.NET Conf TAIWAN 2024

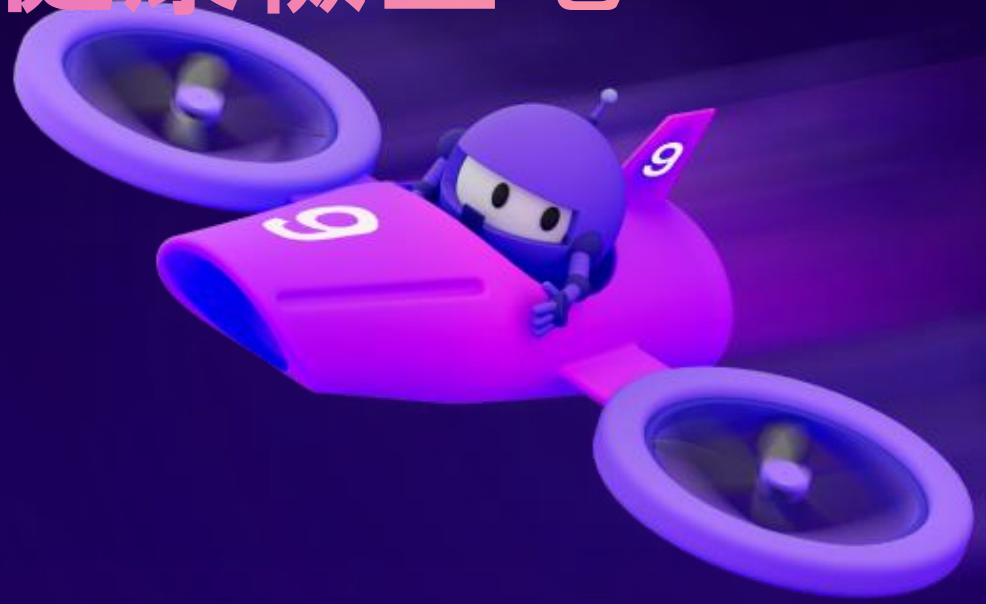
STUDY4

為 學 習 而 生

不時輕聲地以幻覺遮羞的大型語言模型

來幫你的 LLMOps 做個健康檢查吧

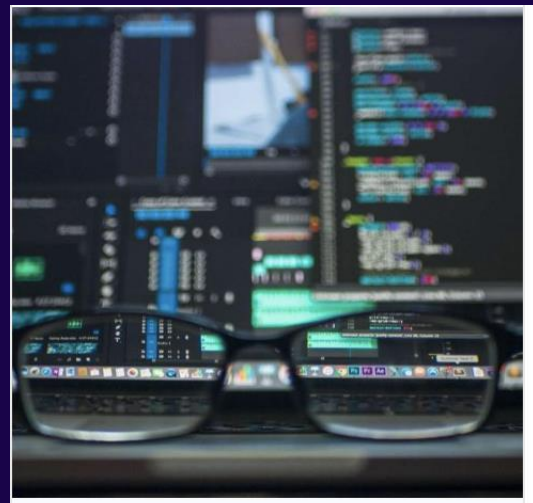
孫玉峰
SUMMIT SUEN
台灣角川數據架構師



.NET Conf TAIWAN 2024

孫玉峰
Summit Suen

台灣角川數據架構師
Microsoft AI MVP
R Ladies Taipei 共同主持人



台灣需要本土AI？中研院大語言模型，竟答國慶為10/1！ | 遠見雜誌
