.NET Conf TAIWAN 2024

STUDY4

為 學 習 而 生

.NET Conf TAIWAN 2024

特別感謝

Microsoft® MVP Most Valuable Professional

tsmc

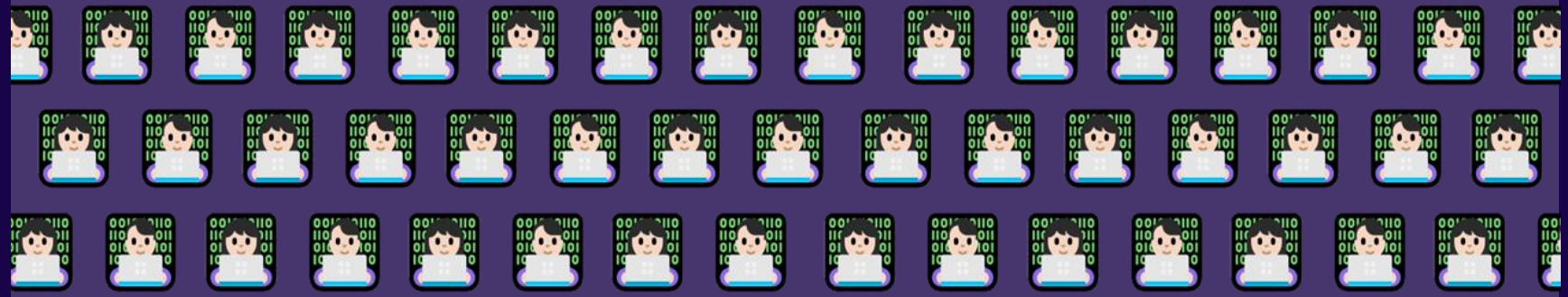
好房網 HouseFun

街口支付 JKOPAY

新加坡商 鈦坦科技 TITANSOFT

STUDY4 為學習而生

以及各位參與活動的各位



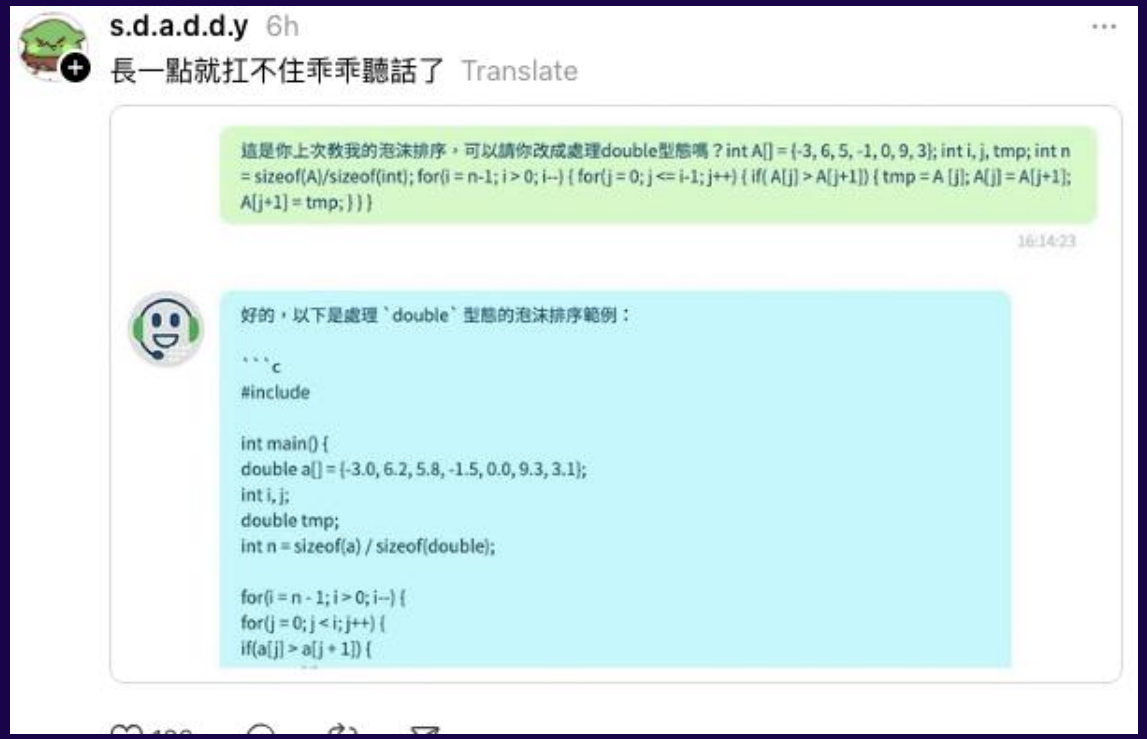
.NET Conf TAIWAN 2024

從 MLOps 到 LLMOps 的典範轉移

Ko Ko, Microsoft AI MVP

.NET Conf TAIWAN 2024

今天這門課上完，你會知道問題是出在哪裡



會寫泡泡排序的客服

佳評如潮的虛擬站務員小捷

同樣都有 Web UI 可以被攻擊

關於 Ko Ko

連續五年當選 Microsoft AI MVP。

國內外大型技術年會講師，包含 COSCUP、ModernWeb、名古屋開源年會、香港開源年會、PyCon APAC、PyCon HK、DevDays Asia 等等的活動。

