SECURAITY

# DEEPSEEK DATA BREACH

FEB 2025

# DeepSeek Data Breach: Report

**Date:** February 10, 2025

## Abstract

This report presents a detailed analysis of the DeepSeek data breach, an incident in which over **one million sensitive records** were exposed due to a *misconfigured ClickHouse database*. Discovered by Wiz Research on January 29, 2025, the breach highlights serious challenges in securing data infrastructures within fast-growing technological environments. This document outlines the timeline of the incident, the technical shortcomings that led to the breach, the nature of the compromised data, and the resulting implications for user privacy, corporate security, and regulatory compliance. In addition, the report provides a series of recommendations to prevent similar future occurrences.

---

## 1. Introduction

In today's technology-driven world, data breaches are becoming increasingly common and can have wide-ranging impacts on privacy, corporate integrity, and regulatory compliance. This report examines the DeepSeek breach—a case that emphasizes the crucial need for secure database management practices. DeepSeek, a Chinese AI startup known for its innovative AI models, experienced a significant security lapse when a ClickHouse database was inadvertently left open to the public internet. By integrating findings from Wiz Research and other industry sources, this analysis explores the technical causes behind the breach, its impacts, and the measures that could help safeguard sensitive information in rapidly evolving environments.

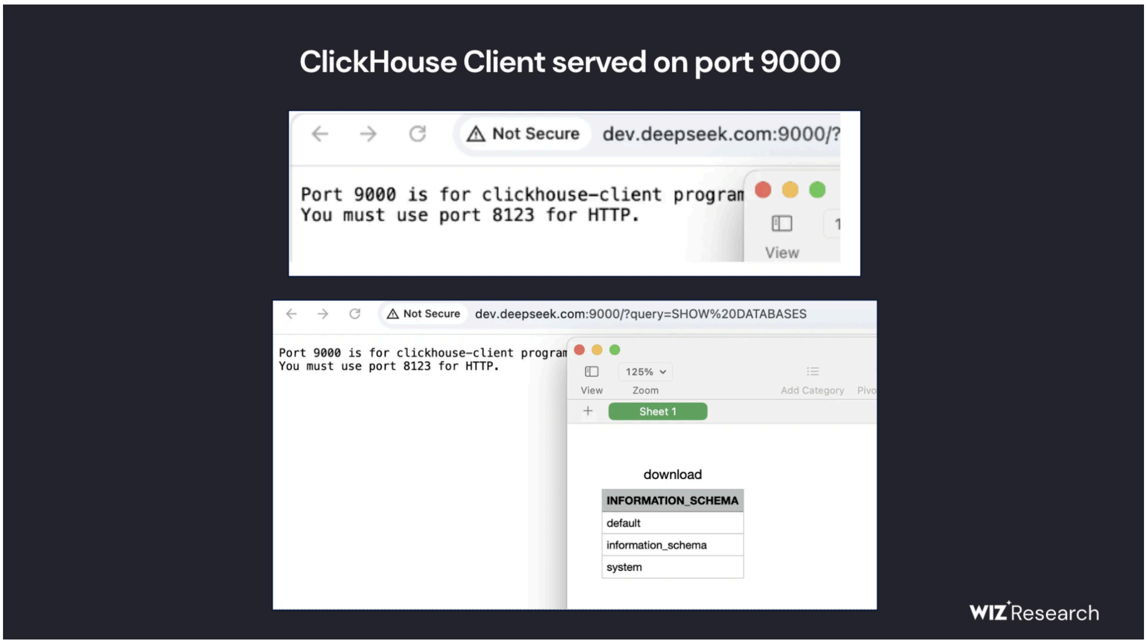
# 2. Incident Overview

## 2.1 Discovery and Timeline

- **Detection :** Wiz Research, while conducting routine security assessments and mapping DeepSeek's internet-facing subdomains, identified an unsecured ClickHouse database on January 29, 2025. The discovery was made during both passive and active reconnaissance, which revealed *unusual open ports (**8123** and **9000**)* on specific endpoints.
- **Exposure Duration**
  Although the exact length of the exposure remains unknown, the misconfiguration was present long enough to potentially allow unauthorized access prior to detection.
- **Mitigation Efforts**
  After being notified by Wiz Research, DeepSeek promptly secured the database within one hour, thereby preventing further unauthorized operations.

