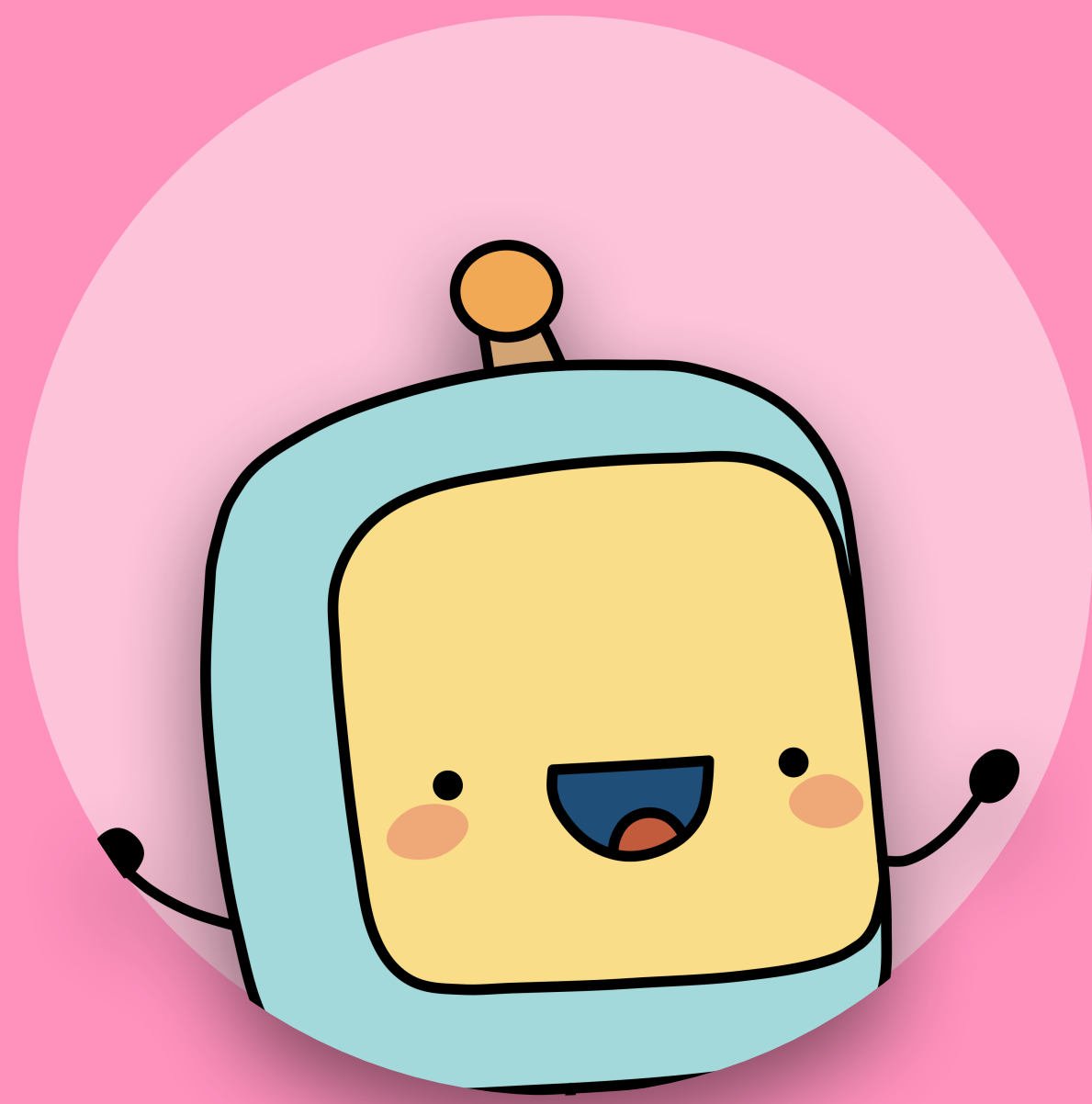


14. Diffusion Models 要被擊潰了嗎？



蔡炎龍

政治大學應用數學系

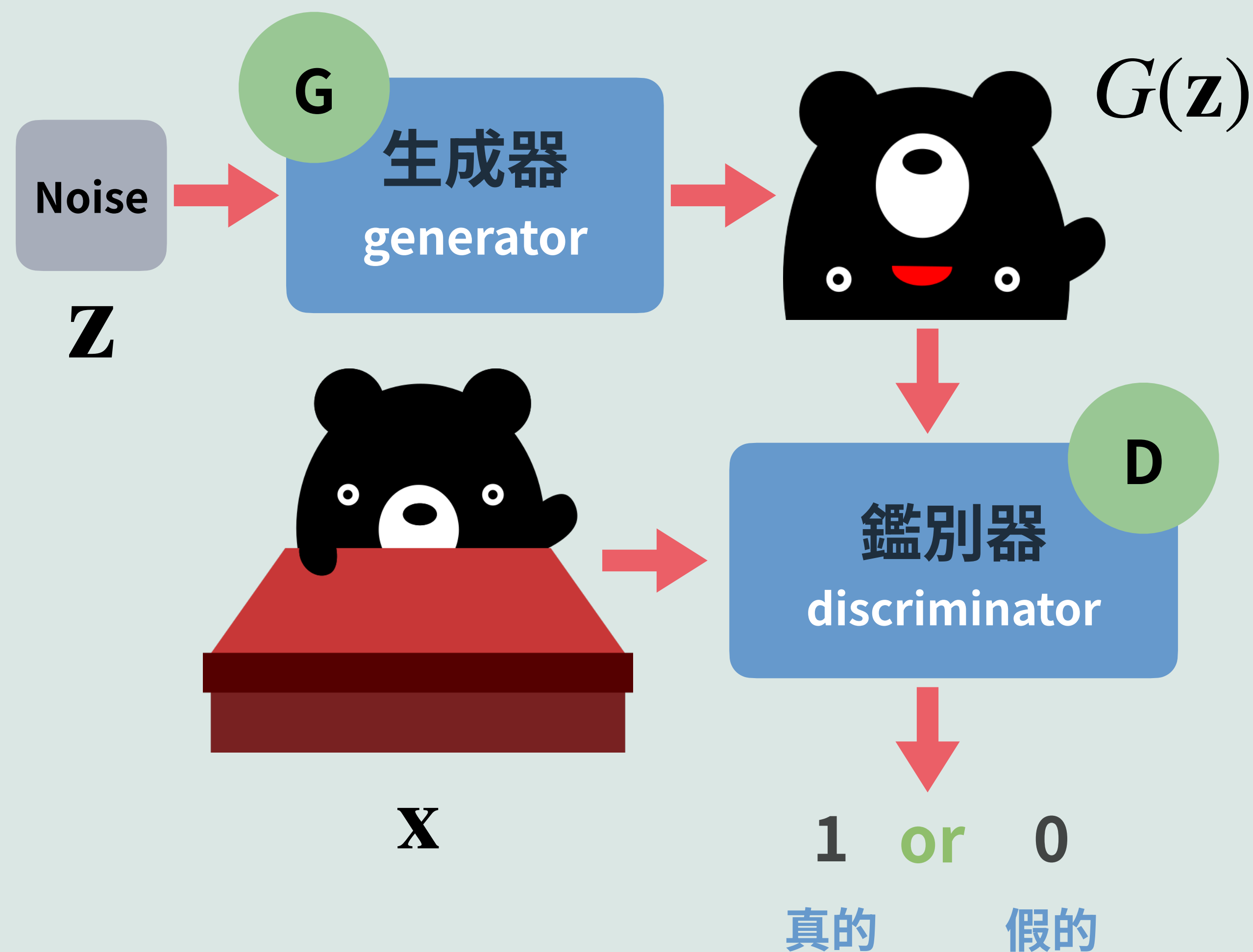


01.