遠見雜誌 - 前進的動力

ChatGPT 有求必應，真要成為「萬事通」仍有哪些挑戰？ |...
Sunrise 旭時報

.NET Conf TAIWAN 2024

特別感謝

Microsoft® MVP Most Valuable Professional

tsmc

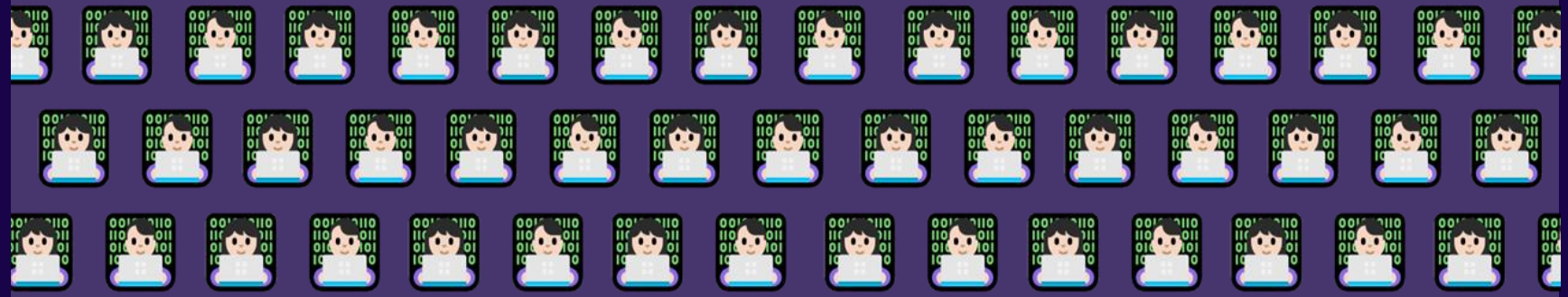
好房網 HouseFun

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

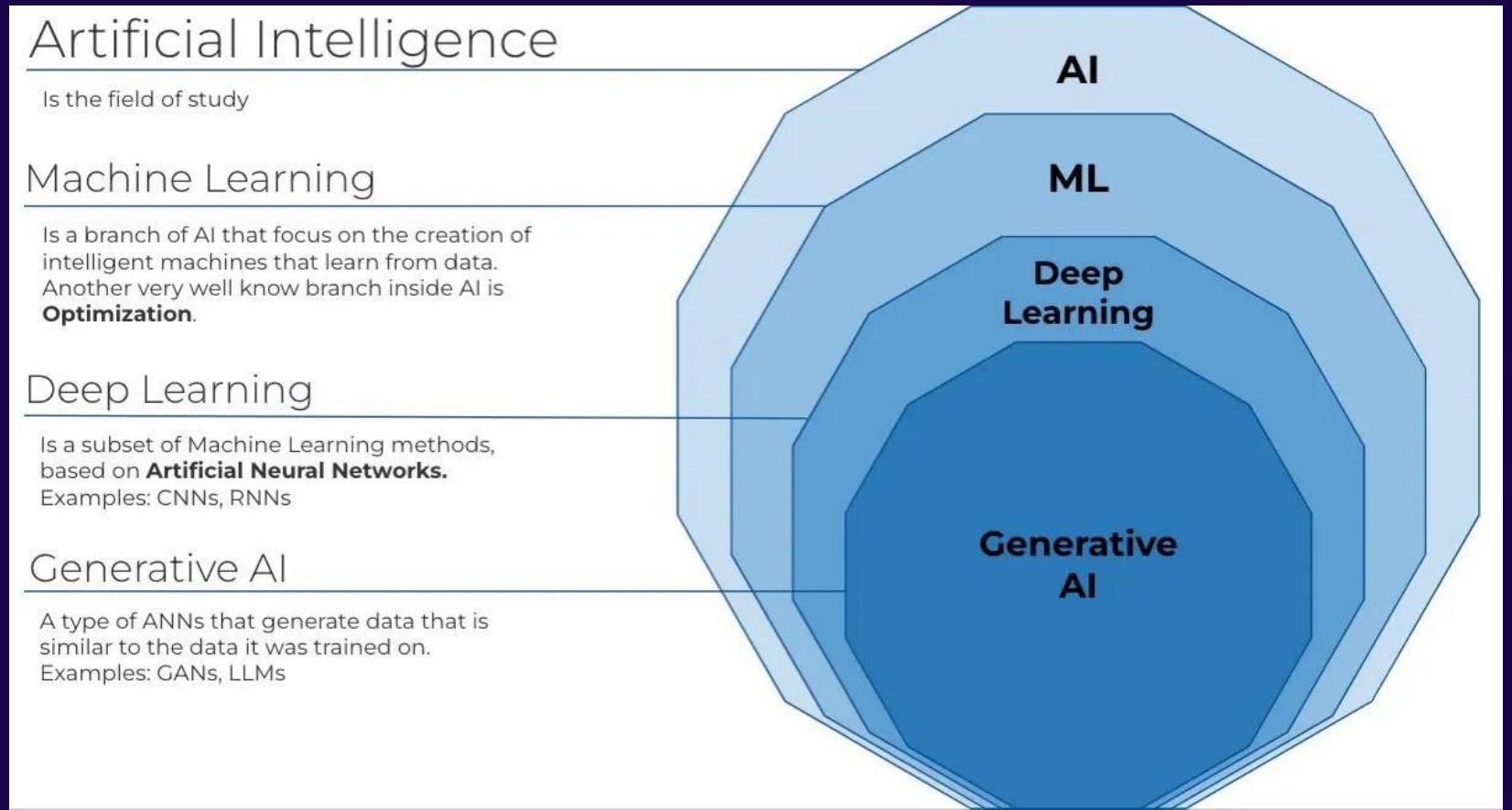