## 2.2 Technical Analysis

- **Database Misconfiguration**
  The breach originated from a misconfigured ClickHouse database, accessible at endpoints such as `oauth2callback.deepseek.com:9000` and `dev.deepseek.com:9000`. This database was left completely open without any authentication protocols, enabling anyone on the internet to interact with it.
- **Exploitation Mechanism**
  By accessing the ClickHouse HTTP interface, attackers could execute arbitrary SQL queries. For example, the use of the `/play` path allowed for the direct execution of commands like `SHOW TABLES;`, which revealed a list of available datasets. This open access included a critical table named `log_stream` containing over one million entries. The table featured sensitive columns such as:
  - **timestamp:** Dates from January 6, 2025, onward.
  - **span_name:** References to internal API endpoints.
  - **string.values:** Plaintext logs including chat histories, API keys, and backend details.
  - **_service and _source:** Information on the originating service and the source of the log requests.



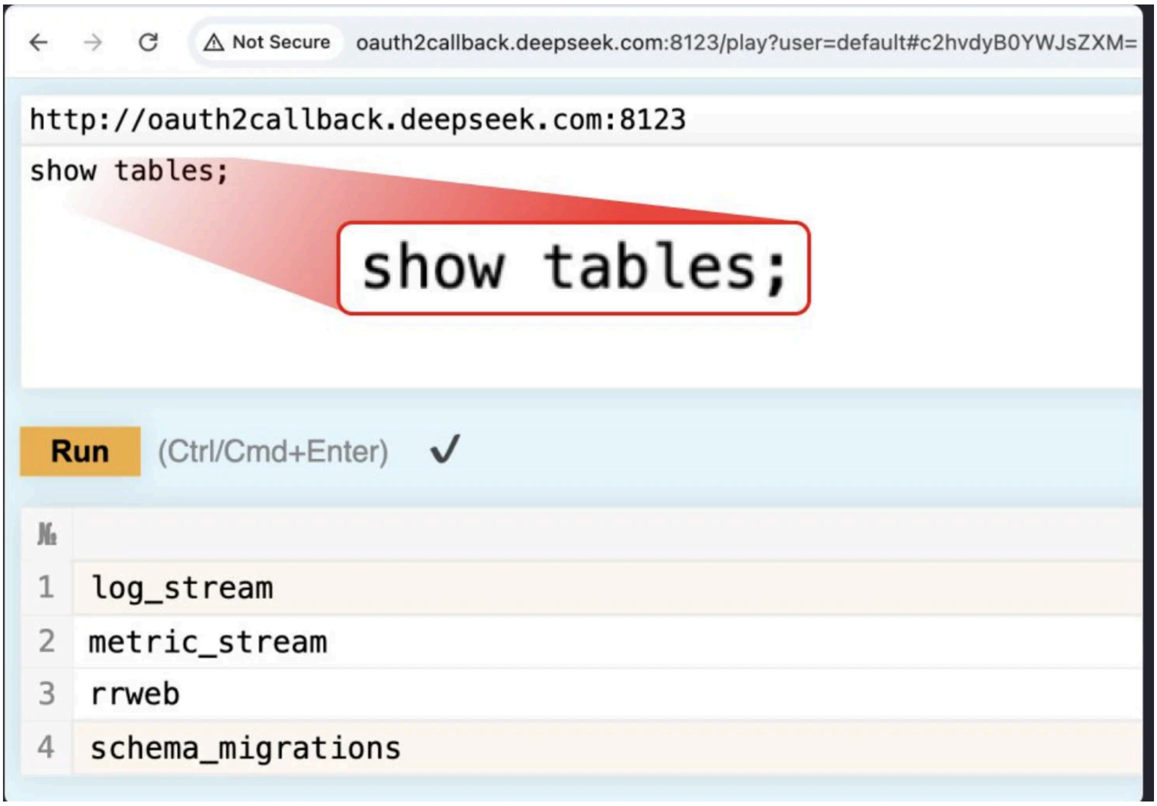
## ClickHouse Client served on port 9000

```
Port 9000 is for clickhouse-client program
You must use port 8123 for HTTP.
```

```
Port 9000 is for clickhouse-client program
You must use port 8123 for HTTP.
```