回顧我們學到的東西



曾經 GAN 被認為生成型 AI 的希望



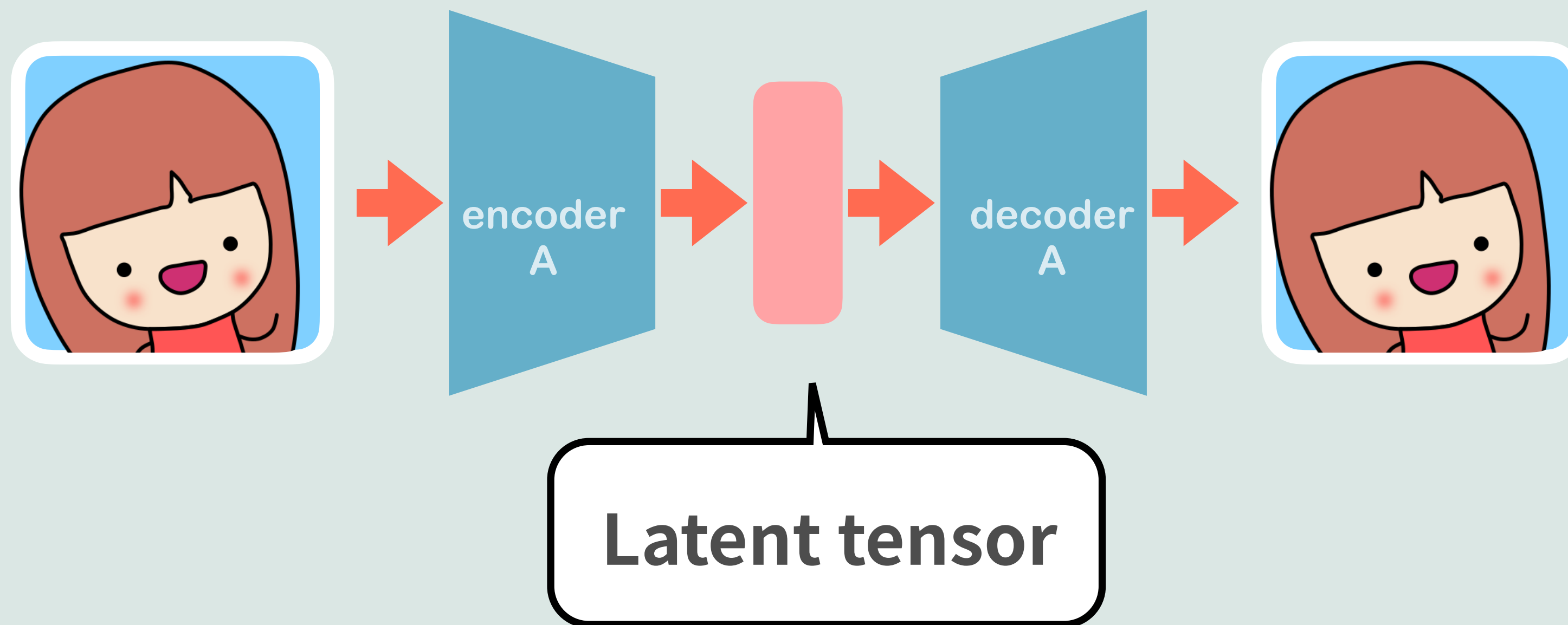
GAN 是兩個神經網路，
一個叫**生成器**、一個叫
鑑別器，相互對抗！



後來被 Diffusion Models 打得快要消失

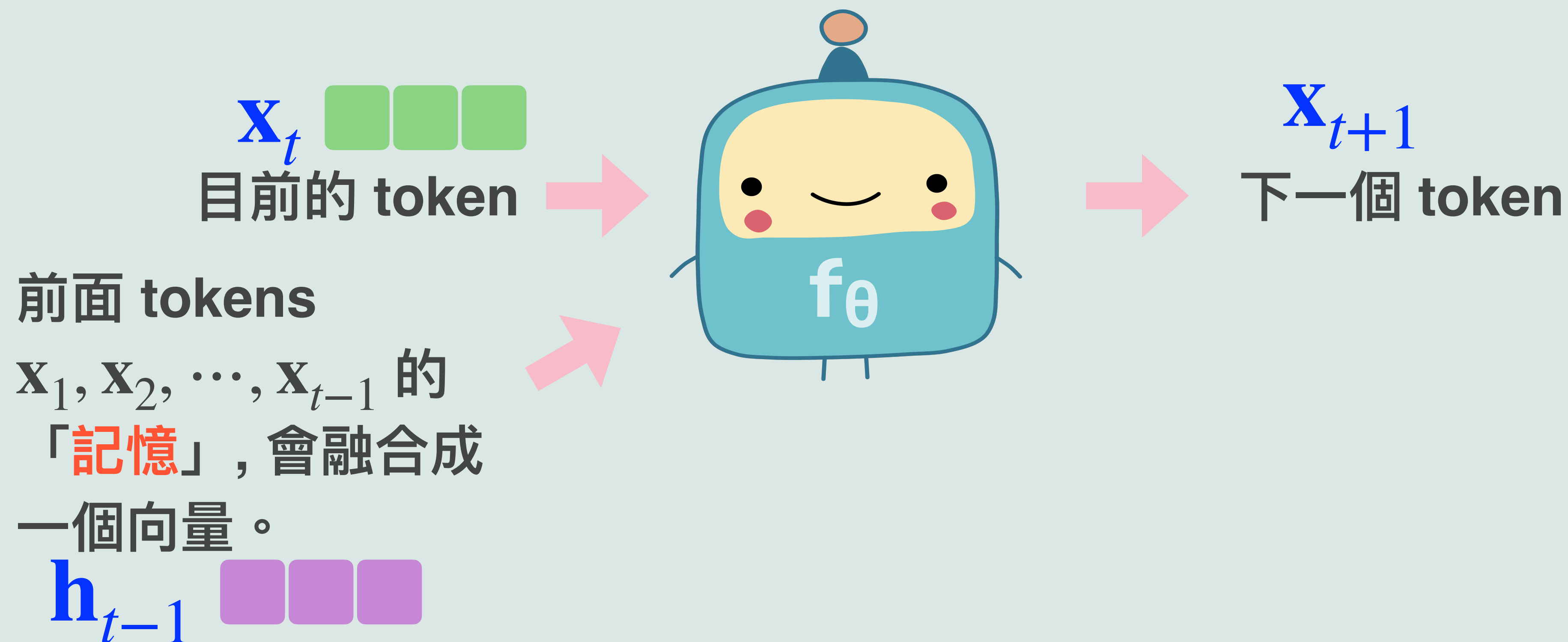
這是算出來的!