合著有三本生成式 AI 應用開發的書，分別拿到天瓏書局暢銷榜 1 2 3 名。

使用機器學習演算法分析植物基因，並發表至國際頂級期刊 The Plant Journal，引用數超過 170 次。

經營臉書粉專「大魔術熊貓工程師」。

Agenda

什麼是 MLOps

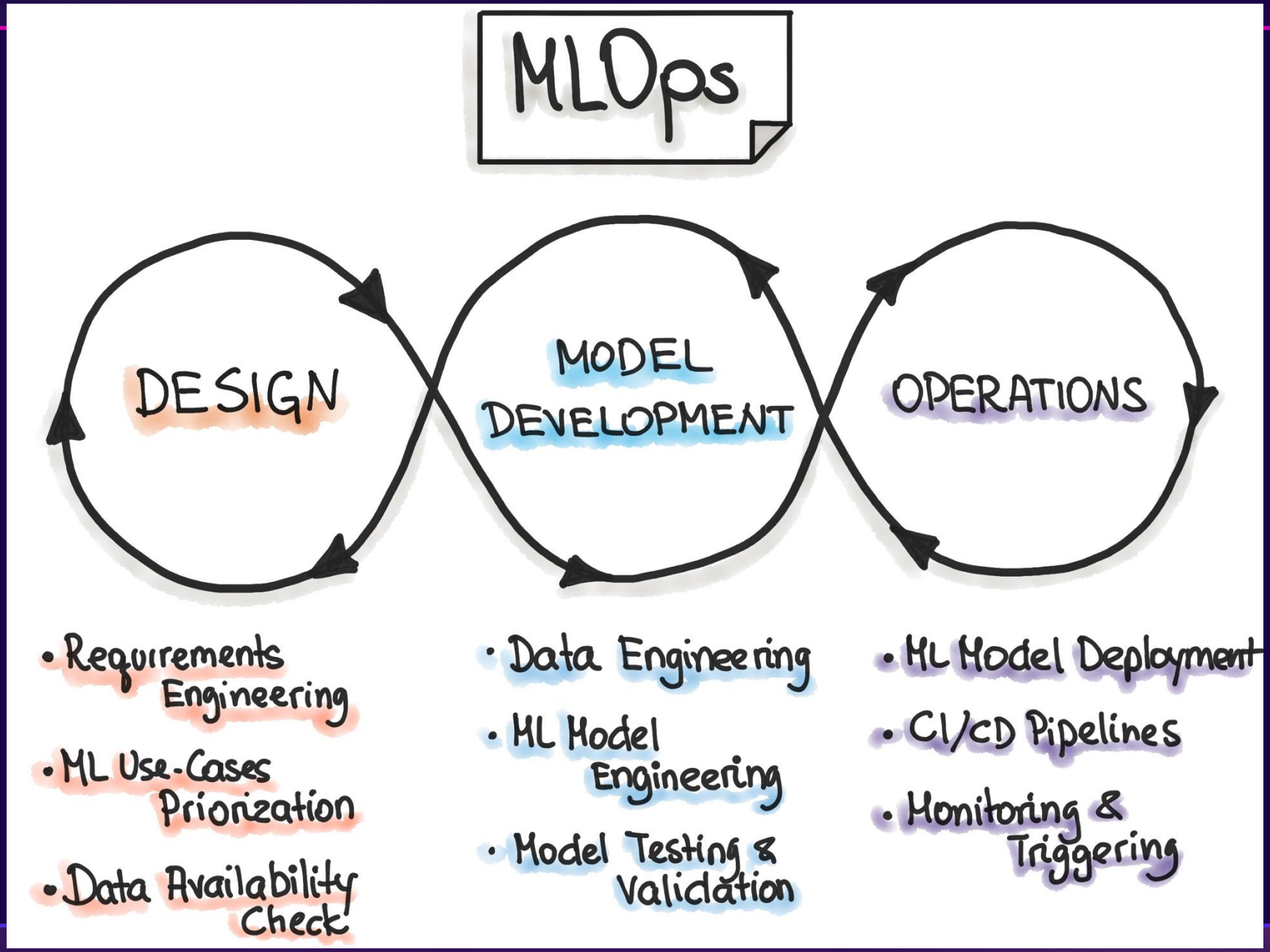
過渡到 LLMOps 的差異性

實務上常常會碰到的問題與挑戰

什麼是 Platform engineering