.NET Conf TAIWAN 2024

幻覺不是 BUG，是 FEATURE！

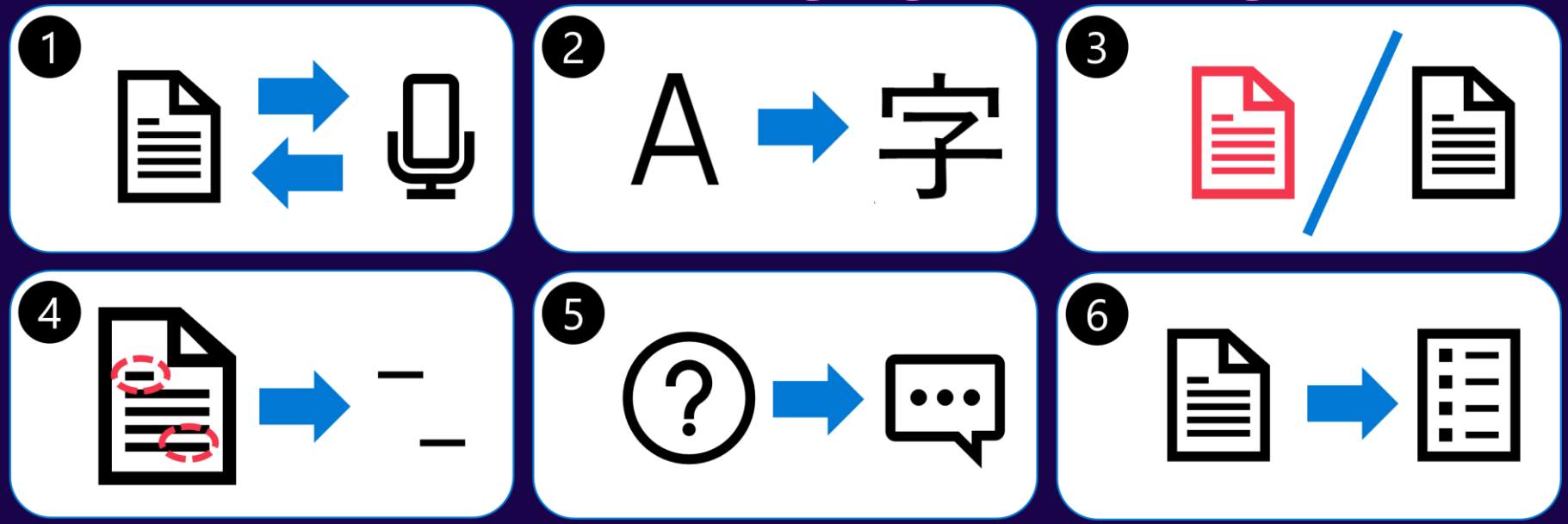
關於 AI 的知識

- GenAI ⊂ DL ⊂ ML ⊂ AI
- 但現在大家在講的 AI 都侷限在 LLMs



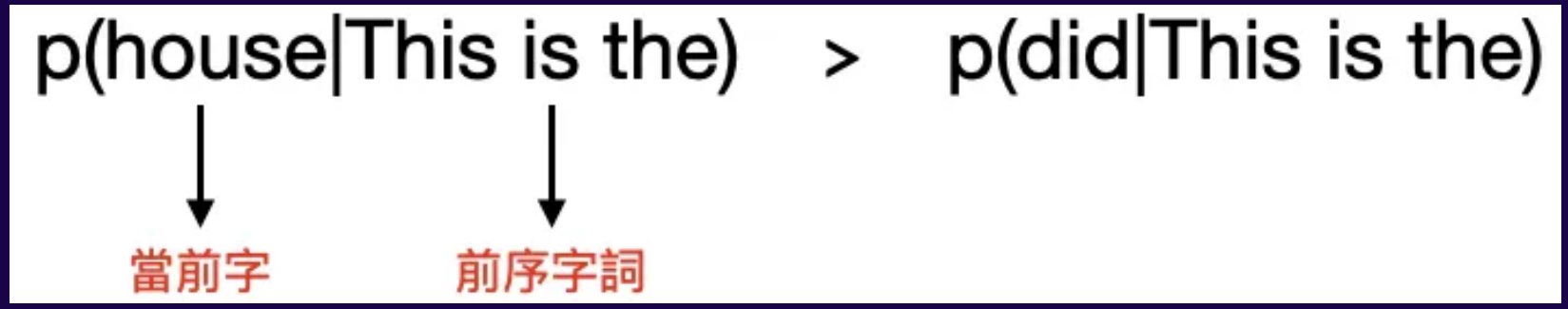
關於 AI 的知識

- 什麼是 LLMs？
- 大型語言模型（Large Language Models）
- 什麼是語言模型？
- 解決自然語言處理（NLP, Natural Language Processing）任務的機率模型