這裡是神經網路



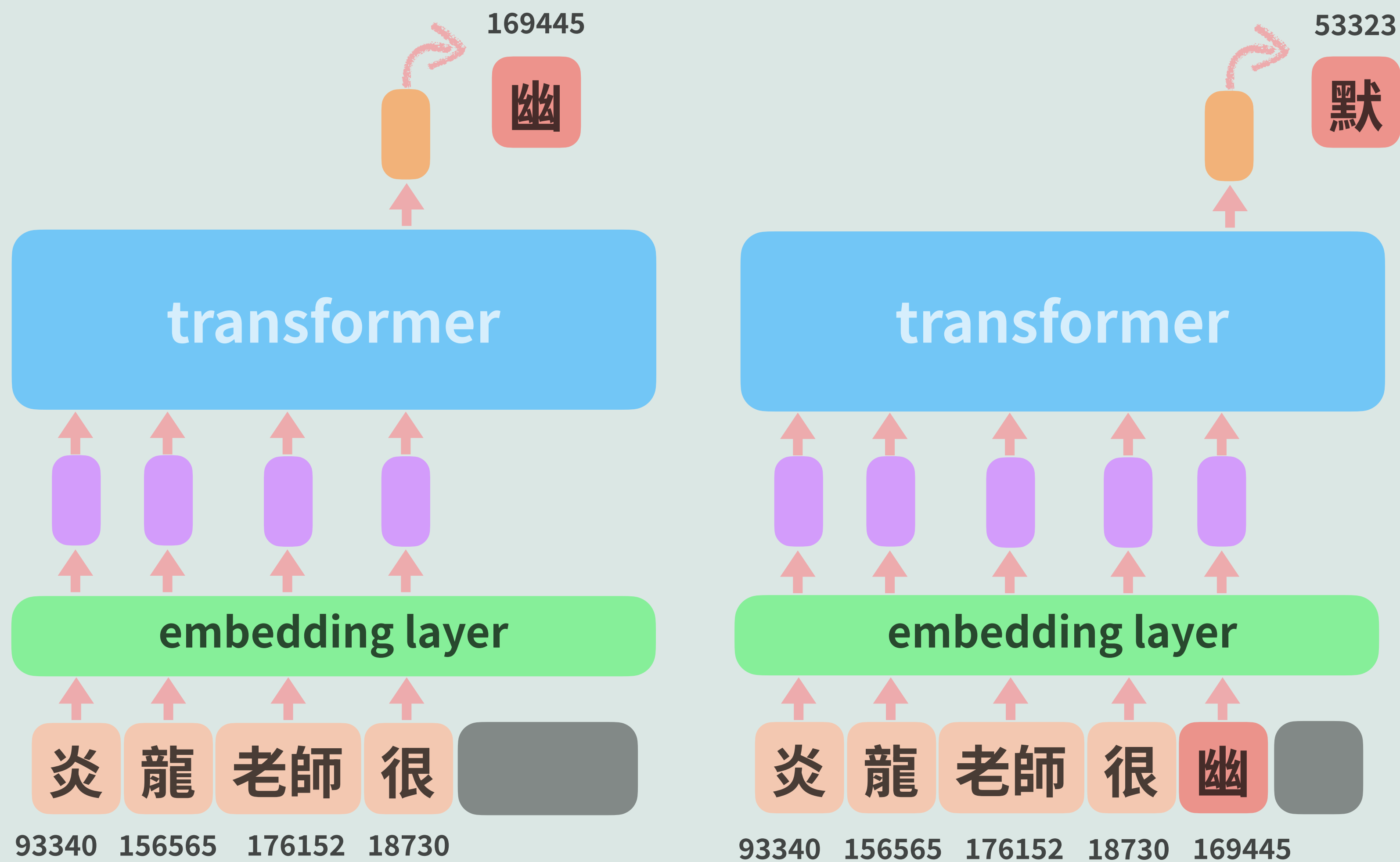


預測下一個 Token 的 LLM



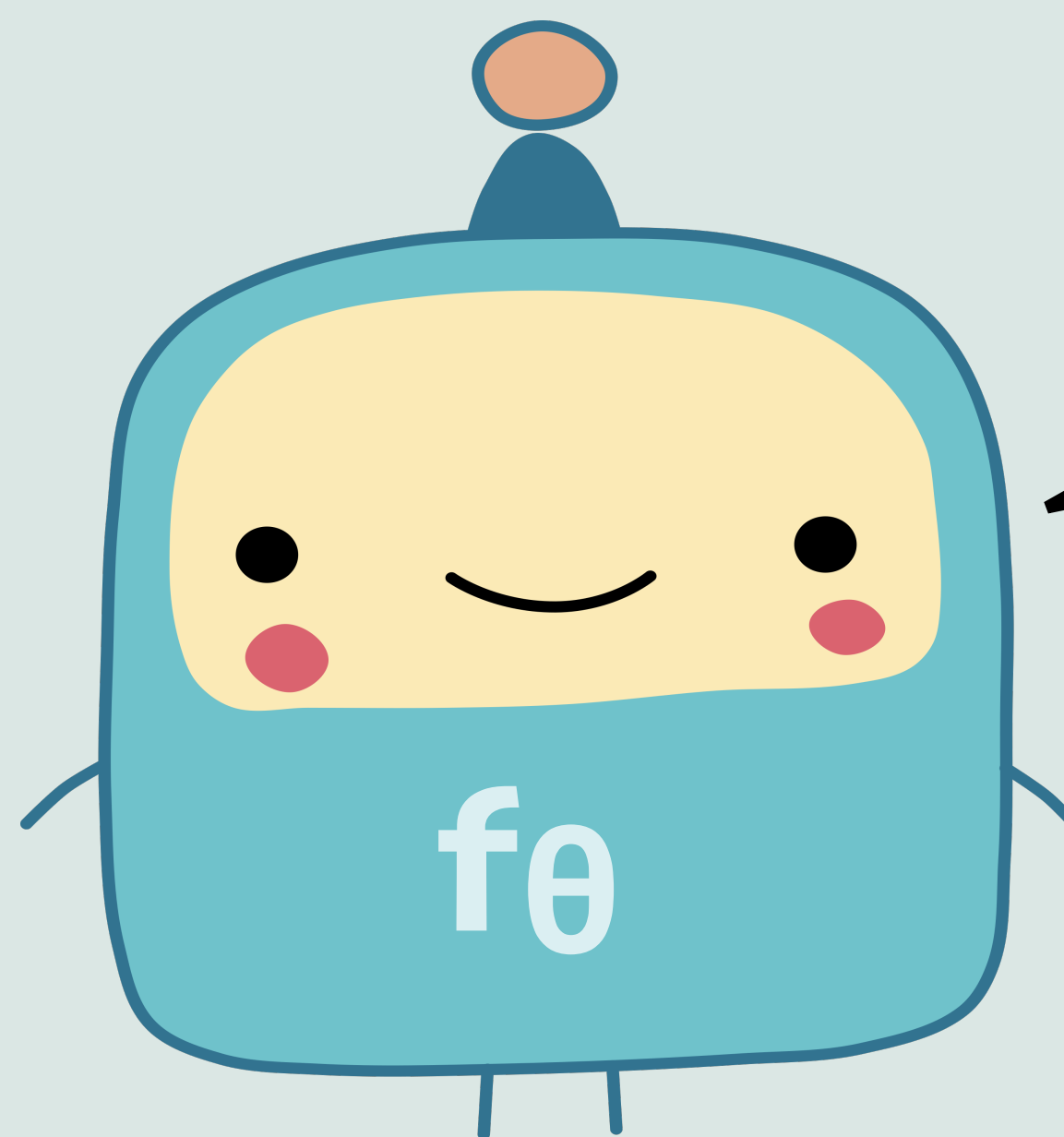


LLM 的原理





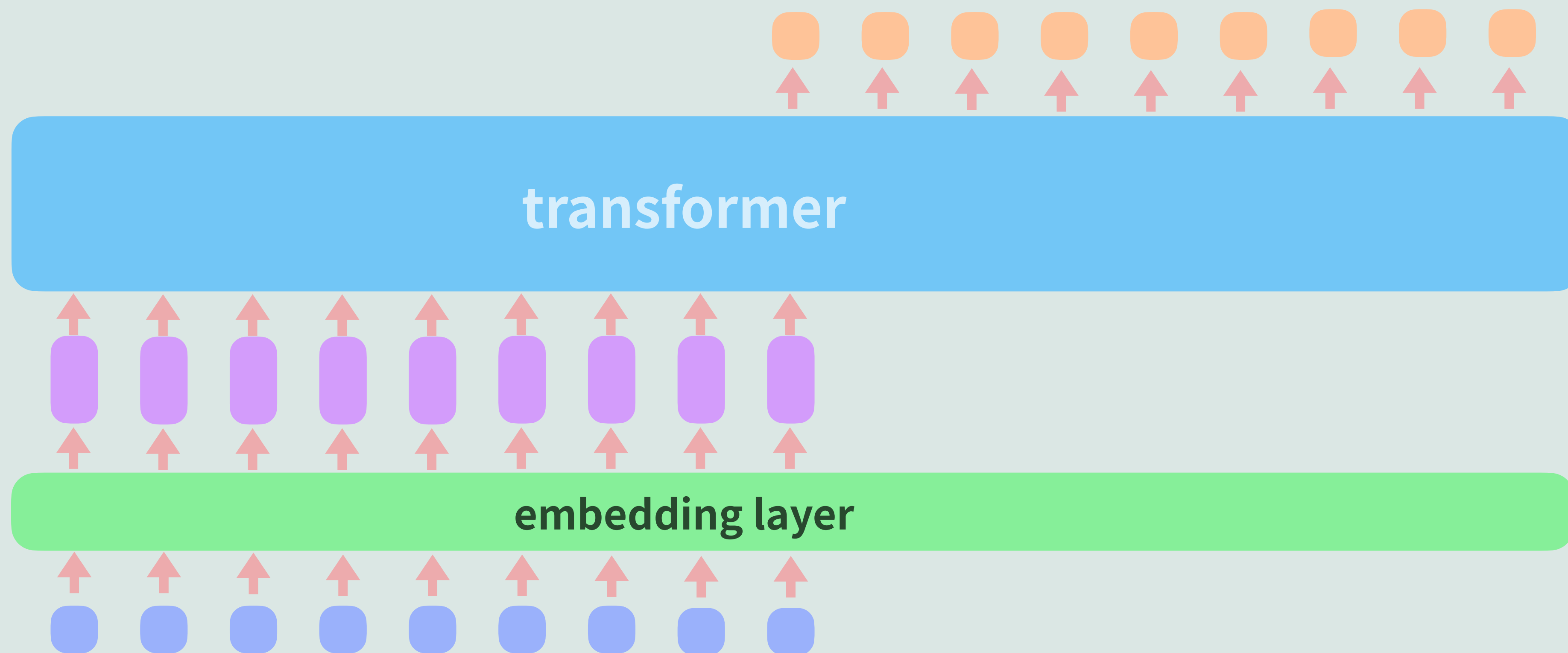
突然發現, 如果把圖像當一種「語言」



如果我會讀, 會
寫圖像這個「語
言」...

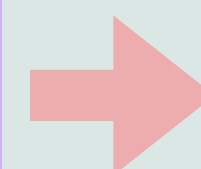
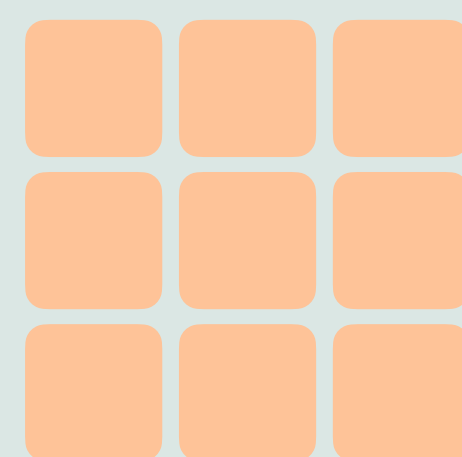


像文字生成一樣預測下一個 token





然後把圖生出來!

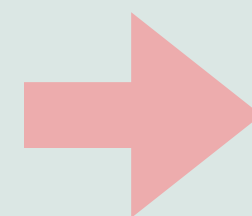


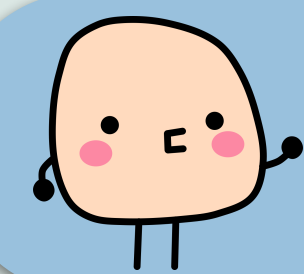


3D 卡通化 (ChatGPT)



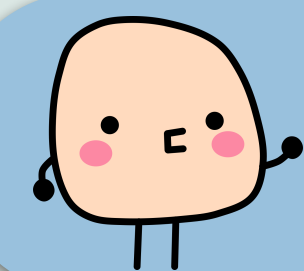
請把我們的吉祥物 "DIVE"
呆萌 AI 機器人也用同樣的
風格, 3D 卡通化。





各種情境都可以產生 (Grok)





還有 Gemini 3 文字產生能力很強





Gemini 學生方案申請快要截止了



Google Gemini：學生可免費 使用 Pro 方案

免費使用 1 年。以更高權限存取 Google 最精準的模型 Gemini 3 Pro，可無限上傳圖片、生成專業級圖像、自訂測驗、使用 NotebookLM 等進階學習工具，並享有 2 TB 儲存空間。優惠兌換期限為 2025 年 12 月 9 日。

[取得優惠](#)[適用相關條款](#)



下週：「你不知道的 Gemini 3 Pro」

鳥巢老師

- 30+ 年視覺設計師
- 22 年人文、婚禮攝影師
- Nikon 官方攝影學校講師
- Adobe 官方特約講師
- 最早運用生成式 AI 的設計專業人士





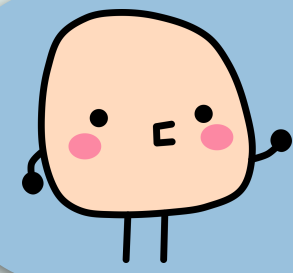
曾經盛極一時的 Diffusion Model, 突然沒人討論...

Diffusion
Models 玩完了!?





02. Diffusion Model 也來 文字生成!?



Inception Labs 的 Mercury



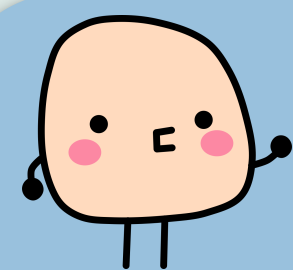
Mercury Coder

Try our first commercial-grade diffusion LLM

|How can I help you code?



<https://chat.inceptionlabs.ai/>



目前看來比較專注在程式生成上面

**Mercury-Coder**

下面是一個簡單的高質感計算機程式，使用 HTML、CSS 和 JavaScript 來實現。這個計算機可以進行基本的算數運算，如加法、減法、乘法和除法。

HTML (index.html)

```
html
1 <!DOCTYPE html>
2 <html lang="en">
3 <head>
4   <meta charset="UTF-8">
5   <meta name="viewport" content="width=device-width, initial-scale=1.0">
6   <title>高質感計算機</title>
7   <link rel="stylesheet" href="styles.css">
8 </head>
9 <body>
10 <div class="calculator">
11   <div class="display" id="display">0</div>
12   <div class="buttons">
13     <button class="btn" onclick="appendNumber(7)">7</button>
14     <button class="btn" onclick="appendNumber(8)">8</button>
```

+ Try another coding prompt

Diffusion Effect

0

7

8

9

/

4

5

6

*

1

2

3

-

0

.