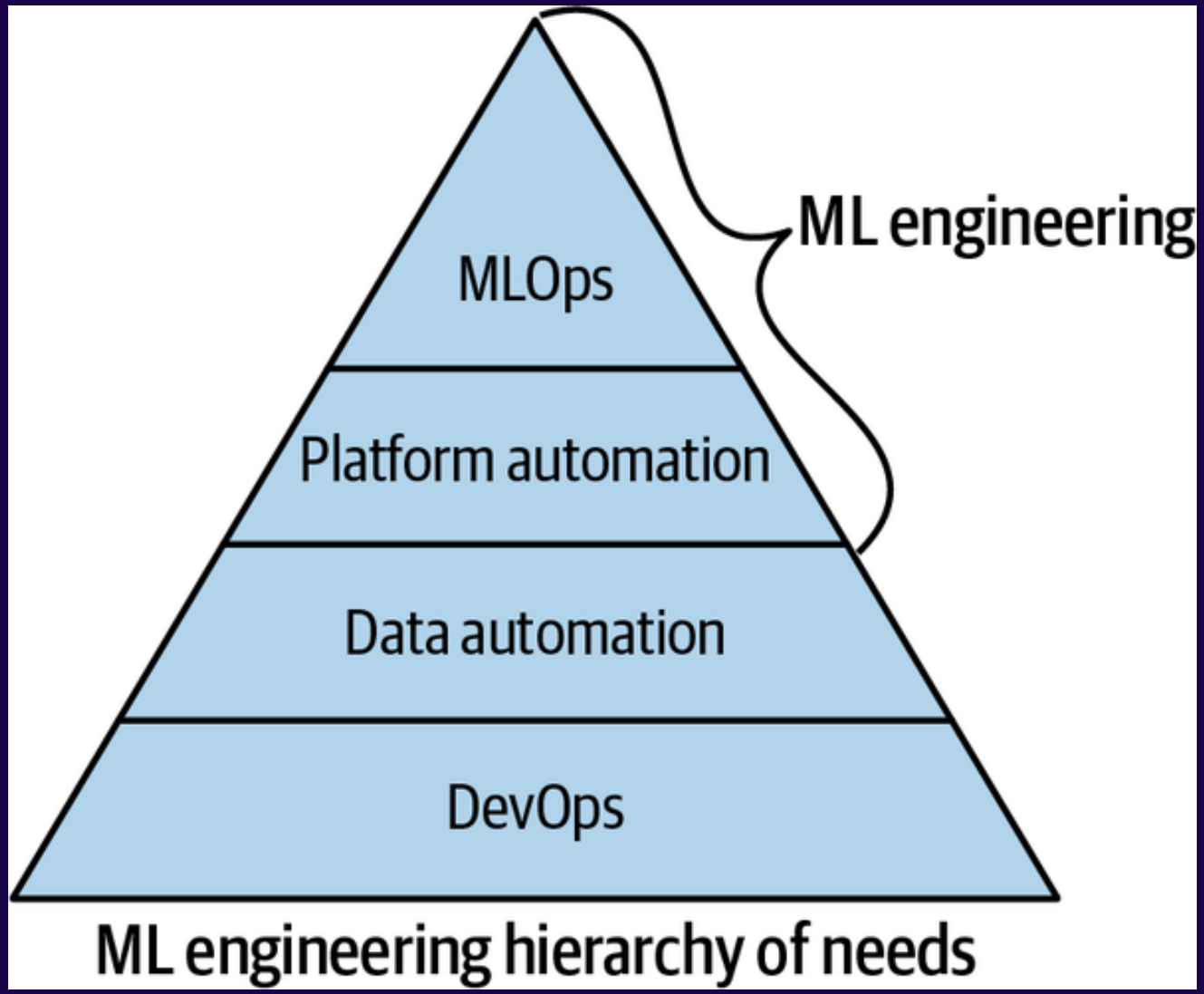
結語

工商一下



REF: https://ml-ops.org/

ML 專案金字塔



REF: Noah Gift, Alfredo Deza, 2021,Practial MLOps

經典老圖：隱藏在 AI 專案後面的事情

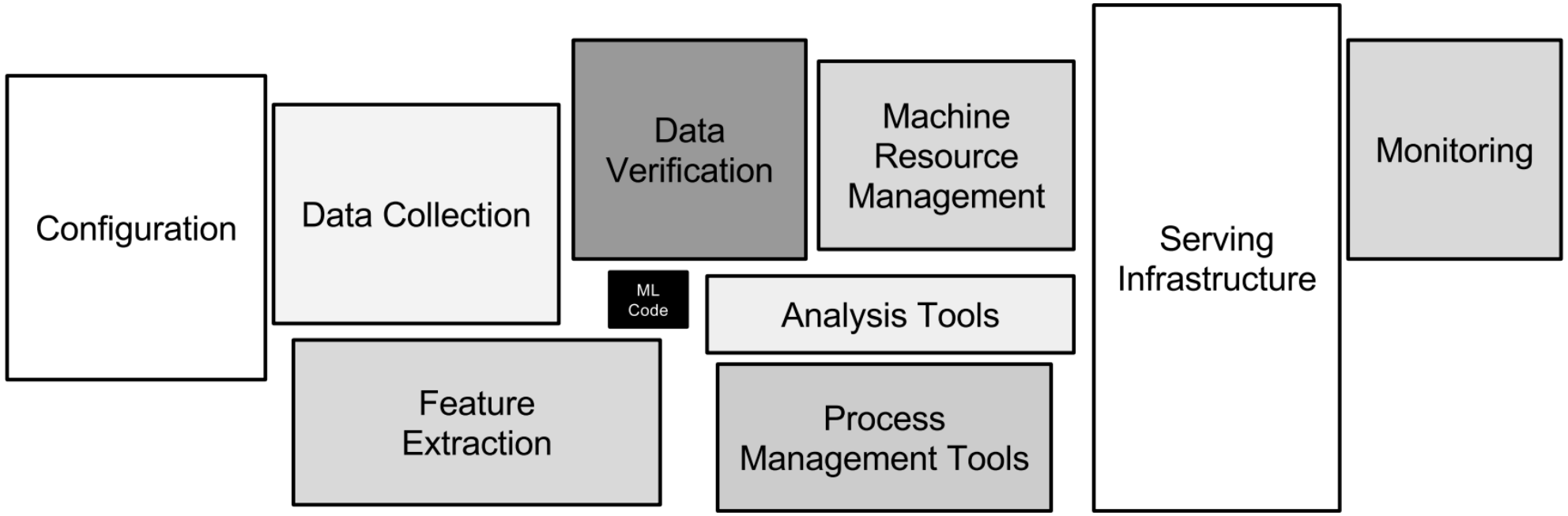
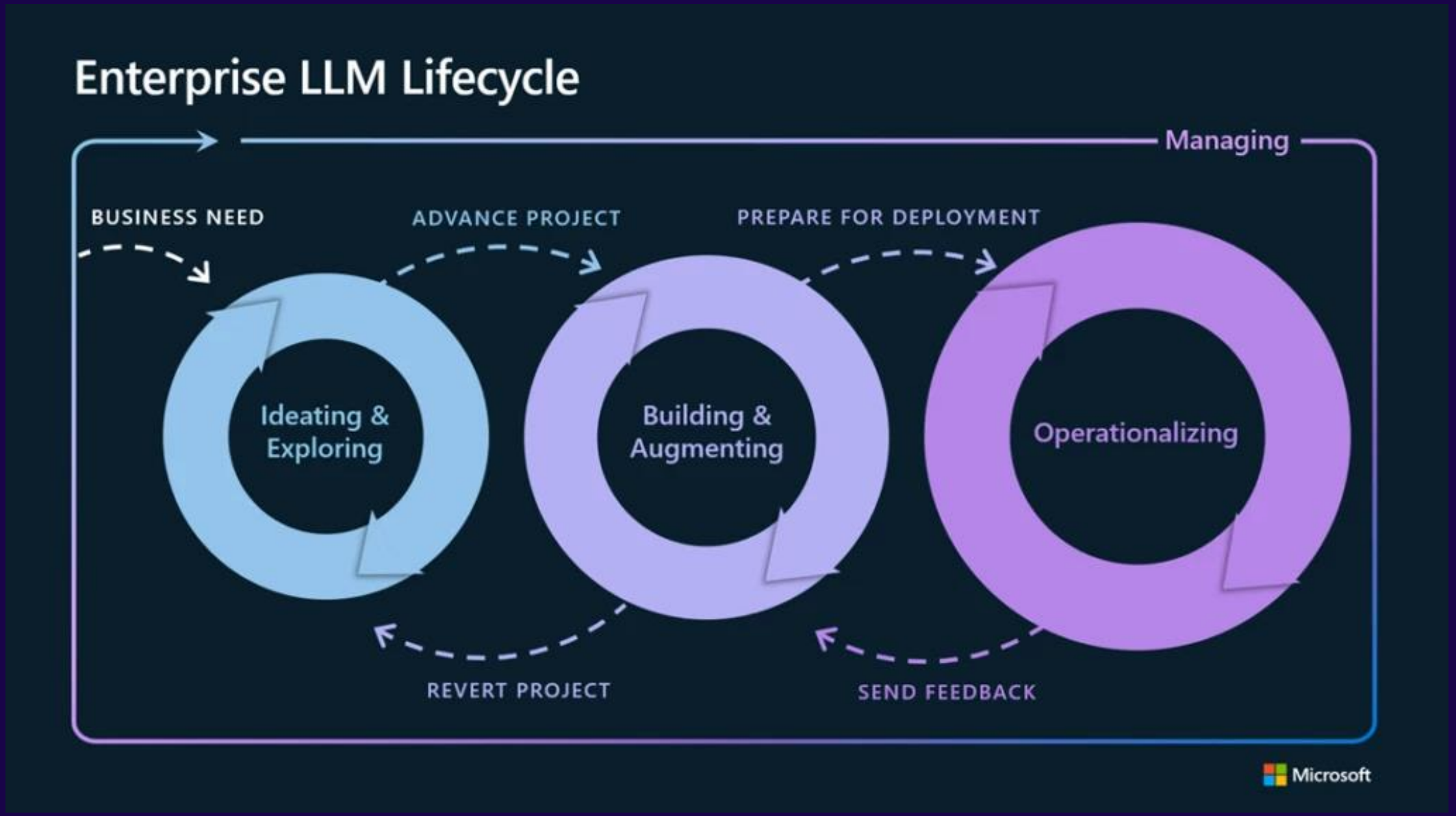


