DEVCORE

紅隊的 AI 視界：攻防演練中的 LLM

鍾孟勳 Alan

戴夫寇爾股份有限公司
alan@devco.re

DEVCORE CONFERENCE 2026 | 2026.03.14

whoami

DEVCORE

鍾孟勳 (Alan / 7.5)

DEVCORE 紅隊主管

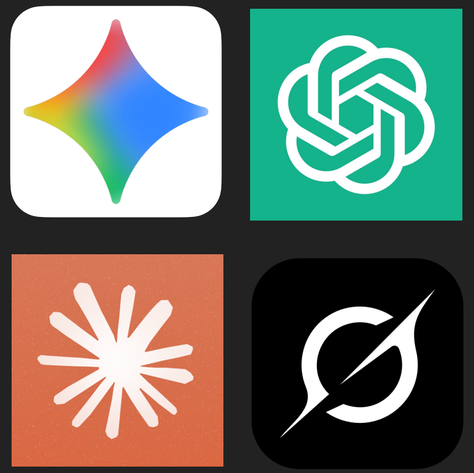
參與超過 35 場紅隊演練專案

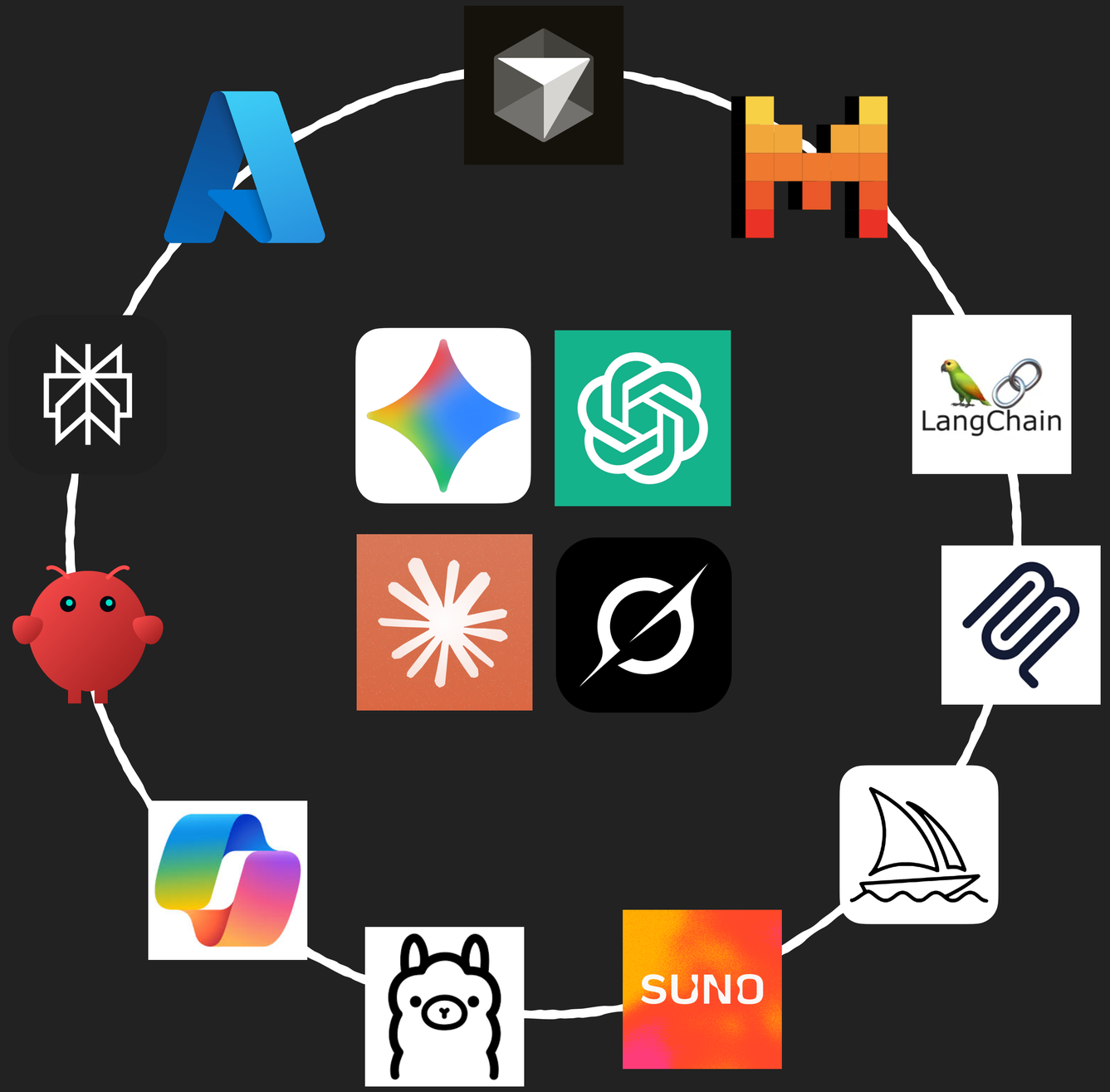
曾任政府、金融紅隊專案負責人

AI

調查：75%中小學生懂AI 42%未想過網路資訊有假

#中小學 #小學生 #學生 #聊天 ...




LangChain
SUNO

McKinsey – State of AI

DEVCORE

https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

McKinsey – State of AI

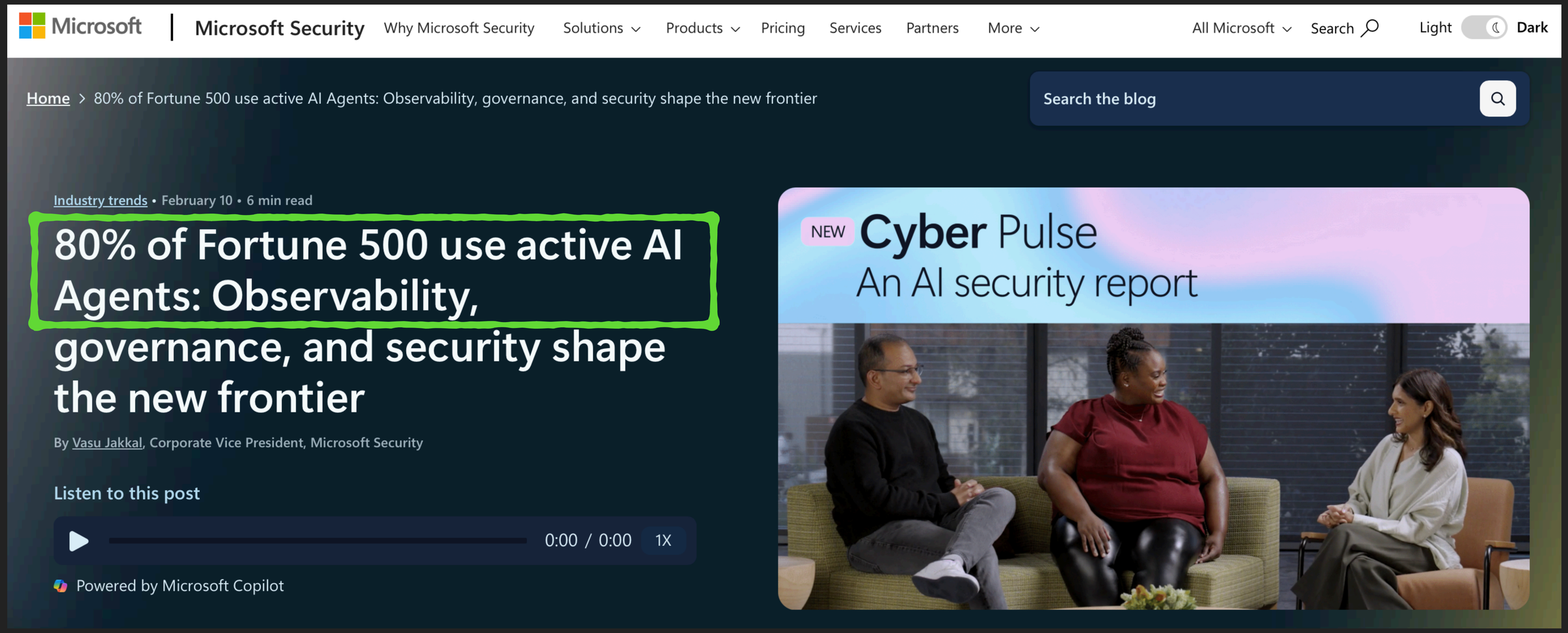
88%

受訪的企業在商業運作中導入 AI

DEVCORE

https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Microsoft 產業趨勢觀察



https://www.microsoft.com/en-us/security/blog/2026/02/10/80-of-fortune-500-use-active-ai-agents-observability-governance-and-security-shape-the-new-frontier/

OpenAI – State of Enterprise AI

https://cdn.openai.com/pdf/7ef17d82-96bf-4dd1-9df2-228f7f377a29/the-state-of-enterprise-ai_2025-report.pdf

OpenAI – State of Enterprise AI

87%

加速排除內部 IT 問題

https://cdn.openai.com/pdf/7ef17d82-96bf-4dd1-9df2-228f7f377a29/the-state-of-enterprise-ai_2025-report.pdf

OpenAI – State of Enterprise AI

87%

加速排除內部 IT 問題

85%

加速行銷計畫上線

https://cdn.openai.com/pdf/7ef17d82-96bf-4dd1-9df2-228f7f377a29/the-state-of-enterprise-ai_2025-report.pdf

OpenAI – State of Enterprise AI

87%

加速排除內部 IT 問題

85%

加速行銷計畫上線

75%

員工參與度提高

https://cdn.openai.com/pdf/7ef17d82-96bf-4dd1-9df2-228f7f377a29/the-state-of-enterprise-ai_2025-report.pdf

OpenAI – State of Enterprise AI

87%

加速排除內部 IT 問題

85%

加速行銷計畫上線

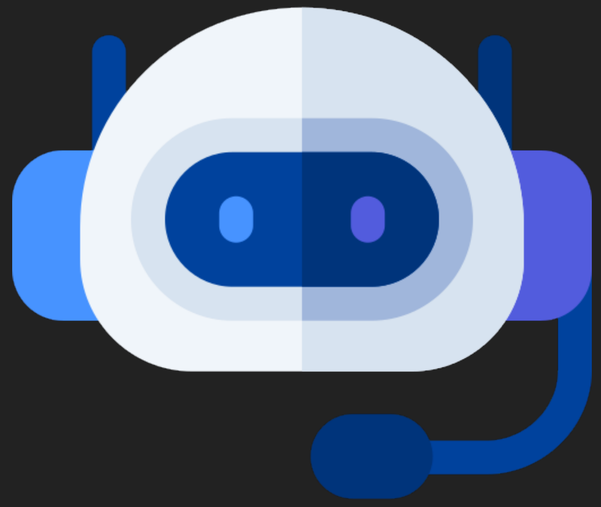
75%

員工參與度提高

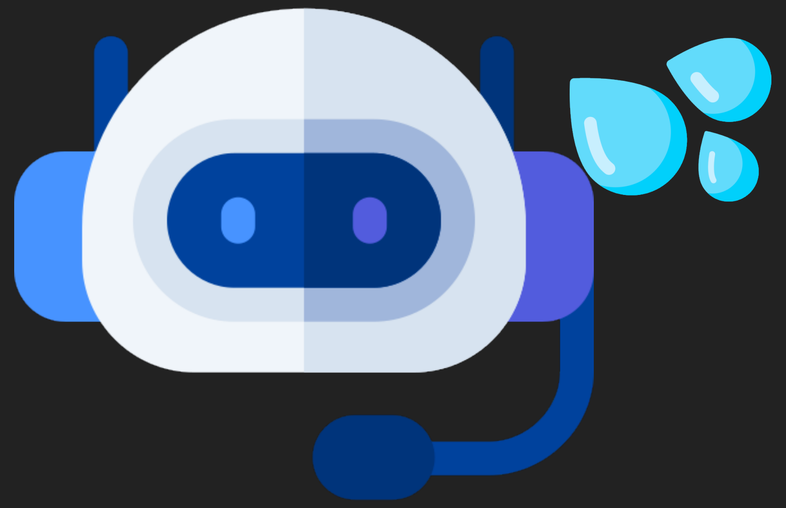
73%

提升程式碼交付速度

https://cdn.openai.com/pdf/7ef17d82-96bf-4dd1-9df2-228f7f377a29/the-state-of-enterprise-ai_2025-report.pdf



Enterprise Chatbot

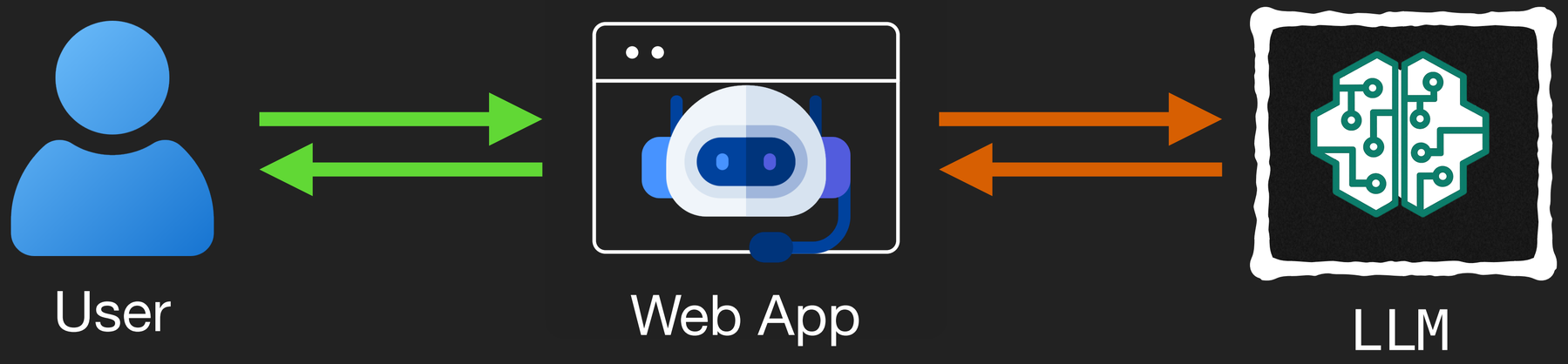


Enterprise Chatbot

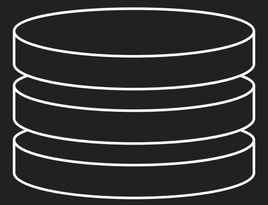

ATTACK VECTORS
LIGHTYEAR
ATTACK VECTORS EVERYWHERE

（以下是虛構情境）

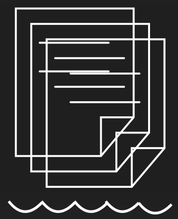
我看友商們都有在內部實作自己的聊天機器人，我們也要做一個


User
Web App
LLM

這個聊天機器人好像不能回答我們公司內部的相關問題，你想想辦法

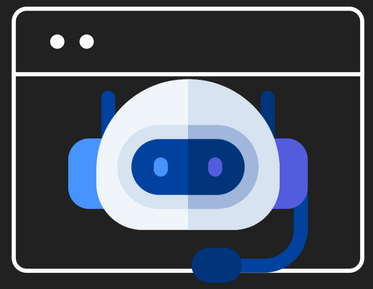


Vector DB

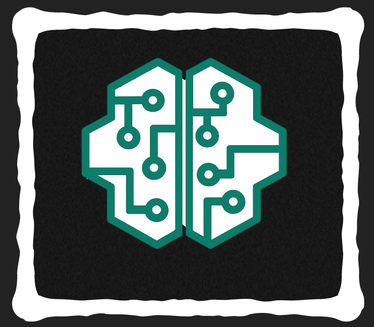


Internal Documents

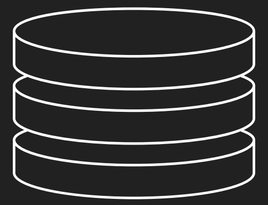
User



Web App

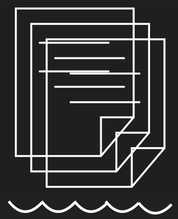


LLM



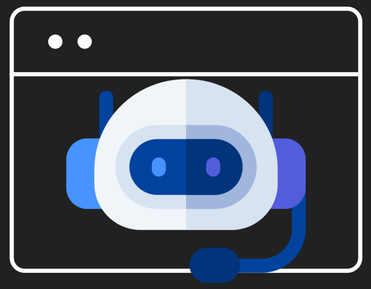
Vector DB

Data Processing

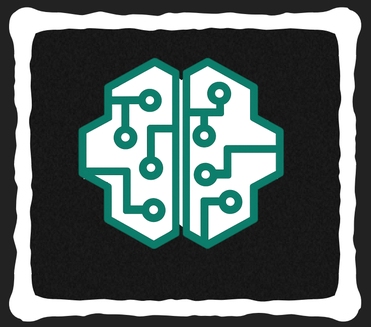


Internal Documents

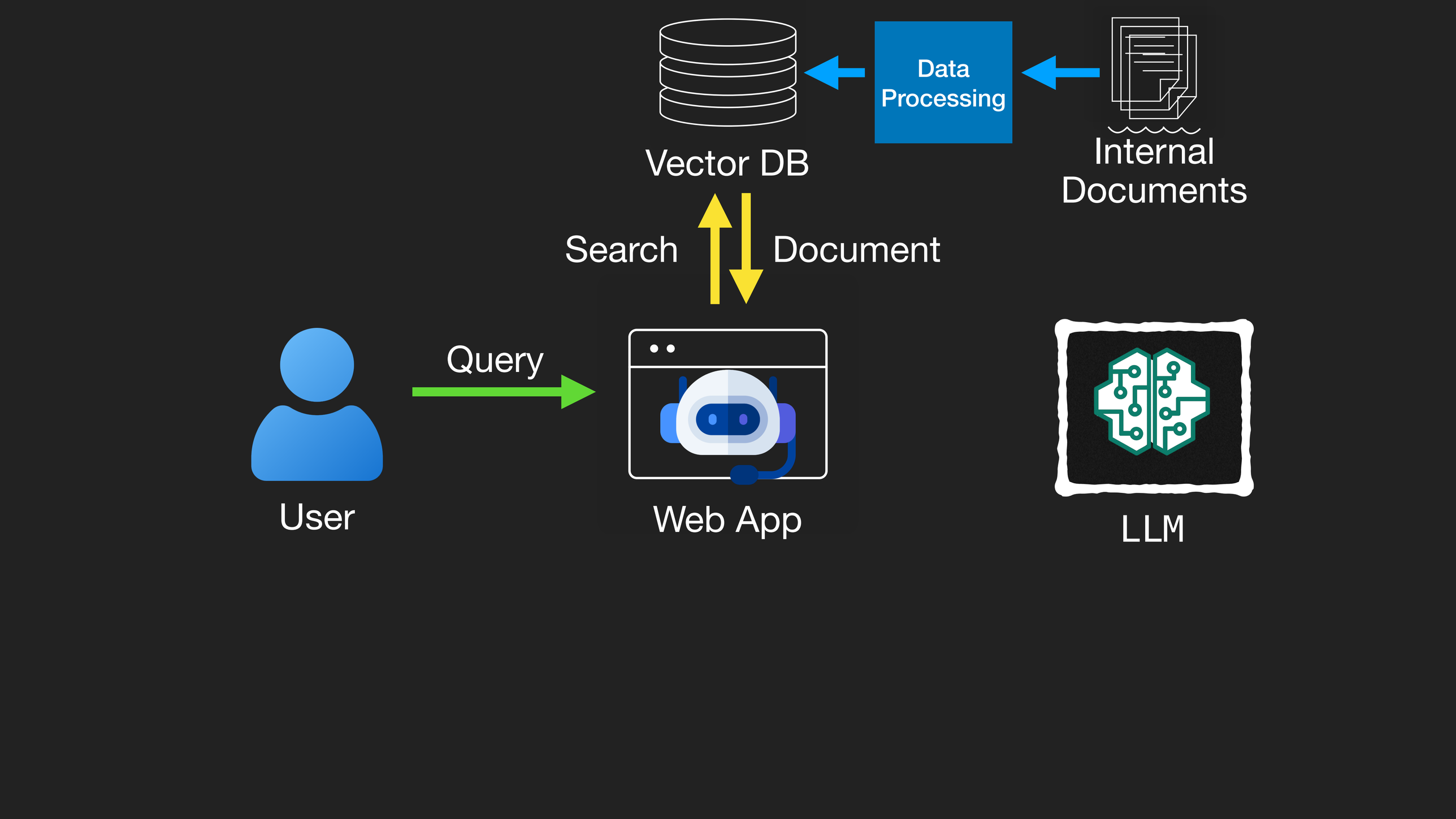
User

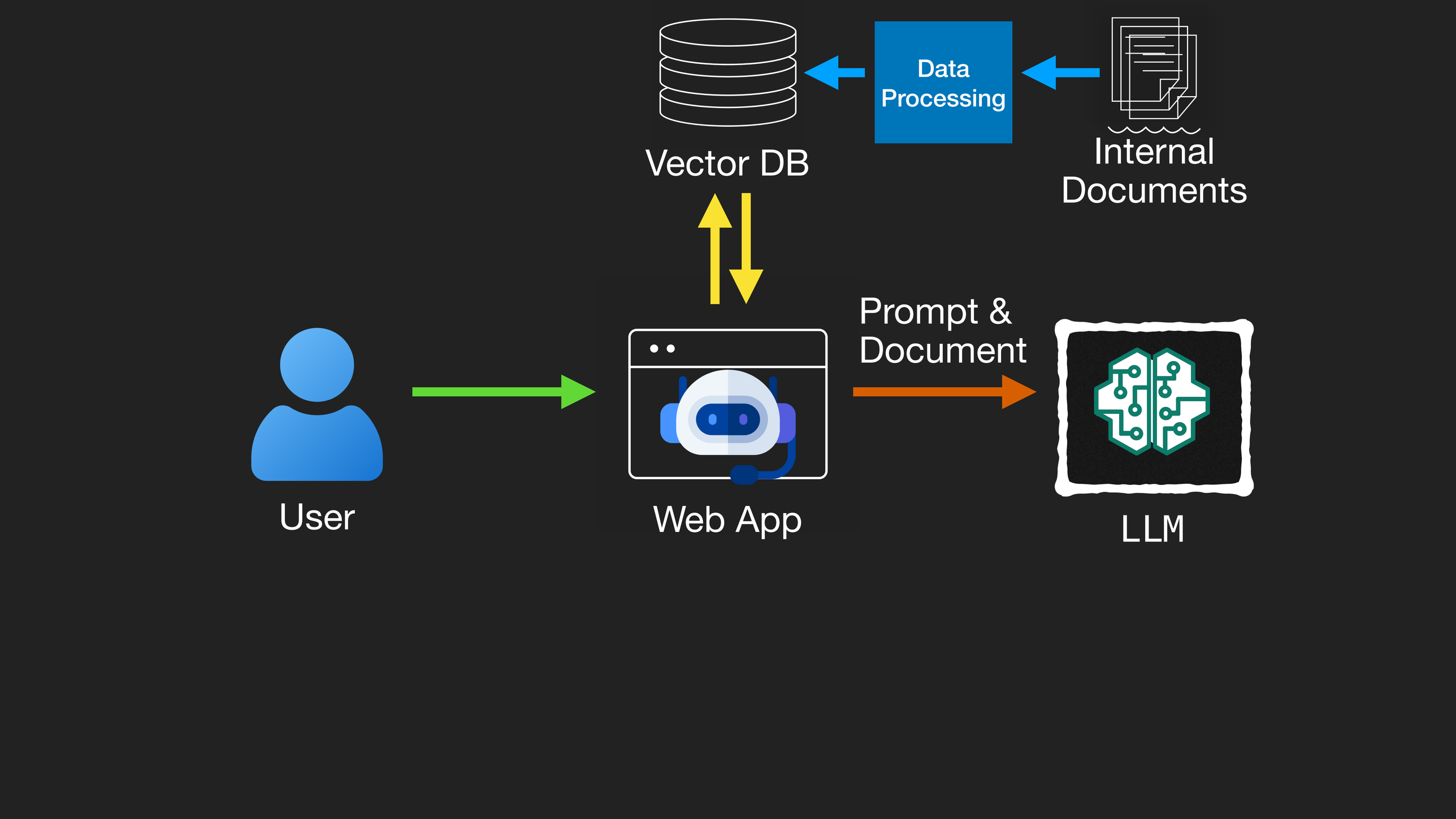


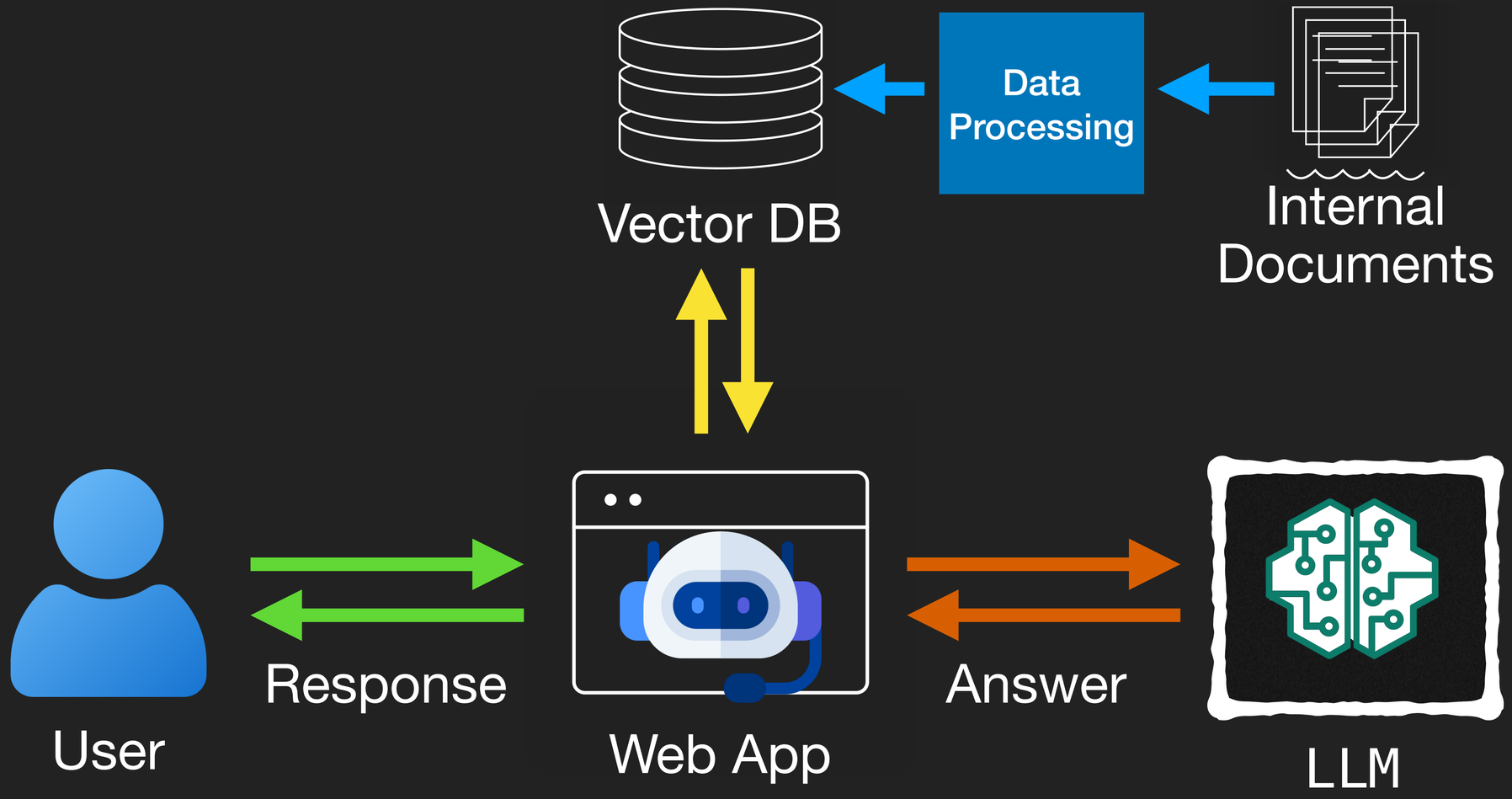
Web App



LLM


Vector DB
Data Processing
Internal Documents
Search
Document
Query
User
Web App
LLM