=

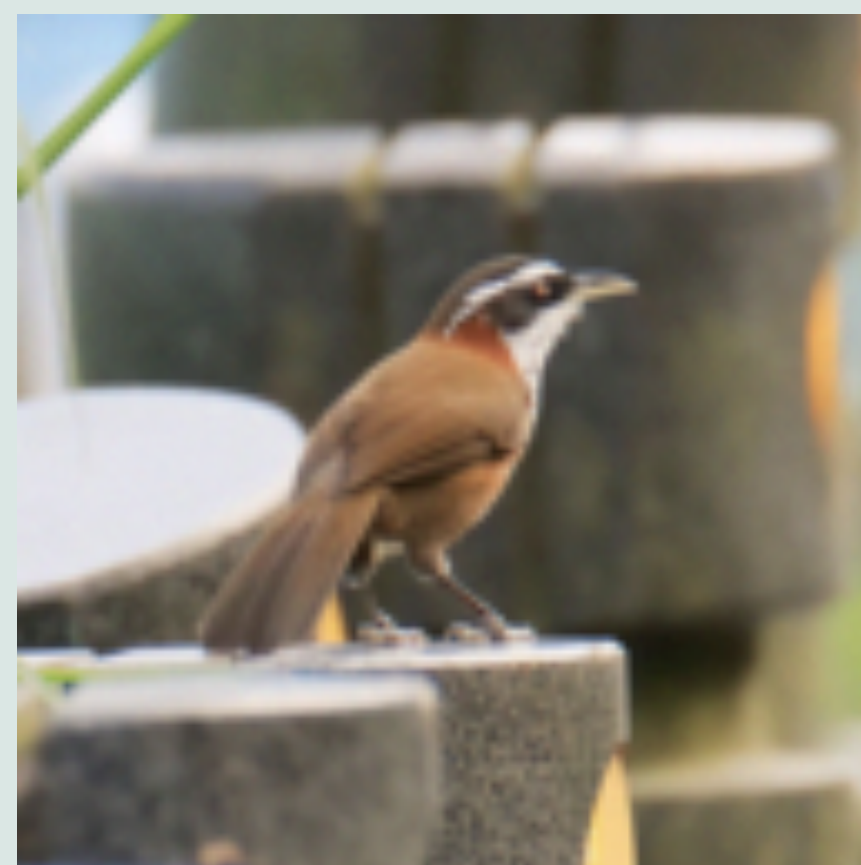
+

C

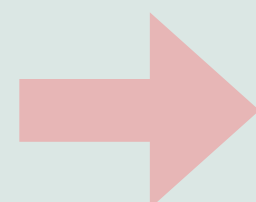


回憶 Diffusion Model (加噪部份)

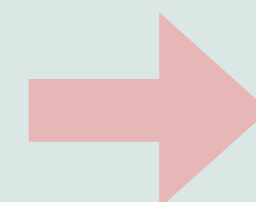
一張清楚的照片，一步步加上高斯雜訊。



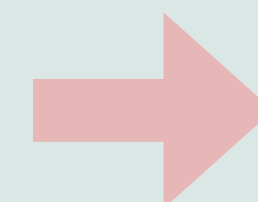
X_0



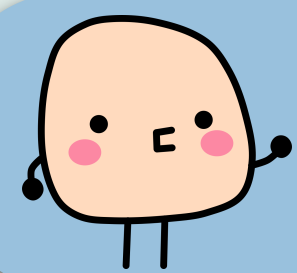
X_{50}



X_{100}



X_{150}

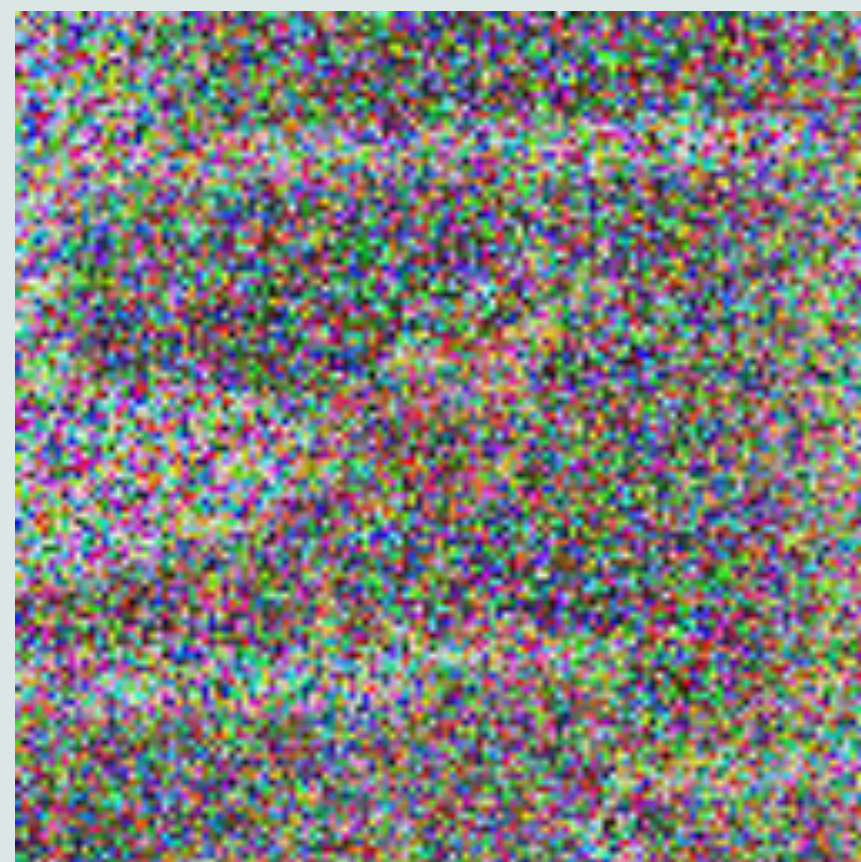
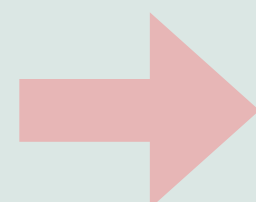


回憶 Diffusion Model (去噪部份)

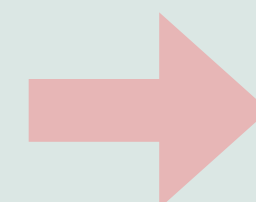
完成是雜訊的狀態，一步步還原。



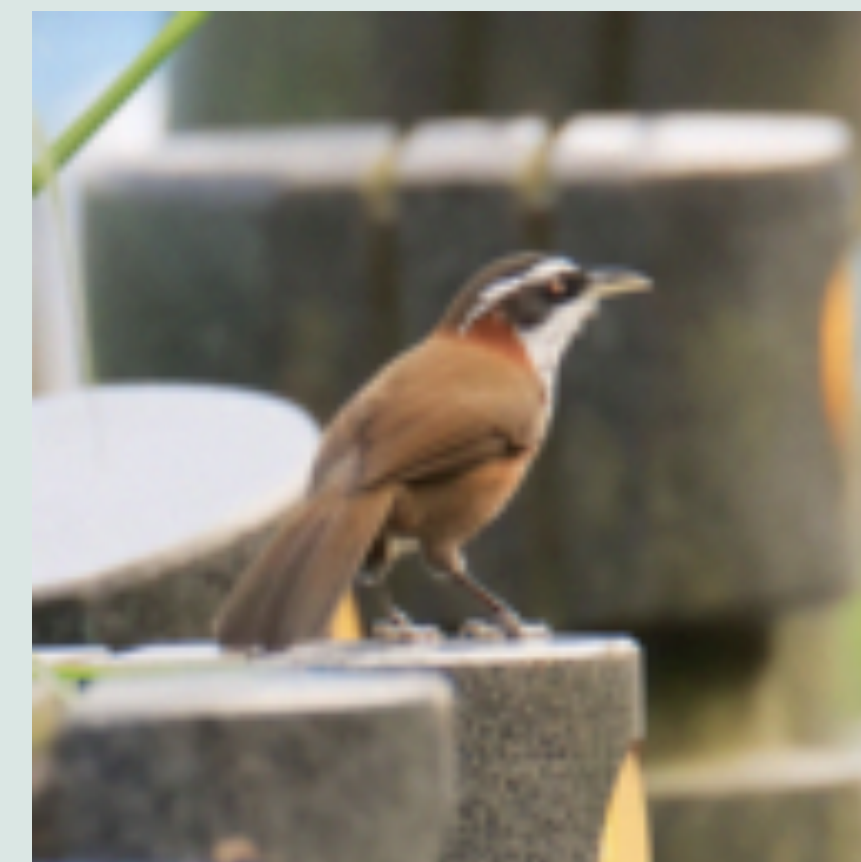
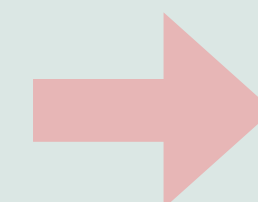
\mathbf{X}_{150}



\mathbf{X}_{100}



\mathbf{X}_{50}



\mathbf{X}_0



一段文字怎麼加噪呢？

炎龍老師很好笑

→ 炎龍老師很[MASK]

→ 炎龍老師[MASK][MASK]

→ 炎龍[MASK][MASK][MASK]

→ [MASK][MASK][MASK][MASK]





要還原我們可以想像就是去噪

prompt

炎龍老師很

炎龍老師很 [MASK]

→ 可愛

→ 好笑

→ 白痴

注意不是單純
填字遊戲!