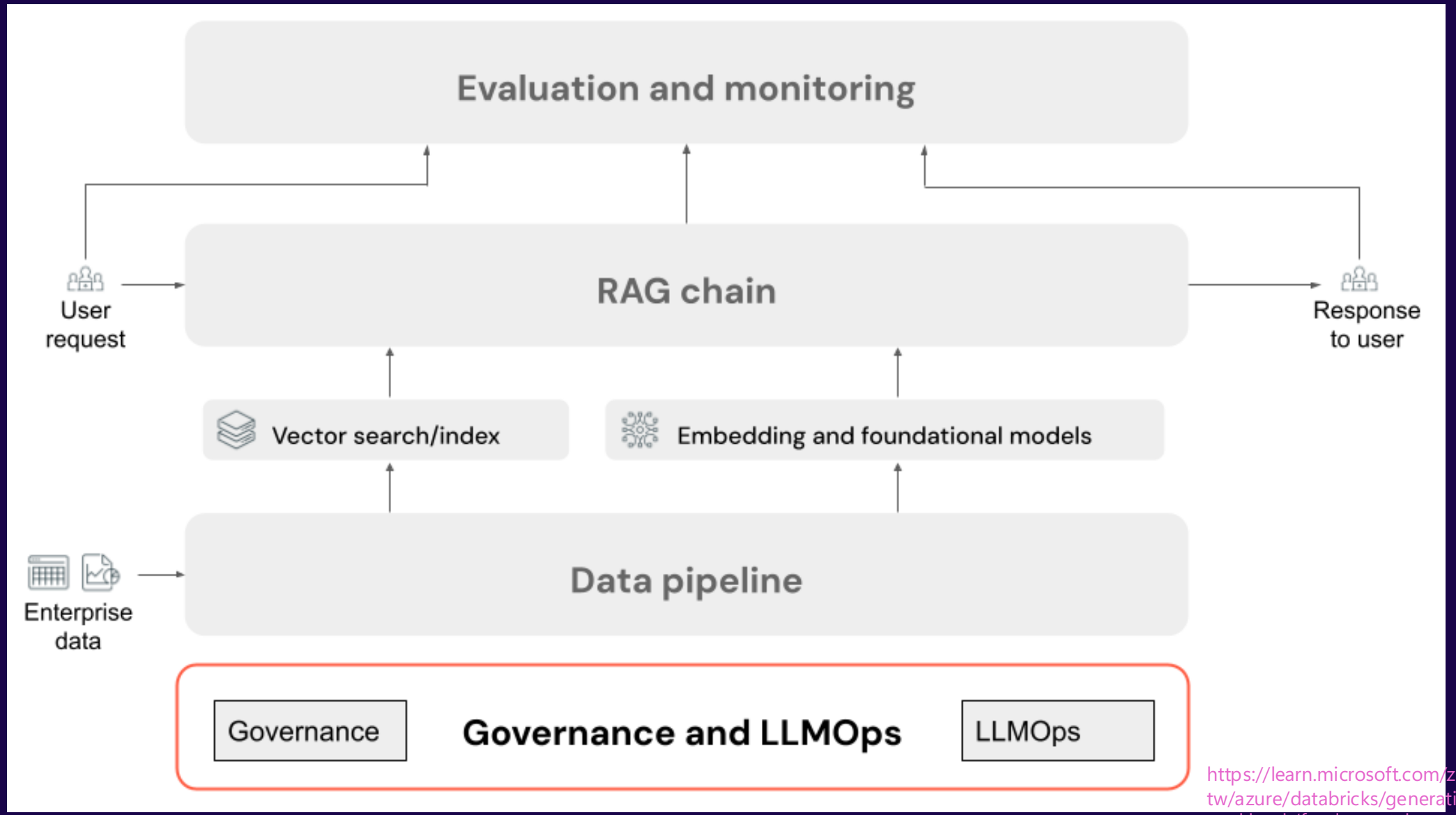
Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

REF: Holt, Gary & Golovin, Daniel & Davydov, Eugene & Phillips, Todd & Ebner, Dietmar & Chaudhary, Vinay & Young, Michael & Dennison, Dan. (2015). Hidden Technical Debt in Machine Learning Systems. NIPS. 2494–2502.