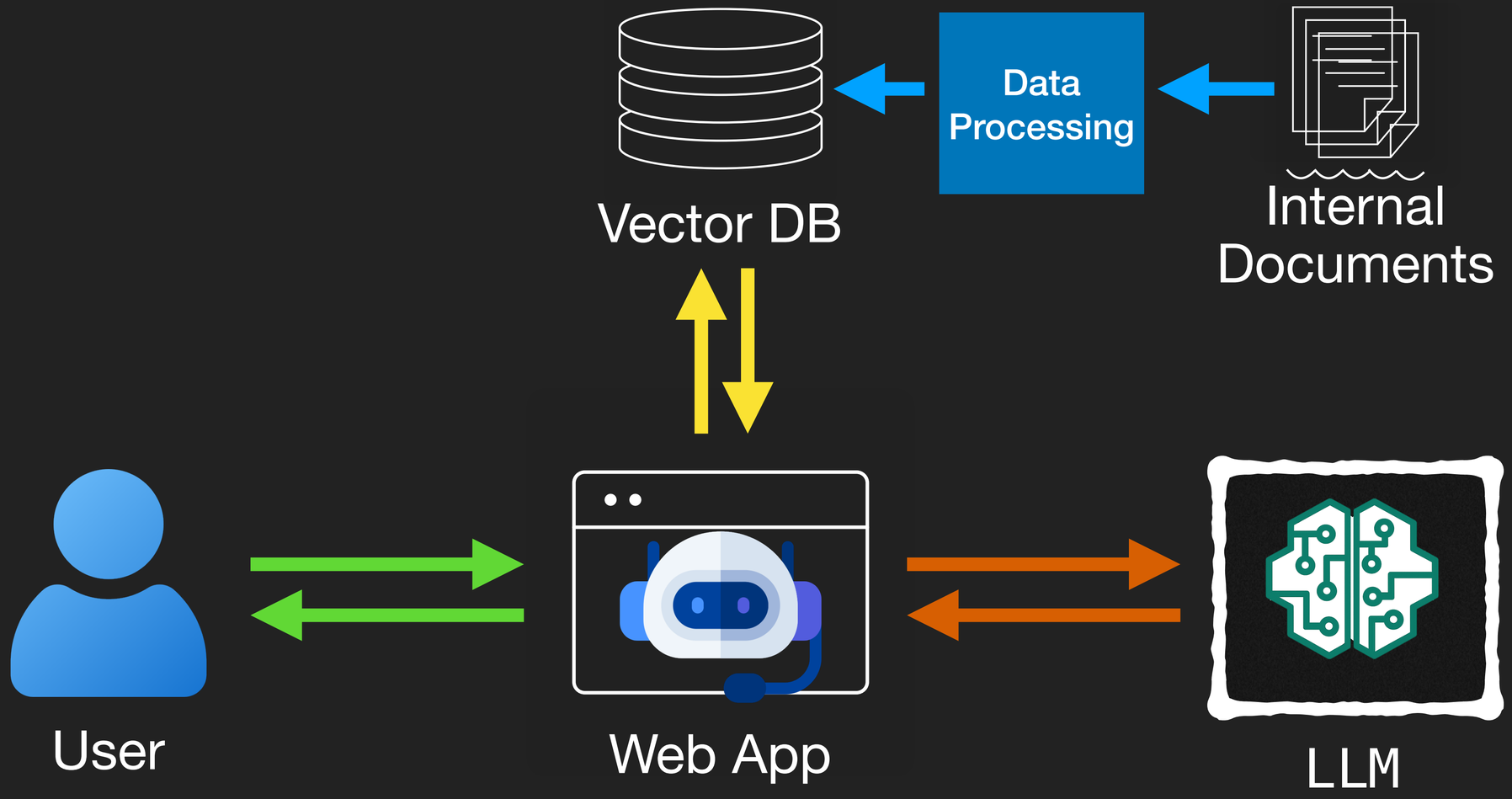

Vector DB
Data Processing
Internal Documents
Prompt & Document
User
Web App
LLM

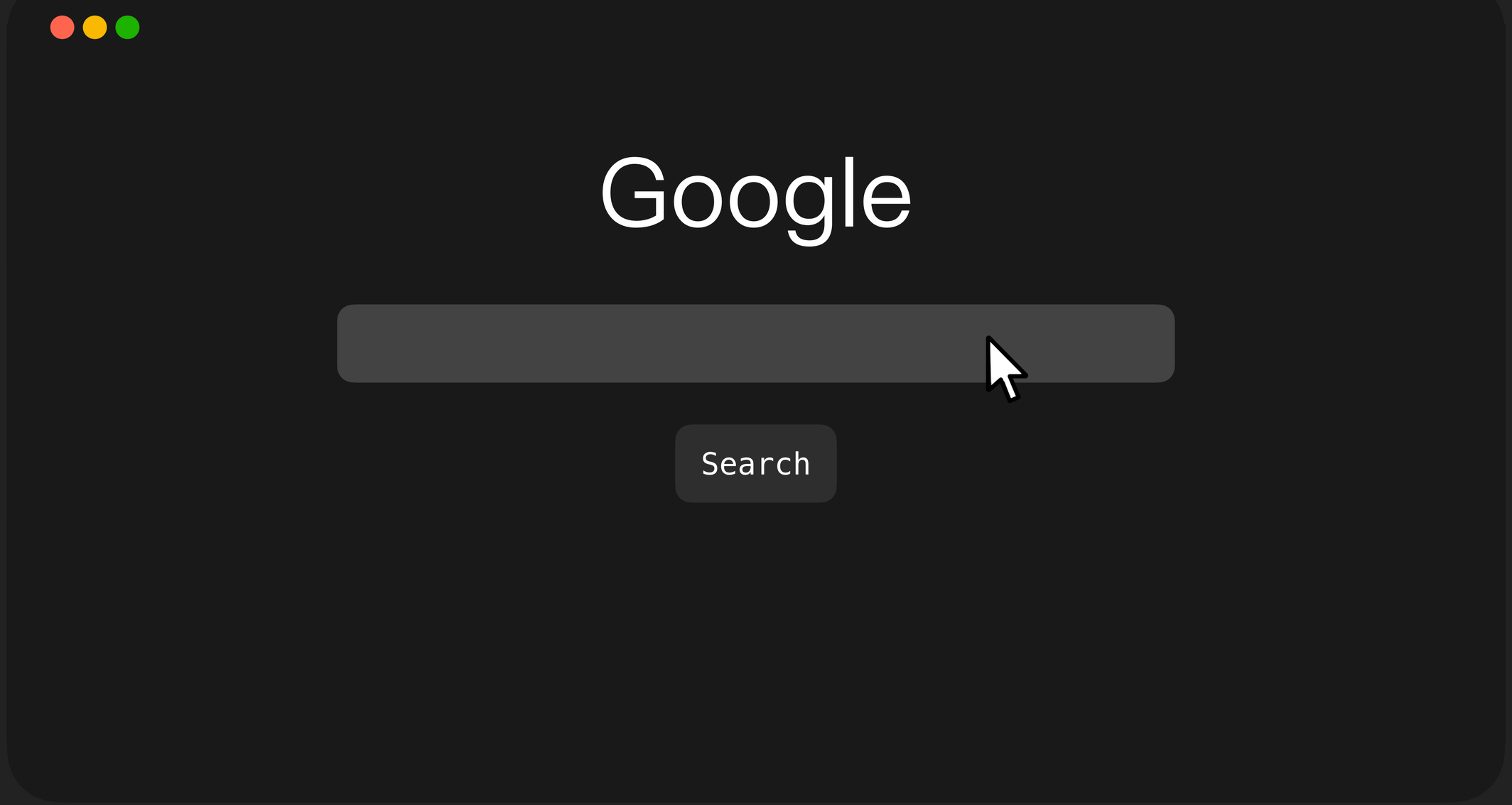

Vector DB
Data Processing
Internal Documents
User
Response
Web App
Answer
LLM

很好，以後新人都可以直接透過這個系統問問題了

這個系統評估一下資安風險，沒問題年底前要上線

我應該留意哪些資安風險？


Vector DB
Data Processing
Internal Documents
User
Web App
LLM


Google
Search


Google
提示注入 攻撃
Search

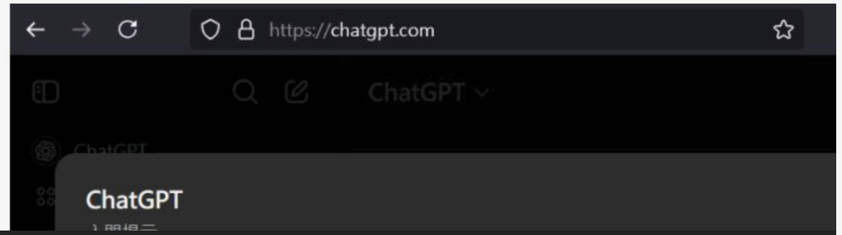
iThome

新聞

【當心提示注入、敏感資訊洩漏、錯誤資訊等問題】已在真實世界發生的LLM資安風險

生成式AI不當使用出現實際案例[redacted]可代寫程式碼[redacted]服務，以及美國律師誤用AI虛構判例打官司，均具體呈現LLM在不同層面的潛在

文/ 羅正漢 | 2025-04-18 發表



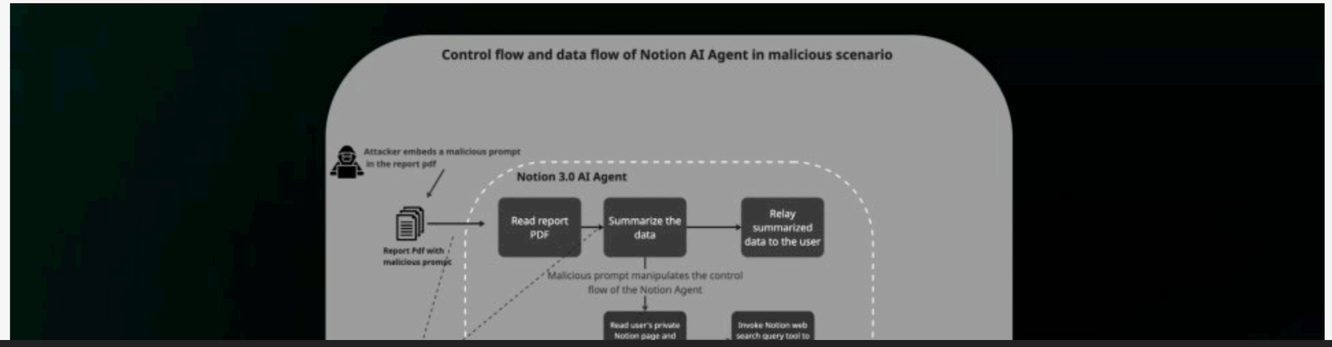
iThome

新聞

[redacted] AI Agents恐遭間接提示注入攻擊洩露機敏資料

[redacted]與MCP整合功能，能跨平臺自動化任務，但研究發現代理可能遭間接提示注入操控，透過Web Search洩露敏感資訊

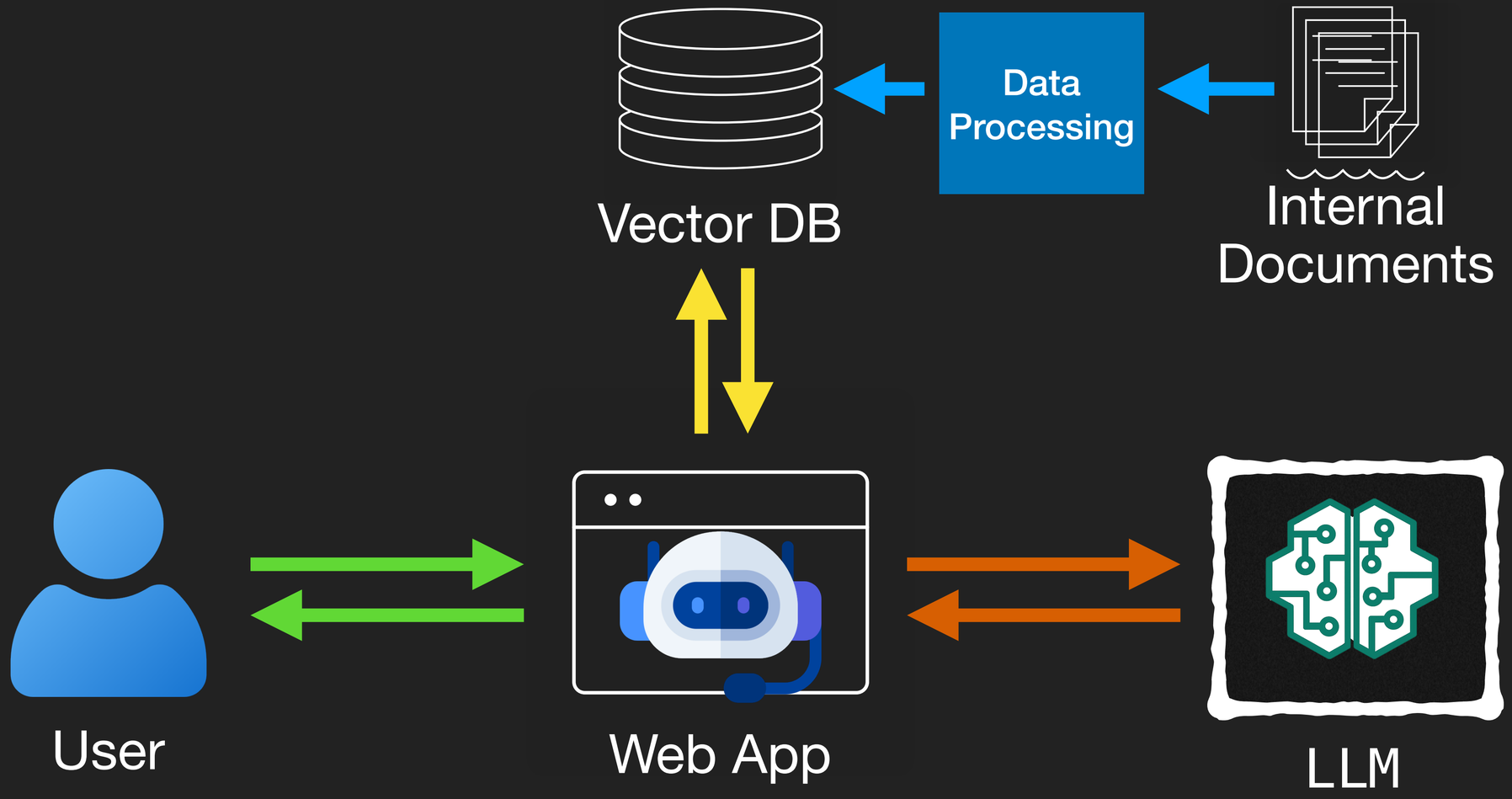
文/ 李建興 | 2025-09-23 發表

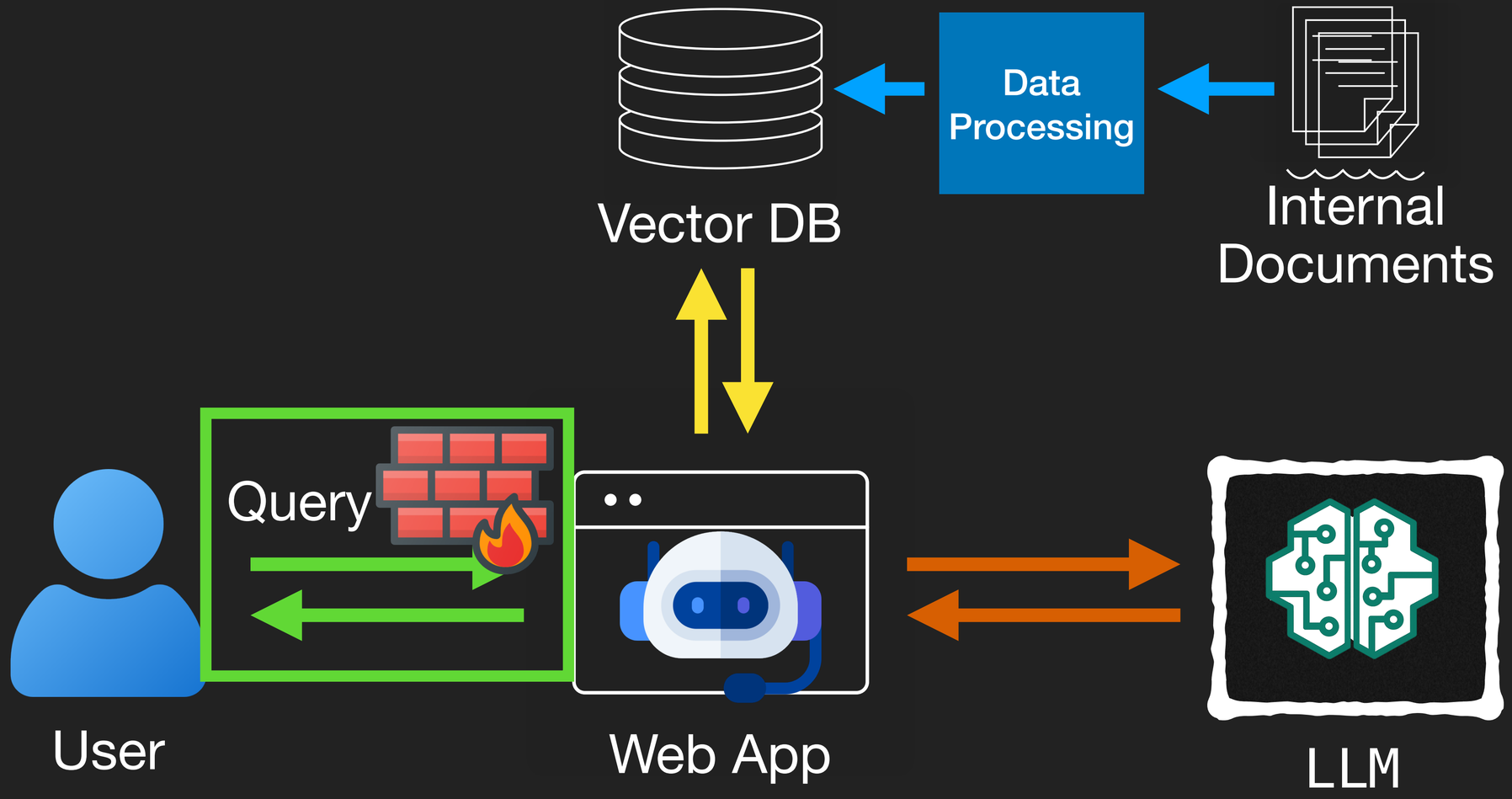


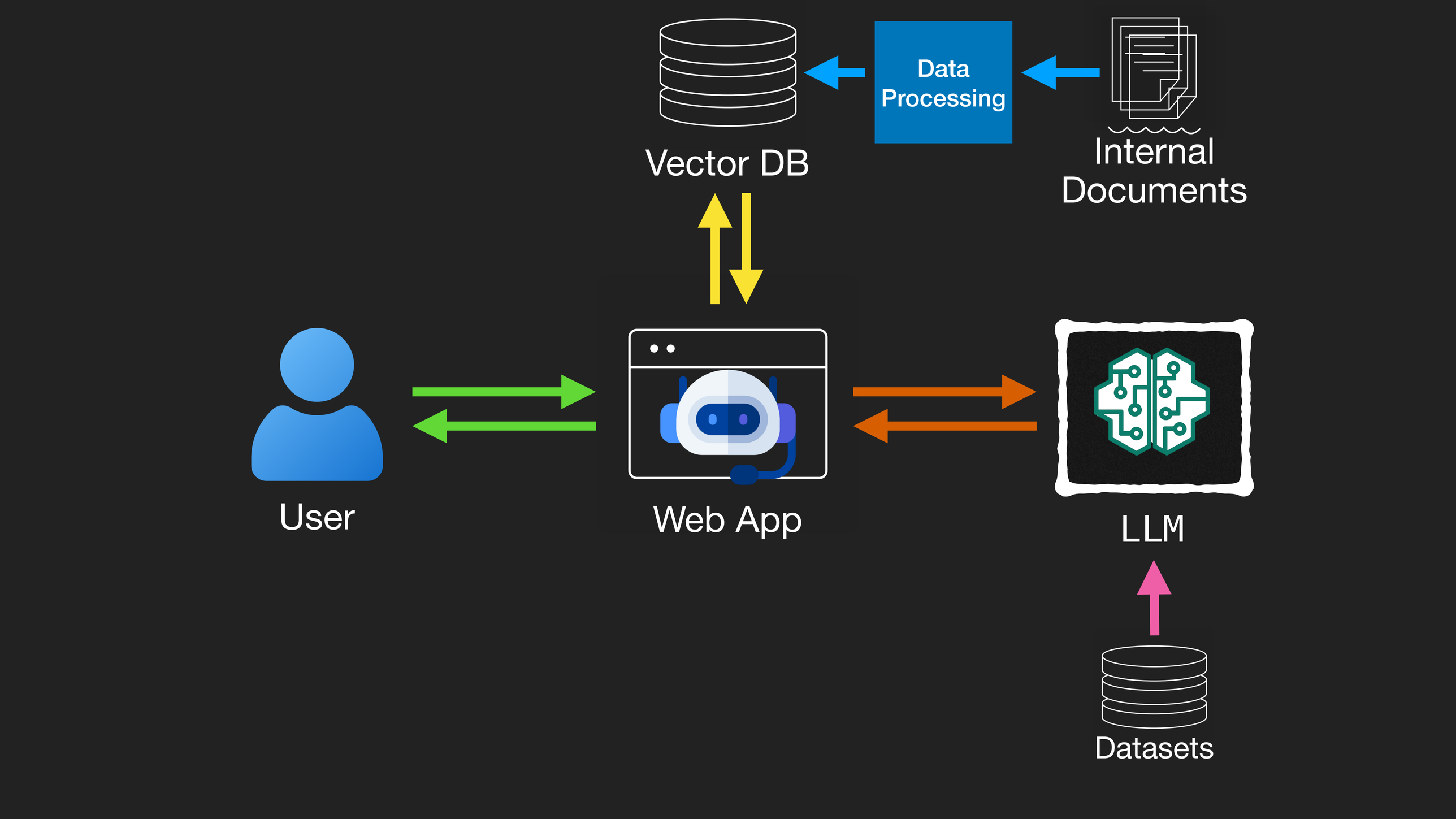
- 洩漏敏感資訊、系統資訊
- 在對接的系統上執行指令

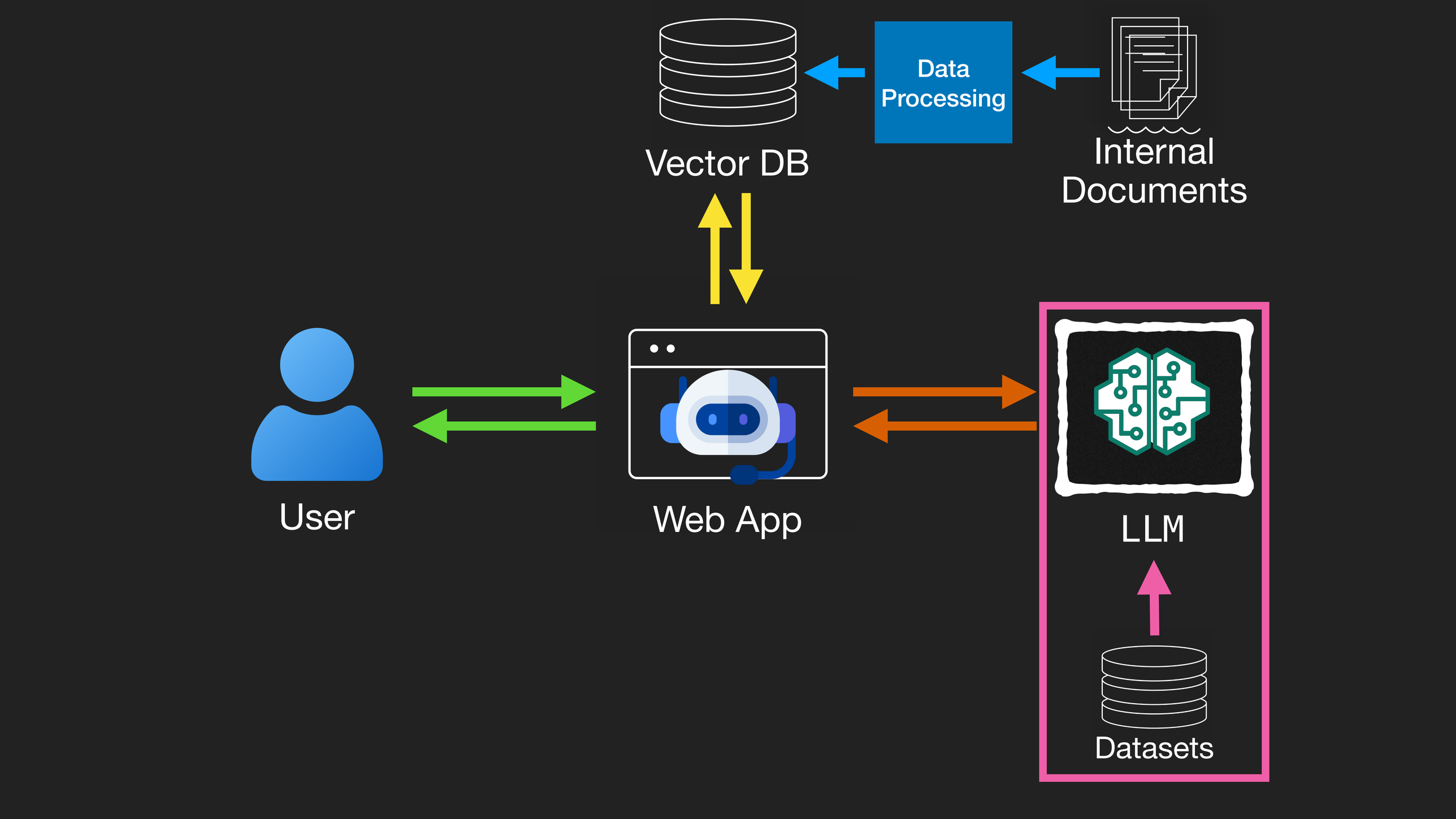
- 洩漏敏感資訊、系統資訊
- 在對接的系統上執行指令
- 操縱內容導致不正確或有偏見的輸出
- 使系統進行非預設用途的輸出