download

| INFORMATION_SCHEMA |
|---|
| default |
| information_schema |
| system |

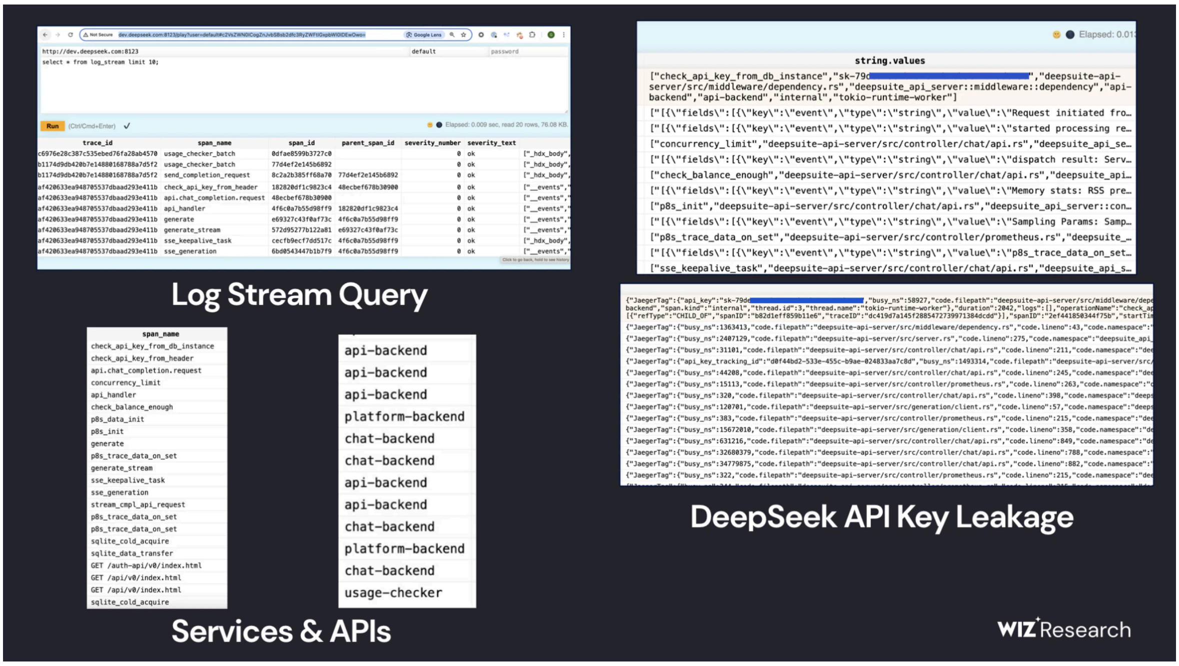


Not Secure
oauth2callback.deepseek.com:8123/play?user=default#c2hvdyB0YWJsZXM=

```
http://oauth2callback.deepseek.com:8123
show tables;
```

show tables;

| № | |
|---|---|
| 1 | log_stream |
| 2 | metric_stream |
| 3 | rrweb |
| 4 | schema_migrations |



Not Secure
dev.deepseek.com:8123/play?user=default#c2VsZWN0ICogZnJvbSBsb2dfc3RyZWFtIGxpbWl0IDEwOwo=
Google Lens

```
http://dev.deepseek.com:8123                                    default          password
select * from log_stream limit 10;
```

Elapsed: 0.009 sec, read 20 rows, 76.08 KB.

