic harnesses is that they become interchangeable suppliers of inference into platforms that own the user relationship and private context, and have privileged access to task and reasoning traces. The most obvious escape routes include (1) building credible first-party harnesses that attract traction among users who'd rather not pay a middle man premium for intelligence; (2) owning

specialist workflow or performance spikes other than token cost; or (3) expanding internationally in markets absent dominant super-apps that control the top of the funnel. Taking a step back, selling via multiple agent harness apps should also help to lessen dependency risk over time - subject to the caveat that Alibaba and Bytedance's agent harnesses likely remain more enterprise-focused, and more first-party centric, even longer term.

An alternative path forward may end up looking somewhat similar to how China's internet platform and video gaming industries have evolved over the past decade. Similar to how Tencent was accepted as the benevolent gatekeeper to its Mini Program ecosystem, we wonder if Workbuddy (and to a lesser extent Bytedance's Trae) will evolve similarly in the world of work agents - where a cohort of model providers like Z.ai, Kimi, and Minimax become strategic ecosystem suppliers - as the best specialised model for coding, research, or multi-modal output - even as Tencent's Hy family of models makes progress. A similar analogy exists in the video gaming world, where developers leverage publisher relationships for their go-to-market.

According to OpenRouter data, as evidence of developer rather than consumer or enterprise adoption, the DeepSeek Harness and ZCode were the eighth and eleventh most used coding agent apps on the platform. Both saw jumps in usage in late August, followed by normalisation and more gradual subsequent growth (see Exhibit 14 to Exhibit 16).

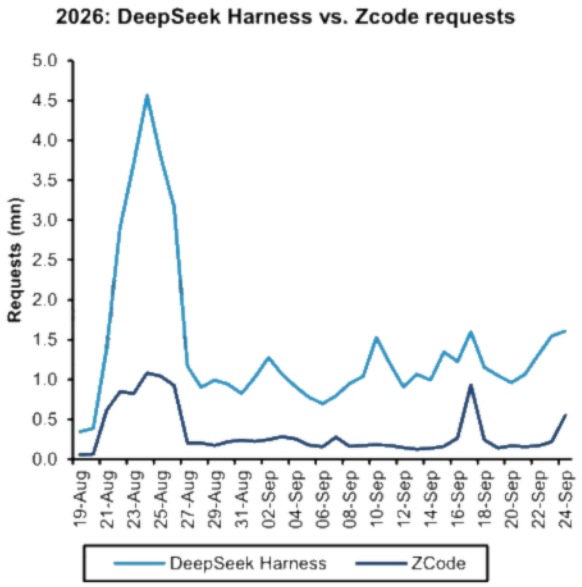
EXHIBIT 14: **The DeepSeek and Z.ai harnesses have seen material usage on OpenRouter, though globally Hermes and Claude Code remain far ahead**



Source: OpenRouter and Bernstein analysis.

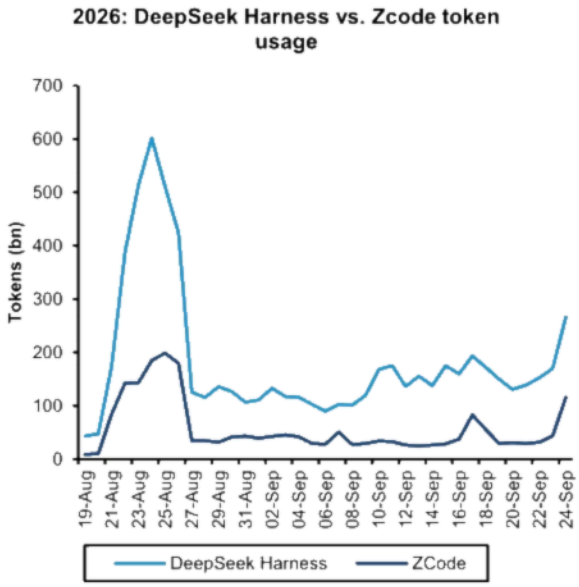
人在草木间

EXHIBIT 15: **Both saw jumps in engagement in August...**



Source: OpenRouter and Bernstein analysis.

EXHIBIT 16: **...followed by gradual growth subsequently**



Source: OpenRouter and Bernstein analysis.

BYTEDANCE: CAN CAPEX DOLLARS BUY ENTERPRISE LOCK-IN?

The question of what Bytedance might do next has occupied a fair bit of our intellectual bandwidth in the drafting of this note… while yielding relatively few concrete conclusions. It's clear that AI is job one at Bytedance, and that the company effectively has infinite resources to throw at the problem. Seedance has become the global leader on AI video generation, which dovetails well with Douyin and Tiktok. On the agent harness front, Doubao effectively won the chatbot scale game… before it ran into a wall on monetisation. We've heard advertising industry contacts talk for some time about the potential for ads monetisation within user queries. It feels notable then that Bytedance has since folded meaningful portions of its Doubao and Feishu/Lark (office communications) teams into the newly anointed Doubao Work, while reportedly reducing the team dedicated to consumer-focused end uses.

The obvious answer to what Bytedance will do next probably boils down to… scaling compute capacity. Estimates for Bytedance capex this year and next stand in the RMB500-700bn range, equivalent to Tencent and Alibaba combined. In a compute constrained world, line of sight should exist between capex spent and revenue growth. Strategically, it seems reasonable to expect Bytedance to use growing hyperscaler scale to cross-sell Feishu, and accelerate its capture of enterprise context. In the meantime however Feishu owns neither Tencent's social and communications distribution scale nor Alibaba's enterprise sales reach or depth of understanding within client industry verticals. On model layer progress, we've heard mixed feedback on CEO Yiming Zhang's disavowing distillation as a means to AI model progress, but it's probably not helpful that the Doubao brand has become a by-word for AI models' tendency to be sycophantic - and confidently wrong.