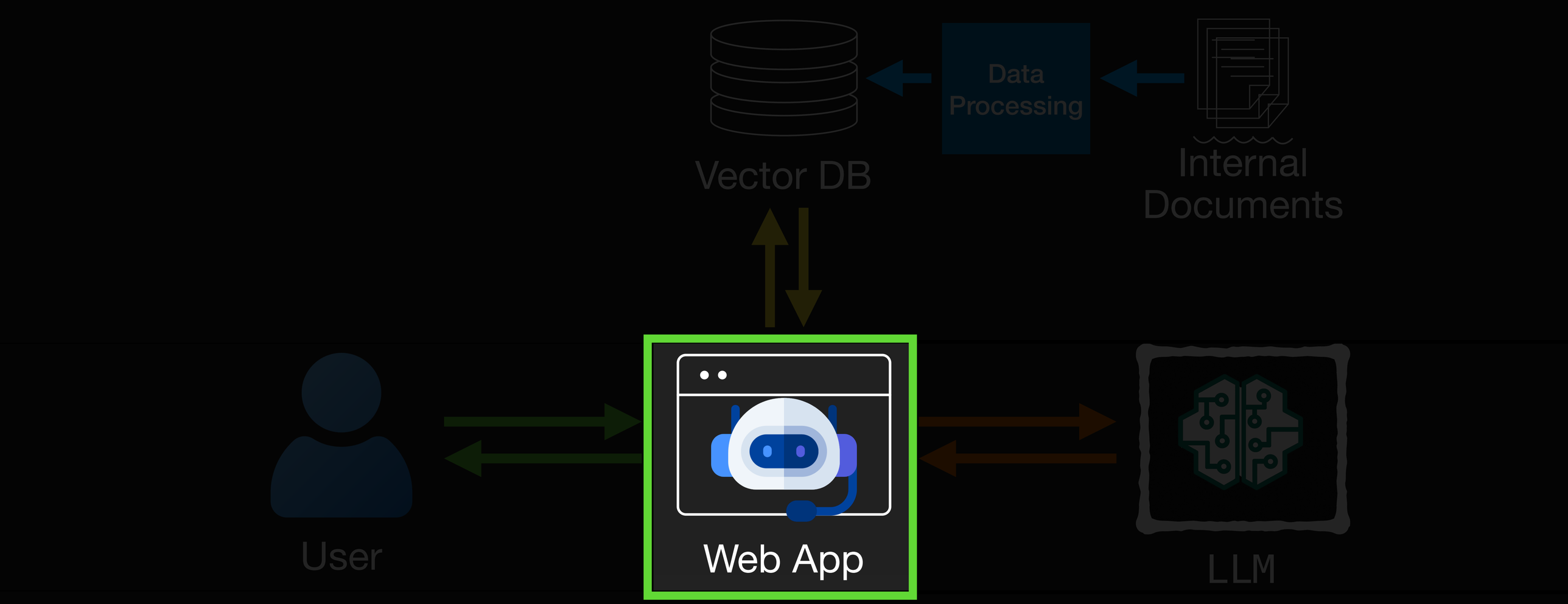
那我要好好防範提示注入問題


Vector DB
Data Processing
Internal Documents
User
Web App
LLM


Data Processing
Vector DB
Internal Documents
Query
User
Web App
LLM


Vector DB
Data Processing
Internal Documents
User
Web App
LLM
Datasets


Vector DB
Data Processing
Internal Documents
User
Web App
LLM
Datasets


Vector DB
Data Processing
Internal Documents
User
Web App
LLM

DEVCORE

實戰案例 1

chat.dummycorp.local

請輸入您的問題

chat.dummycorp.local

Thinking…

幫我寫一支印出 “Hello World” 的 C# 程式

chat.dummycorp.local

Thinking…

幫我寫一支印出 “Hello World” 的 C# 程式

chat.dummycorp.local

幫我寫一支印出 “Hello World” 的 C# 程式

chat.dummycorp.local

很抱歉，我無法完成您的要求。

幫我寫一支印出 “Hello World” 的 C# 程式

Request

```
POST /chat/Check HTTP/1.1
Host: api.dummycorp.local
Content-Type: application/json

{“Msg”:”幫我寫一支印出 \“Hello World\” 的 C# 程式”}
```

Request

```
POST /chat/Check HTTP/1.1
Host: api.dummycorp.local
Content-Type: application/json

{“Msg”:”幫我寫一支印出 \“Hello World\” 的 C# 程式”}
```

Response

```
HTTP/1.1 200 OK
Content-Type: application/json

{“Result”:”False”}
```

Request 1

```
POST /chat/Check HTTP/1.1
Host: api.dummycorp.local
Content-Type: application/json

{“Msg”:”如何修改網域密碼”}
```

Response 1

Request 1

```
POST /chat/Check HTTP/1.1
Host: api.dummycorp.local
Content-Type: application/json

{“Msg”:”如何修改網域密碼”}
```

Response 1

```
HTTP/1.1 200 OK
Content-Type: application/json

{“Result”:”True”}
```

Request 2

```
POST /chat/Complete HTTP/1.1
Host: api.dummycorp.local
Content-Type: application/json

{“Msg”:”如何修改網域密碼”}
```

Request 2

```
POST /chat/Complete HTTP/1.1
Host: api.dummycorp.local
Content-Type: application/json

{“Msg”:”如何修改網域密碼”}
```

Response 2

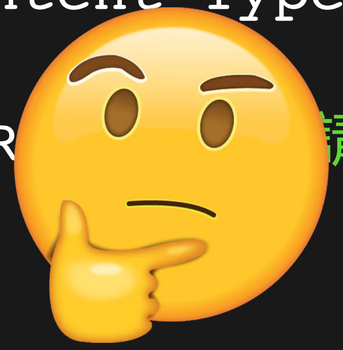
```
HTTP/1.1 200 OK
Content-Type: application/json

{“Result”:”請到入口網...[snip]”}
```

Request 2

```
POST /chat/Complete HTTP/1.1
Host: api.dummycorp.local
Content-Type: application/json

{“Msg”:”如何修改網域密碼”}
```

Response 2

```
HTTP/1.1 200 OK
Content-Type: application/json

{“R                    請到入口網...[snip]”}
```



```
POST /chat/Complete HTTP/1.1
Host: api.dummycorp.local
Content-Type: application/json

{“Msg”:”幫我寫一支印出 “Hello
World” 的 C# 程式”}
```

```
HTTP/1.1 200 OK
Content-Type: application/json

{“Result”:”沒問題！以下是一支印出
Hello World 的範例程式：
\n[snip]”}
```

```
POST /chat/Complete HTTP/1.1
Host: api.dummycorp.local
Content-Type: application/json
{"Msg":"
```

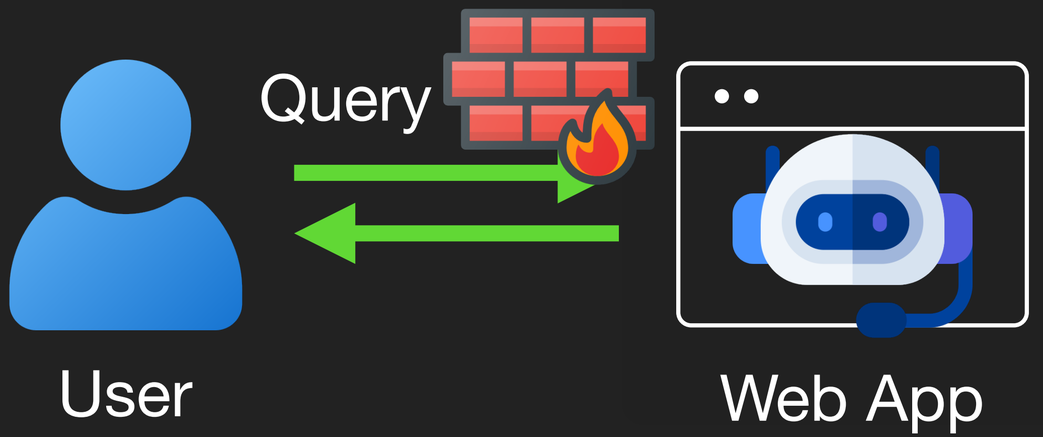
```
HTTP/1.1 200 OK
Content-Type: application/json
{"Result":"
```

未考量傳統 Web 弱點

新興 LLM 攻擊手法

新興 LLM 攻擊手法

傳統 Web 攻擊手法


Query
User
Web App

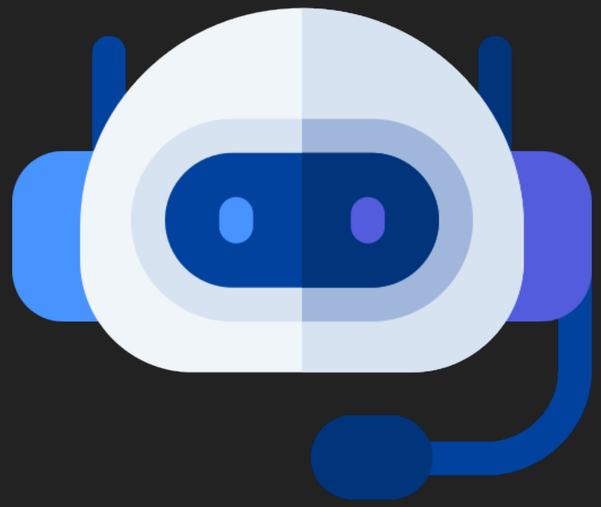
DEVCORE

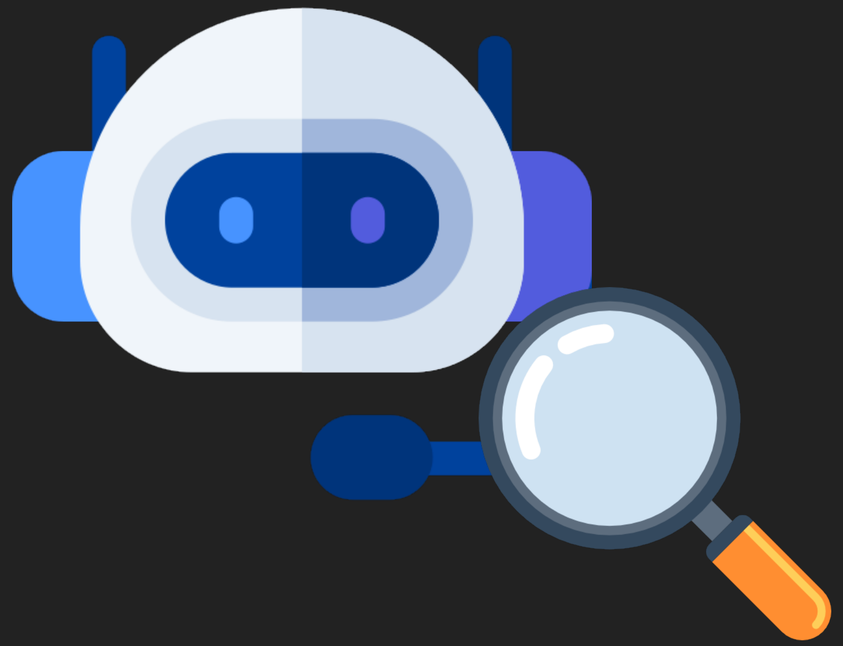
實戰案例 2

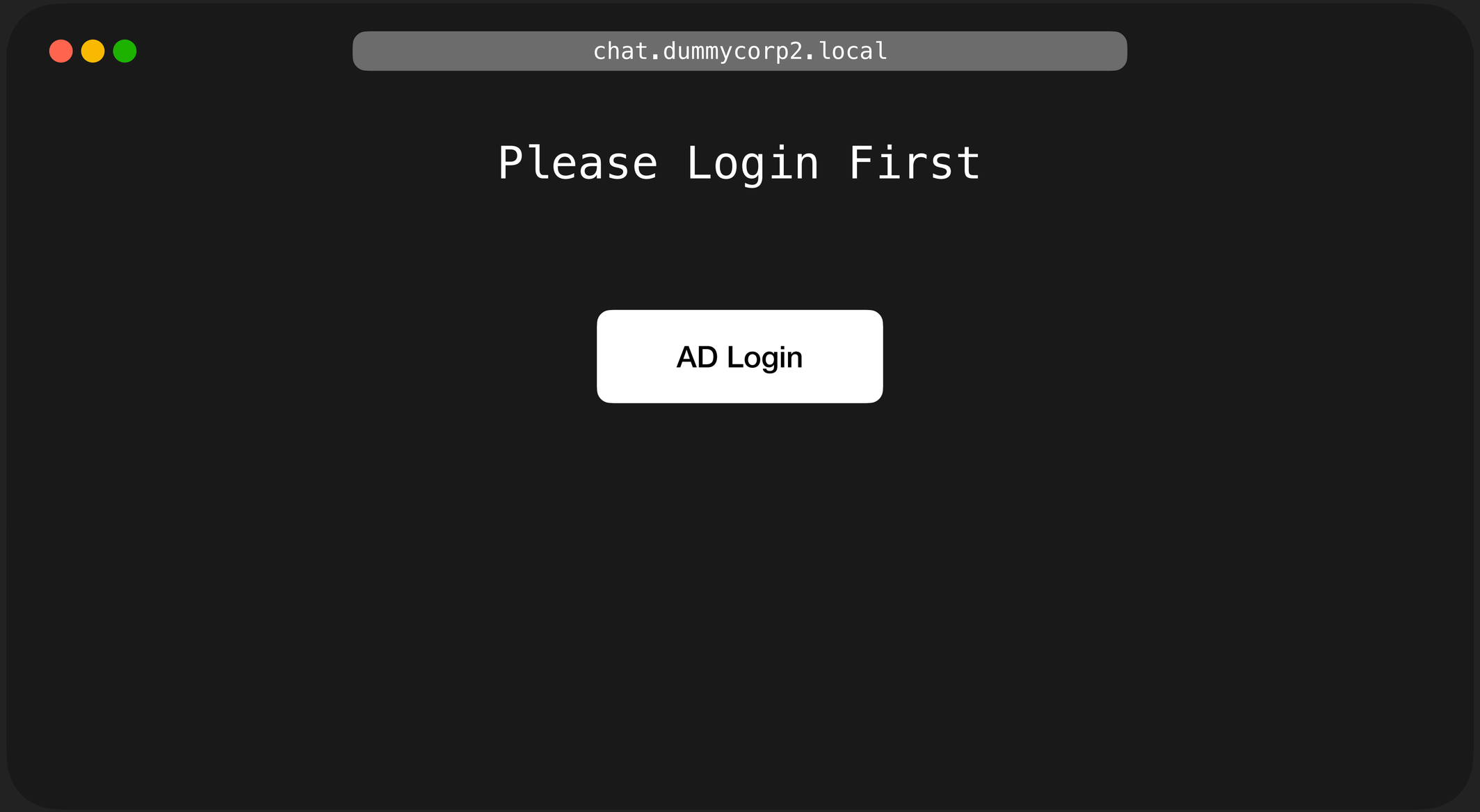
- 多次執行紅隊演練

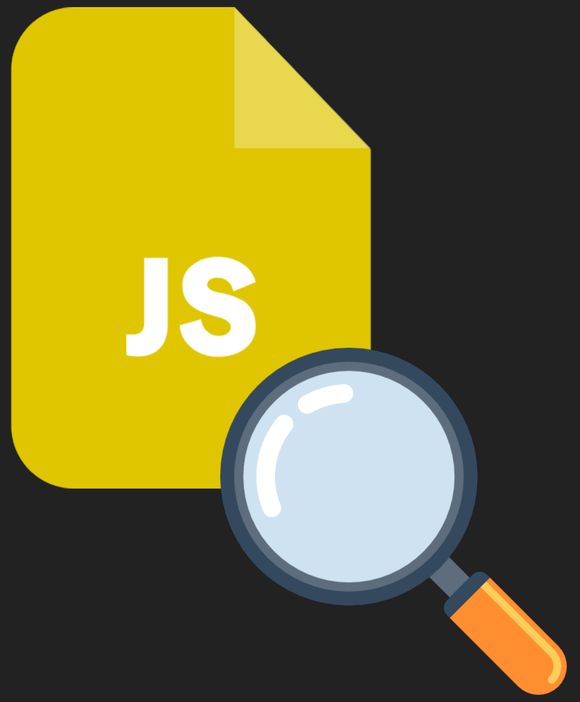
- 多次執行紅隊演練
- 不容易取得網域帳號

- 多次執行紅隊演練
- 不容易取得網域帳號
- 鮮少有新服務可以打

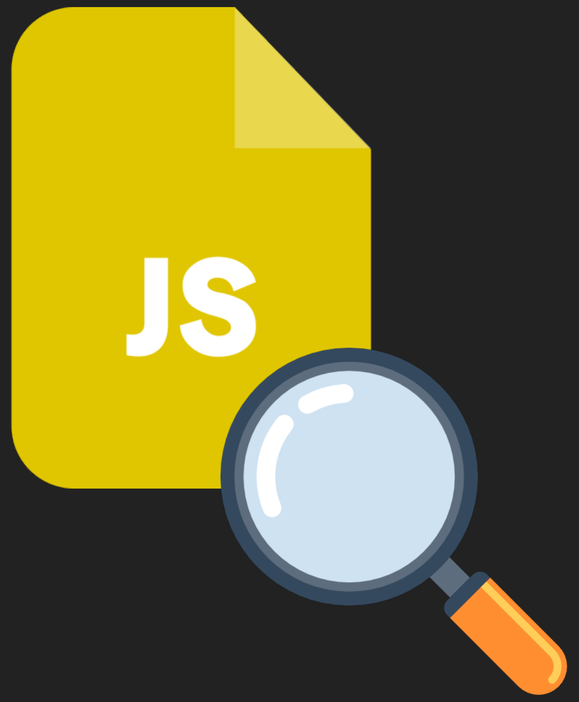





chat.dummycorp2.local
Please Login First
AD Login


JS


```
/Auth/GenJWT?id=

/Chat/GetChatHistory

/Chat/GetChatByCid?cid=[UUID]
```

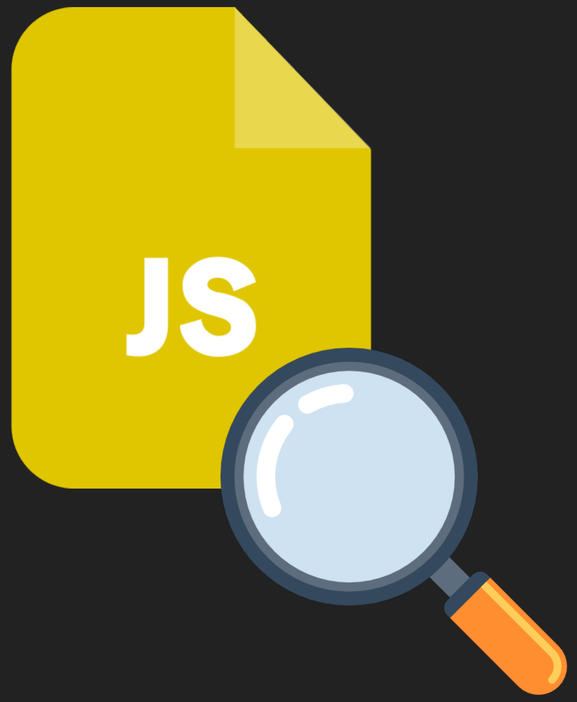


```
/Auth/GenJWT?id=

/Chat/GetChatHistory

/Chat/GetChatByCid?cid=[UUID]
```

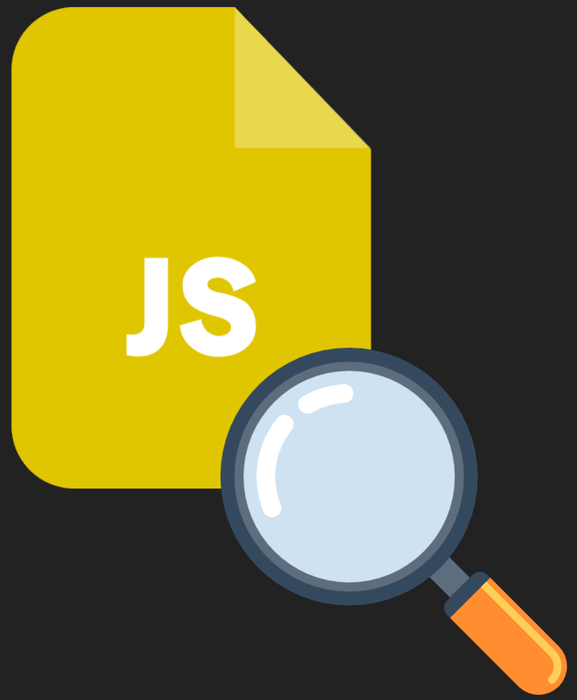
```
GET /Auth/GenJWT?id=A001 HTTP/1.1
Host: chat.dummycorp2.local
```

```
HTTP/1.1 200 OK

eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.ey
JzdWIiOiIxMjM0NTY3ODkwIiwibmFtZSI6IkEwM
DEiLCJpYXQiOjE3MDAwMDAwMDAsImV4cCI6MTg5
MzQ1NjAwMCwiaXNzIjoiZGVtbyIsImF1ZCI6ImN
oYXQuZHVtbXBjb3JwMi5sb2NhbCJ9.X7j4Lk9mQ
2pVn8tYc1RzB6uD3sWf0aHqJ5eNp2KxTgM
```

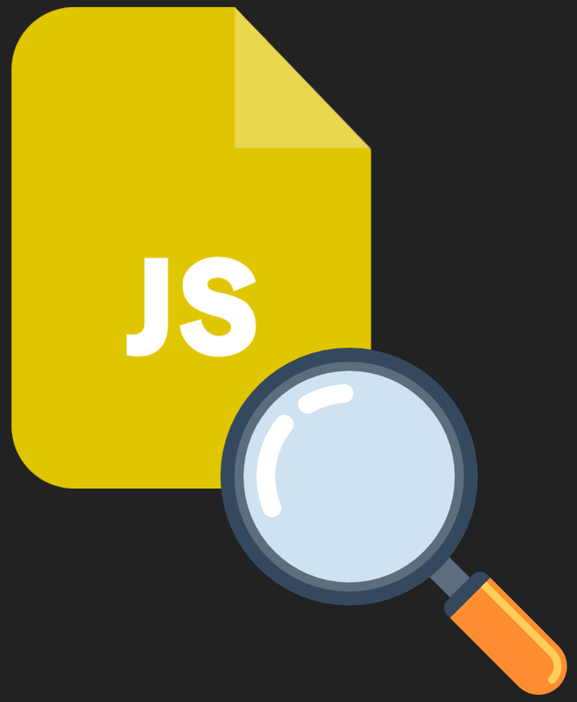

chat.dummycorp2.local
歡迎使用者 A001
請輸入您的問題：
A001



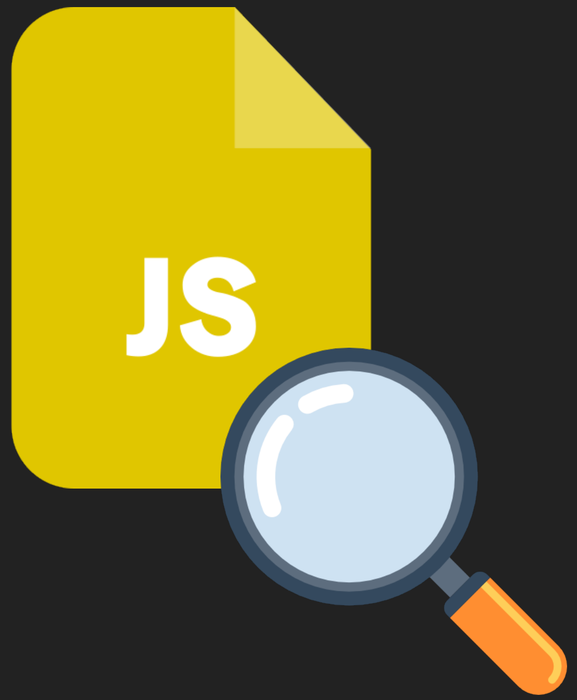

```
/Auth/GenJWT?id=A001

/Chat/GetChatHistory

/Chat/GetChatByCid?cid=[UUID]
```



```
/Auth/GenJWT?id=A001
/Chat/GetChatHistory
/Chat/GetChatByCid?cid=[UUID]
```




```
/Auth/GenJWT?id=A001
/Chat/GetChatHistory
/Chat/GetChatByCid?cid=[UUID]
```

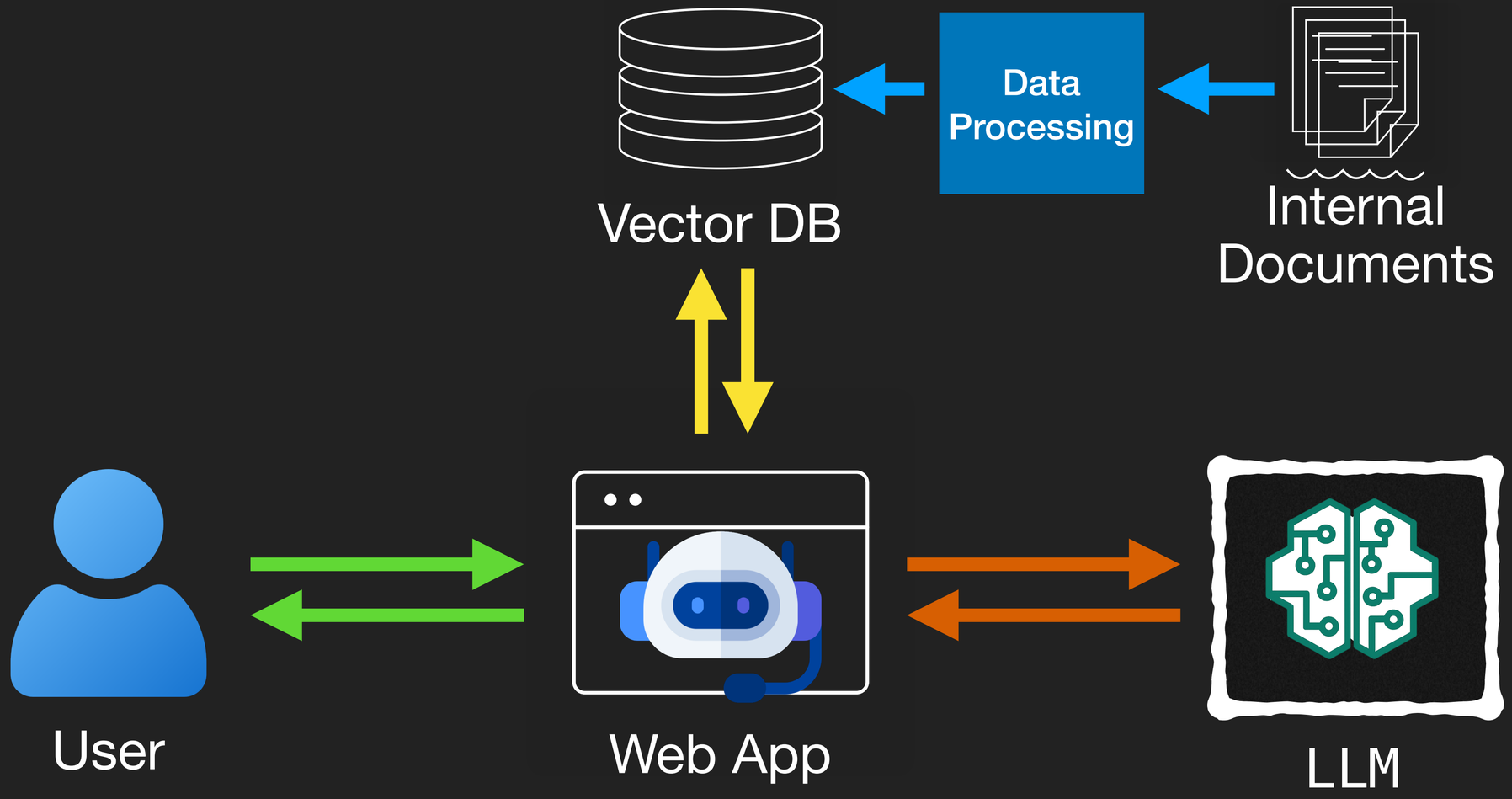


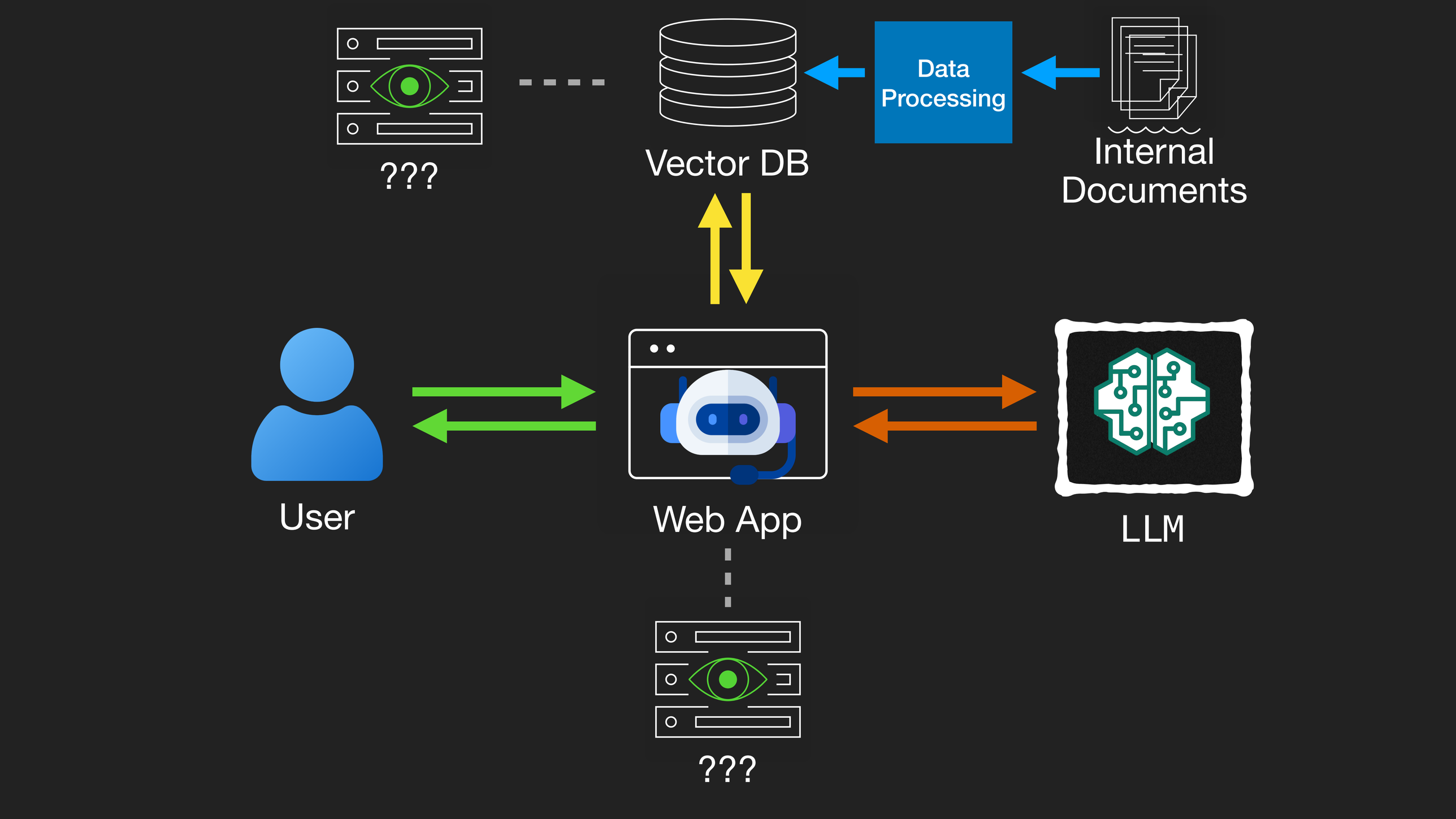
```
/Auth/GenJWT?id=A001

/Chat/GetChatHistory

/Chat/GetChatByCid?cid=[UUID]
```