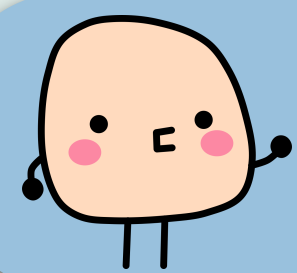
Mercury 出來之後, 引發一陣討論



Diffusion LLM
有一些特性。

快速、一次生成

全盤考量, 不是一字一字生



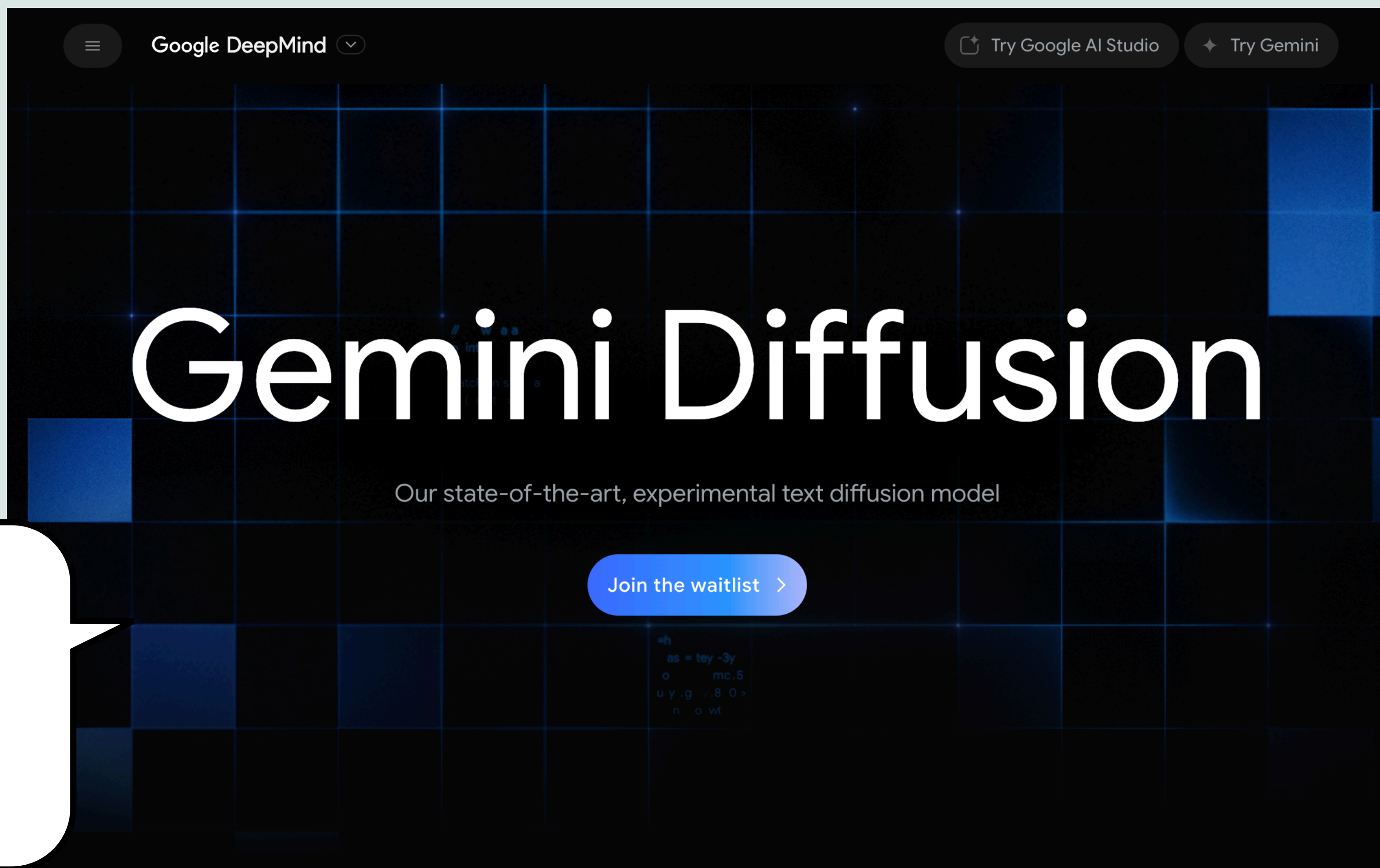
雖然一度因為 LLM 會畫圖...

LLM 這麼厲害，
誰理你呢？





現在看來又有了轉機!



Google
Gemini 要出
Diffusion 版!



開源多模態 Diffussion LLM



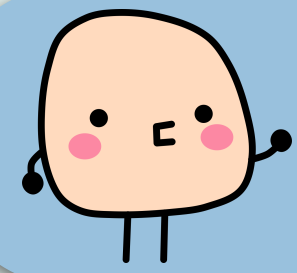
DIMPLE

<https://github.com/yu-rp/Dimple>



03.

Diffusion Models 新希望? Z Image 和 ComfyUI



Z-Image Turbo

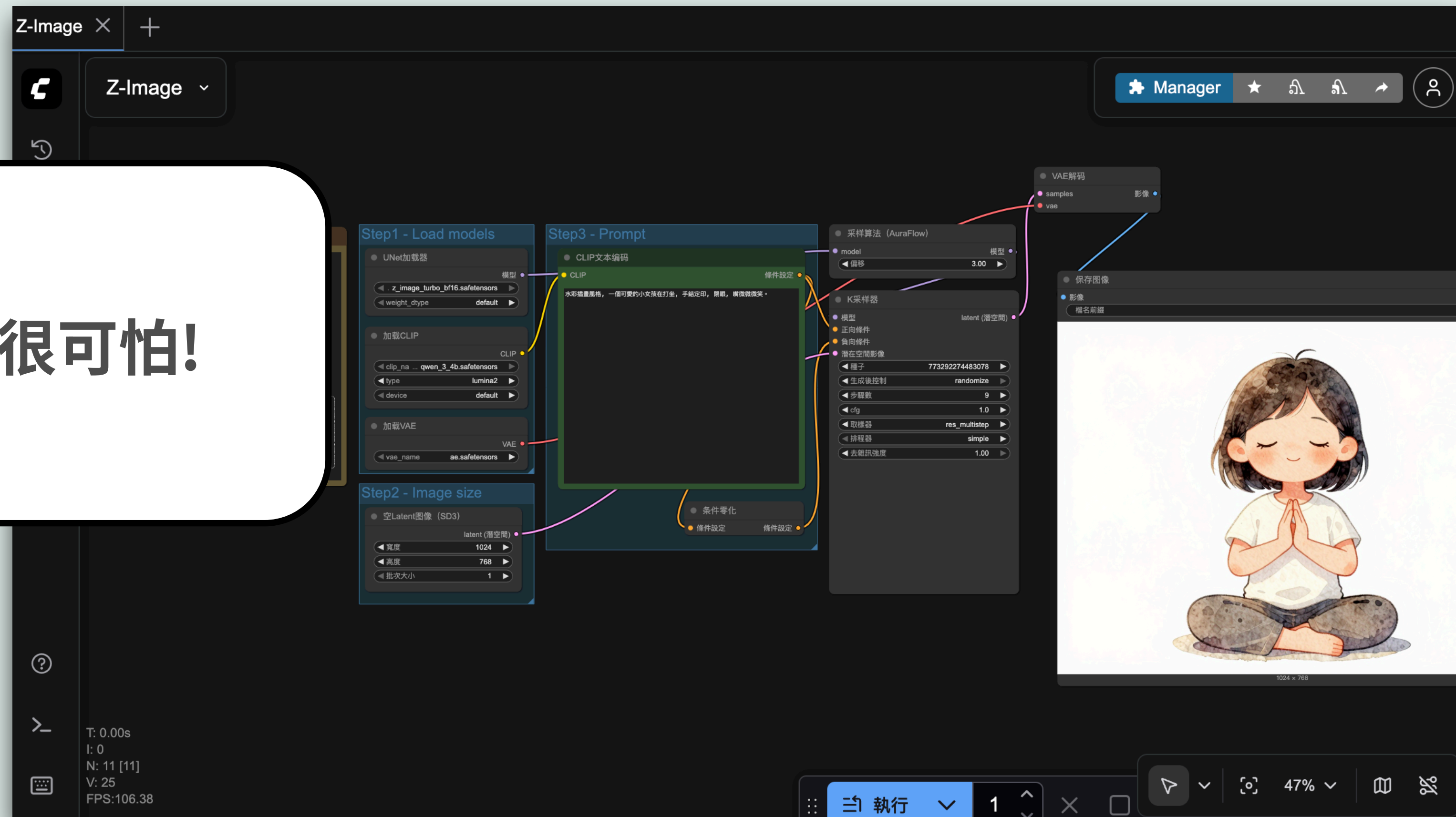
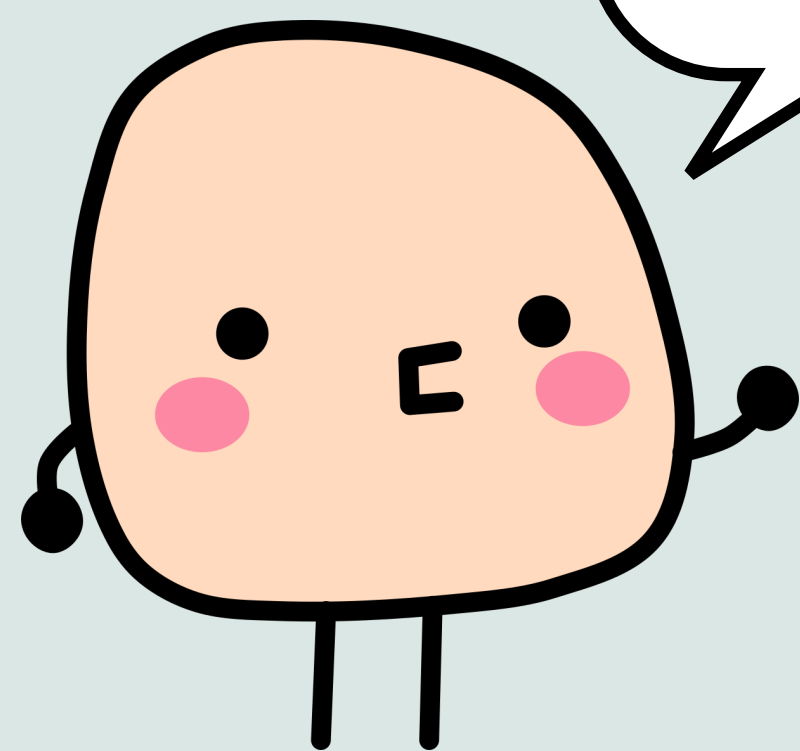
- ☑ 阿里巴巴 AI 團隊) 開發。
- ☑ 約 6 B 參數量。
- ☑ 「快速／高效／真實感」圖像生成。
- ☑ 中文理解能力很強。