關於 AI 的知識

- N-gram：條件機率（上下文）



- RNN：處理連續資料的神經網路
- LSTM：加入短期長期記憶
- Transformer：注意力機制

關於 AI 的知識

- 幻覺不是 BUG，是 FEATURE！

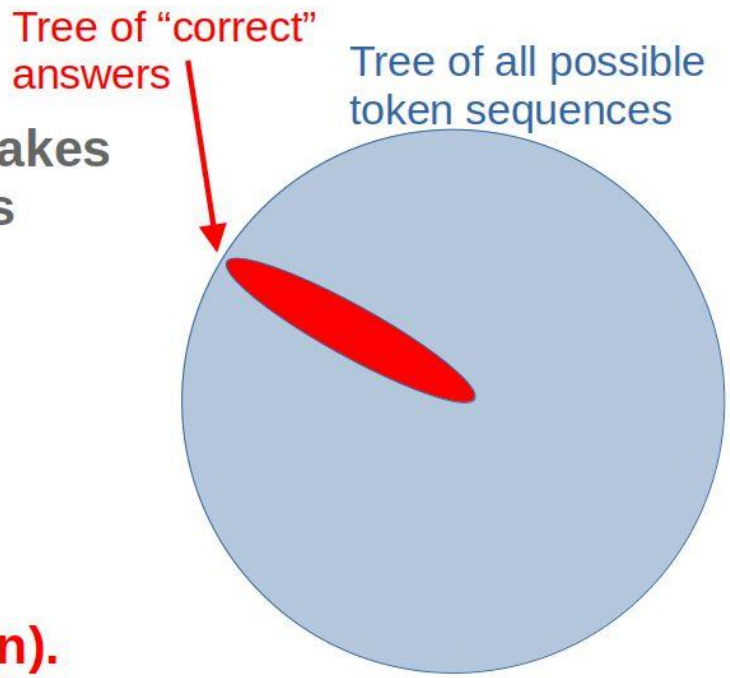
Unpopular Opinion about AR-LLMs

Y. LeCun

- Auto-Regressive LLMs are **doomed**.
- They cannot be made factual, non-toxic, etc.
- They are not controllable

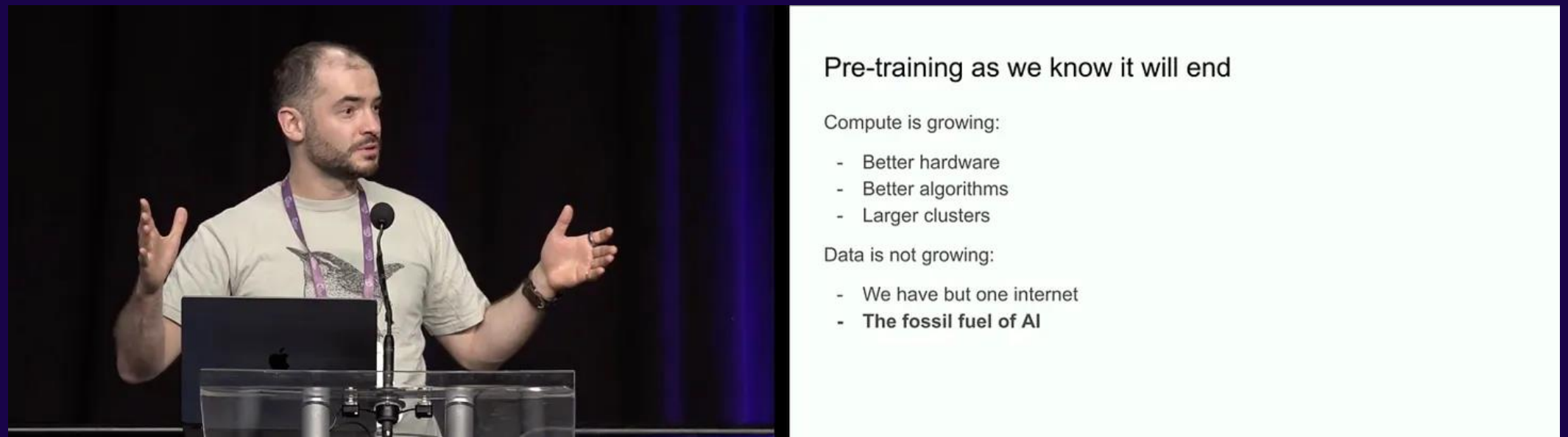
- Probability e that any produced token takes us outside of the set of correct answers
- Probability that answer of length n is correct:
  - $P(\text{correct}) = (1-e)^n$

