← Blog

Wiz Research Uncovers Exposed DeepSeek Database Leaking Sensitive Information, Including Chat History

A publicly accessible database belonging to DeepSeek allowed full control over database operations, including the ability to access internal data. The exposure includes over a million lines of log streams with highly sensitive information.

Gal Nagli
January 29, 2025

3 minute read

Table of contents

Wiz Research has identified a publicly accessible ClickHouse database belonging to DeepSeek, which allows full control over database operations, including the ability to access internal data. The exposure includes over a million lines of log streams containing chat history, secret keys, backend details, and other highly sensitive information. The Wiz Research team immediately and responsibly disclosed the issue to DeepSeek, which promptly secured the exposure.

In this blog post, we will detail our discovery and also consider the broader

implications for the industry at large.

Executive Summary

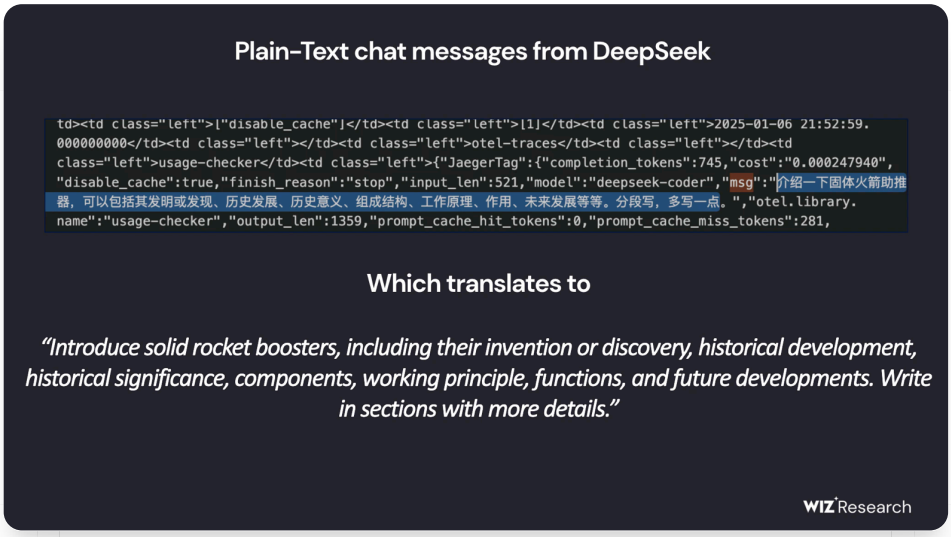
DeepSeek, a Chinese AI startup, has recently garnered significant media attention due to its groundbreaking AI models, particularly the DeepSeek-R1 reasoning model. This model rivals leading AI systems like OpenAI's o1 in performance and stands out for its cost-effectiveness and efficiency.

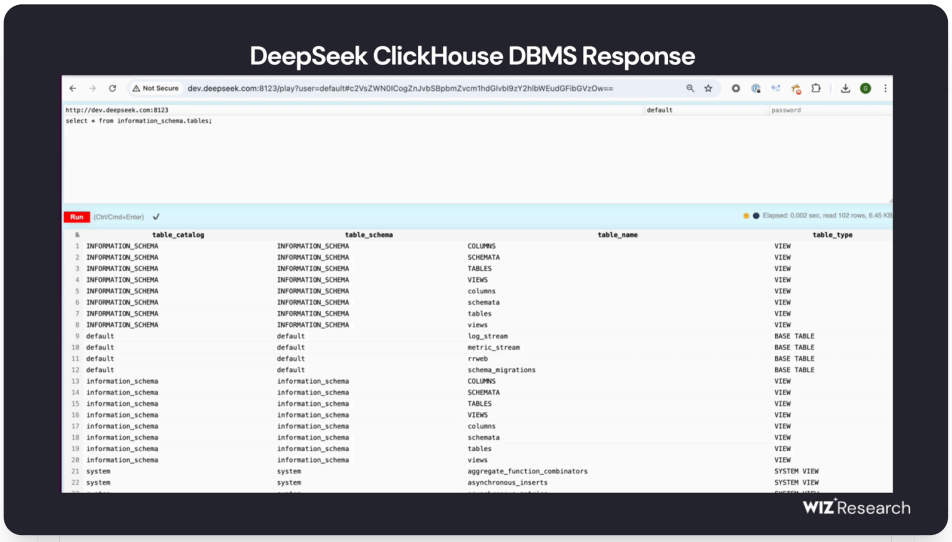
As DeepSeek made waves in the AI space, the Wiz Research team set out to assess its external security posture and identify any potential vulnerabilities.

Within minutes, we found a publicly accessible ClickHouse database linked to DeepSeek, completely open and unauthenticated, exposing sensitive data. It was hosted at oauth2callback.deepseek.com:9000 and dev.deepseek.com:9000.