| trace_id | span_name | span_id | parent_span_id | severity_number | severity_text | |
|---|---|---|---|---|---|---|
| c6976e28c387c535ebed76fa28ab4570 | usage_checker_batch | 0dfae8599b3727c0 | | 0 | ok | ["_hdx_body", |
| b1174d9db420b7e14880168788a7d5f2 | usage_checker_batch | 77d4ef2e145b6892 | | 0 | ok | ["_hdx_body", |
| b1174d9db420b7e14880168788a7d5f2 | send_completion_request | 8c2a2b385ff68a70 | 77d4ef2e145b6892 | 0 | ok | ["_hdx_body", |
| af420633ea948705537dbaad293e411b | check_api_key_from_header | 182820df1c9823c4 | 48ecbef678b30900 | 0 | ok | ["__events"," |
| af420633ea948705537dbaad293e411b | api.chat_completion.request | 48ecbef678b30900 | | 0 | ok | ["__events"," |
| af420633ea948705537dbaad293e411b | api_handler | 4f6c0a7b55d98ff9 | 182820df1c9823c4 | 0 | ok | ["__events"," |
| af420633ea948705537dbaad293e411b | generate | e69327c43f0af73c | 4f6c0a7b55d98ff9 | 0 | ok | ["__events"," |
| af420633ea948705537dbaad293e411b | generate_stream | 572d95277b122a81 | e69327c43f0af73c | 0 | ok | ["__events"," |
| af420633ea948705537dbaad293e411b | sse_keepalive_task | cecfb9ecf7dd517c | 4f6c0a7b55d98ff9 | 0 | ok | ["_hdx_body", |
| af420633ea948705537dbaad293e411b | sse_generation | 6bd0543447b1b7f9 | 4f6c0a7b55d98ff9 | 0 | ok | ["__events"," |

Log Stream Query

| span_name |
|---|
| check_api_key_from_db_instance |
| check_api_key_from_header |
| api.chat_completion.request |
| concurrency_limit |
| api_handler |
| check_balance_enough |
| p8s_data_init |
| p8s_init |
| generate |
| p8s_trace_data_on_set |
| generate_stream |
| sse_keepalive_task |
| sse_generation |
| stream_cmpl_api_request |
| p8s_trace_data_on_set |
| p8s_trace_data_on_set |
| sqlite_cold_acquire |
| sqlite_data_transfer |
| GET /auth-api/v0/index.html |
| GET /api/v0/index.html |
| GET /api/v0/index.html |
| sqlite_cold_acquire |

```
api-backend
api-backend
api-backend
platform-backend
chat-backend
chat-backend
api-backend
api-backend
chat-backend
platform-backend
chat-backend
usage-checker
```