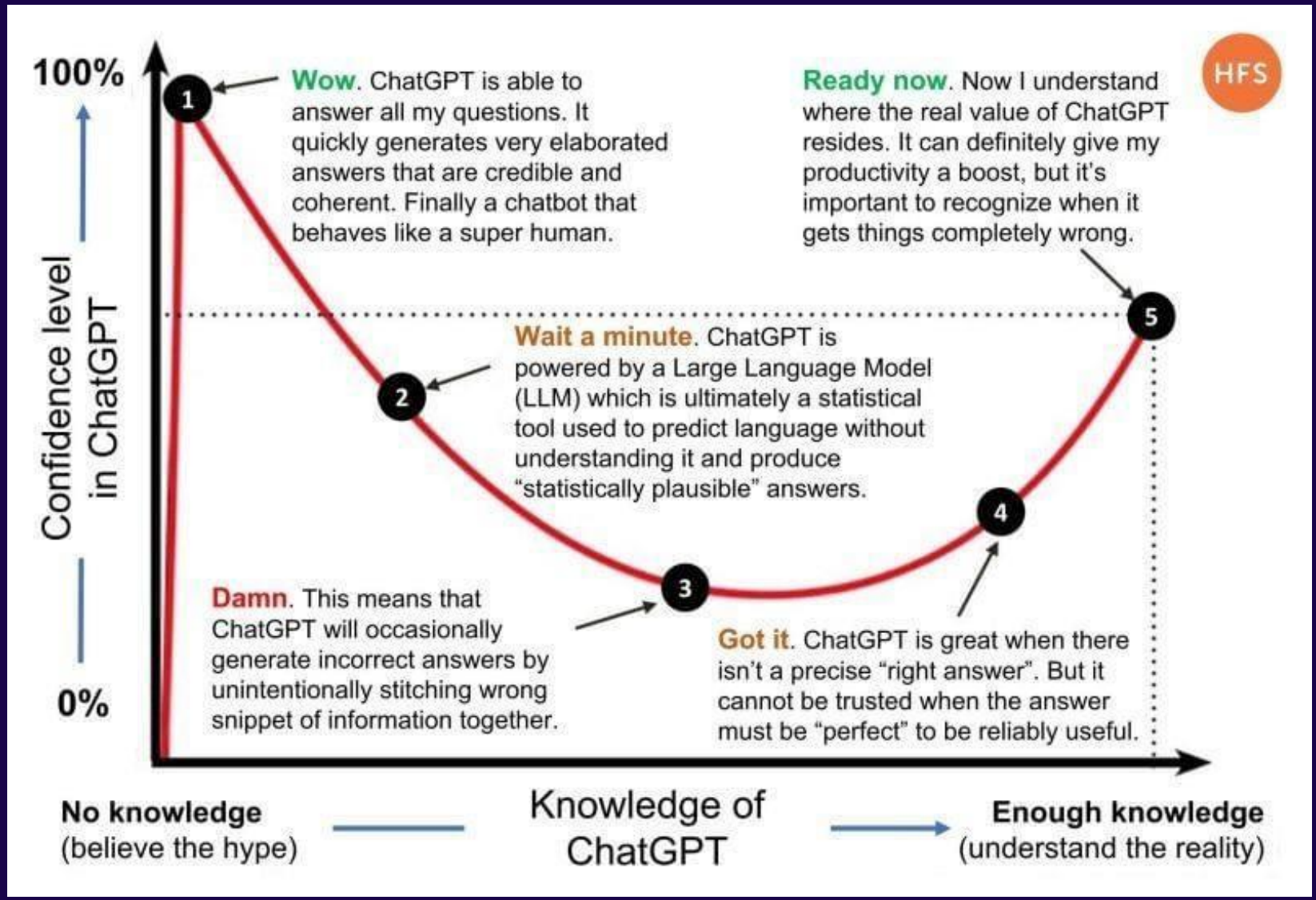
- **This diverges exponentially.**
- **It's not fixable (without a major redesign).**



關於 AI 的知識

- 也不是要黑 LLMs 或是 GenAI
- ChatGPT/LLMs 出圈重點：隨插即用、降低使用門檻
- 只是要知道他的極限在哪裡，或是
- 正視 LLMs 只是處理 NLP tasks 的工具