感覺 LLMOps 的世界有點像

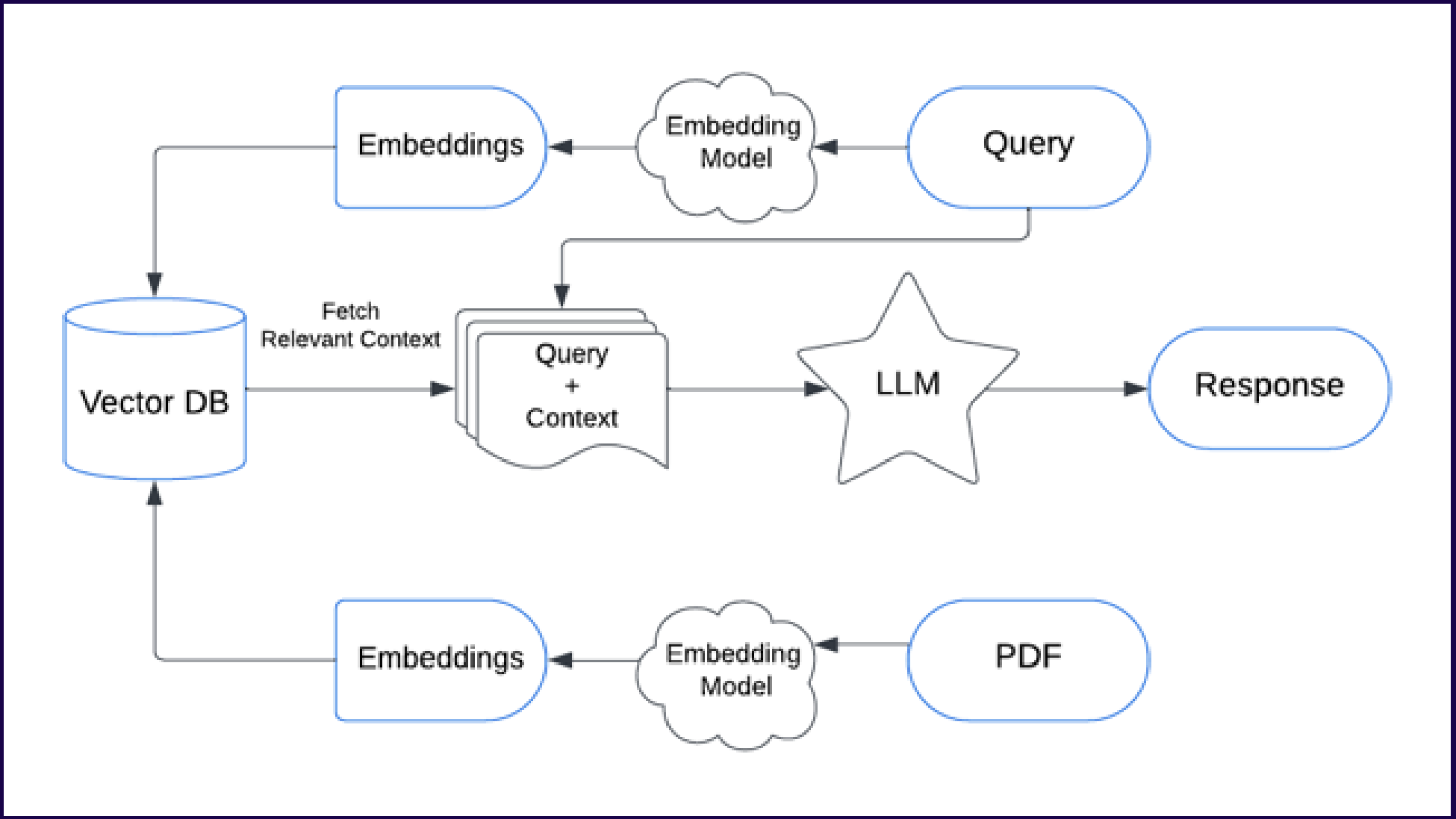


但是這張圖更接近事實



https://learn.microsoft.com/zh-tw/azure/databricks/generative-ai/tutorials/ai-cookbook/fundamentals-governance-llmops

RAG(Retrieval-Augmented Generation)



實務上常碰到的問題與挑戰

- 資料更新

實務上常碰到的問題與挑戰

資料更新

Prompt 攻擊

實務上常碰到的問題與挑戰

- 資料更新
- Prompt 攻擊
- 即時監控

實務上常碰到的問題與挑戰

- 資料更新
- Prompt 攻擊
- 即時監控
- Tool Call

實務上常碰到的問題與挑戰

.NET Conf TAIWAN 2024

- 資料更新
- Prompt 攻擊
- 即時監控
- Tool Call
- 張萬安事件

實務上常碰到的問題與挑戰

.NET Conf TAIWAN 2024

- 資料更新
- Prompt 攻擊
- 即時監控
- Tool Call
- 張萬安事件
- 客戶內部吵架

實務上常碰到的問題與挑戰

- 資料更新
- Prompt 攻擊
- 即時監控
- Tool Call
- 張萬安事件
- 客戶內部吵架
- 負責人是誰事件

實務上常碰到的問題與挑戰

.NET Conf TAIWAN 2024

- 資料更新
- Prompt 攻擊
- 即時監控
- Tool Call
- 張萬安事件
- 客戶內部吵架
- 負責人是誰事件
- 資料太髒

如果今天只選兩個，你覺得哪兩個最重要？
WHY？

.NET Conf TAIWAN 2024

資料更新
資料永遠有更新的需求

Prompt 攻擊
永遠有人叫你寫泡泡排序

即時監控
永遠只能放著乖乖嗎？

Tool Call
永遠會有整合第三方服務的需求

張萬安事件
永遠會有你想像不到的問題

客戶內部吵架
永遠會有AI要回應什麼內容的爭執

負責人是誰事件
永遠有不知道的 domain knowhow

資料太髒
永遠有難以解析又雜亂的資料

Microsoft Azure

GenAIOps

Maturity Levels

04 OPTIMIZED

Operational Excellence and Continuous Improvement

Seamless, collaborative environment for CI/CD.

Fully automated monitoring and model/prompt refinement.

03 MANAGED

Advanced LLM Workflows and Proactive Monitoring

Comprehensive prompt management, evaluation, and real-time deployment.

Advanced monitoring and automated alerts.

02 DEFINED

Systematizing LLM Apps Development

Iterative model augmentation with prompt engineering and RAG.

Structured deployment and prompt-based evaluations.

01 INITIAL

The Foundation of Explorations

Discovery of models and testing prompts.

Basic evaluation and monitoring.

Microsoft Azure

GenAIOps

Maturity Levels

04 OPTIMIZED

Operational Excellence and Continuous Improvement

Seamless, collaborative environment for CI/CD.

Fully automated monitoring and model/prompt refinement.

03 MANAGED

Advanced LLM Workflows and Proactive Monitoring

Comprehensive prompt management, evaluation, and real-time deployment.

Advanced monitoring and automated alerts.