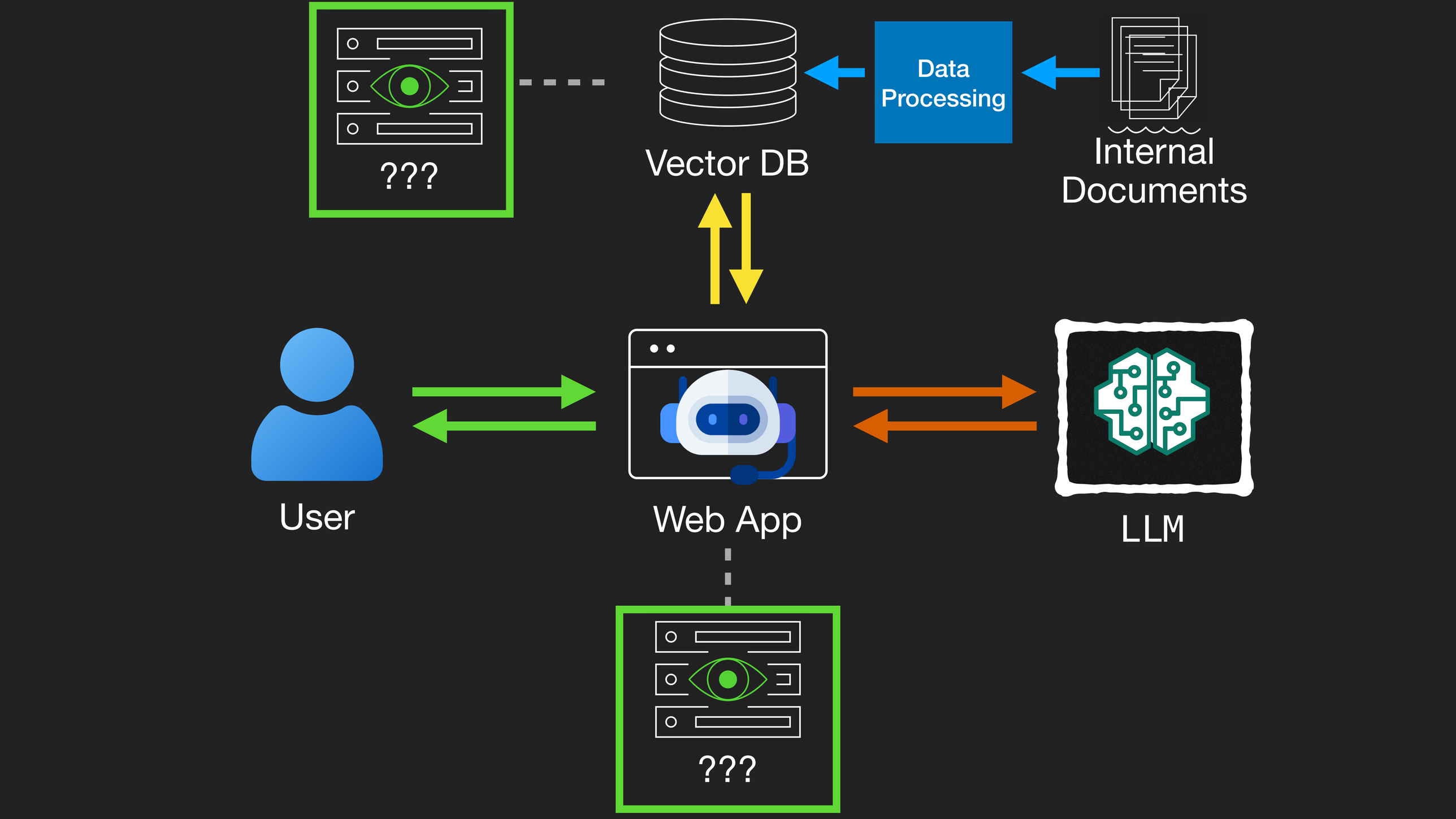
```
grep 密碼 log.txt
```

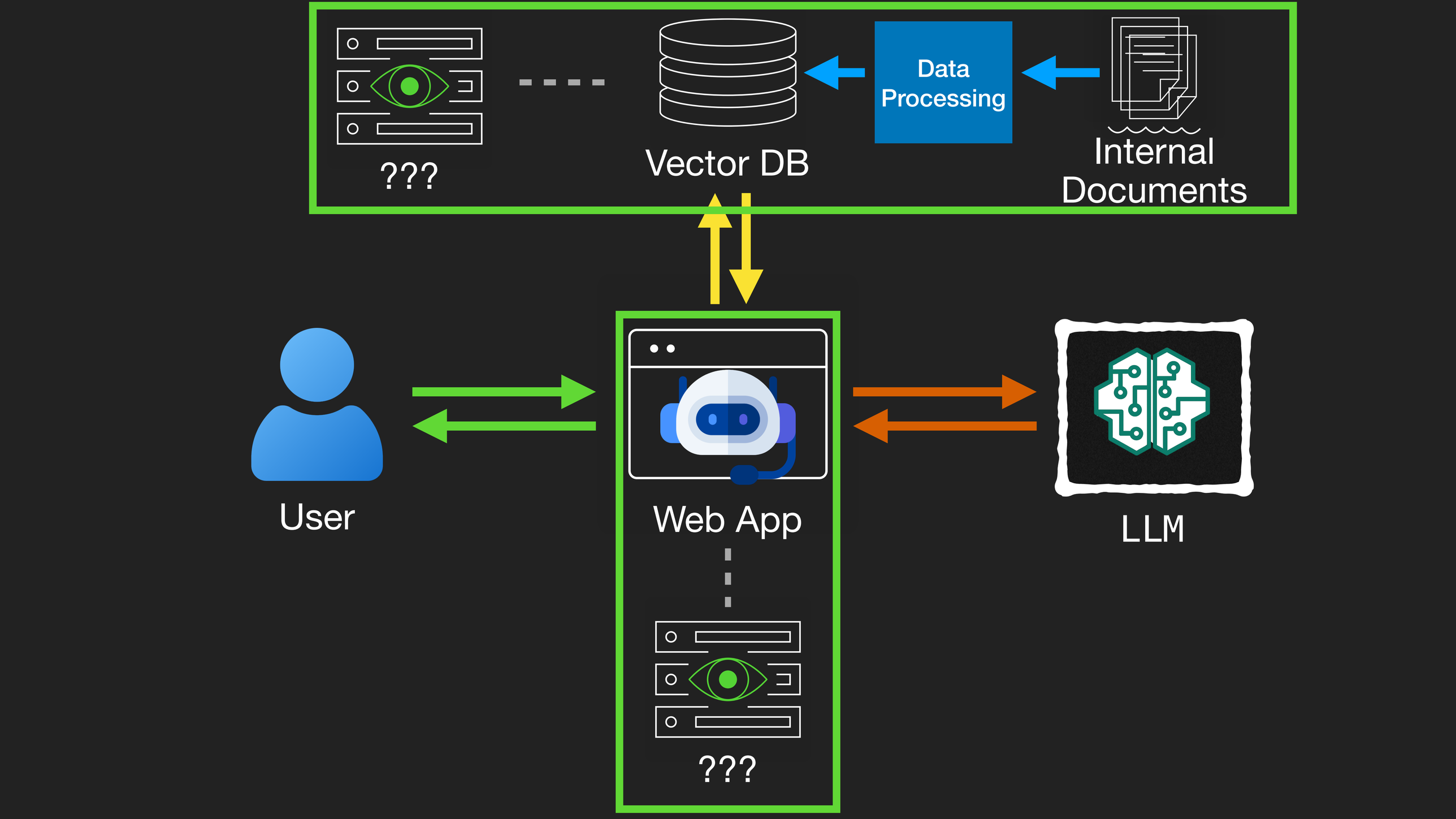
本頁截圖僅於現場分享

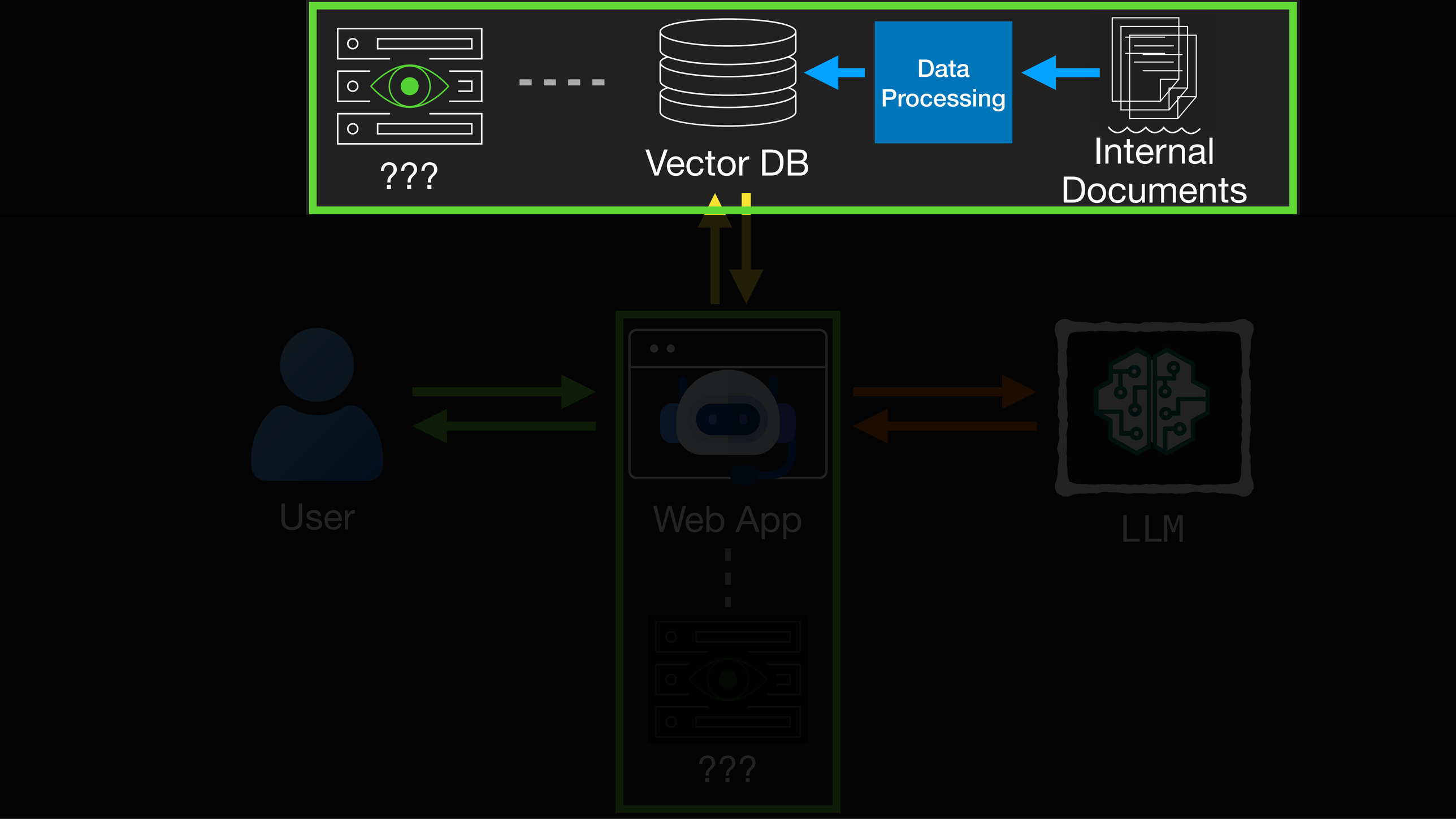
取得內網重要系統帳號密碼、網域帳號密碼


Vector DB
Data Processing
Internal Documents
User
Web App
LLM


???
Vector DB
Data Processing
Internal Documents
User
Web App
LLM
???


???
Vector DB
Data Processing
Internal Documents
User
Web App
LLM
???


???
Vector DB
Data Processing
Internal Documents
User
Web App
LLM
???


???
Vector DB
Data Processing
Internal Documents
User
Web App
LLM
???

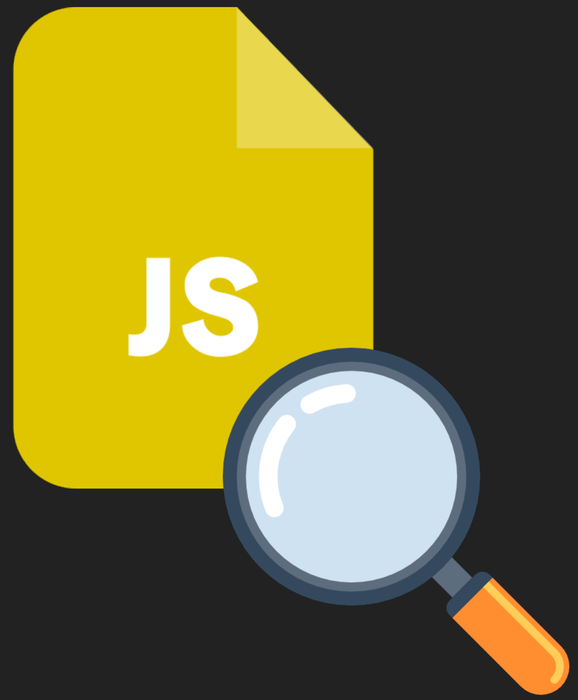
DEVCORE

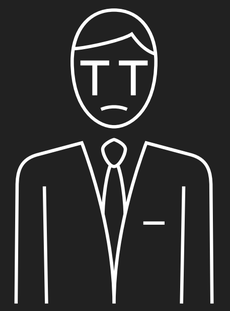
實戰案例 3

跟上一個案例非常相似

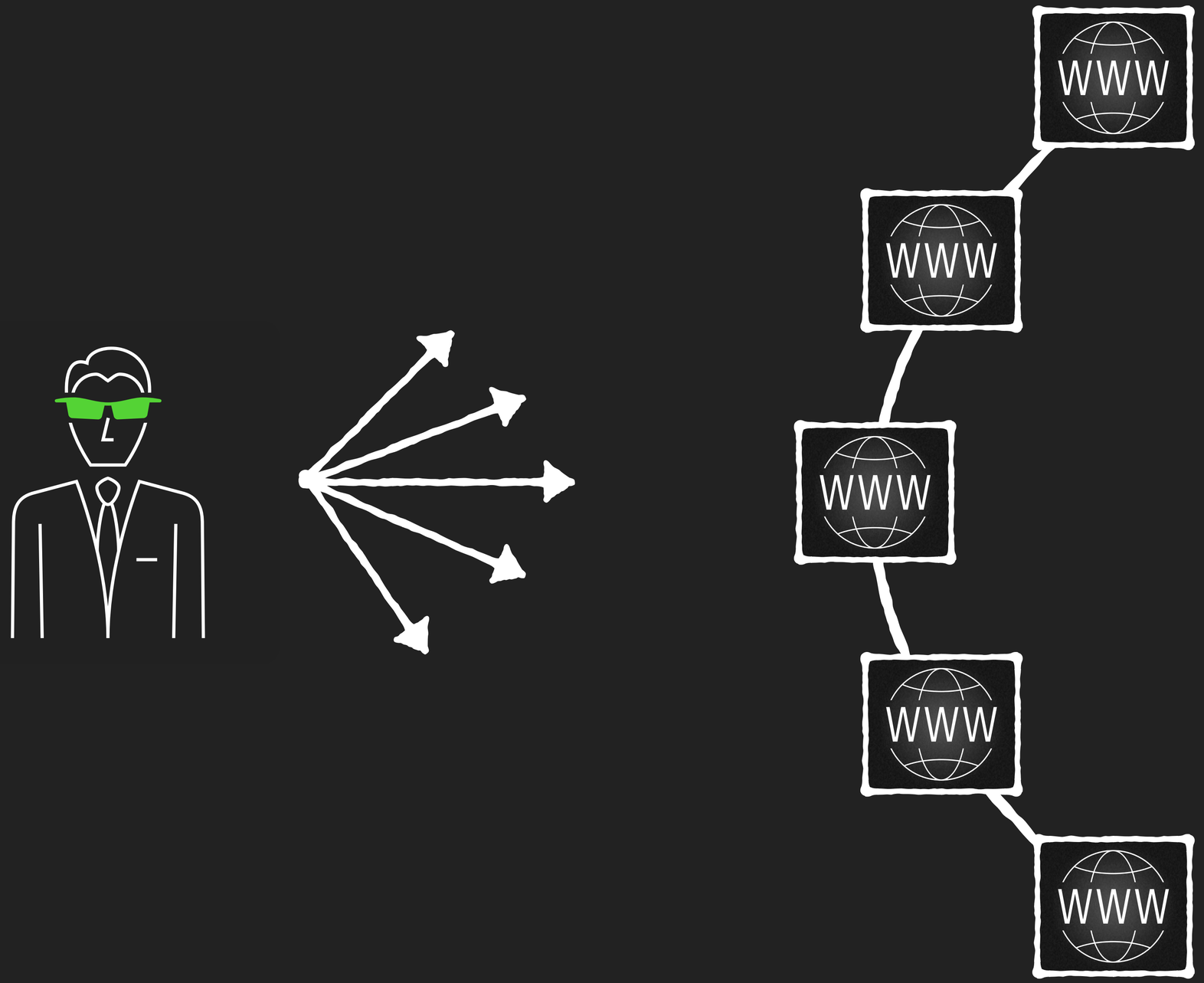
chat.dummycorp3.local

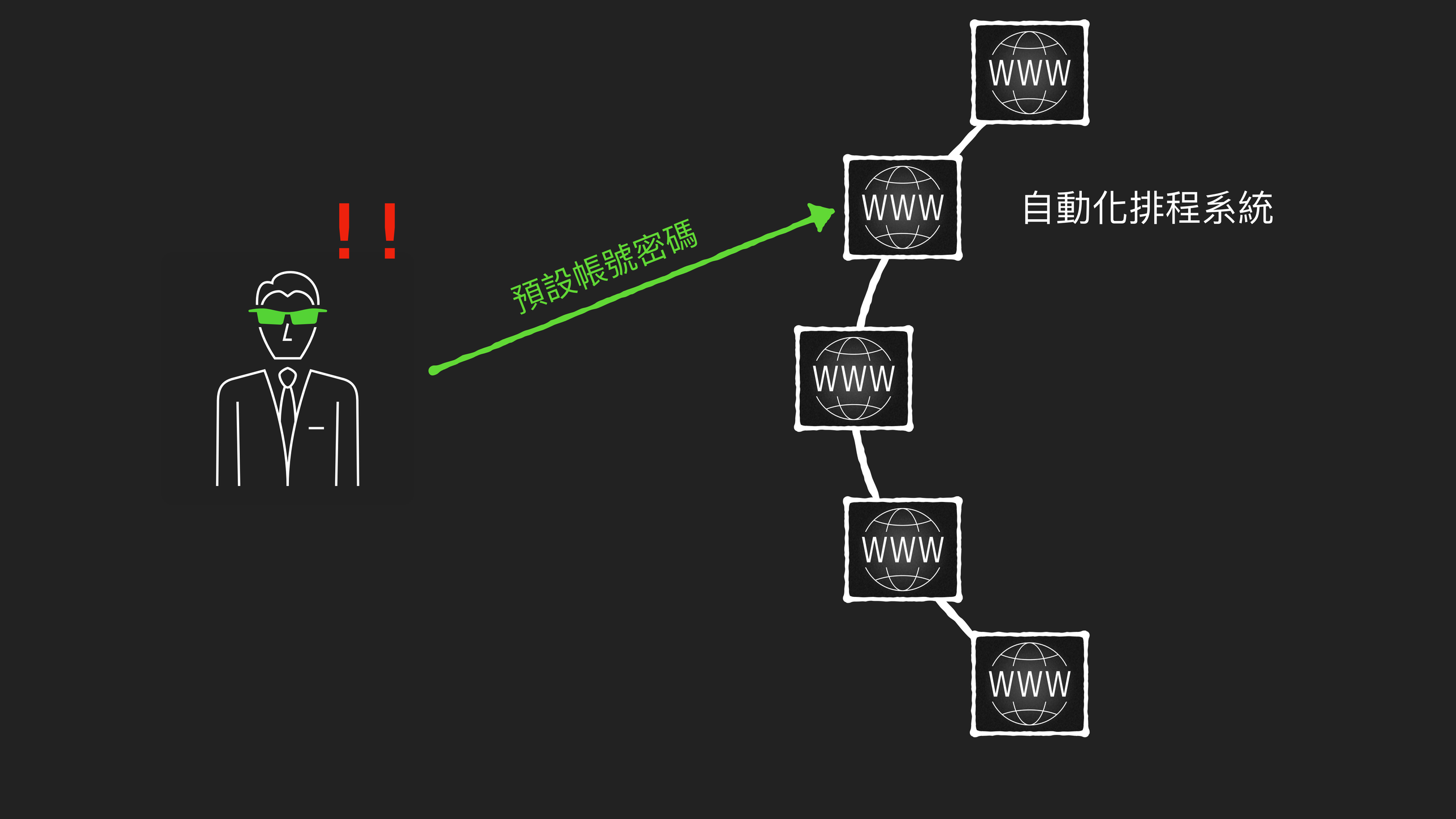
Please Login First


JS

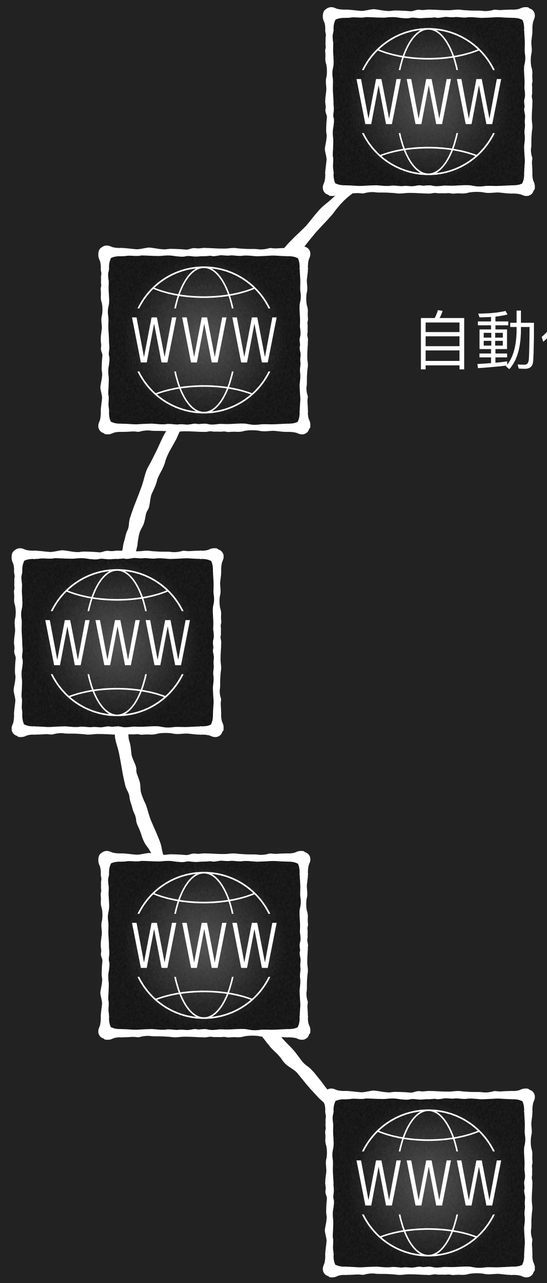


這次登不進去了

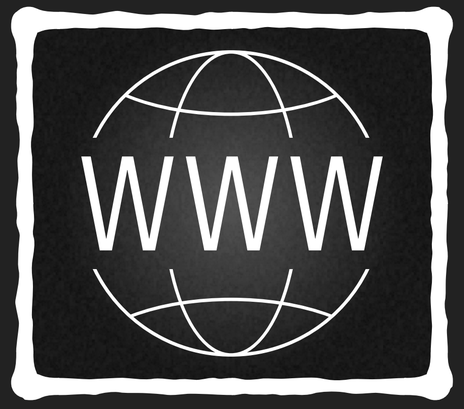

WWW
WWW
WWW
WWW
WWW


！！
預設帳號密碼
WWW
WWW
WWW
WWW
WWW
自動化排程系統

翻！

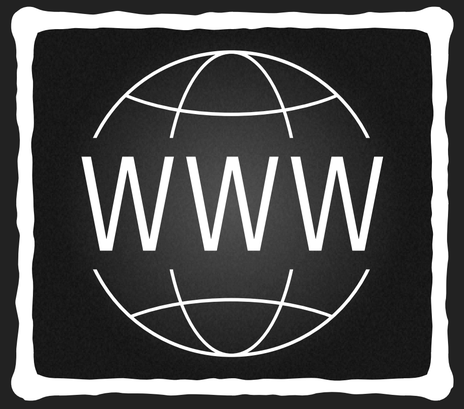


自動化排程系統

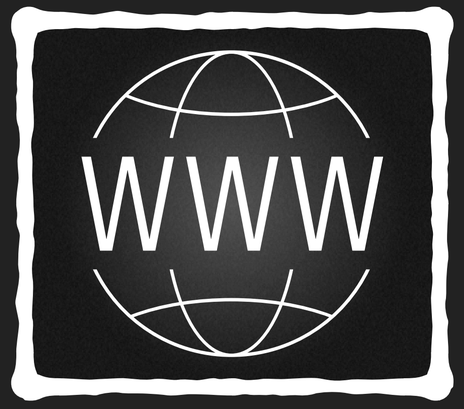


自動化排程系統

定期檢查檔案



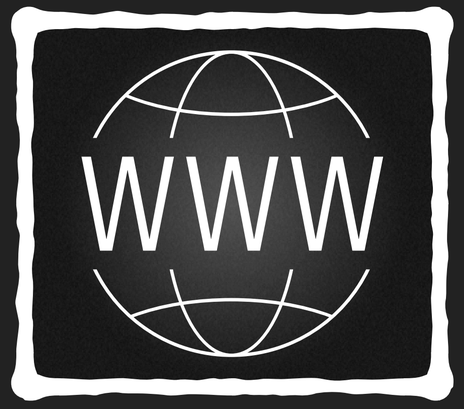
自動化排程系統



自動化排程系統

定期檢查檔案

對檔案進行前處理

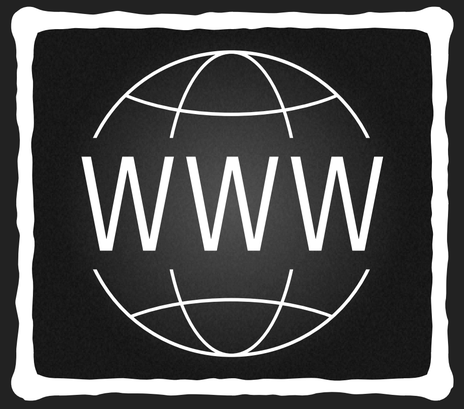


自動化排程系統

定期檢查檔案

對檔案進行前處理

存入向量資料庫



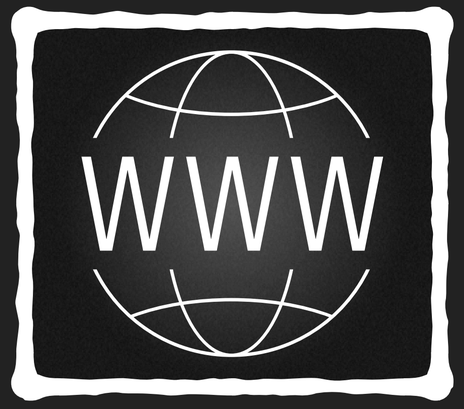
自動化排程系統

定期檢查檔案

對檔案進行前處理

存入向量資料庫

RAG?