HFS
100%
0%
Confidence level in ChatGPT
1
Wow. ChatGPT is able to answer all my questions. It quickly generates very elaborated answers that are credible and coherent. Finally a chatbot that behaves like a super human.
2
Wait a minute. ChatGPT is powered by a Large Language Model (LLM) which is ultimately a statistical tool used to predict language without understanding it and produce "statistically plausible" answers.
3
Damn. This means that ChatGPT will occasionally generate incorrect answers by unintentionally stitching wrong snippet of information together.
4
Got it. ChatGPT is great when there isn't a precise "right answer". But it cannot be trusted when the answer must be "perfect" to be reliably useful.
5
Ready now. Now I understand where the real value of ChatGPT resides. It can definitely give my productivity a boost, but it's important to recognize when it gets things completely wrong.
No knowledge
(believe the hype)
Knowledge of ChatGPT
Enough knowledge
(understand the reality)

使用 AI 的知識

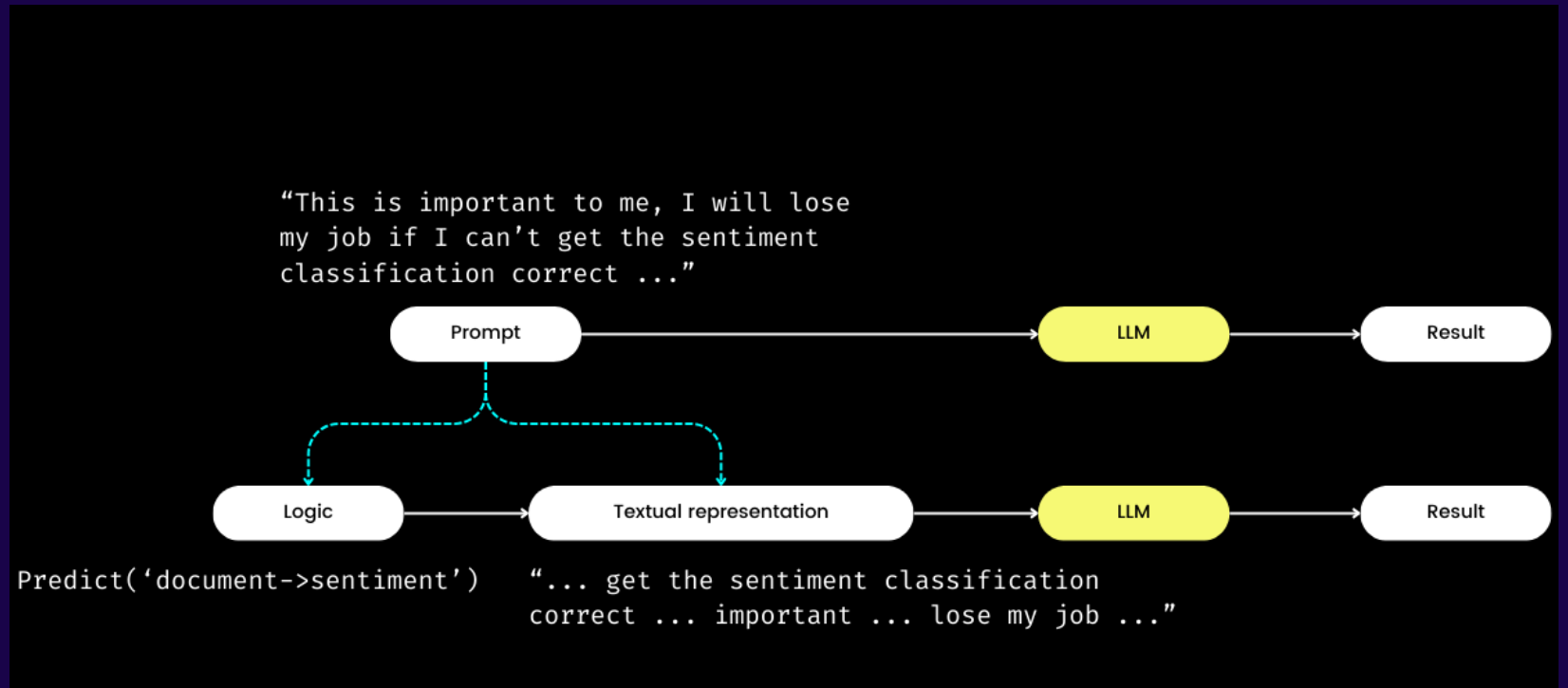
- 單純一點拿來處理 NLP tasks
- 🤗 Transformers

```
from transformers import pipeline
pipe = pipeline("text-classification")
pipe(["This restaurant is awesome", "This restaurant is awful"])

[{'label': 'POSITIVE', 'score': 0.9998743534088135},
 {'label': 'NEGATIVE', 'score': 0.9996669292449951}]
```