This database contained a significant volume of chat history, backend data and sensitive information, including log streams, API Secrets, and operational details.

More critically, the exposure allowed for full database control and potential **privilege escalation** within the DeepSeek environment, without any authentication or defense mechanism to the outside world.





Exposure Walkthrough

Our reconnaissance began with assessing DeepSeek's publicly accessible domains. By mapping the external attack surface with straightforward reconnaissance techniques (passive and active discovery of subdomains), we identified around 30 internet-facing subdomains. Most appeared benign, hosting elements like the chatbot interface, status page, and API documentation—none of which initially suggested a high-risk exposure.

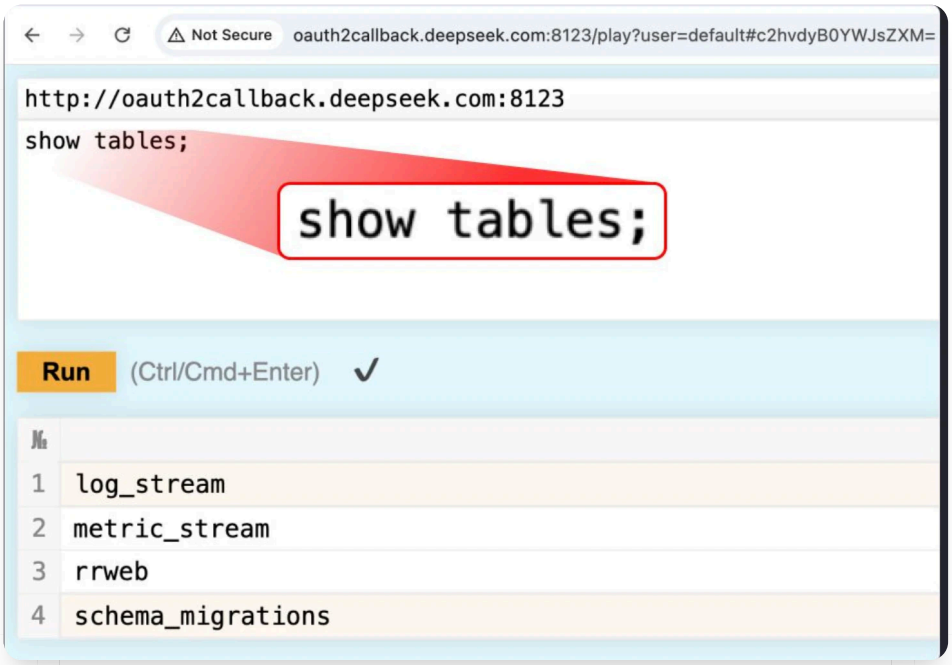
However, as we expanded our search beyond standard HTTP ports (80/443), we detected two **unusual, open ports (8123 & 9000)** associated with the following hosts:

- http://oauth2callback.deepseek.com:8123
- http://dev.deepseek.com:8123
- http://oauth2callback.deepseek.com:9000
- http://dev.deepseek.com:9000

Upon further investigation, these ports led to a **publicly exposed ClickHouse database**, accessible without any authentication at all – immediately raising red flags.

ClickHouse is an open-source, columnar database management system designed for fast analytical queries on large datasets. It was developed by Yandex and is widely used for real-time data processing, log storage, and big data analytics, which indicates such exposure as a very valuable and sensitive discovery.

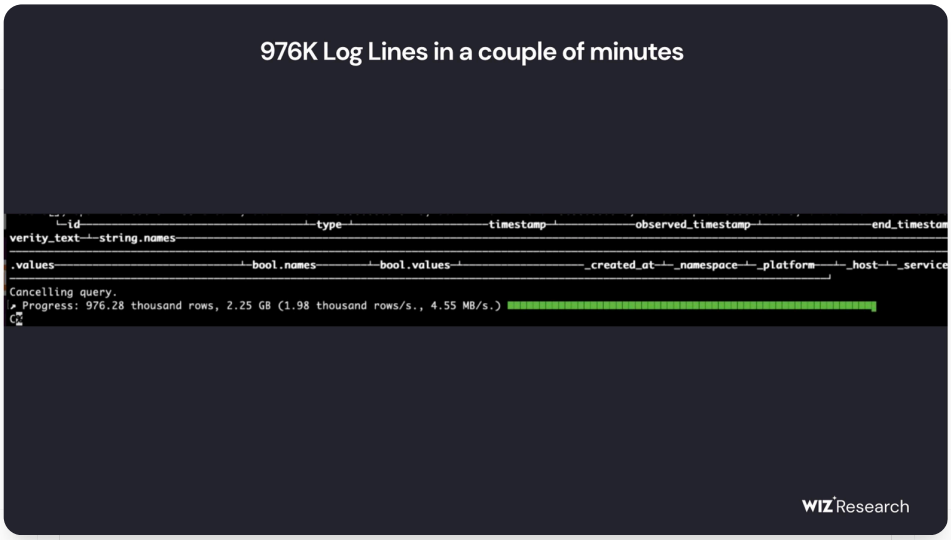
By leveraging ClickHouse's HTTP interface, we accessed the /play path, which **allowed direct execution of arbitrary SQL queries** via the browser. Running a simple SHOW TABLES; query returned a full list of accessible datasets.