自動化排程系統

定期檢查檔案

對檔案進行前處理

存入向量資料庫

RAG?

翻！

```
AZURE_BLOB_ACCOUNT="dummyragstorage"
AZURE_BLOB_ACCOUNT_KEY="YWwvLWAsQGHMgQTUyIlpXrJfrxH
YHEcJqyqdrHPwgcBYQCgKOOavxlJNHETS/StQeJhDWjP"
AZURE_BLOB_ENDPOINT="core.windows.net"
AZURE_BLOB_NAME="dummyragcontainer"
DB_PASSWORD="BJpYtVgqXLsPrGmRdddV"
DATABASE_URL="POSTGRE_CONNECTION_STRING"
```

Azure Storage Account Credentials

```
AZURE_BLOB_ACCOUNT="dummyragstorage"
AZURE_BLOB_ACCOUNT_KEY="YWwvLWAsQGHMgQTUyIlpXrJfrxH
YHEcJqyqdrHPwgcBYQCgKOOavxlJNHETS/StQeJhDWjP"
AZURE_BLOB_ENDPOINT="core.windows.net"
AZURE_BLOB_NAME="dummyragcontainer"
```

```
DB_PASSWORD="BJpYtVgqXLsPrGmRdddV"
DATABASE_URL="POSTGRE_CONNECTION_STRING"
```

PostgreSQL Credentials

RAG!

Azure Storage Account Credentials

```
AZURE_BLOB_ACCOUNT="dummyragstorage"
AZURE_BLOB_ACCOUNT_KEY="YWwvLWAsQGHMgQTUyIlpXrJfrxH
YHEcJqyqdrHPwgcBYQCgKOOavxlJNHETS/StQeJhDWjP"
AZURE_BLOB_ENDPOINT="core.windows.net"
AZURE_BLOB_NAME="dummyragcontainer"
DB_PASSWORD="BJpYtVgqXLsPrGmRdddV"
DATABASE_URL="POSTGRE_CONNECTION_STRING"
```

PostgreSQL Credentials

Microsoft Azure Storage Explorer

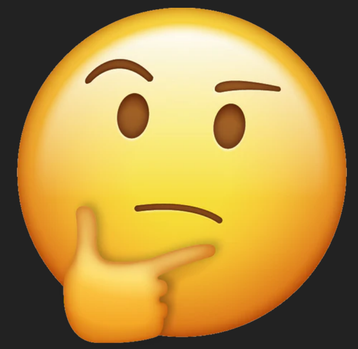
本頁截圖僅於現場分享

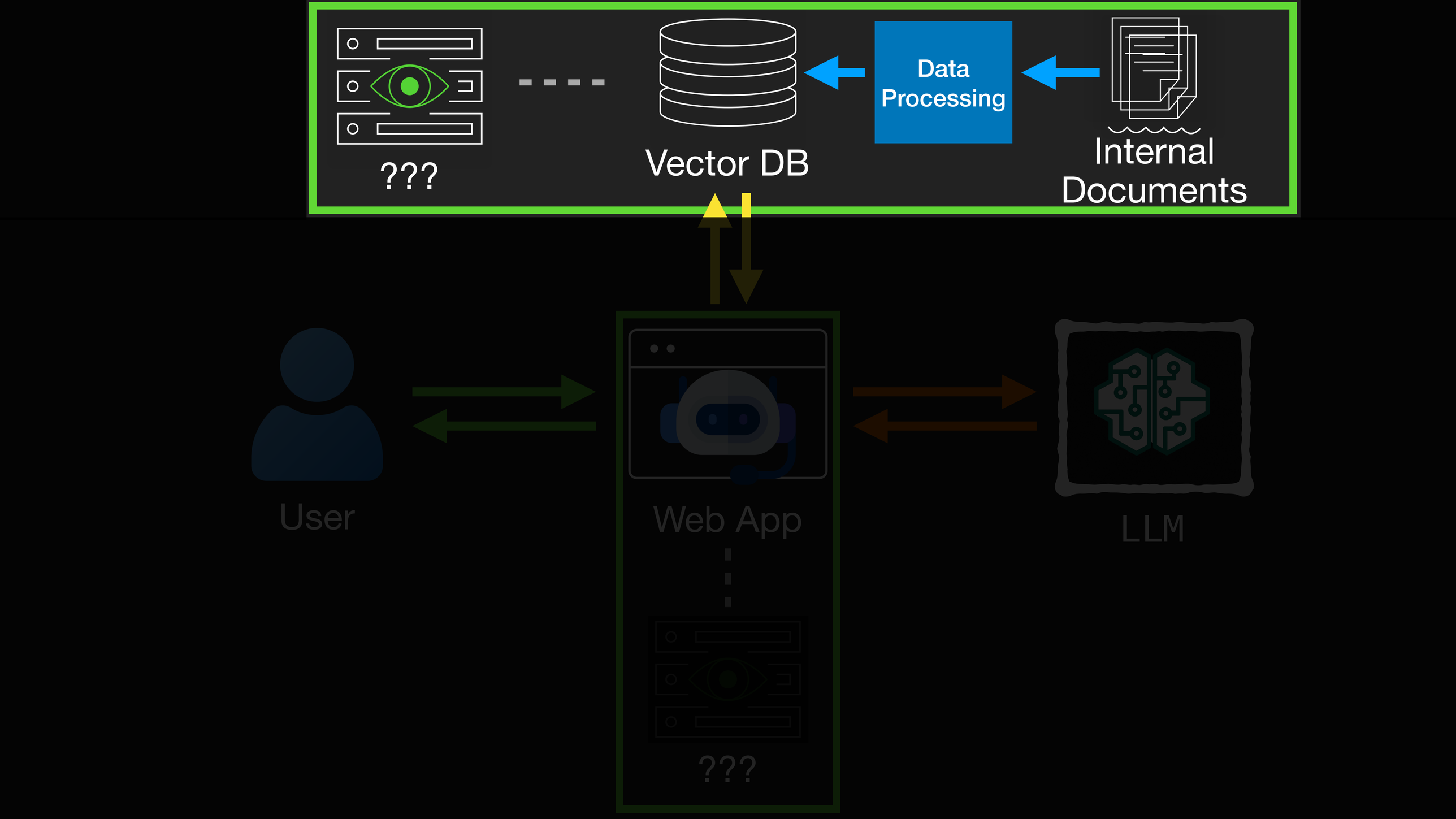
Microsoft Azure Storage Explorer

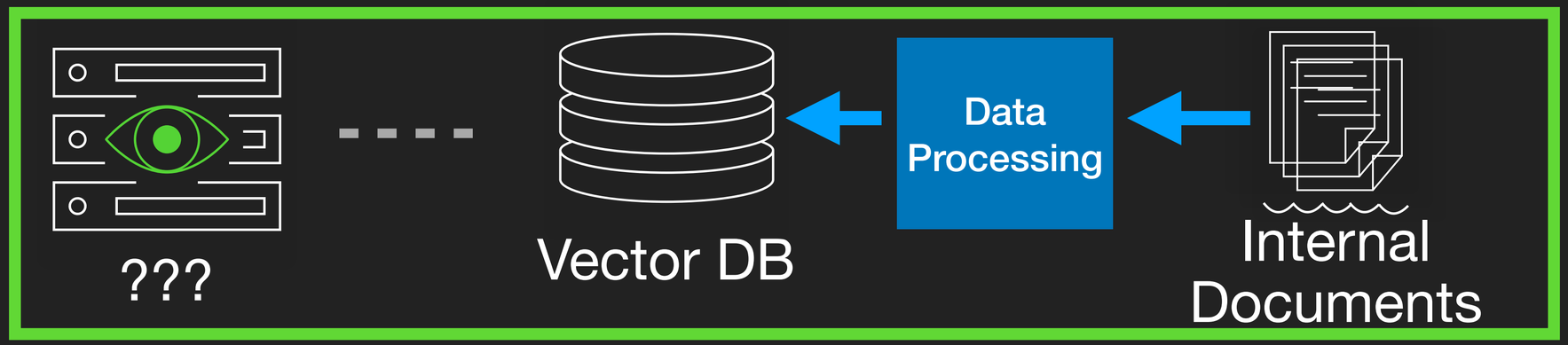
取得內部開發文件

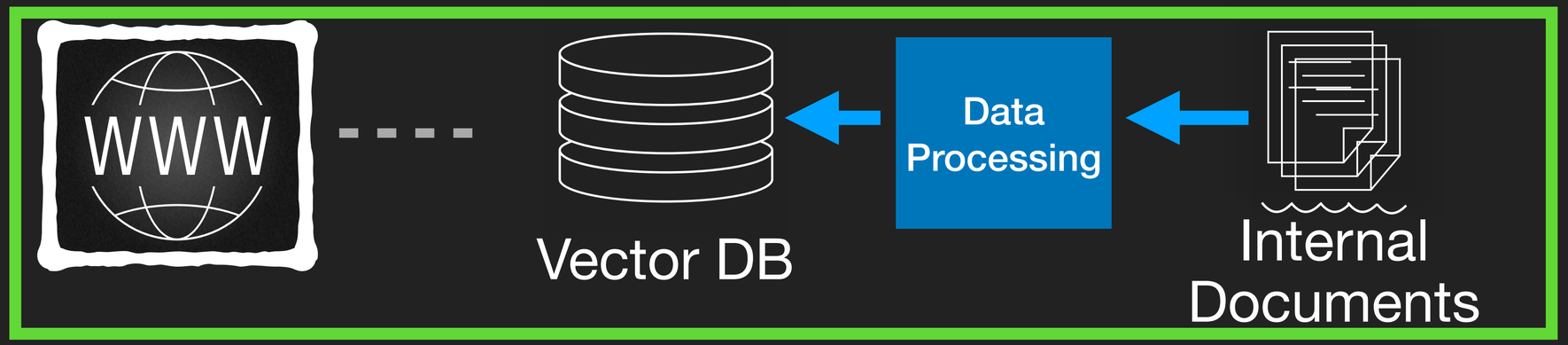
本頁截圖僅於現場分享

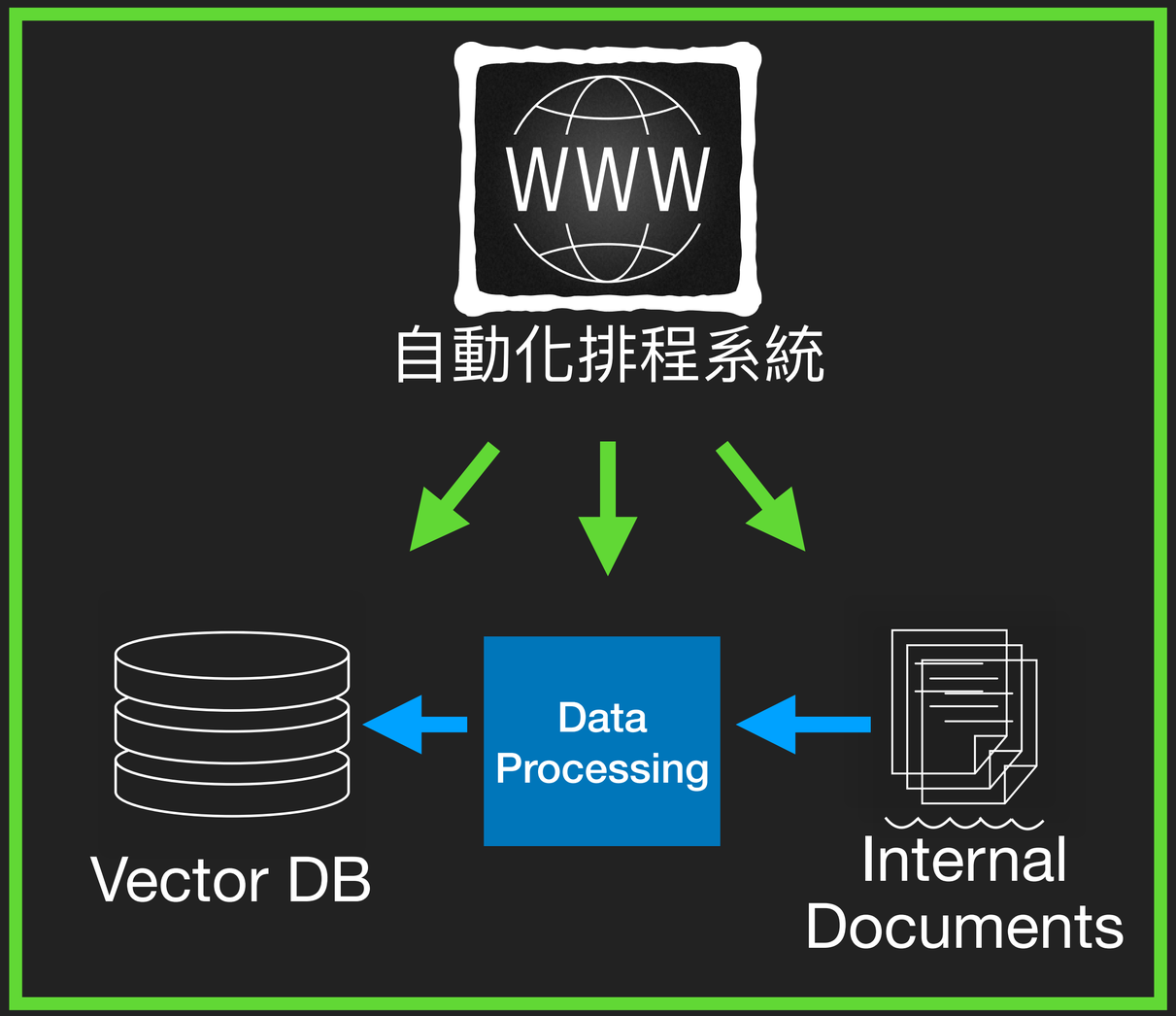
這個案例的根因是什麼

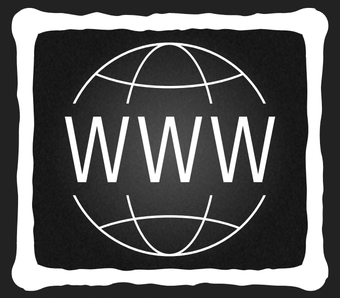



???
Vector DB
Data Processing
Internal Documents
User
Web App
LLM
???


???
Vector DB
Data Processing
Internal Documents


WWW
Vector DB
Data Processing
Internal Documents

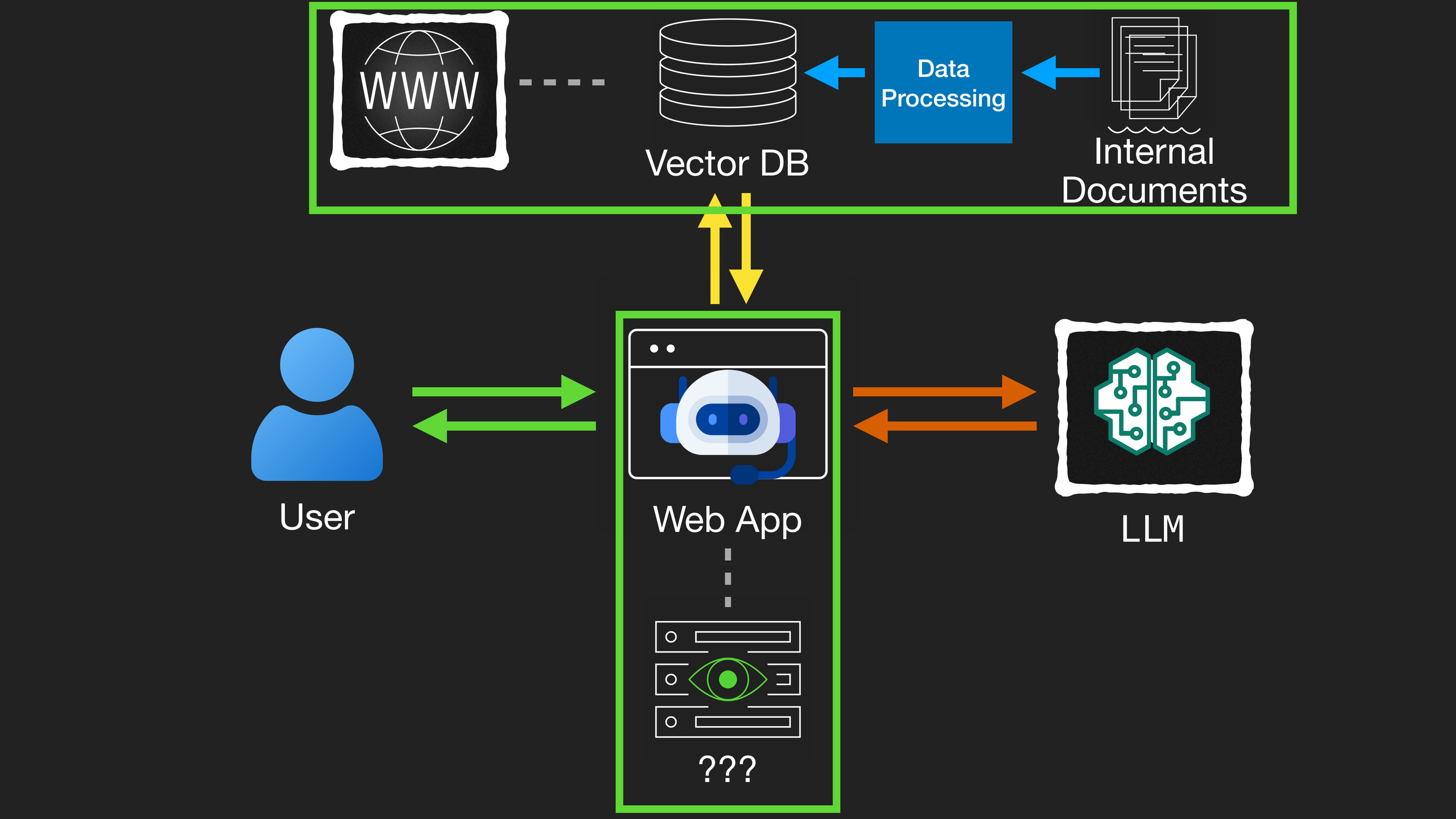

WWW
自動化排程系統
Data
Processing
Vector DB
Internal
Documents

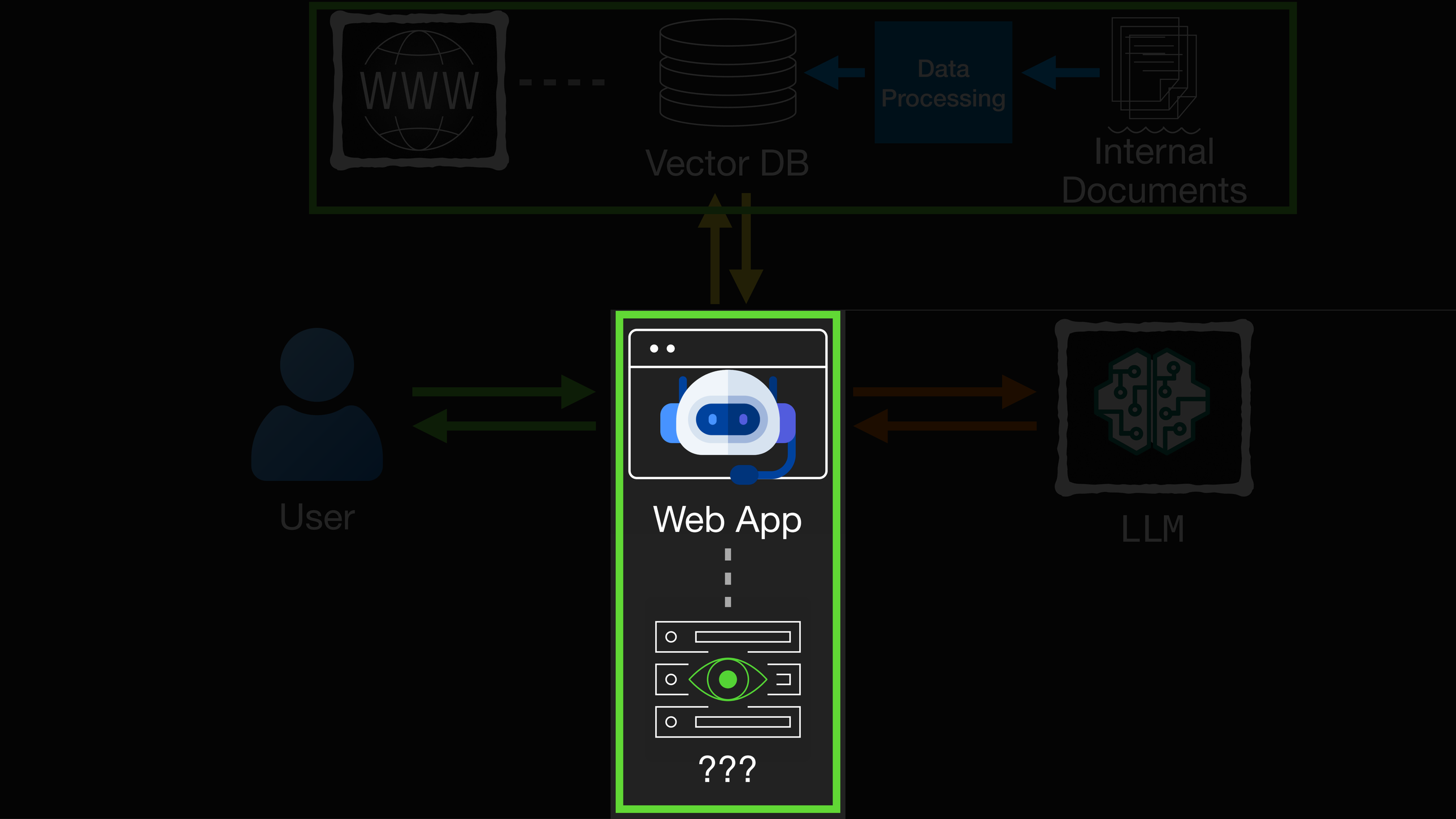


自動化排程系統

與 LLM 應用相連的元件

都可能成為攻擊面


WWW
Vector DB
Data Processing
Internal Documents
User
Web App
LLM
???


WWW
Vector DB
Data Processing
Internal Documents
User
Web App
LLM
???