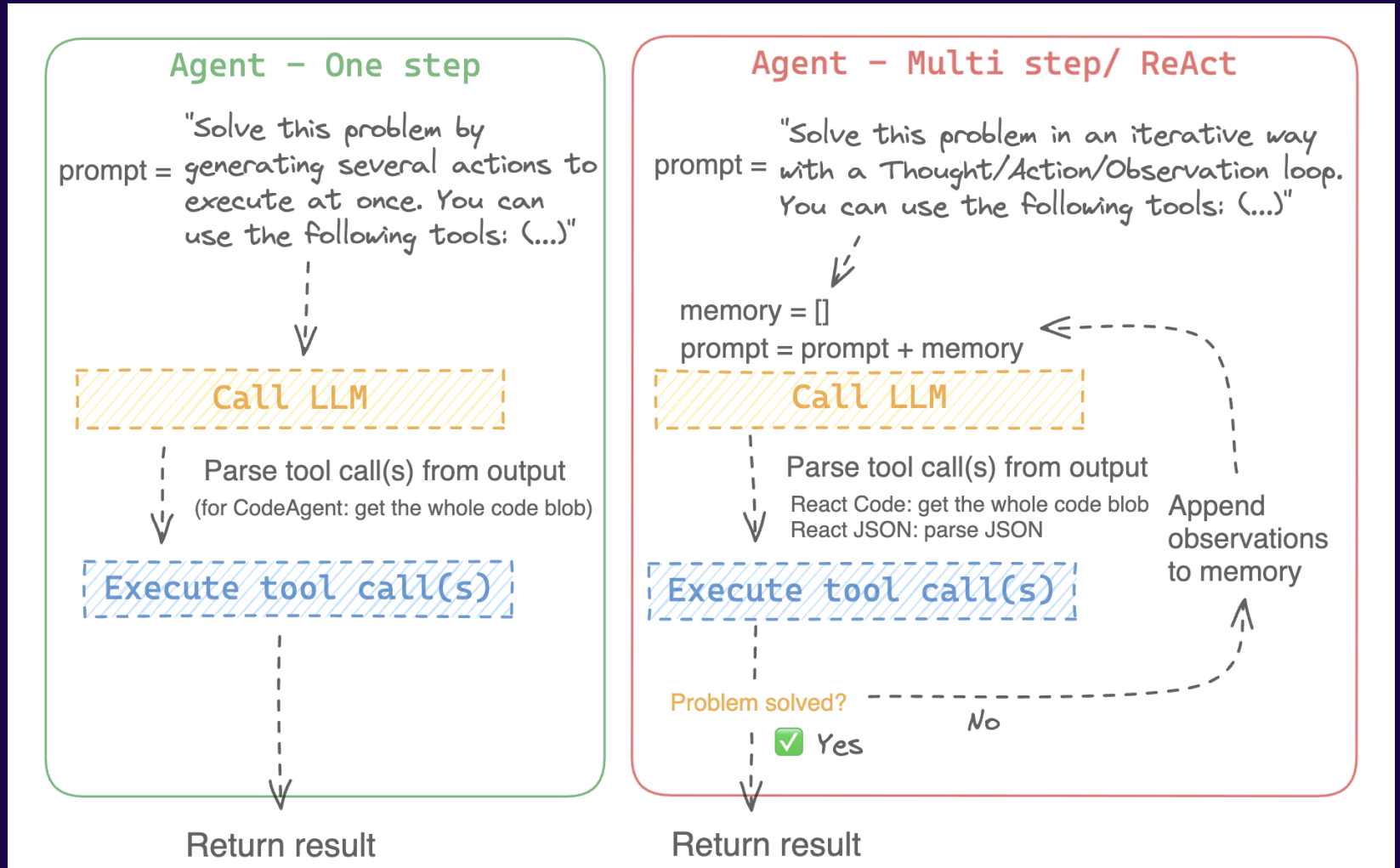
使用 AI 的知識

- DSPy: Programming—not prompting—LMs
- 用 LMs 來寫程式的 framework（而不是 tune prompt）
- https://ithelp.ithome.com.tw/articles/10348919