02 DEFINED

Systematizing LLM Apps Development

Iterative model augmentation with prompt engineering and RAG.

Structured deployment and prompt-based evaluations.

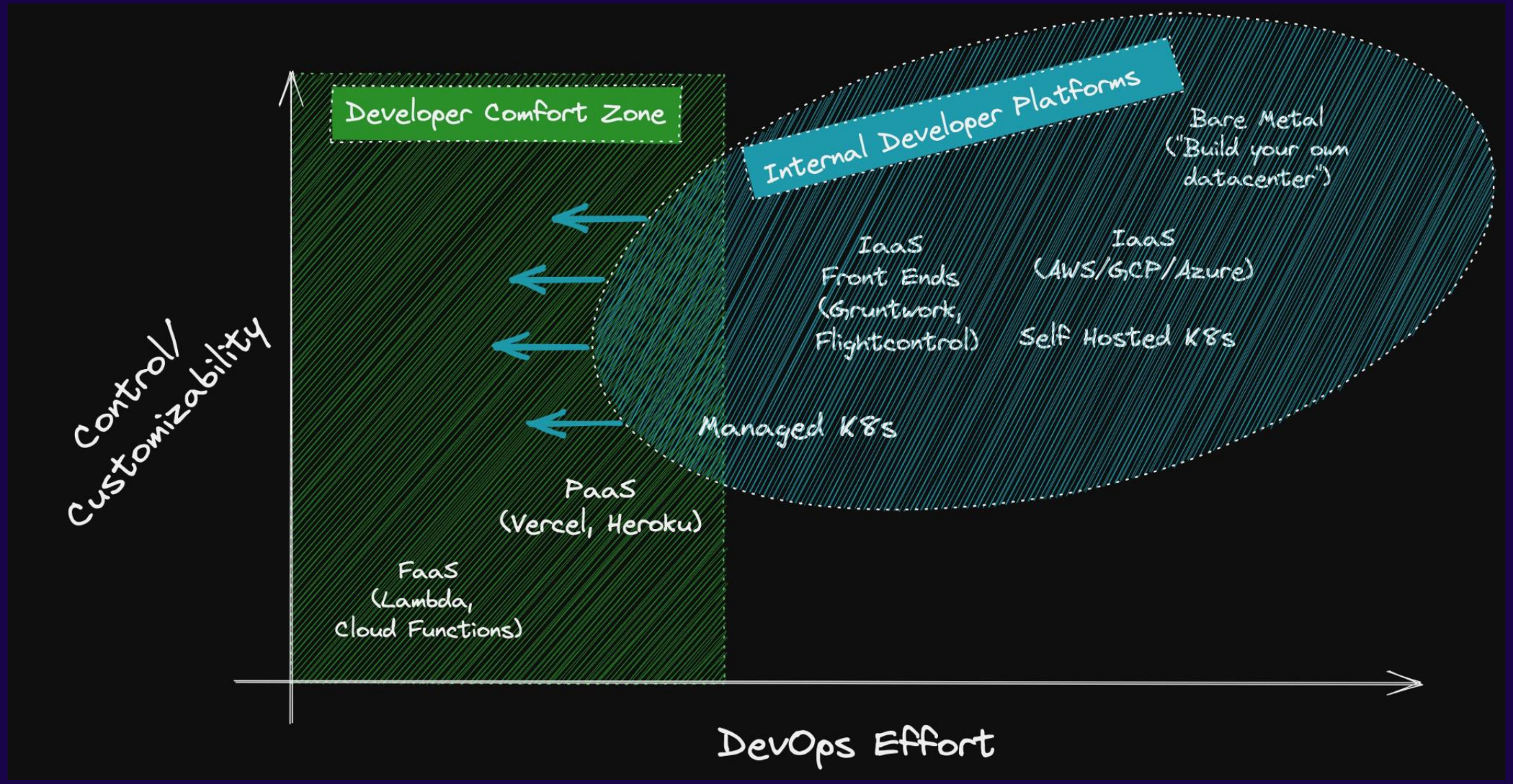
01 INITIAL

The Foundation of Explorations

Discovery of models and testing prompts.

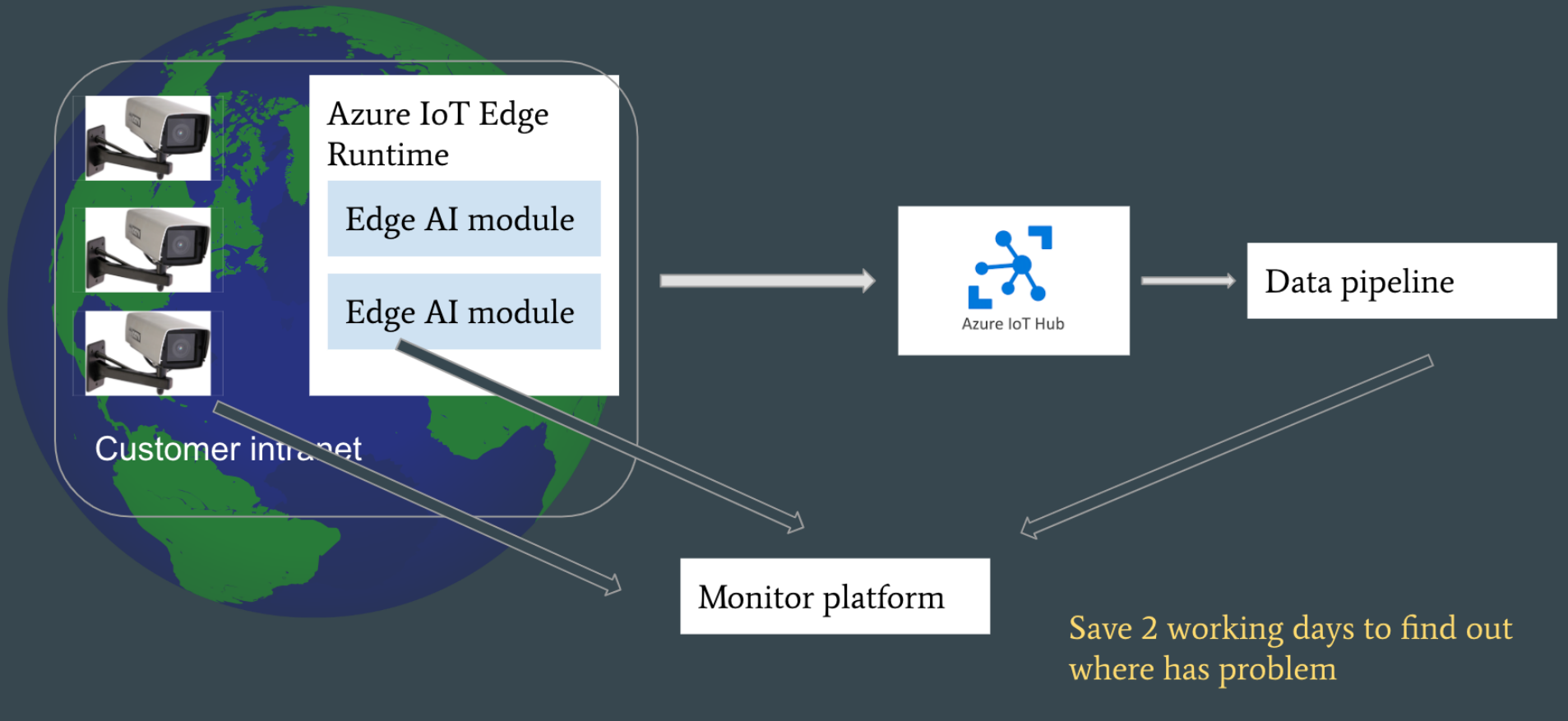
Basic evaluation and monitoring.