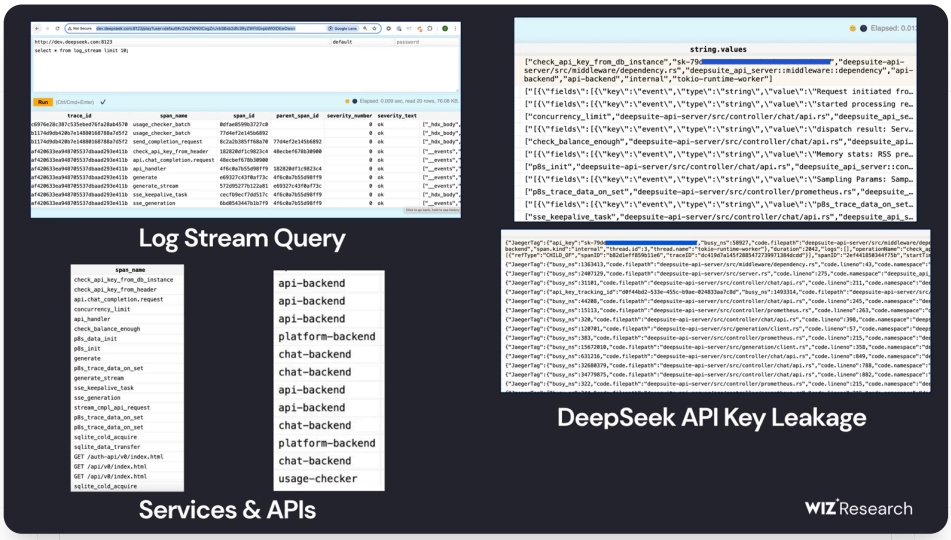
Tables output from ClickHouse Web UI

Among them, one table stood out: log_stream, which contained extensive logs with **highly sensitive data**.

The log_stream table contained **over 1 million log entries**, with particularly revealing columns:



- timestamp – Logs dating from **January 6, 2025**
- span_name – References to various internal **DeepSeek API endpoints**
- string.values – **Plaintext logs**, including **Chat History**, **API Keys, backend details, and operational metadata**
- _service – Indicating which **DeepSeek service** generated the logs
- _source – Exposing the **origin of log requests**, containing **Chat History, API Keys, directory structures, and chatbot metadata logs**



This level of access posed a critical risk to DeepSeek's own security and for its end-users. Not only an attacker could retrieve sensitive logs and actual plain-text chat messages, but they could also potentially exfiltrate plaintext passwords and local files along propriety information directly from the server using queries like: SELECT * FROM file('filename') depending on their ClickHouse configuration.

(Note: We did not execute intrusive queries beyond enumeration to preserve ethical research practices.)

Key Takeaways

The rapid adoption of AI services without corresponding security is inherently risky. This exposure underscores the fact that the immediate **security risks for AI applications** stem from the infrastructure and tools supporting them.

While much of the attention around AI security is focused on futuristic threats, the real dangers often come from basic risks—like accidental external exposure of databases. These risks, which are fundamental to security, should remain a top priority for security teams.

As organizations rush to adopt AI tools and services from a growing number of startups and providers, it's essential to remember that by doing so, we're entrusting these companies with sensitive data. The rapid pace of adoption often leads to overlooking security, but protecting customer data must remain the top priority. It's crucial that security teams work closely with AI engineers to ensure visibility into the architecture, tooling, and models being used, so we can safeguard data and prevent exposure.

Conclusion

The world has never seen a piece of technology adopted at the pace of AI. Many AI companies have rapidly grown into critical infrastructure providers without the security frameworks that typically accompany such widespread adoptions. As AI becomes deeply integrated into businesses worldwide, the industry must recognize the risks of handling sensitive data and enforce security practices on par with those required for public cloud providers and major infrastructure providers.

Tags #Research

PLATFORM

- Wiz CNAPP
- Wiz Code
- Wiz Cloud
- Wiz Defend
- Integrations
- Environments
- Documentation

LEARN

- Customer stories
- Train Your Team For Cloud
- Blog
- CloudSec Academy
- Resources Center
- Cloud threat landscape
- Cloud Security Assessment

COMPANY

- About Wiz
- Join the team
- Newsroom
- Events
- Contact us
- Trust Center
- Our partners

English (US)