使用 AI 的知識

- AI Agent, or Agentic AI



使用 AI 的知識

- **如何 bound 住一個會發散的東西？**
  - 限制範圍：RAGs
  - 切小 task，每個 task/step 都要 evaluate

SuperAnnotate

Type	Description	Example metric
Diversity	How does the model respond to different types of queries?	Fluency, Perplexity, ROUGE scores
User feedback	Uses human feedback to check preference alignment and accuracy	Coherence, Quality, Relevance
Correctness	Compares RAG's responses to a set of predefined, correct answers	Binary classification (Correct/Incorrect)
Relevance	How relevant is the LLM's response to a given user's query?	Binary classification (Relevant/Irrelevant)
Toxicity	Are the responses racist, biased, or toxic?	Disparity Analysis, Fairness Scoring, Binary classification (Non-Toxic/Toxic)

使用 AI 的知識

- 連 MLOps 都沒聽過的我就要用 LLMOps 是不是搞錯了什麼？
  https://dotnetconf.study4.tw/Speaker#Summit

- 與其想方設法消除幻覺，不如去用把握度較高／對症下藥的方法
  （通才的大模型 v.s. 專精的小模型）

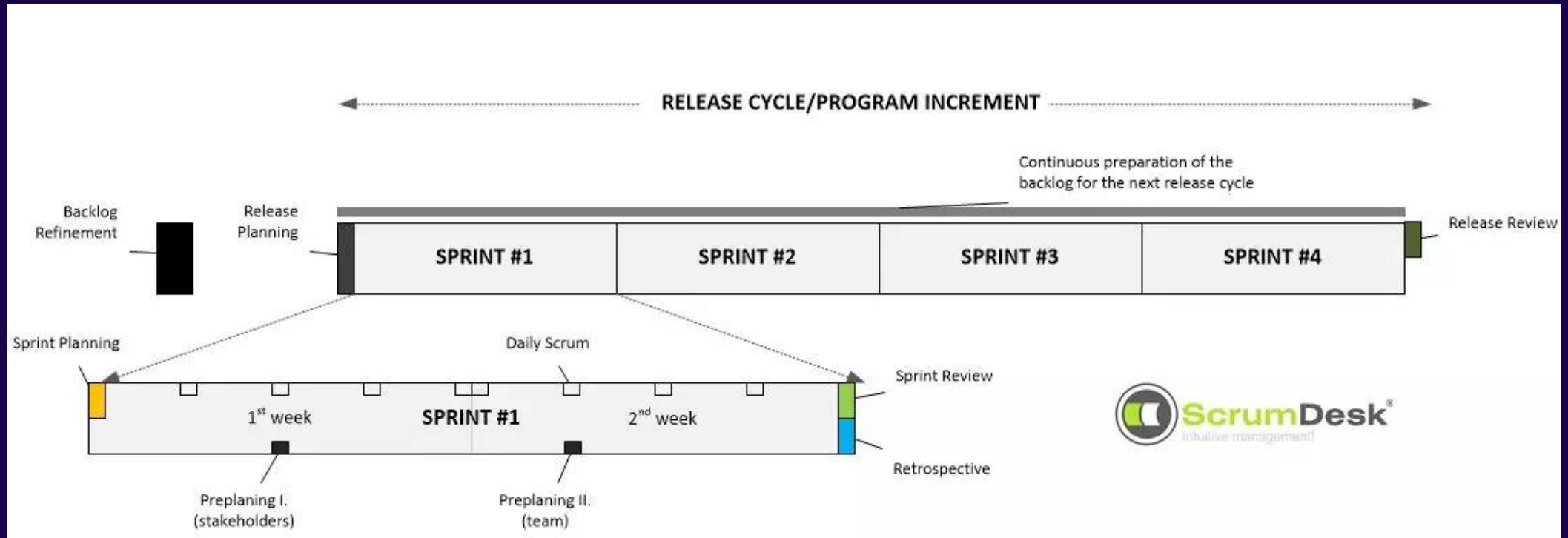
- 評估驅動開發 Eval-Driven Development (EDD): 生成式 AI 軟體不確定性的解決方法 from Wen-Tien Chang

Agile is not enough!
Be Anti-FrAgile

.NET Conf TAIWAN 2024

LLMOps vs. Scrum

- AI Agent AS (Junior) Engineer
- AI Engineer AS Scrum Master
- Stack Holder AS Product Owner



Case Study

Kadokado
角角者輕小說平台

推薦系統

詐騙文章偵測

……

你好像會喜歡這些作品(ゝ∀･)　換一換

人們將那抹藍稱為神、希…

科幻

下屬出租方案

現代社會

原來穿書是為了談戀愛

BL

配角全詣

BL

.NET Conf TAIWAN 2024

Thank you