OpenRouter removes three AI providers after test prompts and decoy credentials leaked



OpenRouter dropped Together AI, Phala and Novita after Dragos found test prompts had leaked. Organisations that sent sensitive data or secrets through the platform should rotate credentials and review exposure.

- Dragos sent two million seeded test prompts through OpenRouter and saw data from at least three AI providers exposed.
- OpenRouter says its own systems were not breached; the leaks occurred at the provider level.
- Together AI, Phala and Novita were removed, and Atlas Cloud withdrew its models voluntarily pending security testing.
- ZDR eligibility was revoked for 125 endpoints, and potentially affected customers were notified.
- Providers advise users to rotate any secrets that may have been in the leaked requests.

OpenRouter, a broker that lets developers reach AI models from many companies through a single API key, has removed three inference providers after finding they had breached their stated data-handling policies.

What happened

Dragos, an operational technology security company, sent two million test prompts through the platform. Each held a unique tracking link or decoy credential. Dragos then saw activity showing data from at least three providers had been exposed, including attempts to use the decoy credentials against its controlled test systems and canary AWS accounts.

OpenRouter says its own systems were not breached. It removed Together AI, Phala and Novita, and revoked Zero Data Retention (ZDR) eligibility for 125 endpoints as a precaution. Atlas Cloud said it would withdraw its models voluntarily pending security testing after a third-party incident.

Provider explanations

- Together AI: an AI coding agent exposed a repository credential on GitHub, letting an unauthorized party download stored prompt samples from June and July.
- Phala: an attacker compromised a server handling requests routed through OpenRouter on October 21 and 22, possibly up to November 8, and obtained application credentials.
- Novita: a routing misconfiguration sent some MiniMax requests to MiniMax's official API between May 28 and November 11.

All four providers are tightening controls and data-deletion policies. OpenRouter will adapt the Dragos tests for continuing use and require validation before restoring ZDR eligibility.

Why it matters

Prompts often carry credentials, internal data or personal information. A broker model adds a chain of third parties whose handling is hard to see from the application side.

What to do now

1. Identify applications and teams that use OpenRouter or similar AI brokers, and which providers and models they route to.
2. Rotate API keys, tokens and any other secrets that may have appeared in prompts sent through the affected providers.
3. Review prompt logs for personal data and classified information, and assess PDPL notification duties if exposure is plausible.
4. Add AI brokers and model providers to third-party risk reviews, requiring evidence of data retention and deletion controls.

5. Prohibit credentials and sensitive data in prompts, and use secret-scanning on code and AI agent output.