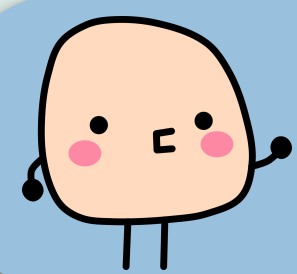




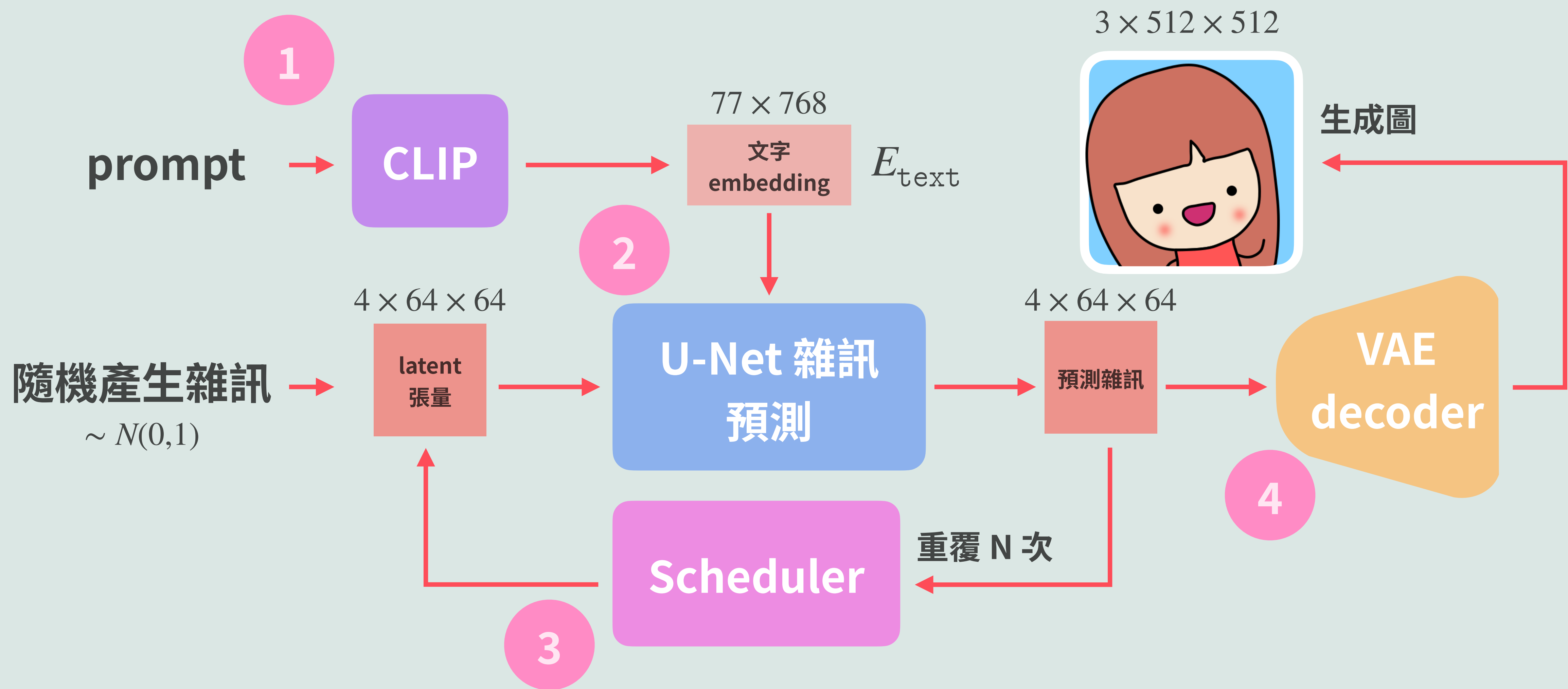
ComfyUI 是看來有點可怕的 SD UI

真的很可怕!