Services & APIs

Elapsed: 0.01

```
string.values
["check_api_key_from_db_instance","sk-79d[redacted]","deepsuite-api-server/src/middleware/dependency.rs","deepsuite_api_server::middleware::dependency","api-backend","api-backend","internal","tokio-runtime-worker"]
["[{\"fields\":[{\"key\":\"event\",\"type\":\"string\",\"value\":\"Request initiated fro…
["[{\"fields\":[{\"key\":\"event\",\"type\":\"string\",\"value\":\"started processing re…
["concurrency_limit","deepsuite-api-server/src/controller/chat/api.rs","deepsuite_api_se…
["[{\"fields\":[{\"key\":\"event\",\"type\":\"string\",\"value\":\"dispatch result: Serv…
["check_balance_enough","deepsuite-api-server/src/controller/chat/api.rs","deepsuite_api…
["[{\"fields\":[{\"key\":\"event\",\"type\":\"string\",\"value\":\"Memory stats: RSS pre…
["p8s_init","deepsuite-api-server/src/controller/chat/api.rs","deepsuite_api_server::con…
["[{\"fields\":[{\"key\":\"event\",\"type\":\"string\",\"value\":\"Sampling Params: Samp…
["p8s_trace_data_on_set","deepsuite-api-server/src/controller/prometheus.rs","deepsuite_…
["[{\"fields\":[{\"key\":\"event\",\"type\":\"string\",\"value\":\"p8s_trace_data_on_set…
["sse_keepalive_task","deepsuite-api-server/src/controller/chat/api.rs","deepsuite_api_s…
```

```
{"JaegerTag":{"api_key":"sk-79de[redacted],"busy_ns":58927,"code.filepath":"deepsuite-api-server/src/middleware/depe
backend","span.kind":"internal","thread.id":3,"thread.name":"tokio-runtime-worker"},"duration":2042,"logs":[],"operationName":"check_ap
[{"refType":"CHILD_OF","spanID":"b82d1eff859b11e6","traceID":"dc419d7a145f2885472739971384dcdd"}],"spanID":"2ef441850344f75b","startTim
{"JaegerTag":{"busy_ns":1363413,"code.filepath":"deepsuite-api-server/src/middleware/dependency.rs","code.lineno":43,"code.namespace":"
{"JaegerTag":{"busy_ns":2407129,"code.filepath":"deepsuite-api-server/src/server.rs","code.lineno":275,"code.namespace":"deepsuite_api_
{"JaegerTag":{"busy_ns":31101,"code.filepath":"deepsuite-api-server/src/controller/chat/api.rs","code.lineno":211,"code.namespace":"dee
{"JaegerTag":{"api_key_tracking_id":"d0f44bd2-533e-455c-b9ae-024833aa7c8d","busy_ns":1493314,"code.filepath":"deepsuite-api-server/src/
{"JaegerTag":{"busy_ns":44208,"code.filepath":"deepsuite-api-server/src/controller/chat/api.rs","code.lineno":245,"code.namespace":"dee
{"JaegerTag":{"busy_ns":15113,"code.filepath":"deepsuite-api-server/src/controller/prometheus.rs","code.lineno":263,"code.namespace":"d
{"JaegerTag":{"busy_ns":320,"code.filepath":"deepsuite-api-server/src/controller/chat/api.rs","code.lineno":398,"code.namespace":"deeps
{"JaegerTag":{"busy_ns":120701,"code.filepath":"deepsuite-api-server/src/generation/client.rs","code.lineno":57,"code.namespace":"deeps
{"JaegerTag":{"busy_ns":383,"code.filepath":"deepsuite-api-server/src/controller/prometheus.rs","code.lineno":215,"code.namespace":"dee
{"JaegerTag":{"busy_ns":15672010,"code.filepath":"deepsuite-api-server/src/generation/client.rs","code.lineno":358,"code.namespace":"de
{"JaegerTag":{"busy_ns":631216,"code.filepath":"deepsuite-api-server/src/controller/chat/api.rs","code.lineno":849,"code.namespace":"de
{"JaegerTag":{"busy_ns":32680379,"code.filepath":"deepsuite-api-server/src/controller/chat/api.rs","code.lineno":788,"code.namespace":"
{"JaegerTag":{"busy_ns":34779875,"code.filepath":"deepsuite-api-server/src/controller/chat/api.rs","code.lineno":882,"code.namespace":"
{"JaegerTag":{"busy_ns":322,"code.filepath":"deepsuite-api-server/src/controller/prometheus.rs","code.lineno":215,"code.namespace":"dee
```

DeepSeek API Key Leakage

WIZ Research

## 2.3 Data Compromised

The exposed database contained multiple layers of sensitive information, including:

- **User Data**
  - Plaintext chat histories, which can directly reveal personal interactions.
- **System Credentials and Metadata**
  - API keys and backend infrastructure details that are critical to the operation of DeepSeek's systems.
  - Operational metadata that could be used to understand and potentially disrupt internal processes.
- **Internal Logs**
  - Detailed log streams that provide a window into internal system operations and can even reveal plaintext passwords and file structures if further exploited.

These exposures not only compromise user privacy but also provide an attacker with potential tools for privilege escalation and further exploitation within the DeepSeek environment.