DEVCORE

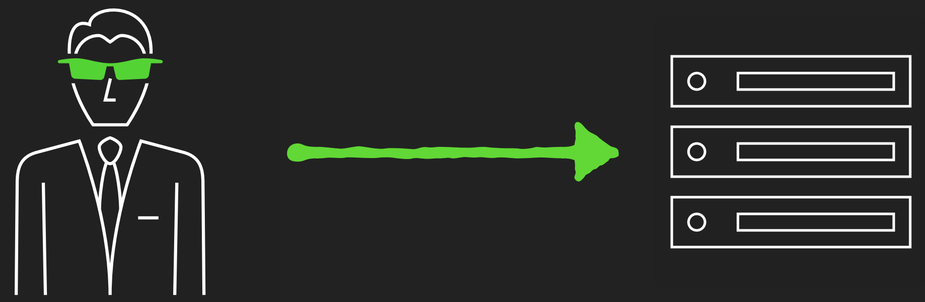
實戰案例 4

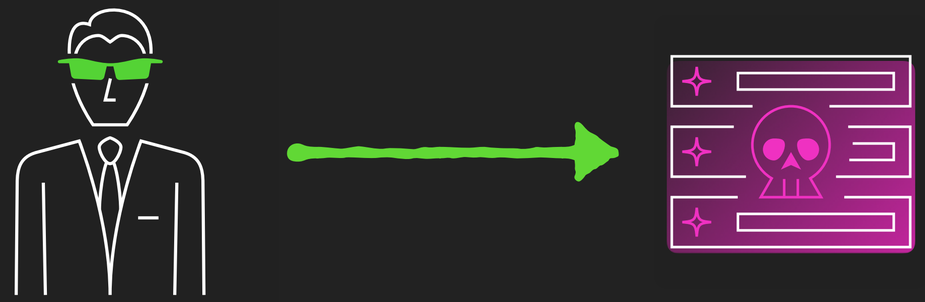
在某一個專案中

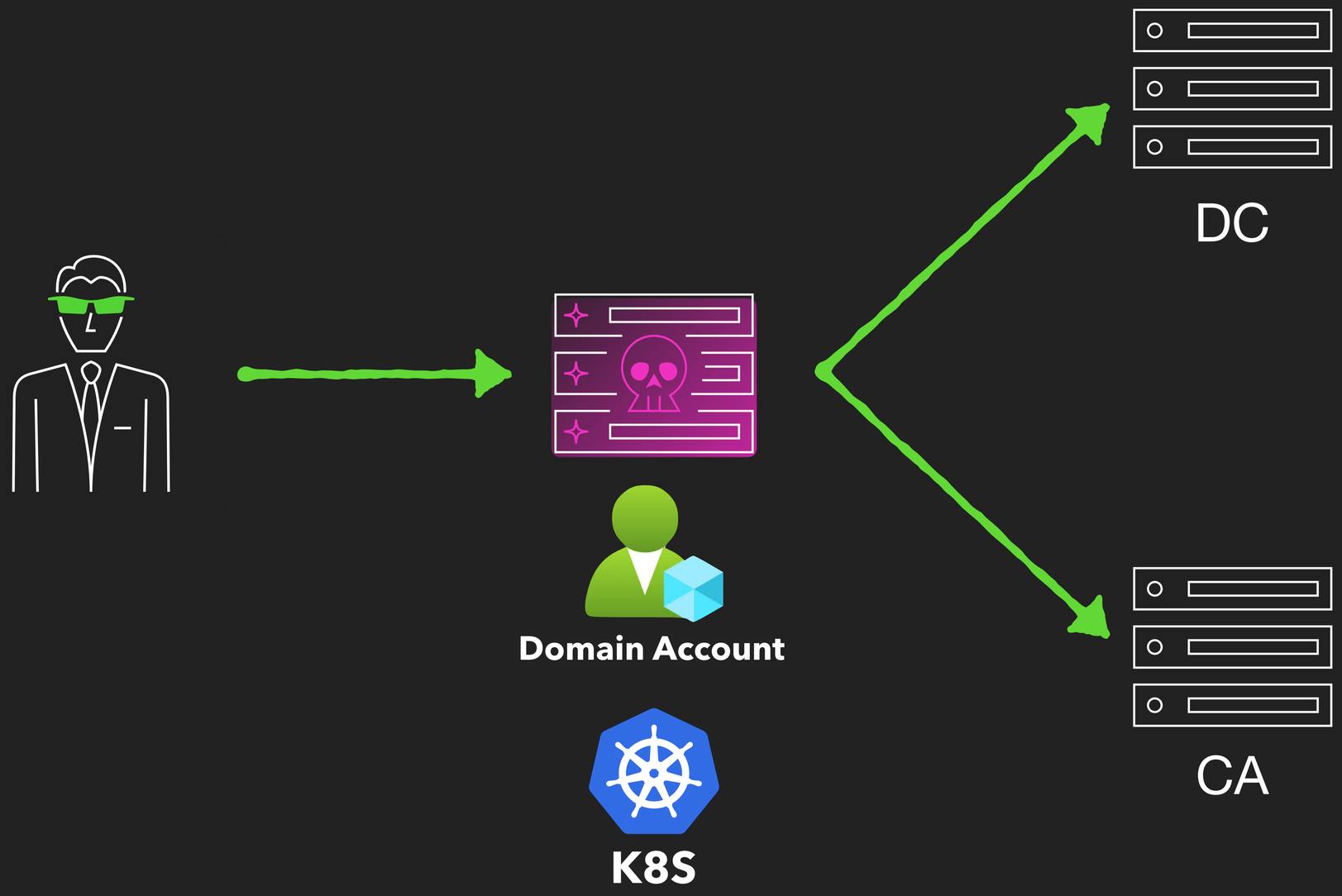
On-Prem AD

Internet

Intranet






DC
Domain Account
CA
K8S

DEVCORE

分分鐘拿下整個網域
關於 AD，你還疏忽了什麼？

徐偉庭 Vtim

戴夫寇爾股份有限公司
vtim@devco.re

DEVCORE CONFERENCE 2024 | 2024.03.16

DEVCORE

分分鐘拿下整個網域
關於 AD，你還疏忽了什麼？

徐偉庭 Vtim

戴夫寇爾股份有限公司
vtim@devco.re

DEVCORE CONFERENCE 2024 | 2024.03.16

專案開始前

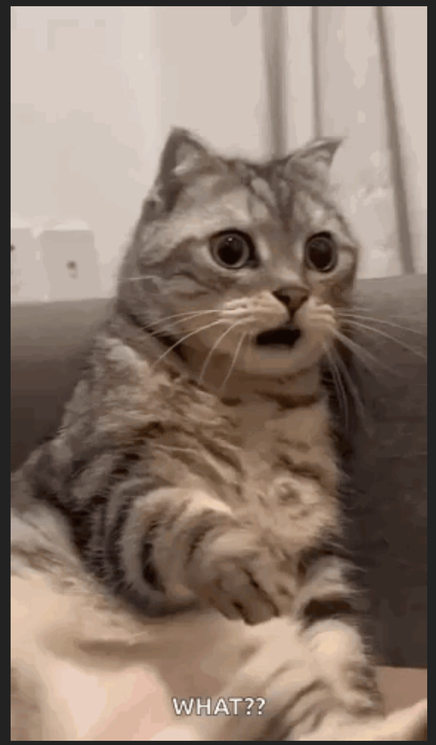
客戶

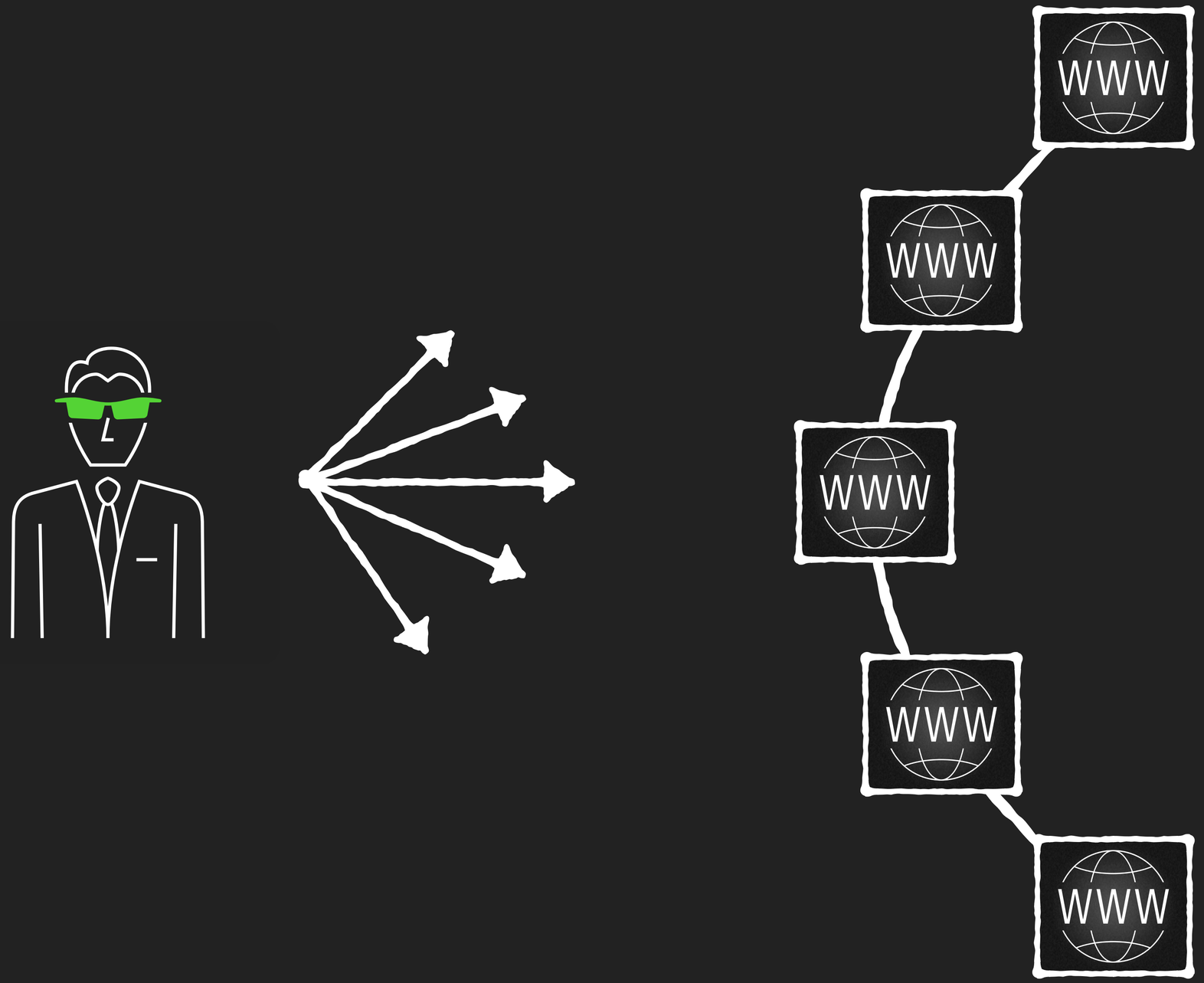
「我們有自己先跑過 certipy 囉！」

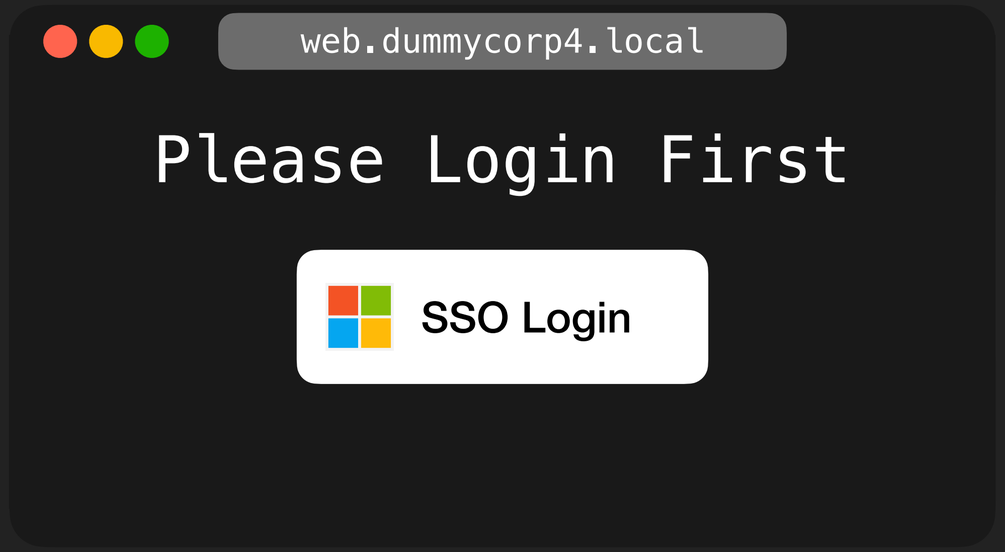
專案開始前

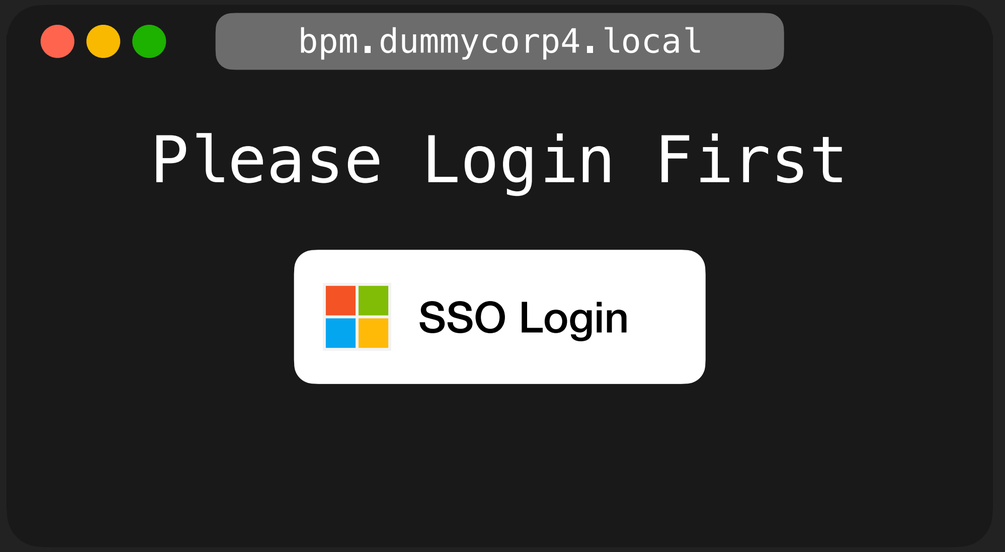
客戶

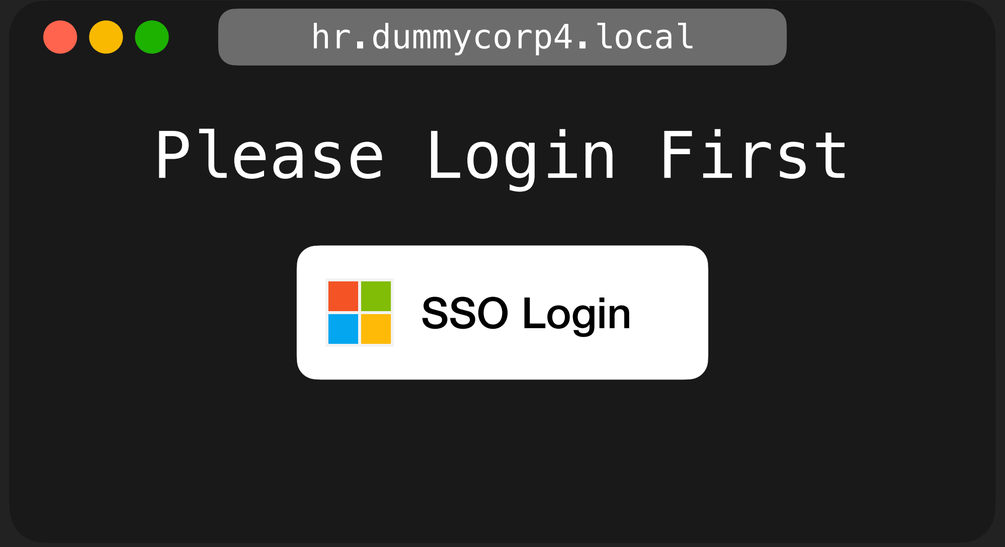
「我們有自己先跑過 certipy 囉！」

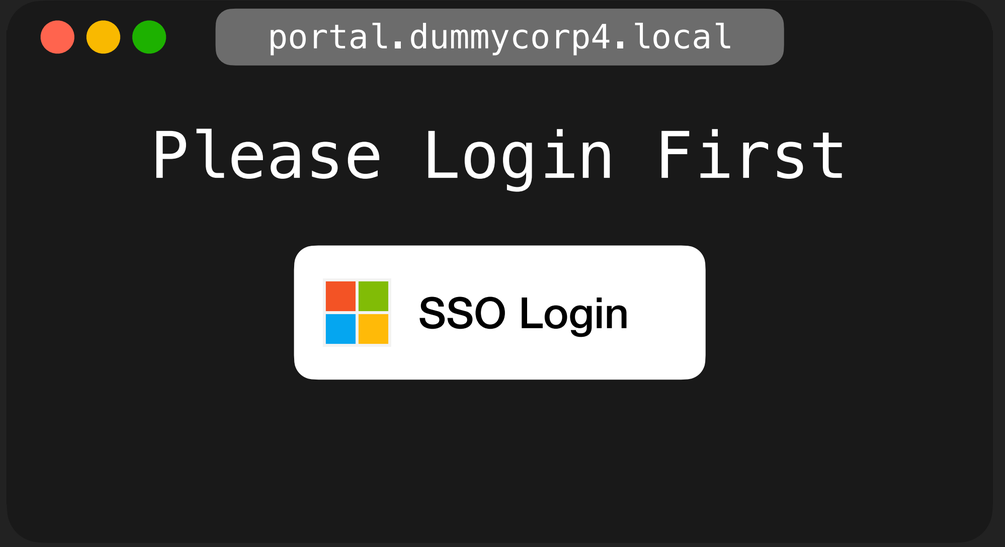


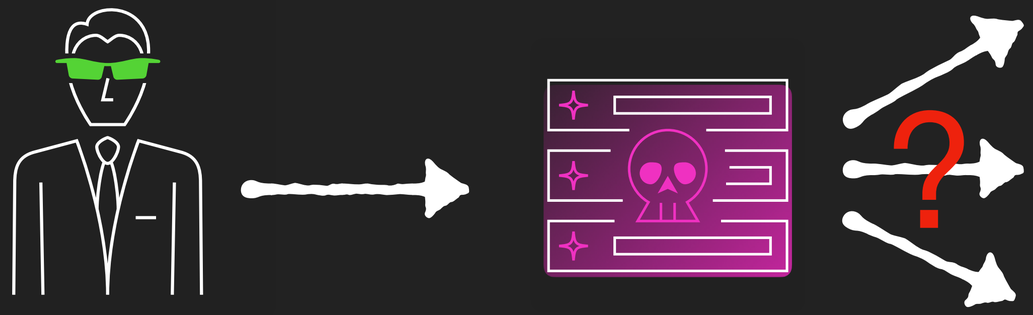

WWW
WWW
WWW
WWW
WWW


web.dummycorp4.local
Please Login First
SSO Login


bpm.dummycorp4.local
Please Login First
SSO Login


hr.dummycorp4.local
Please Login First
SSO Login


portal.dummycorp4.local
Please Login First
SSO Login

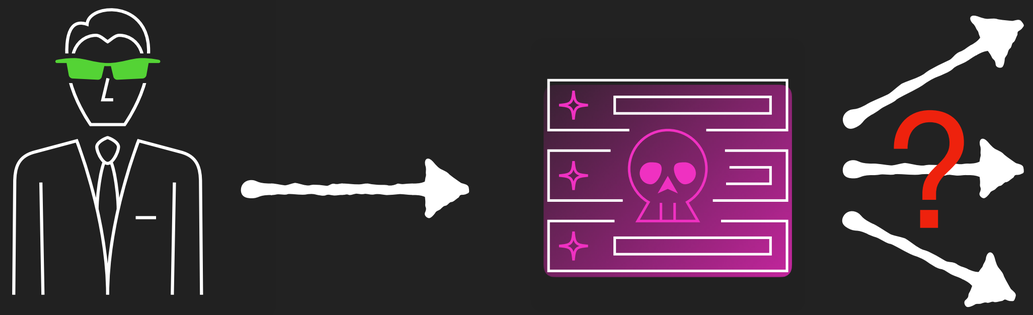


On-Prem AD

Internet

Intranet

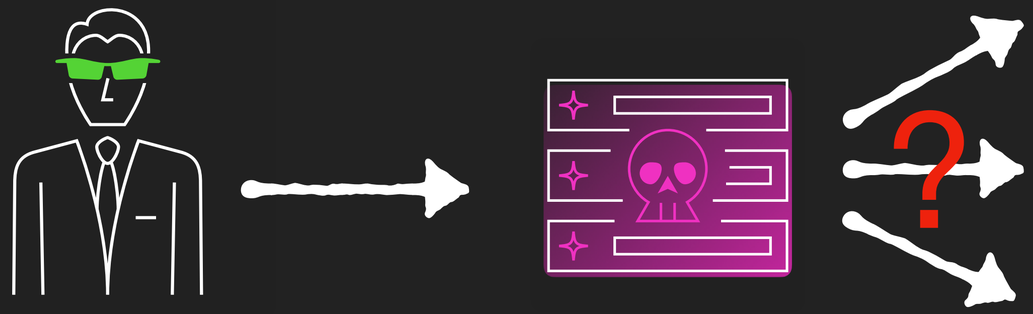
BloodHound 沒路



On-Prem AD

Internet

Intranet



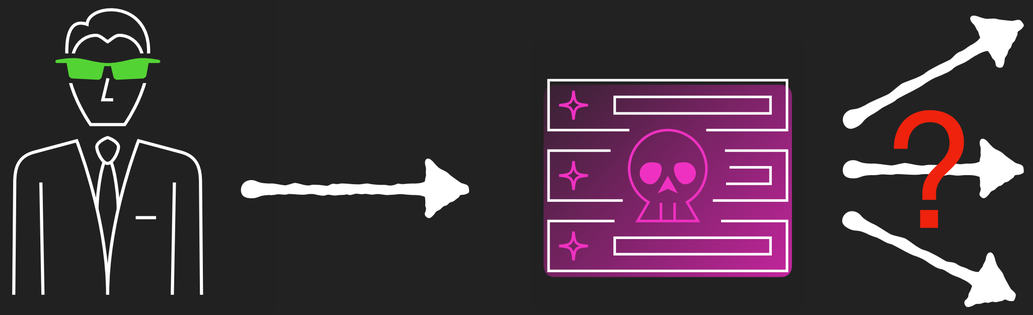
BloodHound 沒路

AD CS 沒招

On-Prem AD

Internet

Intranet



BloodHound 沒路

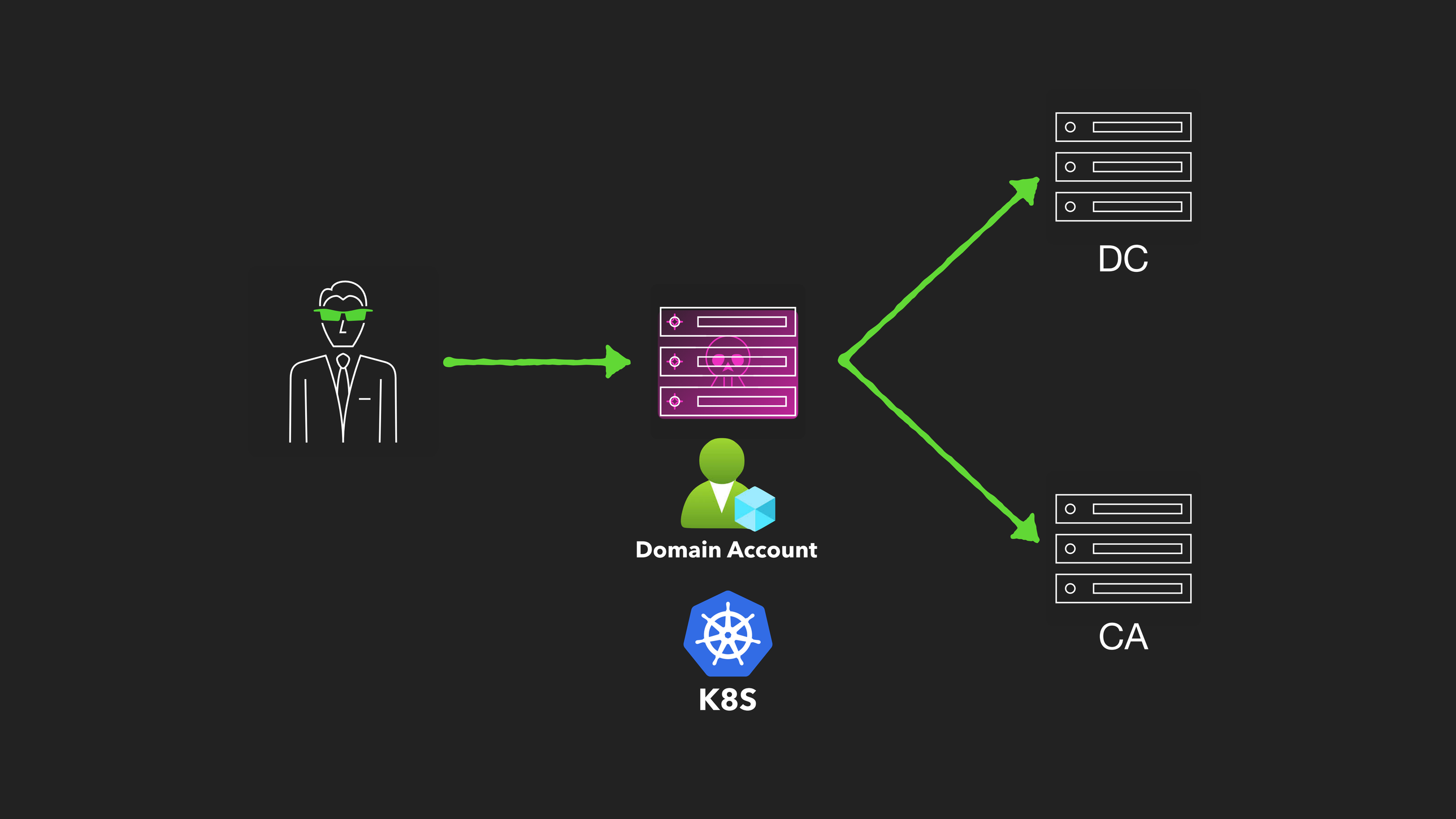
AD CS 沒招

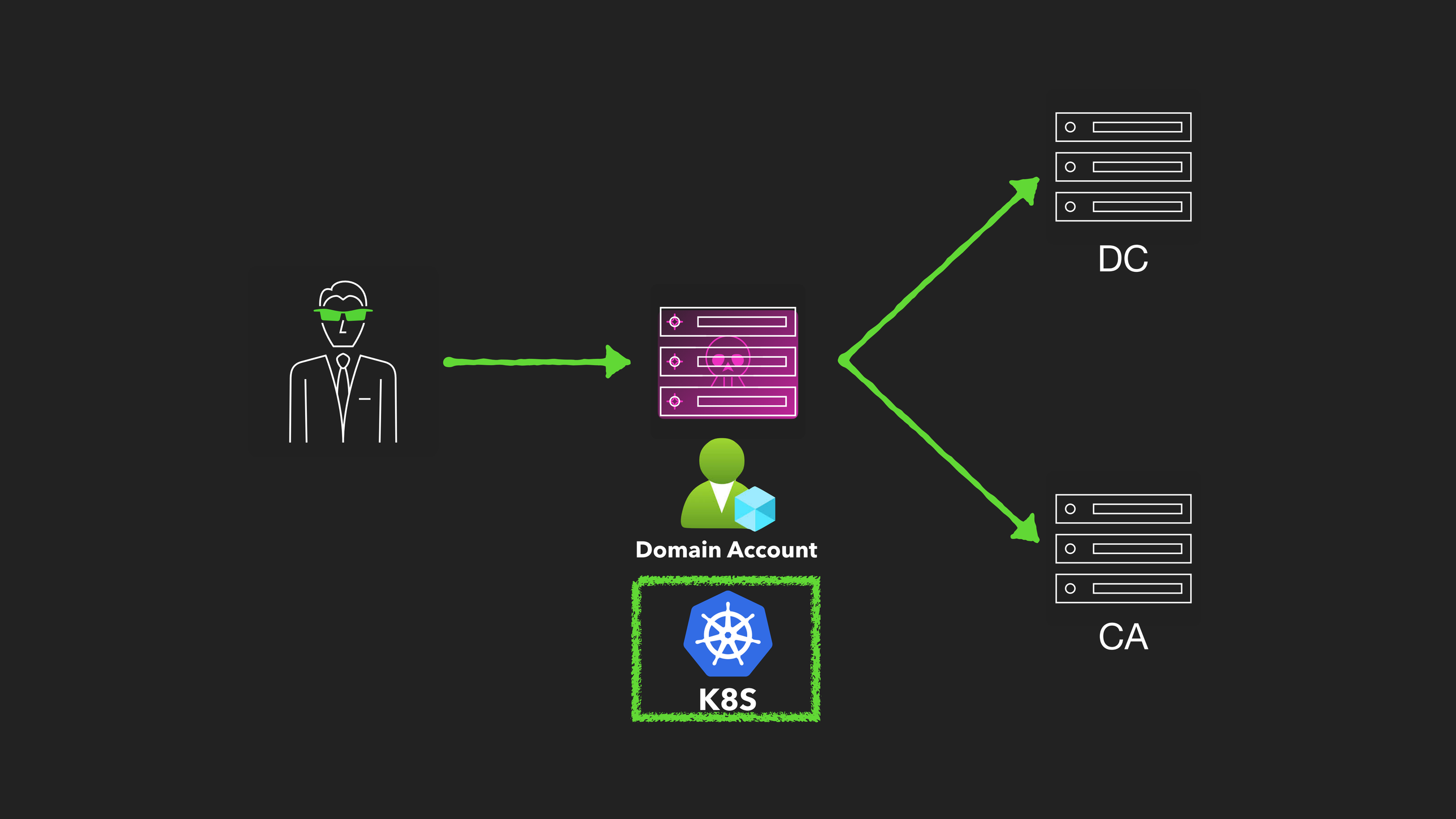
檢測 Web
要繞 MFA

On-Prem AD

Internet

Intranet


DC
Domain Account
K8S
CA


DC
Domain Account
K8S
CA

K8S

```
dummy-web-server-c9561
dummy-chroma-prod-4b5cb
dummy-crawler-dev-de62a
dummy-cronjob-uat-d01b9
dummy-webchat-server-e4560
dummy-vector-backend-d95bcf
dummy-meeting-c6442
dummy-airflow-dev-993c2
dummy-portal-server-7df17
```

