

Saturday, October 3, 2026

Tech Mornings • Sat, 03 Oct 2026 14:28:21 GMT • 3:48

Maya: Good morning, and welcome to the Tech Morning Briefing for Saturday, October third, twenty twenty-six. I'm Maya.

Devon: And I'm Devon. Happy Saturday. We have two new A I models competing on price, a very quiet week for Muse, and a security beat where the suspected culprits may be A I agents themselves.

Maya: Let's start with DeepSeek. The Chinese lab shipped DeepSeek V four point one Flash last month. It is a five hundred fifty-two billion parameter model, but it only activates about eight billion of those at a time, thanks to a new architecture that shrinks the memory it needs per token to roughly a quarter of the previous version.

Devon: And the pricing is the real headline. Thirty cents per million input tokens, one dollar twenty cents for output, and half that during off-peak hours. DeepSeek is so confident in it that since mid-September it has been quietly routing requests meant for its flagship Pro model to Flash instead, while charging Flash rates. When a lab undercuts its own premium model, you know the efficiency story is real.

Maya: Meanwhile, a startup called Inception released Mercury two point five, which takes a completely different route to speed. Instead of writing text one token at a time like most models, it generates many tokens at once and then refines them, a technique called diffusion.

Devon: The payoff is eleven hundred tokens per second on ordinary Nvidia graphics cards, at a launch price of four cents per million input tokens. For voice agents and search products where latency piles up every time the model is called, that is a very tempting number.

Maya: On the Muse front, it is a quiet week. No new Muse launch since the small business connectors on September twenty-ninth, and we checked both the product docs and this week's announcements. The Connect roadmap items, the real-time avatar, the Mac app, the smart glasses integration, and the Muse Charm pocket device, are all still slated for the coming months. When there is something genuinely new, we will say so.

Devon: Around the corner, watch the price war. Coinbase says it cut its A I spending

in the coming weeks.

Devon: Now the security news, and it is a busy one. In South Korea, Shinhan Bank says an attacker got into services used by its loan recruiters and exposed information on about twenty-five thousand customers, including names, phone numbers, annual income, and borrowing limits. Yonhap reports that investigators suspect the hackers used sophisticated A I agents to hunt for the weak spots.

Maya: Two other Korean banks were hit the same week. K B Kookmin says personal data from one hundred nineteen customers leaked, and Hana Bank says eighty-nine customers were affected. The financial regulator has started emergency on-site inspections and called the banks in for a meeting.

Devon: Over in health tech, iRhythm Holdings began notifying people on October second about a breach it first detected in June. Between June third and June eighth, an intruder got into some of its third-party hosted business applications and downloaded patient data, including names, addresses, insurance numbers, and device serial numbers.

Maya: And the China-linked ransomware crew called Warlock is back in the news. Over the past two months it has broken into a water utility, a telecom operator, a regional government body, and a university, all through old SharePoint vulnerabilities, with a focus on Portuguese and Spanish speaking countries across Europe, Africa, and Latin America.

Devon: That's the briefing for today. Stay patched, stay skeptical, and we will see you tomorrow morning.

Maya: See you tomorrow.