# 3. Impact Analysis

## 3.1 Immediate Risks

- **User Privacy Concerns**
  The exposure of **plaintext chat logs** and **API keys** poses immediate risks, including phishing, identity theft, and unauthorized access to personal accounts.
- **Corporate Vulnerabilities**
  With full control over the exposed database operations, an attacker could manipulate data, escalate privileges, or even engage in corporate espionage. The ability to access detailed internal logs and backend configurations significantly increases the risk to DeepSeek's infrastructure.

## 3.2 Regulatory and Legal Implications

- **Investigative Actions**
  Regulatory bodies such as *Ireland's Data Protection Commission (DPC)*, *Italy's Garante*, and the *U.S. National Security Council (NSC)* have launched investigations into the breach. The inquiries focus on whether DeepSeek's practices comply with international data protection laws like the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA).
- **Potential Penalties**
  Failure to adhere to these regulations could result in substantial fines and legal sanctions, further damaging DeepSeek's reputation and operational capabilities.

## 3.3 Reputational and Operational Repercussions

- **Erosion of Trust**
  The breach has the potential to significantly undermine confidence in DeepSeek's security protocols. Users and potential partners may reconsider their engagement with the company if data protection is not assured.
- **Service Disruptions**
  In the aftermath of the breach, DeepSeek had to temporarily suspend new user registrations, thereby affecting the company's growth and operational stability.

### 3.4 Broader Industry Implications

- **Security Best Practices**
  The incident underscores the importance of robust security measures, particularly as AI services continue to scale rapidly. Even advanced systems can be vulnerable if foundational security practices are neglected.
- **Enhanced Regulatory Oversight**
  With increased media attention and regulatory interest, the industry may see a move toward stricter data localization and encryption standards. This breach serves as a reminder that basic security oversights can have far-reaching consequences.

## 4. Recommendations for Future Prevention

### 4.1 Technical Measures

- **Enhanced Database Security:**
  - **Authentication and Access Controls:**
    Implement robust, role-based access controls to ensure that sensitive databases are not accessible without proper credentials.
  - **Data Encryption:**
    Enforce encryption for data at rest and in transit, including sensitive user data, API keys, and internal logs.

### 4.2 Operational Enhancements

- **Routine Security Audits:**
  Regular vulnerability assessments and penetration tests should be conducted to identify and address potential weaknesses before they can be exploited.
- **Real-Time Monitoring:**
  Deploy comprehensive monitoring systems to detect any unauthorized access or anomalous activities immediately, allowing for prompt intervention.

### 4.3 Compliance with Regulatory Standards

- **Adherence to Global Regulations:**
  Align data security practices with international standards such as GDPR and CCPA. Transparency in data storage, processing, and breach notification is essential to maintain trust and comply with legal requirements.

## 5. Conclusion

The DeepSeek data breach serves as a critical case study on the importance of securing data infrastructures, especially in fast-growing technological environments like those found in the AI industry. The misconfigured ClickHouse database allowed unauthorized access to sensitive user and corporate data, highlighting vulnerabilities that can have far-reaching implications for privacy, security, and regulatory compliance.

Despite the swift remediation efforts by DeepSeek, the incident reveals *how basic security oversights—such as improper database configuration—can expose an organization to significant risks.* As AI companies continue to scale rapidly, it is imperative that they adopt a security-first approach. This involves integrating rigorous technical safeguards, continuous operational oversight, and strict adherence to international regulatory standards to protect sensitive data and maintain public trust.

The insights provided by Wiz Research and other industry sources underscore the need for ongoing vigilance and proactive measures to safeguard against similar incidents in the future.

---

**Sources:**

https://www.wiz.io/blog/wiz-research-uncovers-exposed-deepseek-database-leak

DeepSeek leaks one million sensitive records in a major data breach | CSO Online

DeepSeek Data Breach Exposes Over a Million Sensitive Records

DeepSeek Database Leaked - Database Secret keys, Logs & Chat History Exposed

---

*End of Report*