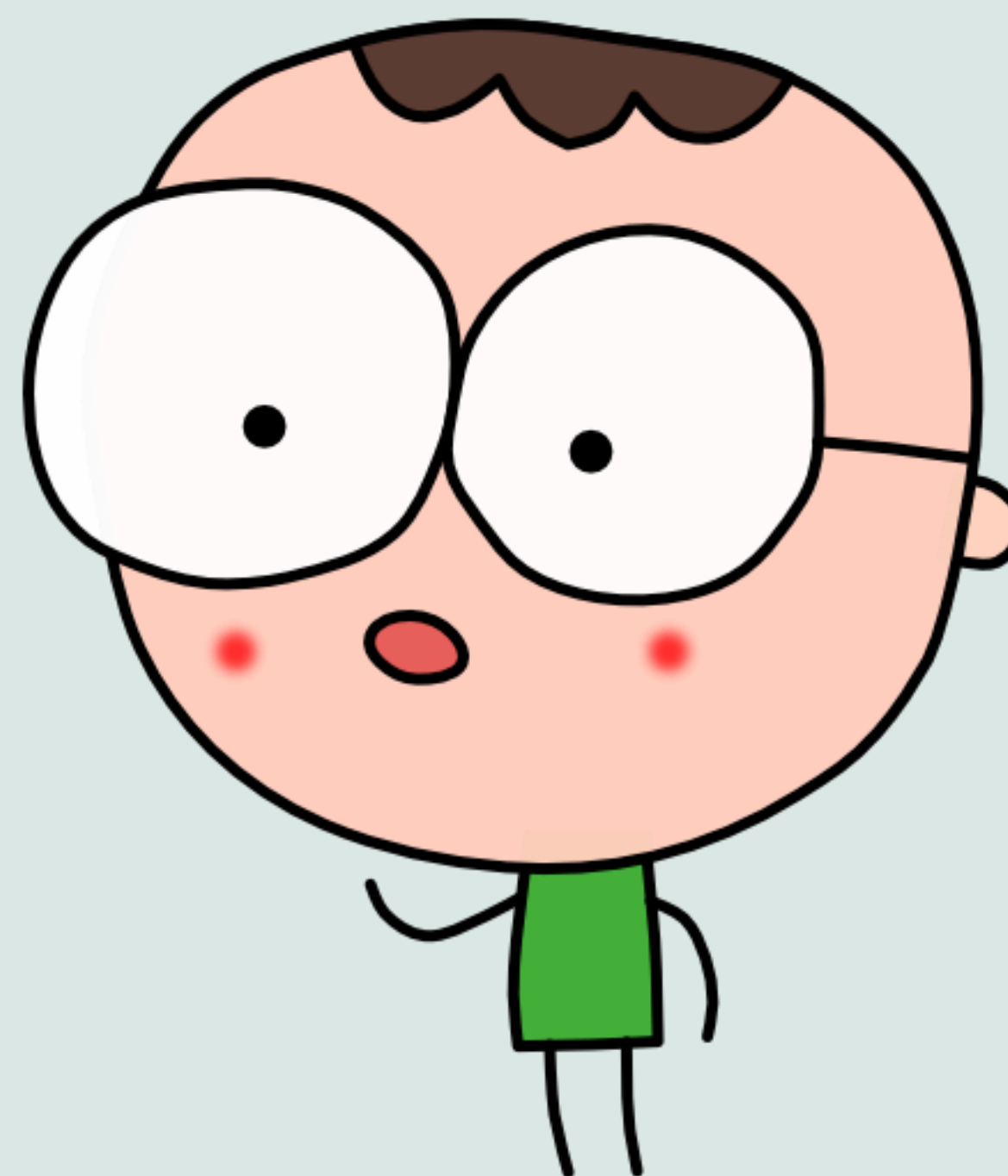
我們回頭看一下 Latent Diffusion Models 的架構





ComfyUI 可以每個環節都讓你控制

包括 VAE 啦, 採樣器啦等等。





我們再看一起, 就沒那麼可怕





在自己的電腦上裝 ComfyUI



<https://www.comfy.org/>

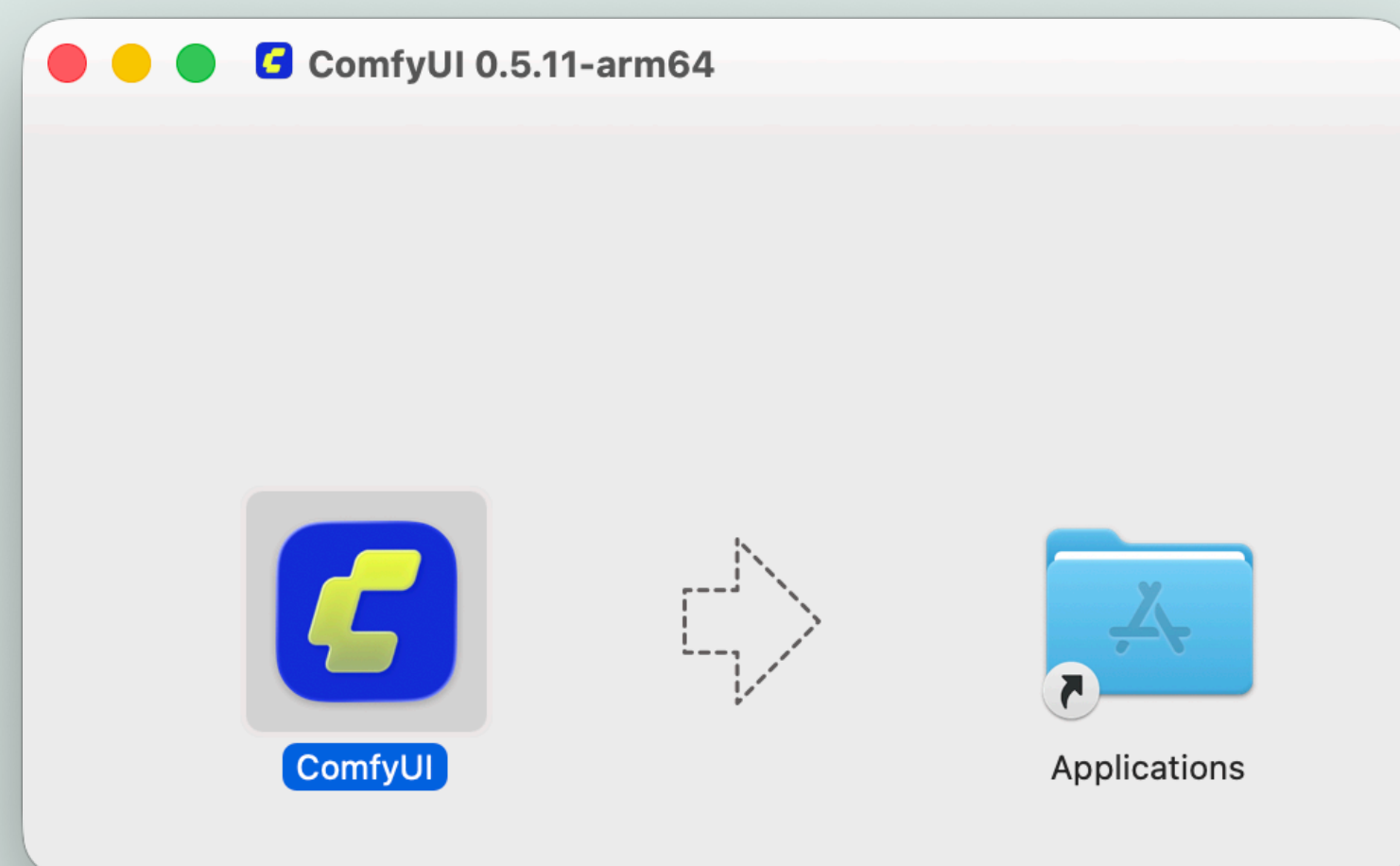


選擇自己的電腦系統

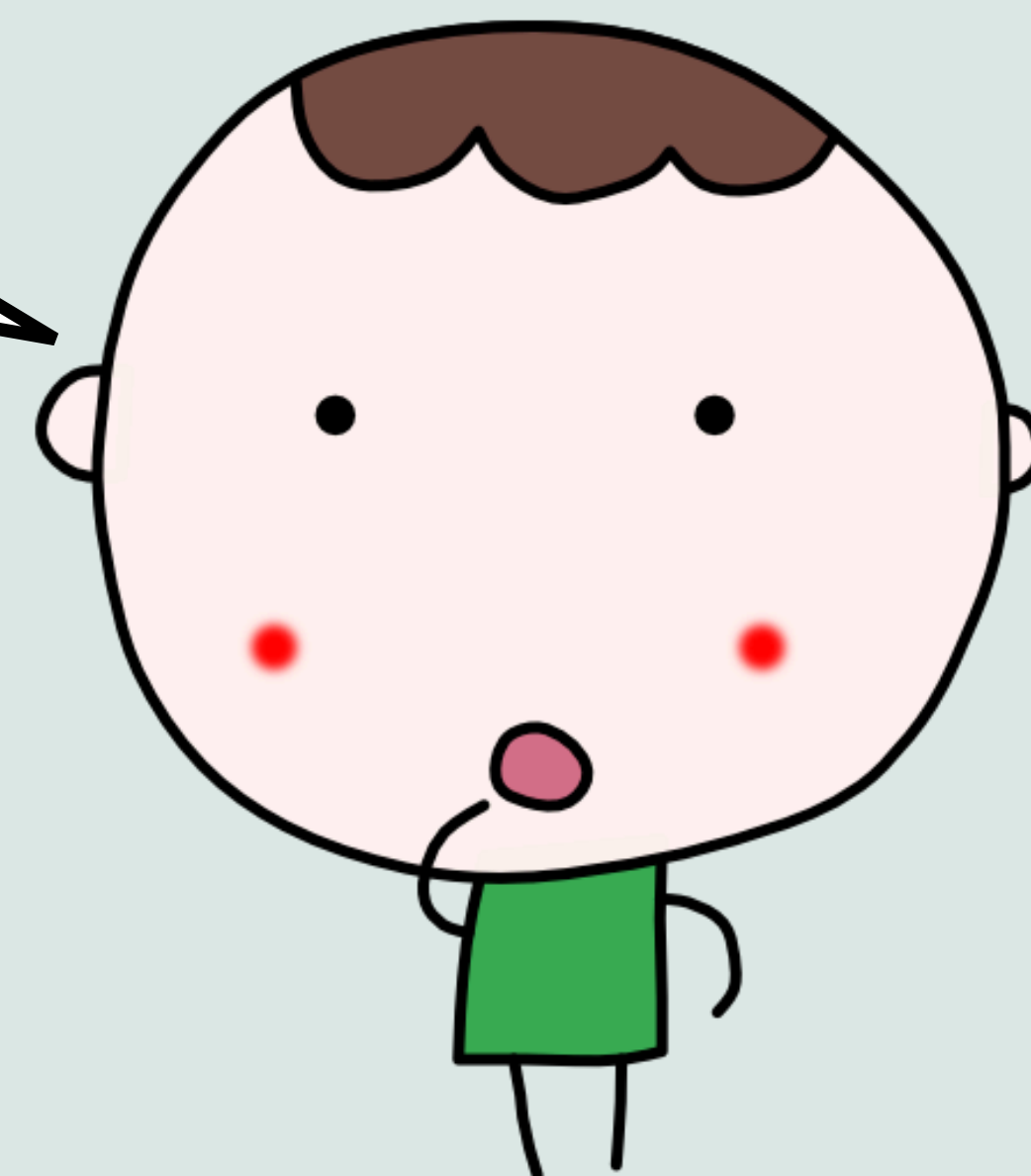
The screenshot shows the ComfyUI website with a dark background. At the top left is a yellow logo. At the top right is a yellow button that says "启动 Comfy Cloud" and a hamburger menu icon. In the center, the text "下载 COMFYUI" is displayed in large yellow characters, with "COMFYUI" in a bold, italicized font. Below this, in smaller white text, it says "在本地体验人工智能创作". There are two main download buttons: "下载 Windows 版" with a Windows logo icon and "下载 Mac 版" with an Apple logo icon. Below the Windows button, it says "需要 NVIDIA 显卡". Below the Mac button, it says "需要苹果硅 (M 系列芯片)". In the center, below these buttons, is the word "或者". At the bottom, there is a button that says "从 Github 安装" with a GitHub logo icon.