K8S

```
dummy-web-server-c9561
dummy-chroma-prod-4b5cb
dummy-crawler-dev-de62a
dummy-cronjob-uat-d01b9
dummy-webchat-server-e4560
dummy-vector-backend-d95bcf
dummy-meeting-c6442
dummy-airflow-dev-993c2
dummy-portal-server-7df17
```

K8S

```
dummy-web-server-c9561
dummy-chroma-prod-4b5cb
dummy-crawler-dev-de62a
dummy-cronjob-uat-d01b9
dummy-webchat-server-e4560
dummy-portal-server-7df17
```

都聊些什麼呢

```
cat /app/history/chats
```

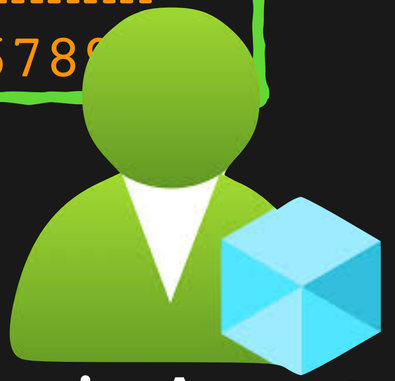
```
CLIENT_ID = “12345678-90ab-cdef-1234-567890abcdef”
REFRESH_TOKEN =
“0.AAAA_fakeRefreshTokenForDemoOnly_xxxxxxxxxxxxxxxxxxxxx”
CLIENT_SECRET = “fake-client-secret-demo-value-1234567890”
URL_BASE = “https://graph.microsoft.com/v1.0"
SCOPE = “User.Read Demo.Fake.Permission”

class GraphAPI:
    def __init__(self,client_id, client_secret …)

…
[code snip]
…
```

要怎麼在 Scope 中，把 chat 的權限設成像我上面那樣的完整權限？

```
CLIENT_ID = “12345678-90ab-cdef-1234-567890abcdef”
REFRESH_TOKEN =
“0.AAAA_fakeRefreshTokenForDemoOnly_xxxxxxxxxxxxxxxxxxxxx”
CLIENT_SECRET = “fake-client-secret-demo-value-1234567890”
URL_BASE = “https://graph.microsoft.com/v1.0"
SCOPE = “User.Read Demo.Fake.Permission”

class GraphAPI:
    def __init__(self,client_id, client_secret …)

…
[code snip]
…
```

要怎麼在 Scope 中，把 chat 的權限設成像我上面那樣的完整權限？

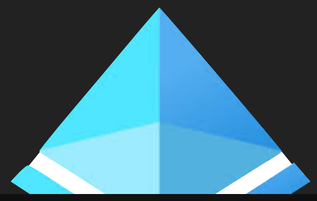
```
CLIENT_ID = “12345678-90ab-cdef-1234-567890abcdef”
REFRESH_TOKEN =
“0.AAAA_fakeRefreshTokenForDemoOnly_xxxxxxxxxxxxxxxxxxxxx”
CLIENT_SECRET = “fake-client-secret-demo-value-12345678
URL_BASE = “https://graph.microsoft.com/v1.0"
SCOPE = “User.Read Demo.Fake.Permission”

class GraphAPI:
    def __init__(self,client_id, client_secret …)

…
[code snip]
…
```



要怎麼在 Scope 中，把 chat 的權限設成像我上面那樣的完整權限？

Service Account

Tenant Listing



Tenant Listing

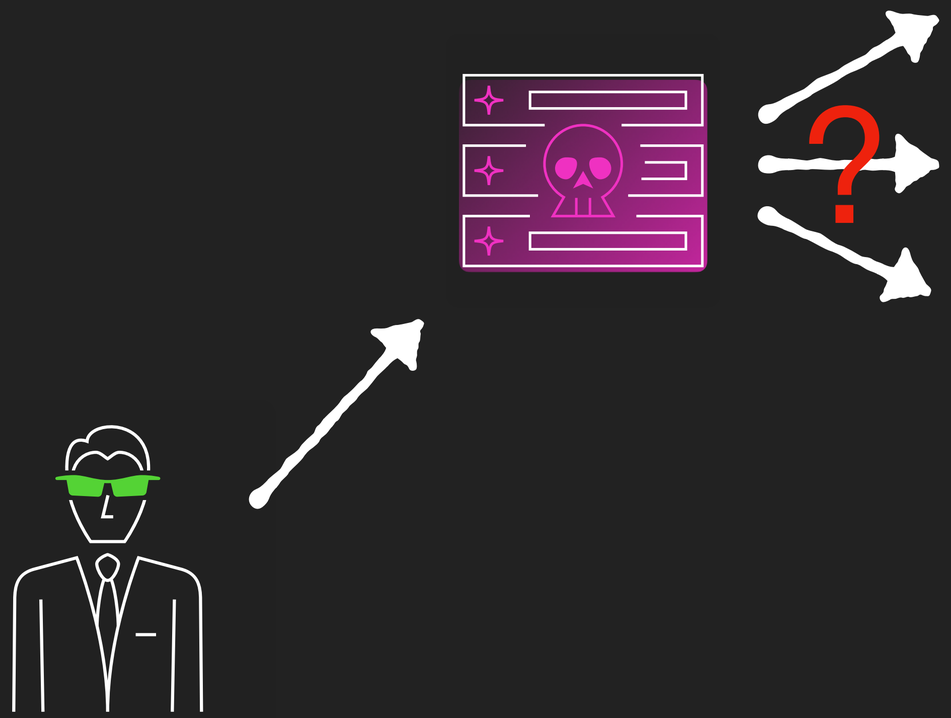
開啟檢測 Entra ID 的大門

紅隊成功在上面提升權限

On-Prem AD

Internet

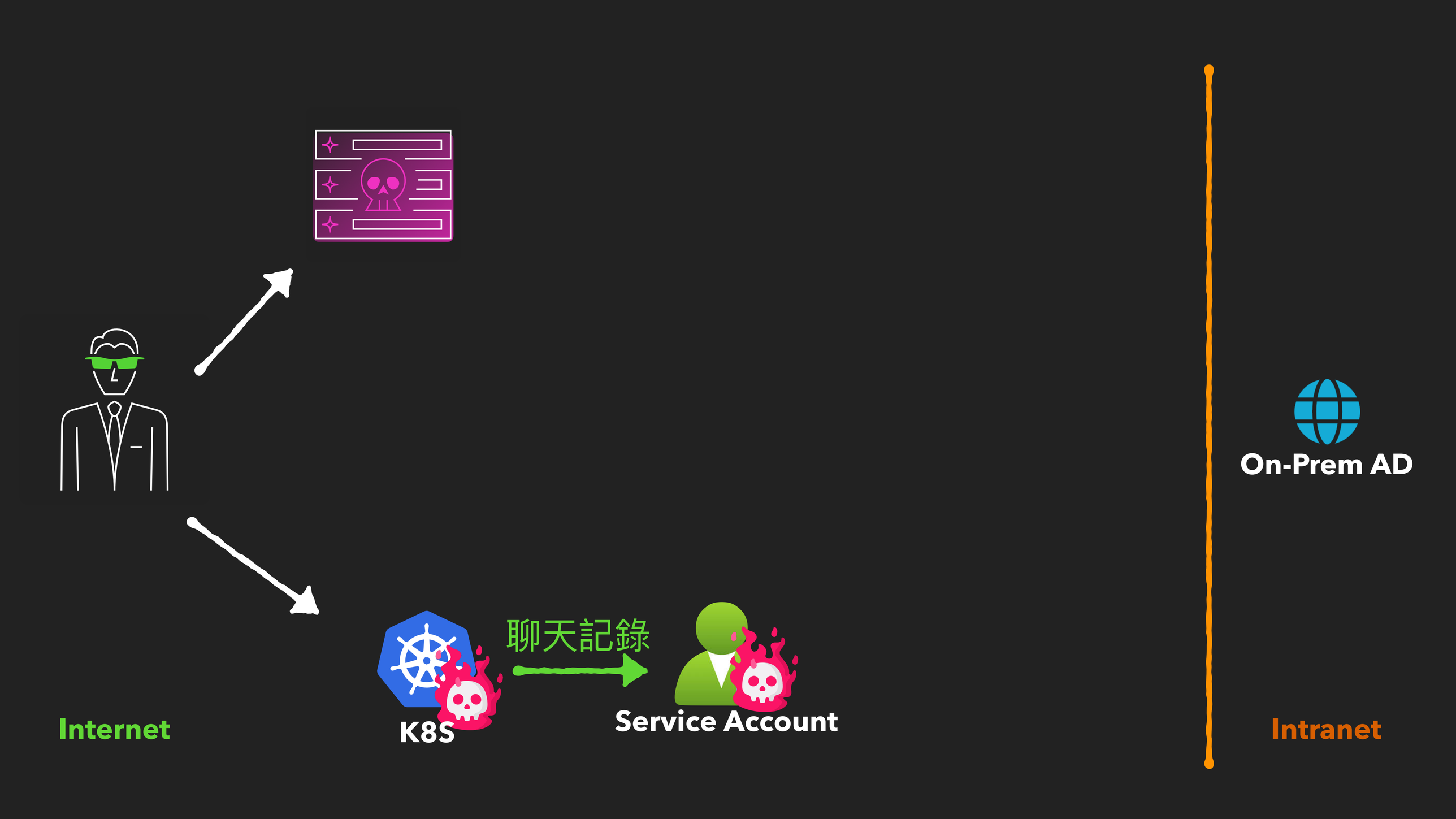
Intranet

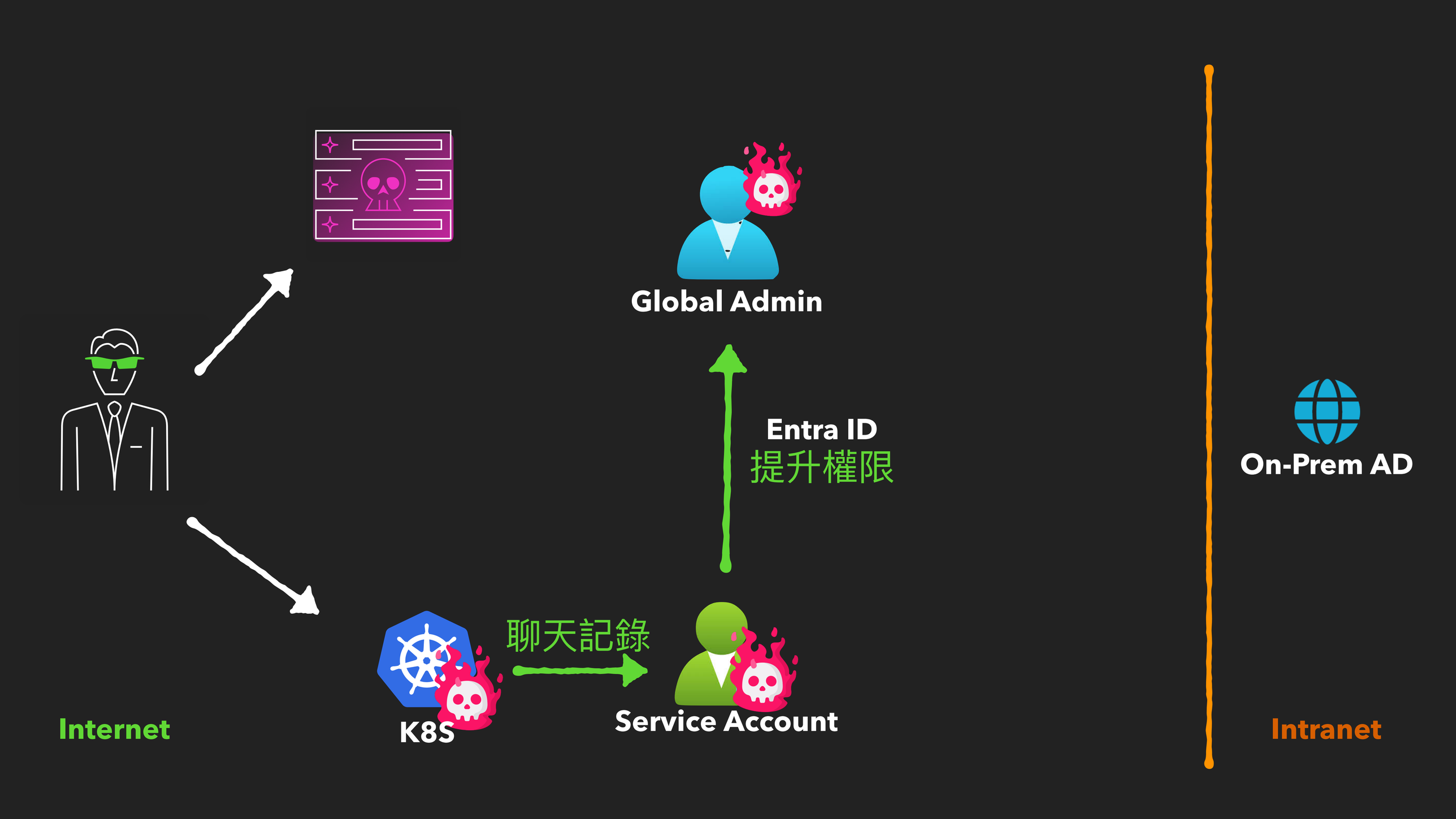


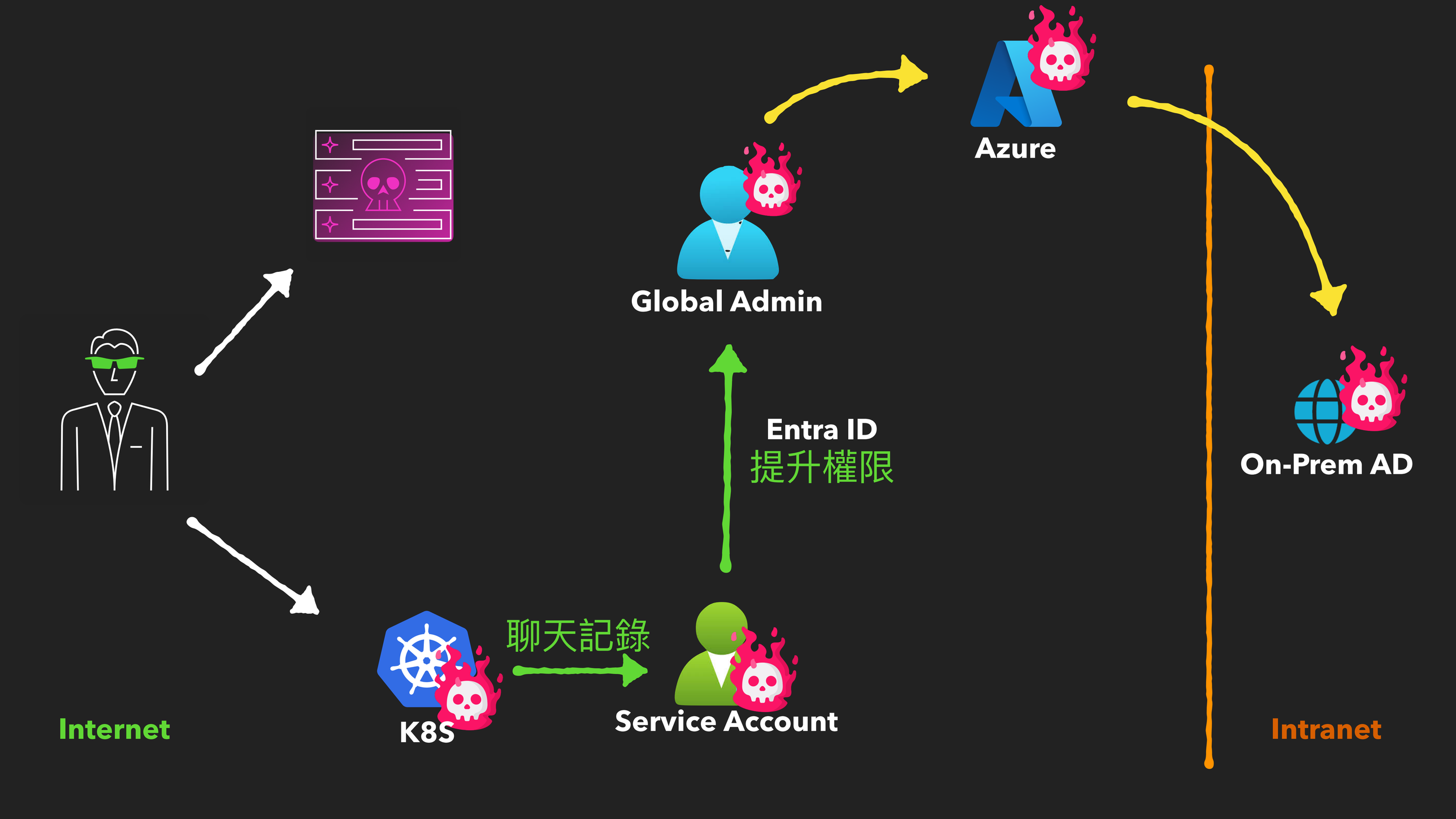
On-Prem AD

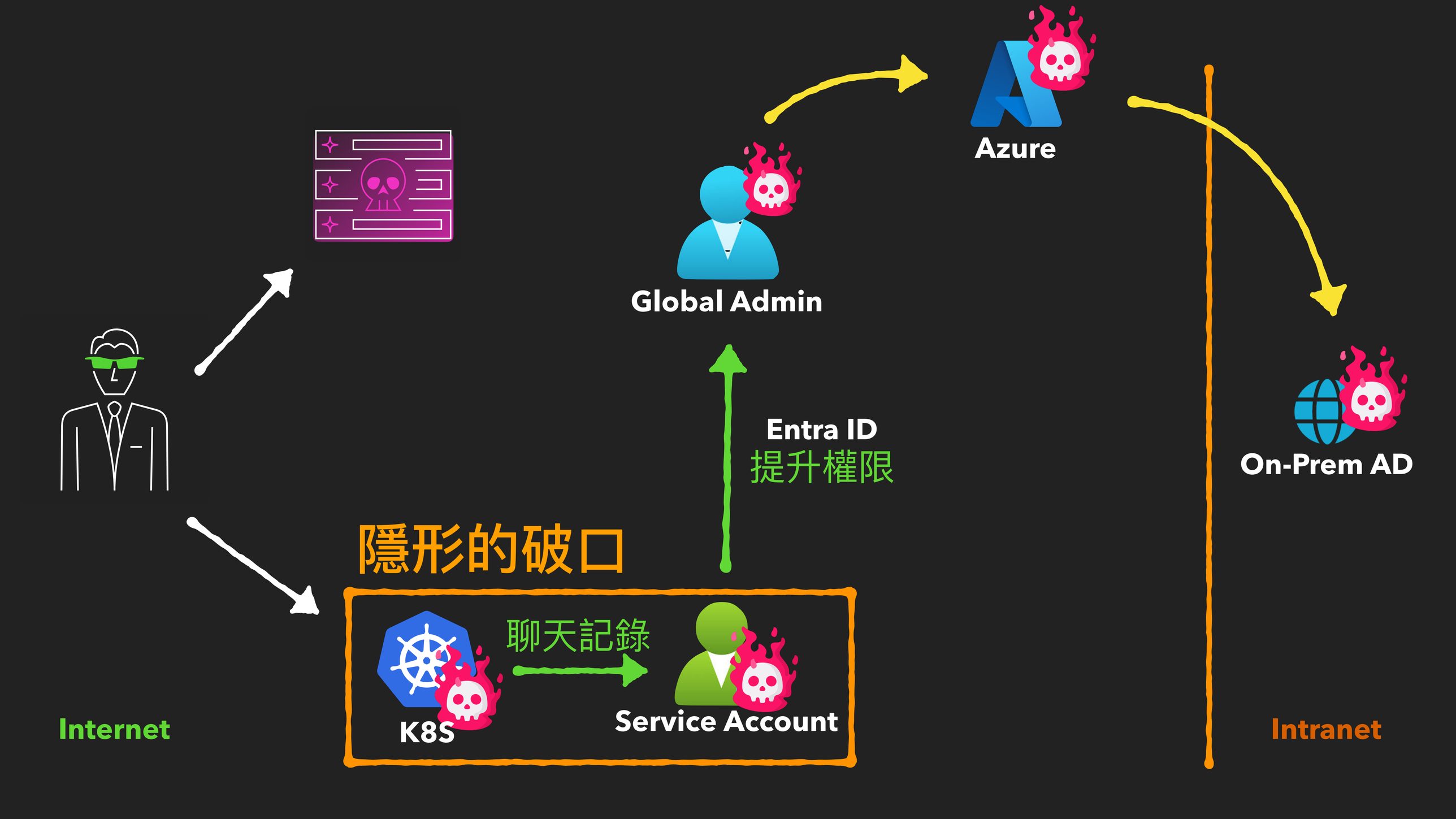
Internet

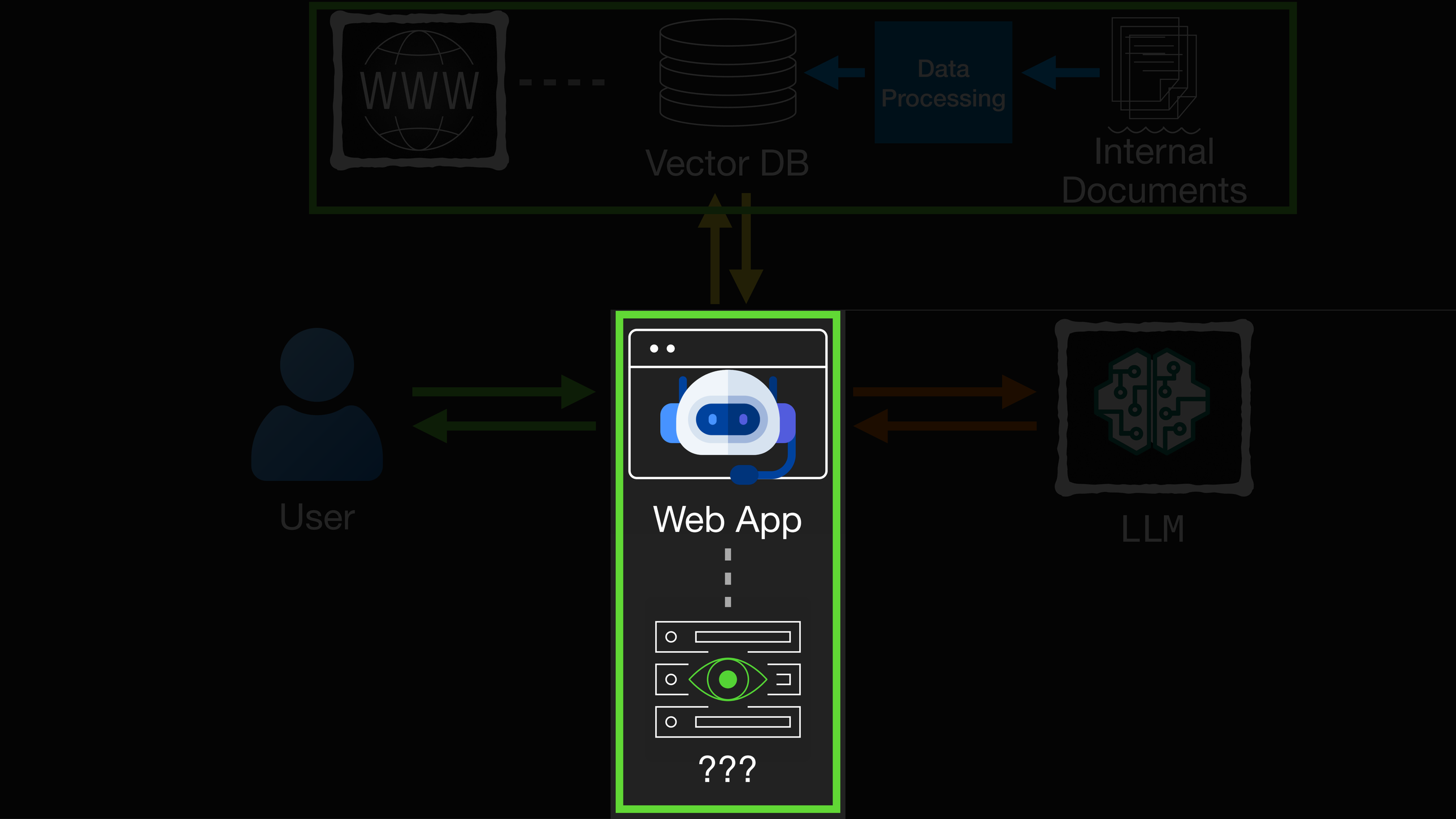
Intranet

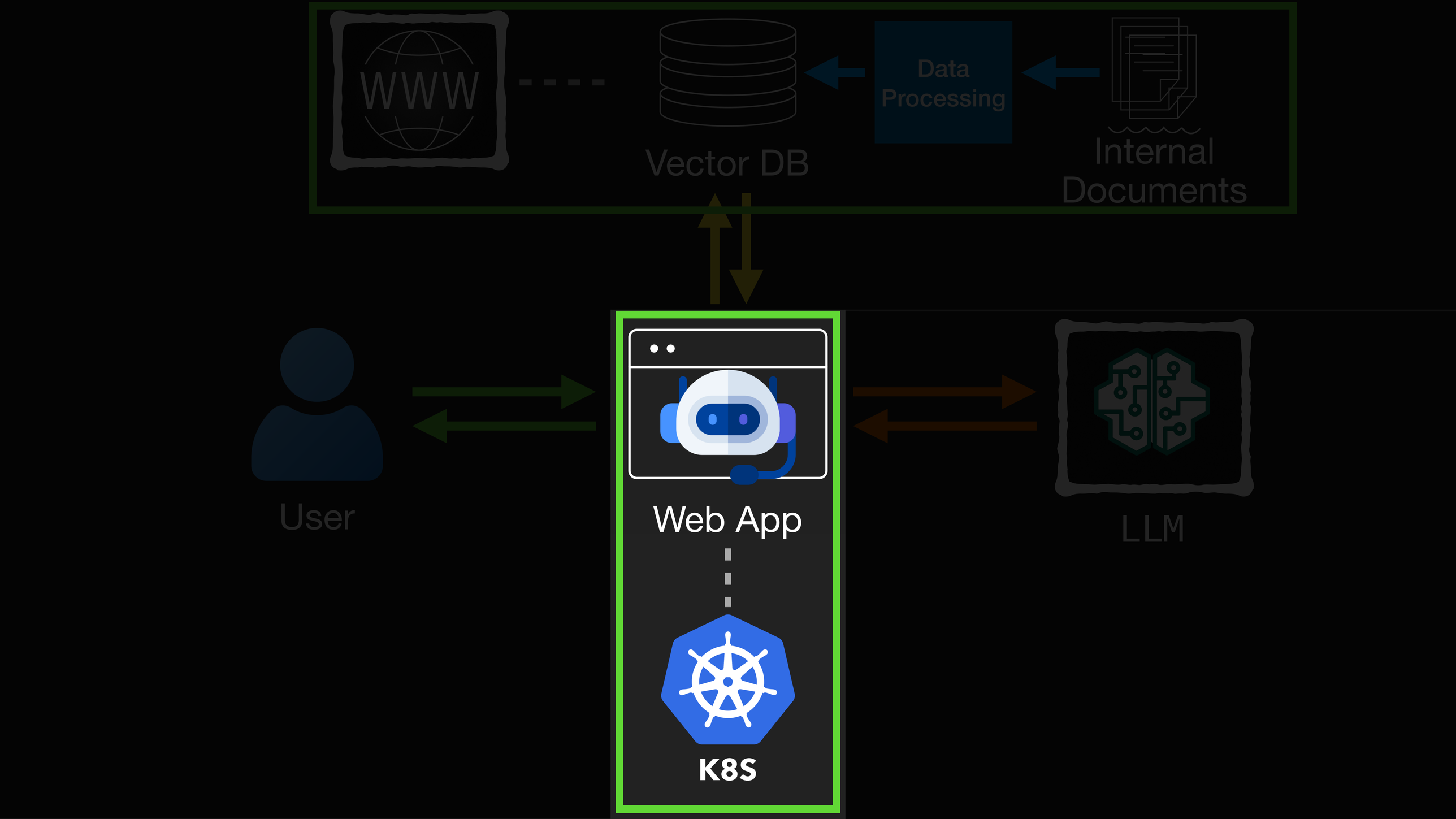

聊天記錄
K8S
Service Account
Internet
On-Prem AD
Intranet


Global Admin
Entra ID
提升權限
On-Prem AD
聊天記錄
K8S
Service Account
Internet
Intranet


Azure
Global Admin
Entra ID
提升權限
On-Prem AD
聊天記錄
K8S
Service Account
Internet
Intranet