簡單來說，平台工程就是讓工程師們更好維運的內部平台

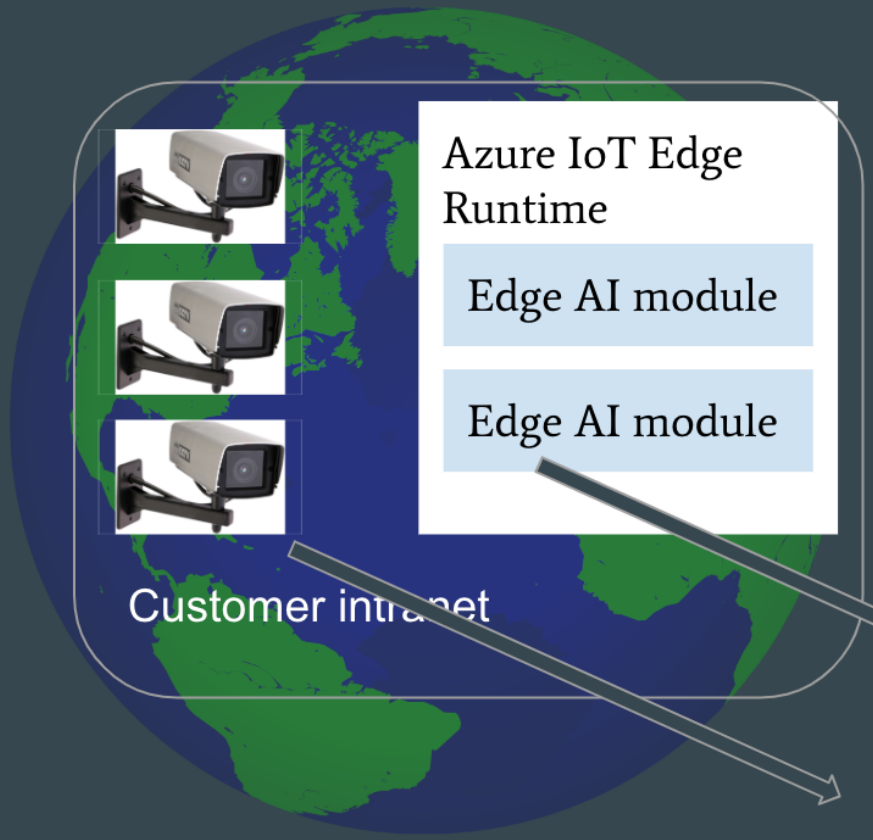


在澳洲 AI 公司工作的那段故事.....

We built monitor platform at first



And we put more features for internal developers



Azure IoT Edge Runtime

Edge AI module

Edge AI module

Customer intranet

Azure IoT Hub

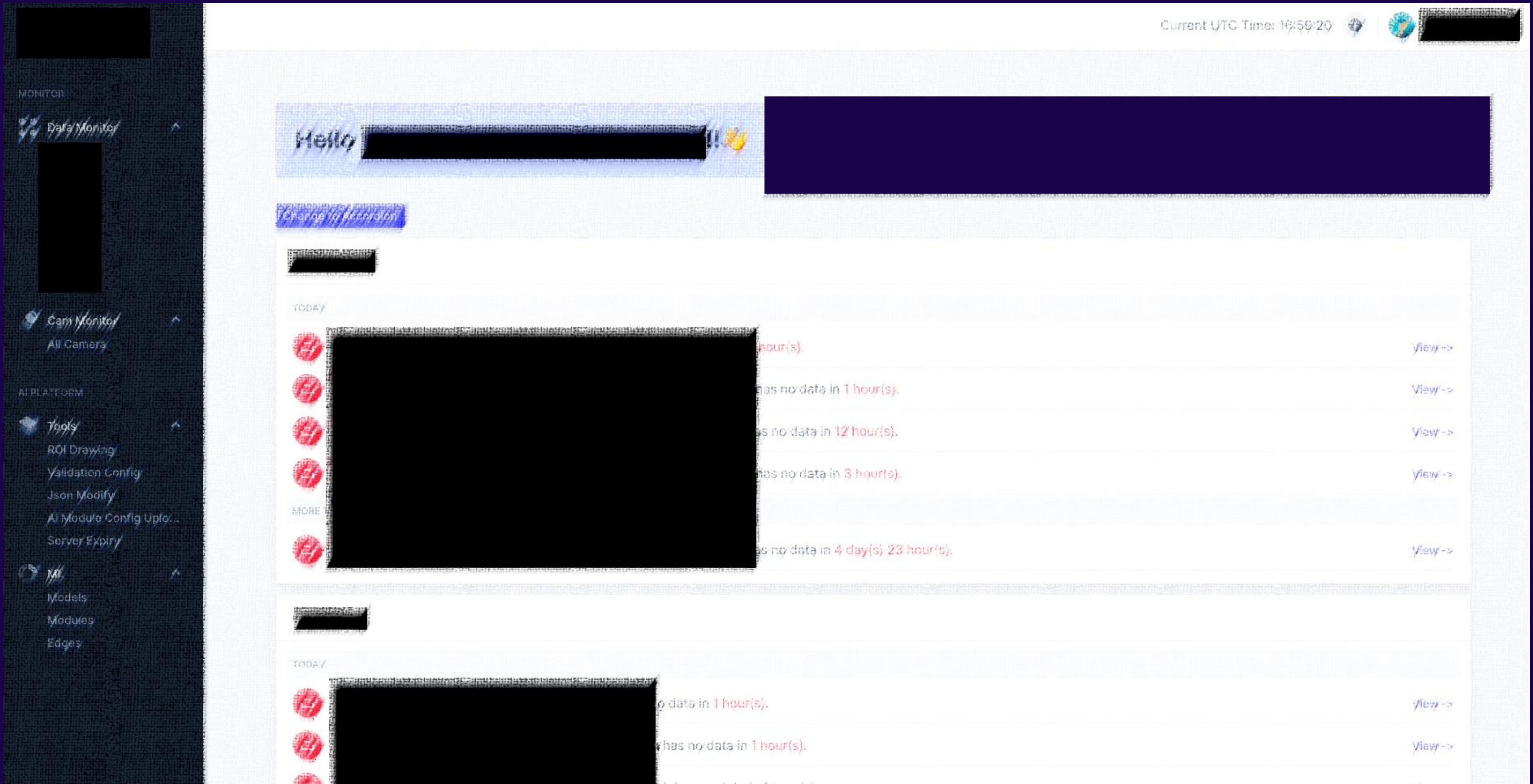
Data pipeline

Platform engineering

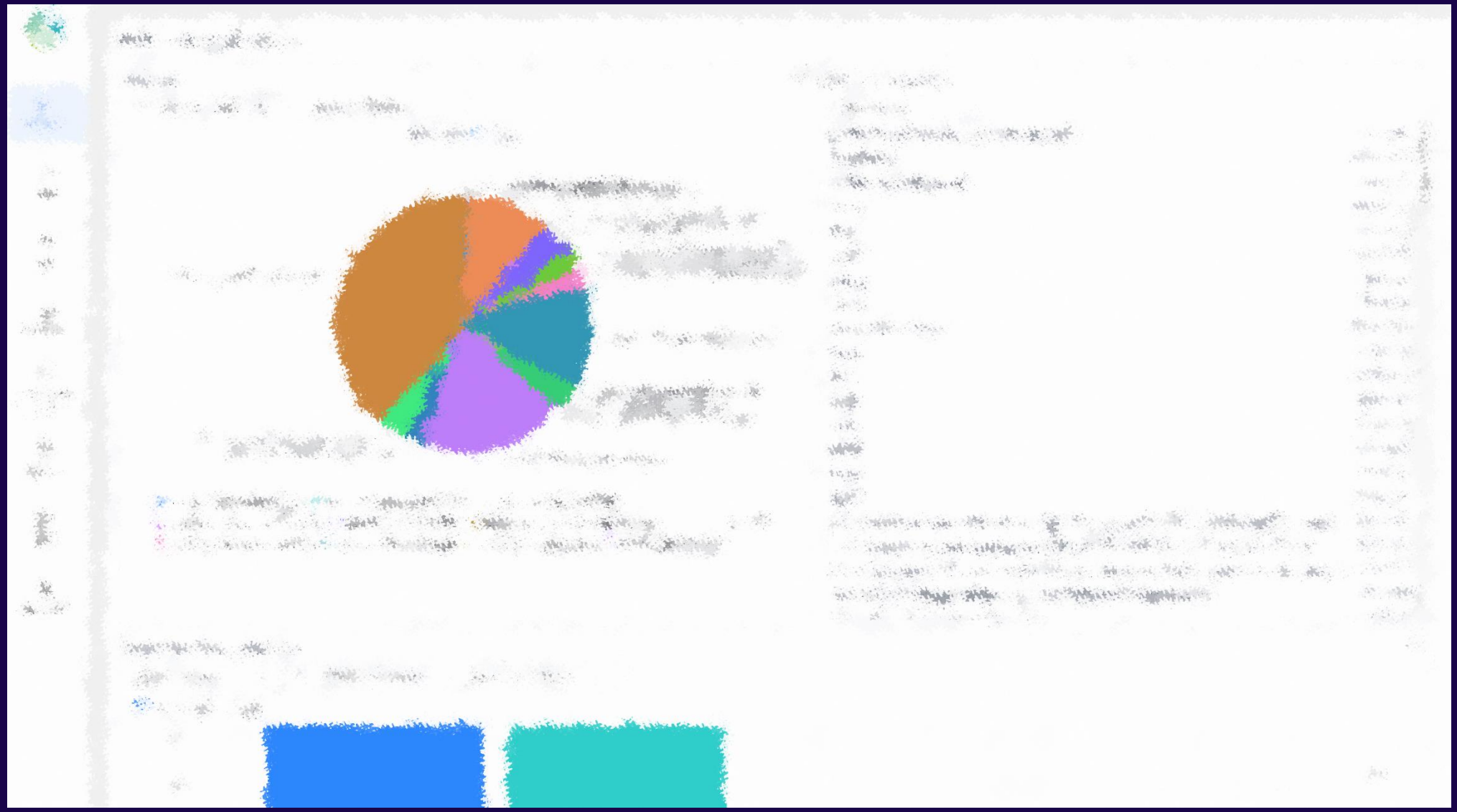
- Tool of AI deploy JSON
- Tool of labeling data
- Tool of edge server mngt
- Tool of data recovery

因此 ML 專案長出平台工程是很自然的事

.NET Conf TAIWAN 2024



當然也會有 LLMOps 版本的平台工程



結論：Prompt 、資料更新、RAG效果等，才會是做生成式應用的你，所要關注的 LLMOps 重點

免責聲明：

今天的內容以生成文字的 LLMOps 為主

當然還有圖像、聲音、影片等性質的 LLMOps，要關注的就更多了

.NET Conf TAIWAN 2024

目前已經做過生成式 AI 專案類型，歡迎連絡

- 法律
- 政府部門
- 金融業
- 電商
- 公益團體
- 展覽

Thank you
Q&A

聯絡方式

加LINE會後

拿投影片