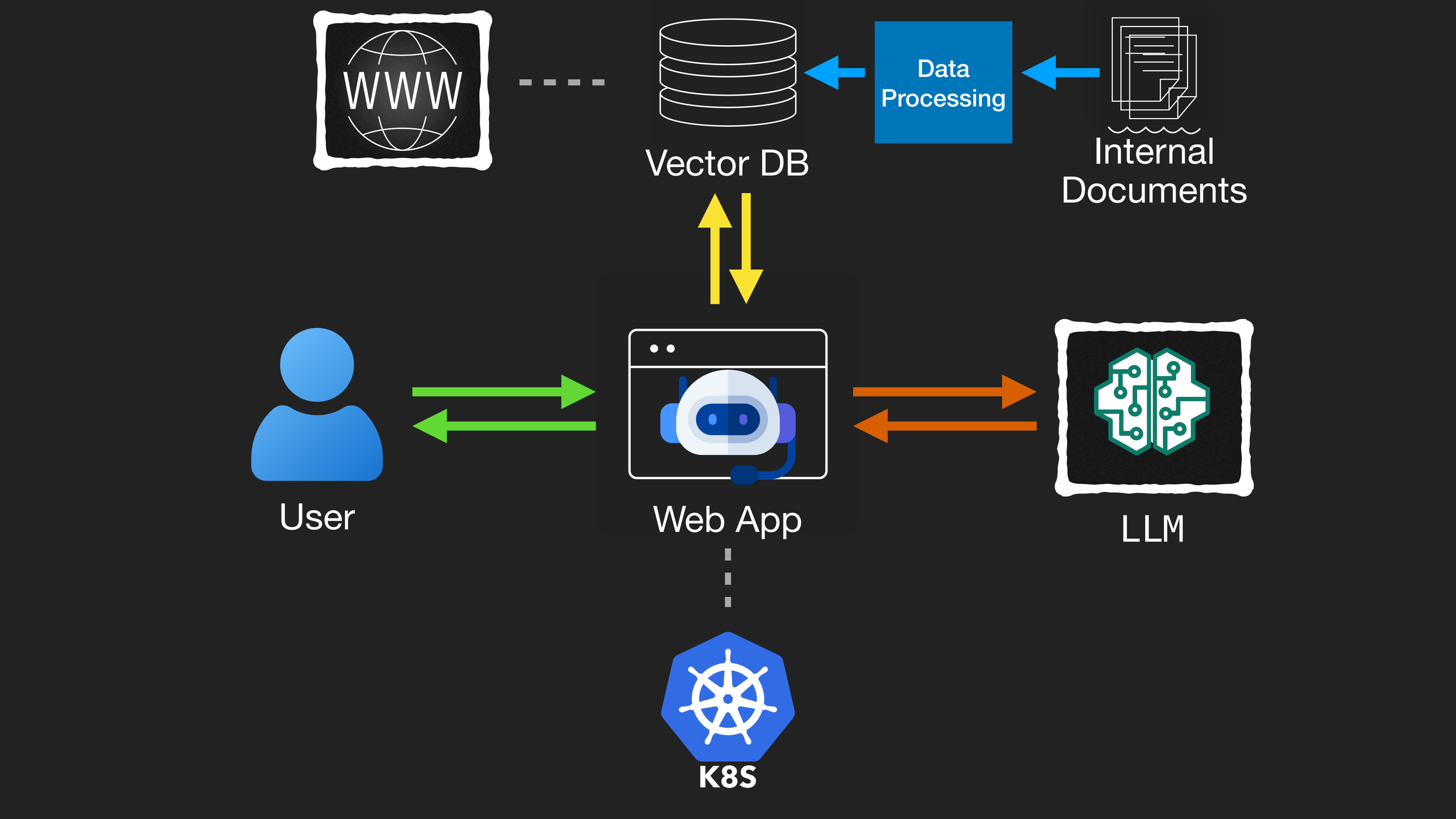

Azure
Global Admin
Entra ID
提升權限
On-Prem AD
隱形的破口
聊天記錄
K8S
Service Account
Internet
Intranet


WWW
Vector DB
Data Processing
Internal Documents
User
Web App
LLM
???


WWW
Vector DB
Data Processing
Internal Documents
User
Web App
LLM
K8S

DEVCORE

案例回顧與建議


WWW
Vector DB
Data Processing
Internal Documents
User
Web App
LLM
K8S

Vector DB

Data Processing

Internal Documents

POST /chat/Complete HTTP/1.1
Host: api.dummycorp.local

HTTP/1.1 200 OK
Content-Type: application/json

未考量傳統 Web 弱點

BYPASSED

User

/Auth/GenJWT?id=A001

/Chat/GetChatHistory

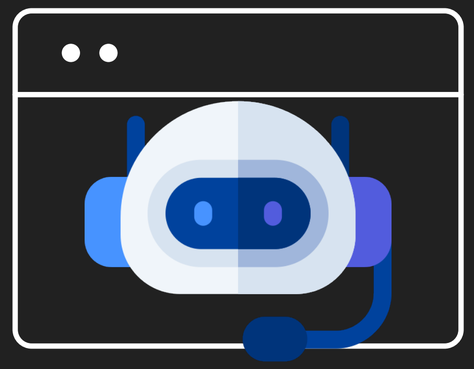
/Chat/GetChatByCid?cid=[UUID]

透過 JWT 與 UUID 讀取對話內容

LLM

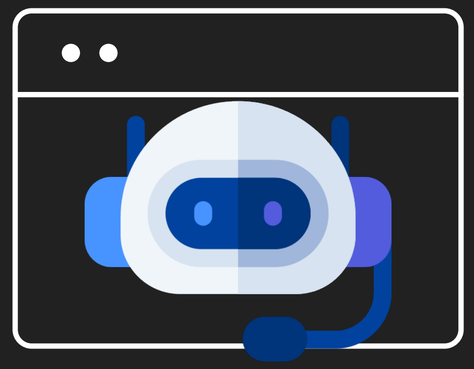
K8S

兼顧傳統 Web 安全性



Web App

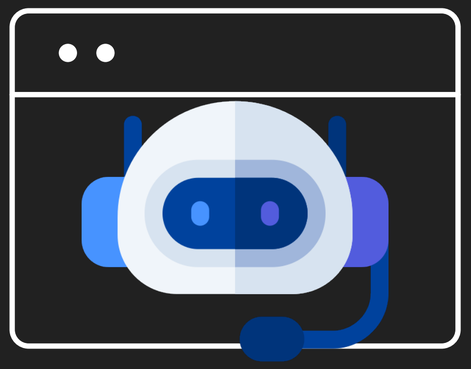
兼顧傳統 Web 安全性



Web App

檢查邏輯

兼顧傳統 Web 安全性

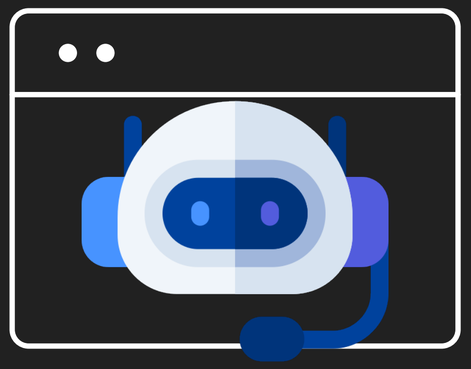


Web App

檢查邏輯

資訊洩漏

兼顧傳統 Web 安全性

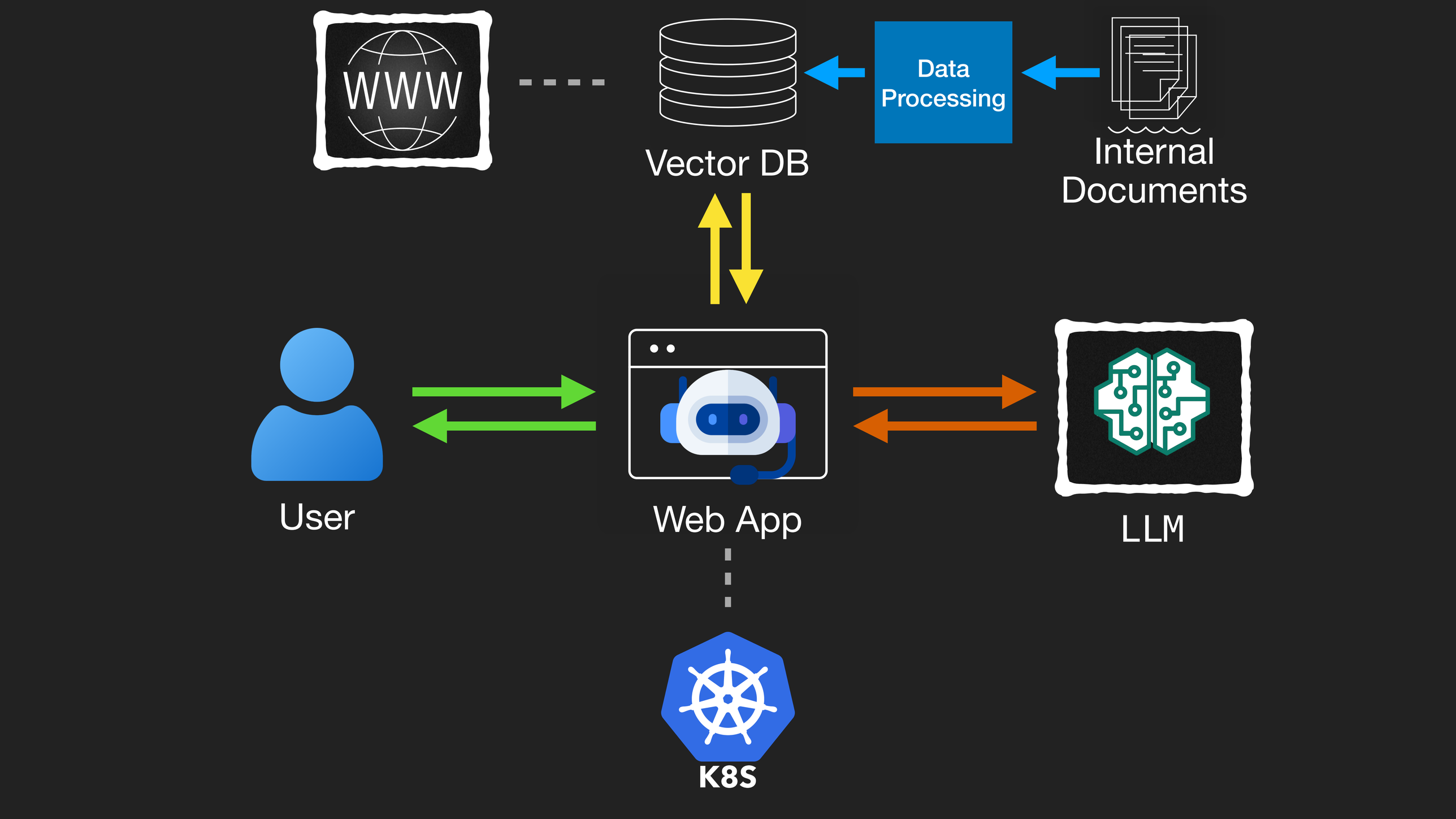


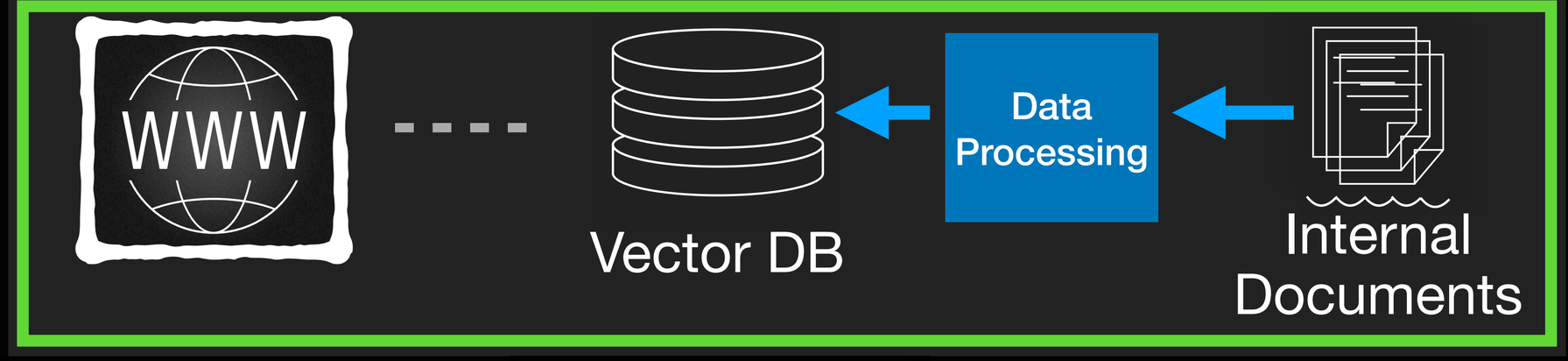
Web App

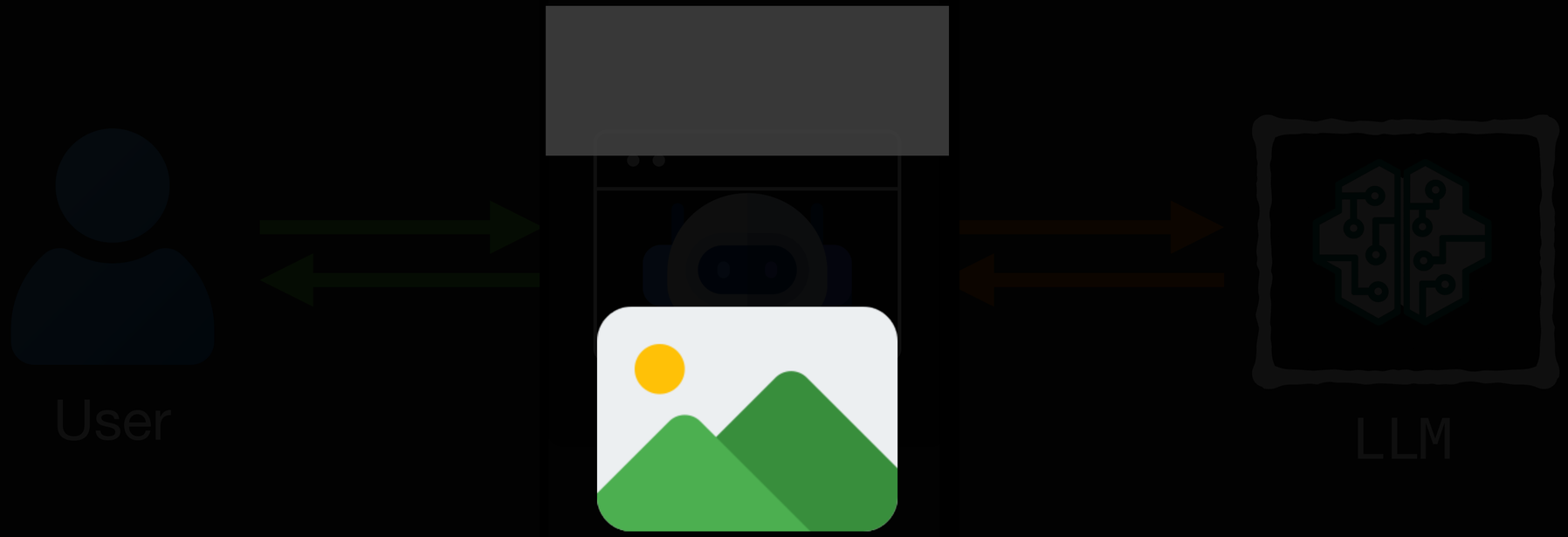
檢查邏輯

資訊洩漏

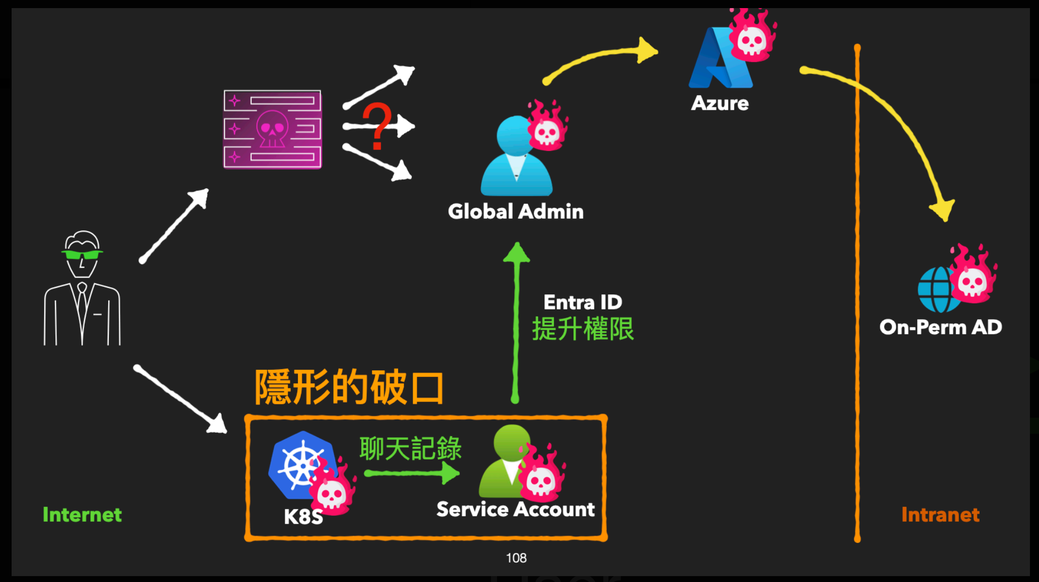
權限控管

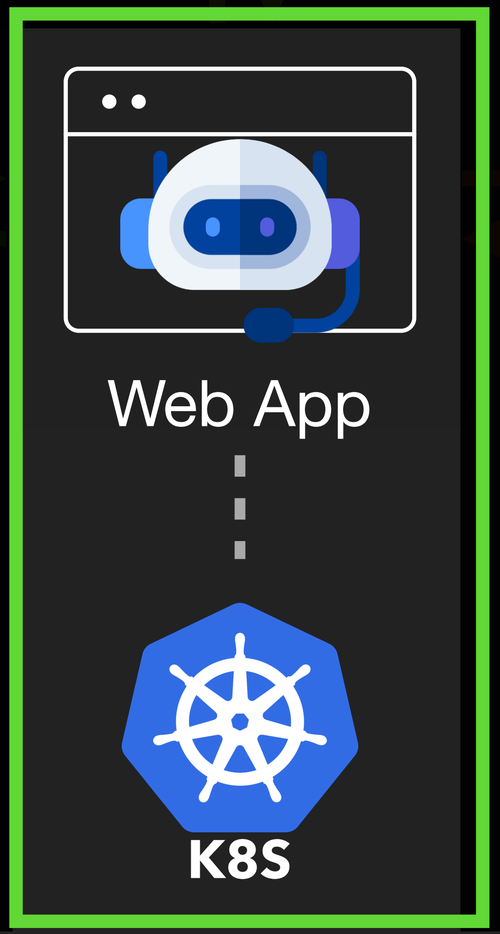

WWW
Vector DB
Data Processing
Internal Documents
User
Web App
LLM
K8S

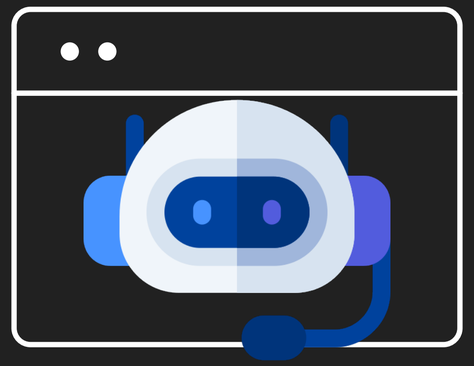

WWW
Vector DB
Data Processing
Internal Documents


User
LLM
K8S

本頁截圖僅於現場分享

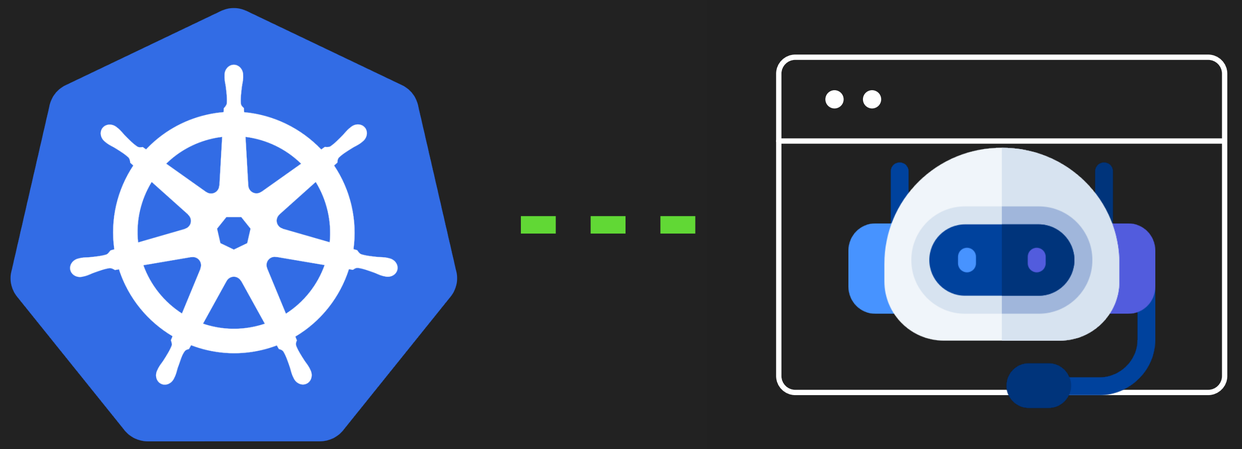

Azure
Global Admin
Entra ID
提升權限
On-Perm AD
隱形的破口
聊天記錄
K8S
Service Account
Internet
Intranet
108


Web App
K8S



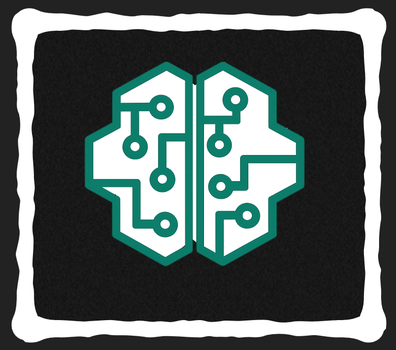
Web App

完整盤點系統架構，將相依的主機與帳號列入風險評估



完整盤點系統架構，將相依的主機與帳號列入風險評估

完整盤點系統架構，將相依的主機與帳號列入風險評估

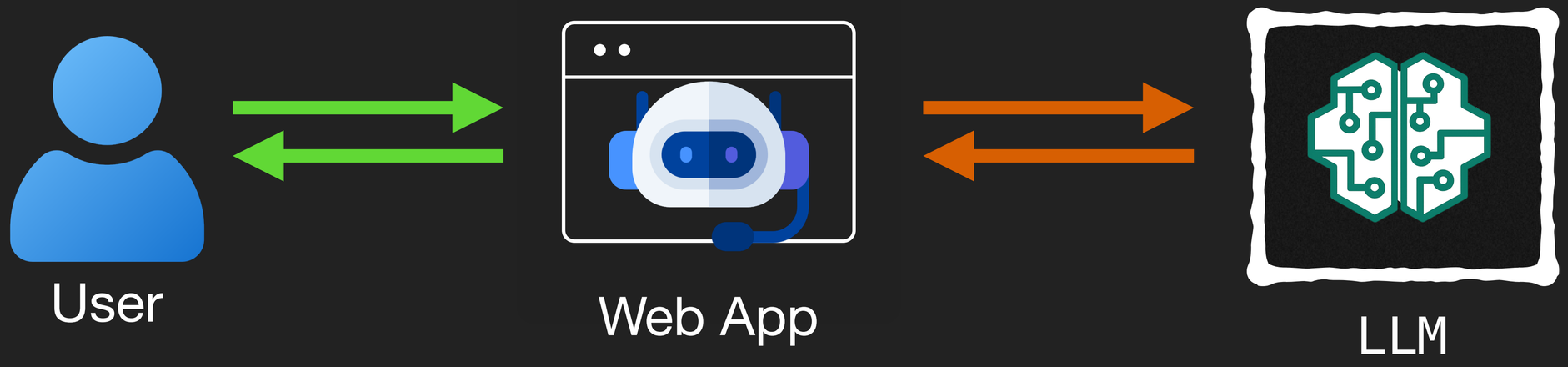


LLM

防禦新型態攻擊手法

OWASP | TOP 10 LLM APPLICATIONS & GENERATIVE AI

MITRE ATLAS™



使用者可能因信任內部系統而上傳機敏資訊，建議

定期清查對話與系統紀錄檔案，並評估潛在風險

總結

DEVCORE

總結

DEVCORE

- 自建 LLM 服務時，需同時兼顧模型本身及其 Web 應用服務的安全性

總結

DEVCORE

- 自建 LLM 服務時，需同時兼顧模型本身及其 Web 應用服務的安全性
- 盤點 LLM 服務完整的系統架構，將檢測範圍涵蓋所有關鍵元件與資料流。

總結

DEVCORE

- 自建 LLM 服務時，需同時兼顧模型本身及其 Web 應用服務的安全性
- 盤點 LLM 服務完整的系統架構，將檢測範圍涵蓋所有關鍵元件與資料流。
- 使用者可能因信任內部網站而上傳機敏資訊; 建議定期檢視對話紀錄、系統日誌與附件檔案的留存與存取權限，並持續評估風險。

DEVCORE

Q&A

戴夫寇爾股份有限公司
contact@devco.re
02-2577-0925