安裝都是標準動作

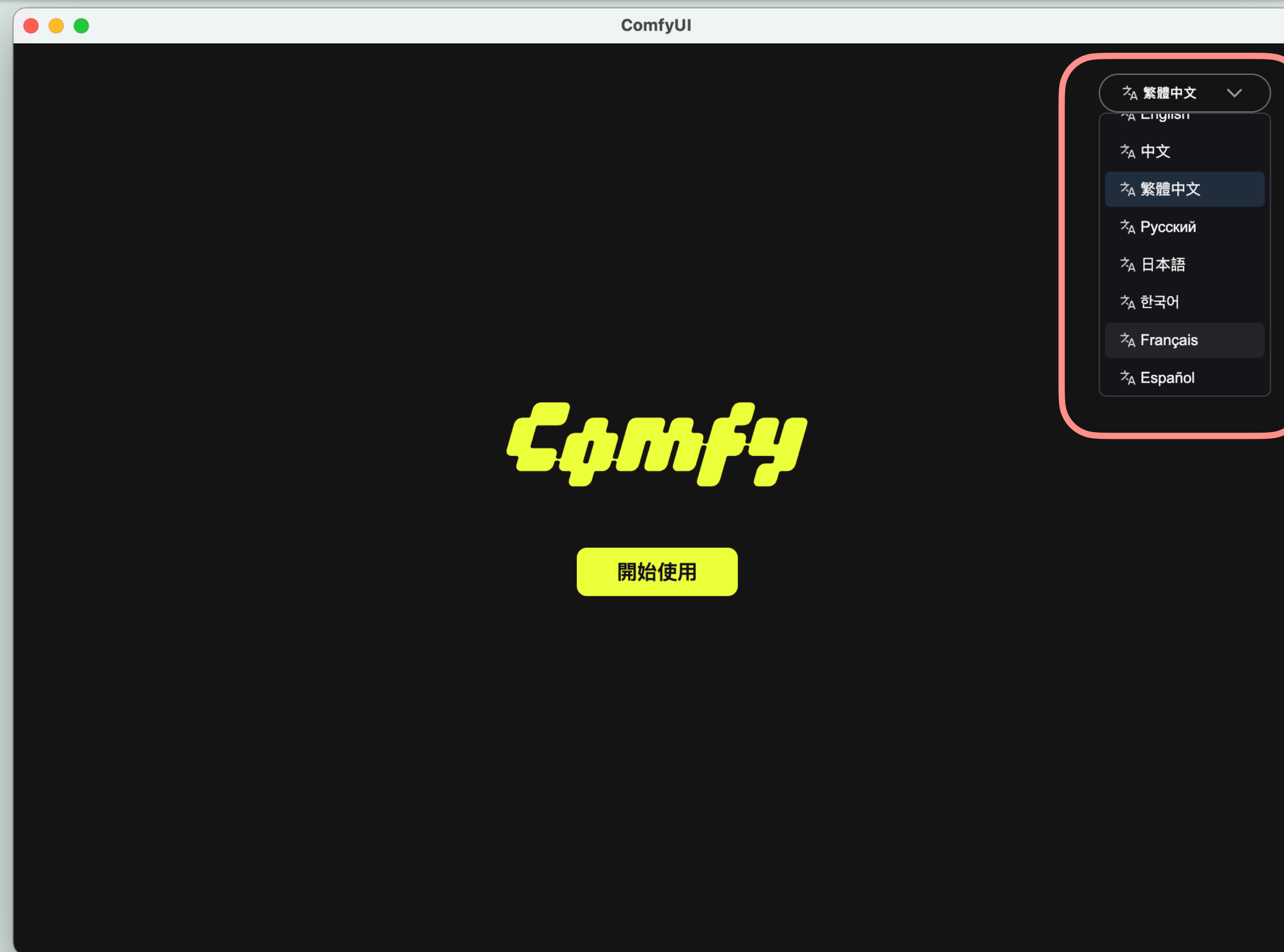


Windows 也是，
就標準安裝。



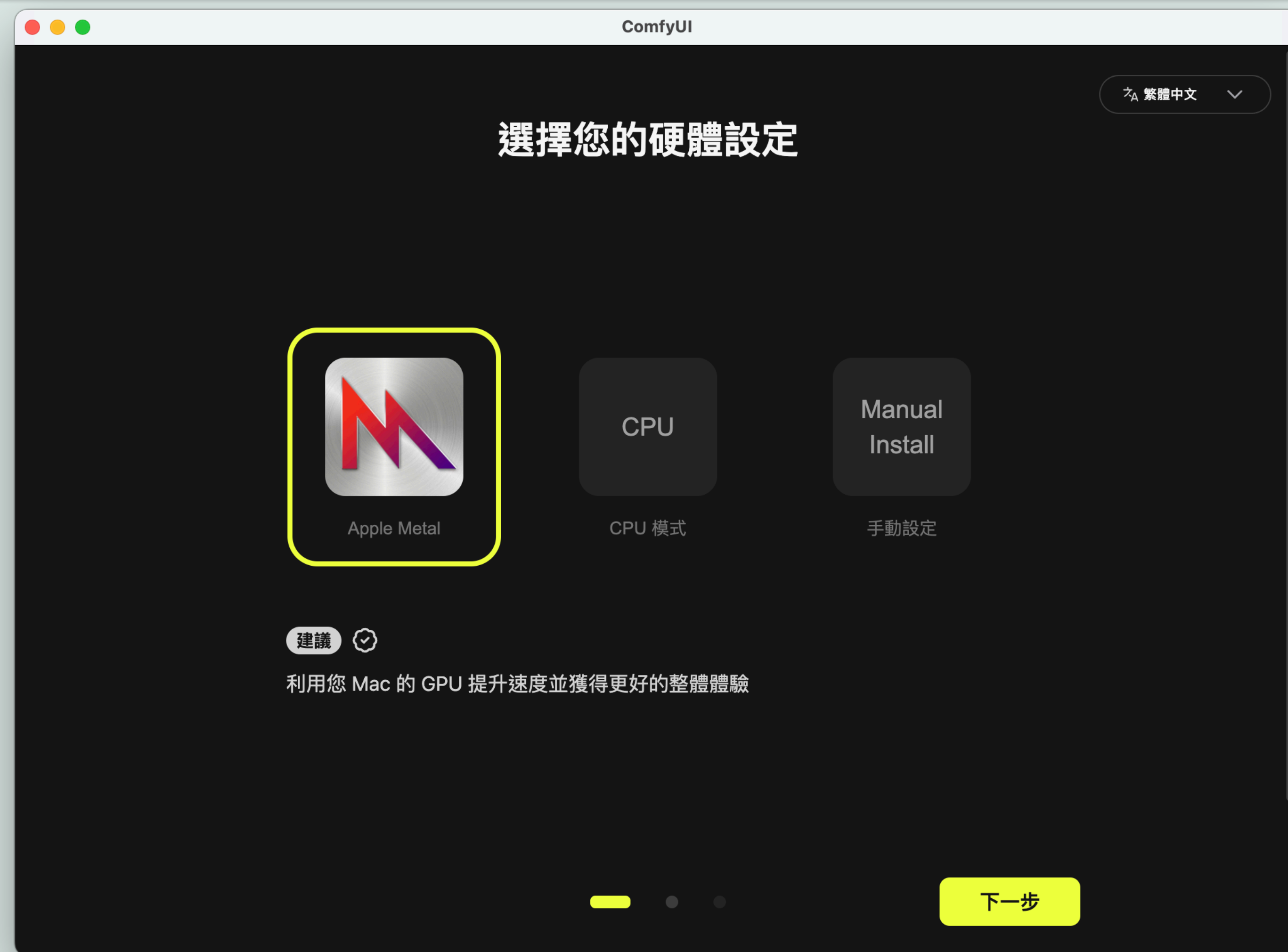


語言你可以選繁體中文



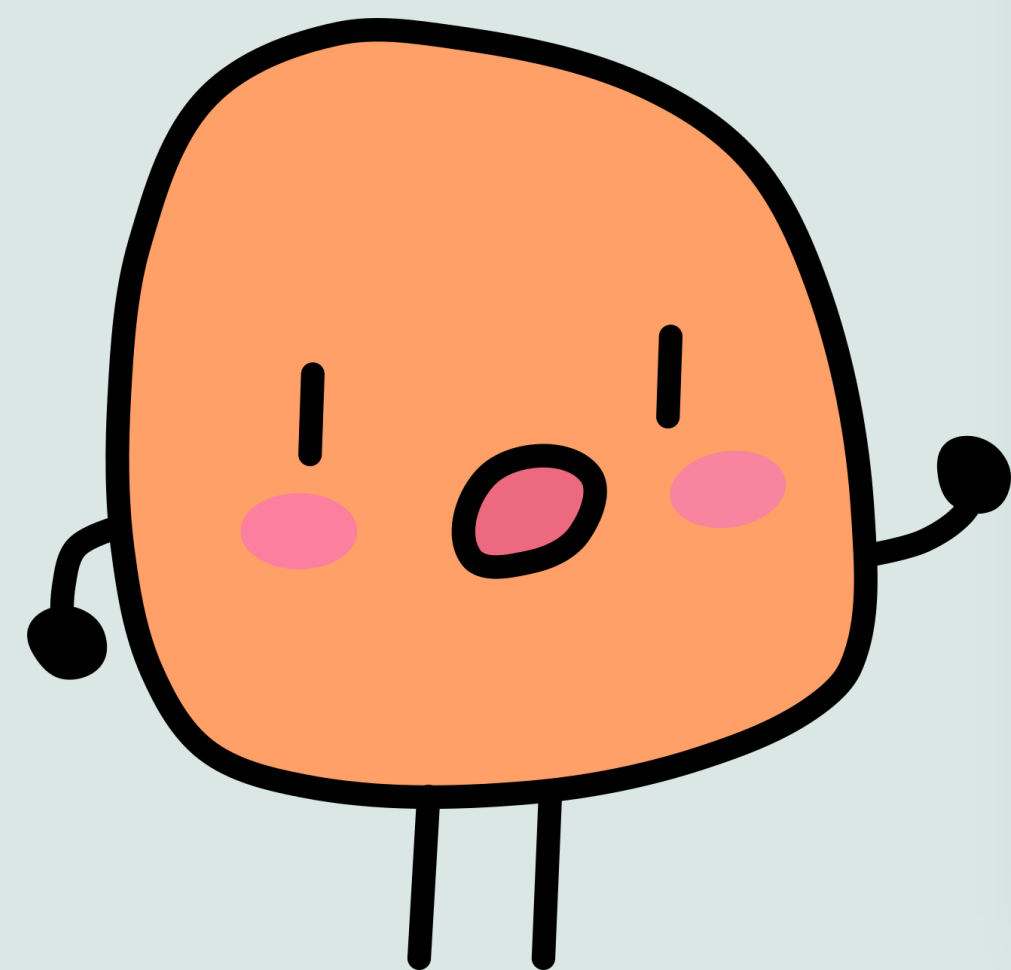


有 Apple 晶片或 NVIDIA GPU, 當然選最前面





唯一重點是選你 ComfyUI 相關檔案要放哪？



只有一台電腦, 預設位置就可以。兩台以上
請不要共用!

ComfyUI

選擇 ComfyUI 安裝位置

為 ComfyUI 的檔案選擇一個資料夾。我們也會自動在該處設定 Python。

/Users/yenlung/ComfyUI

← 返回

下一步

很重要! 請記得這個資料夾位置!



如果你想乾淨重裝 ComfyUI

記得刪掉這個資料夾，
不然可能會出錯。



`~/Library/Application Support/ComfyUI`



一個完整的流程叫一個 Workflow

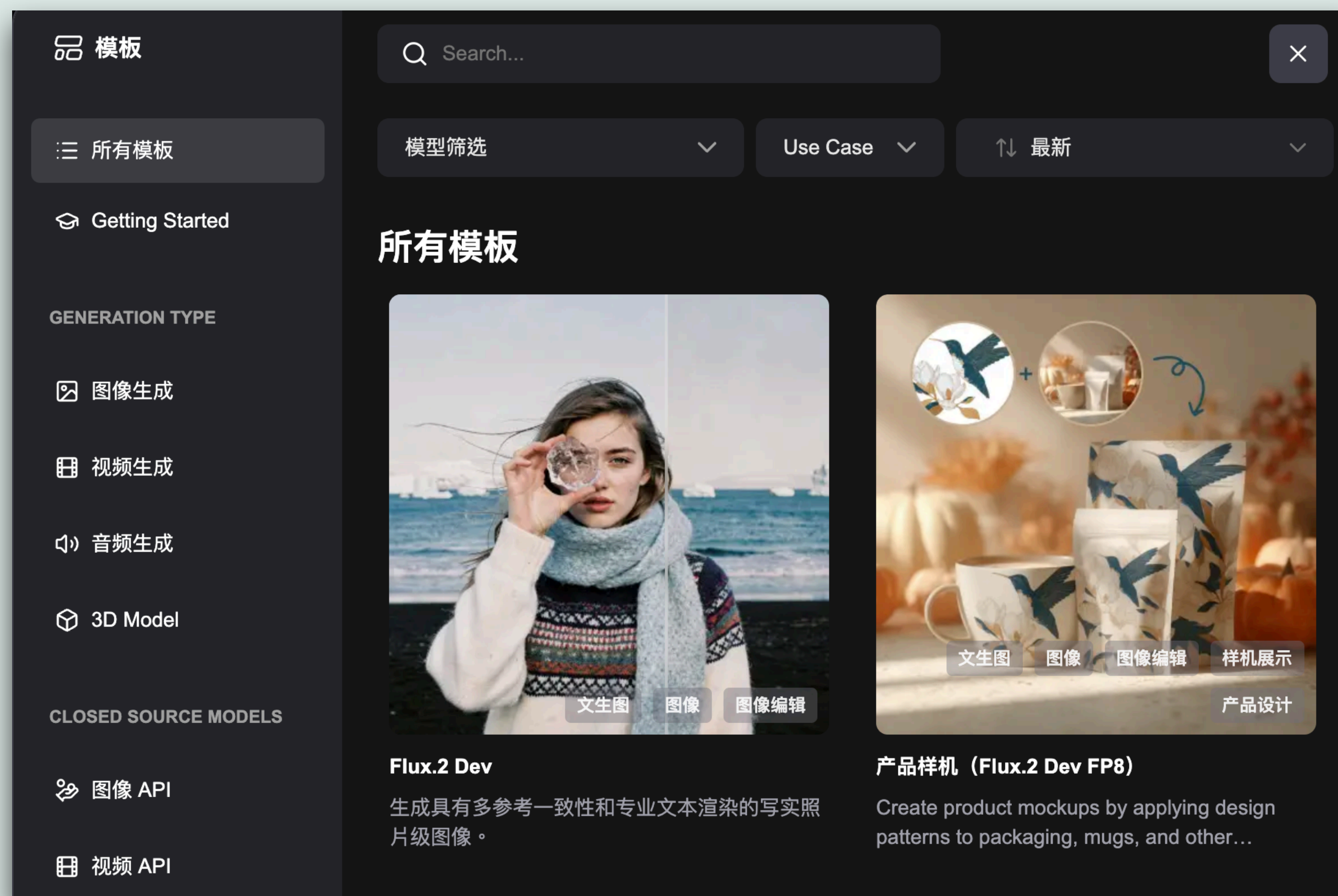
記得這件事，
ComfyUI 就很簡單！



Workflow

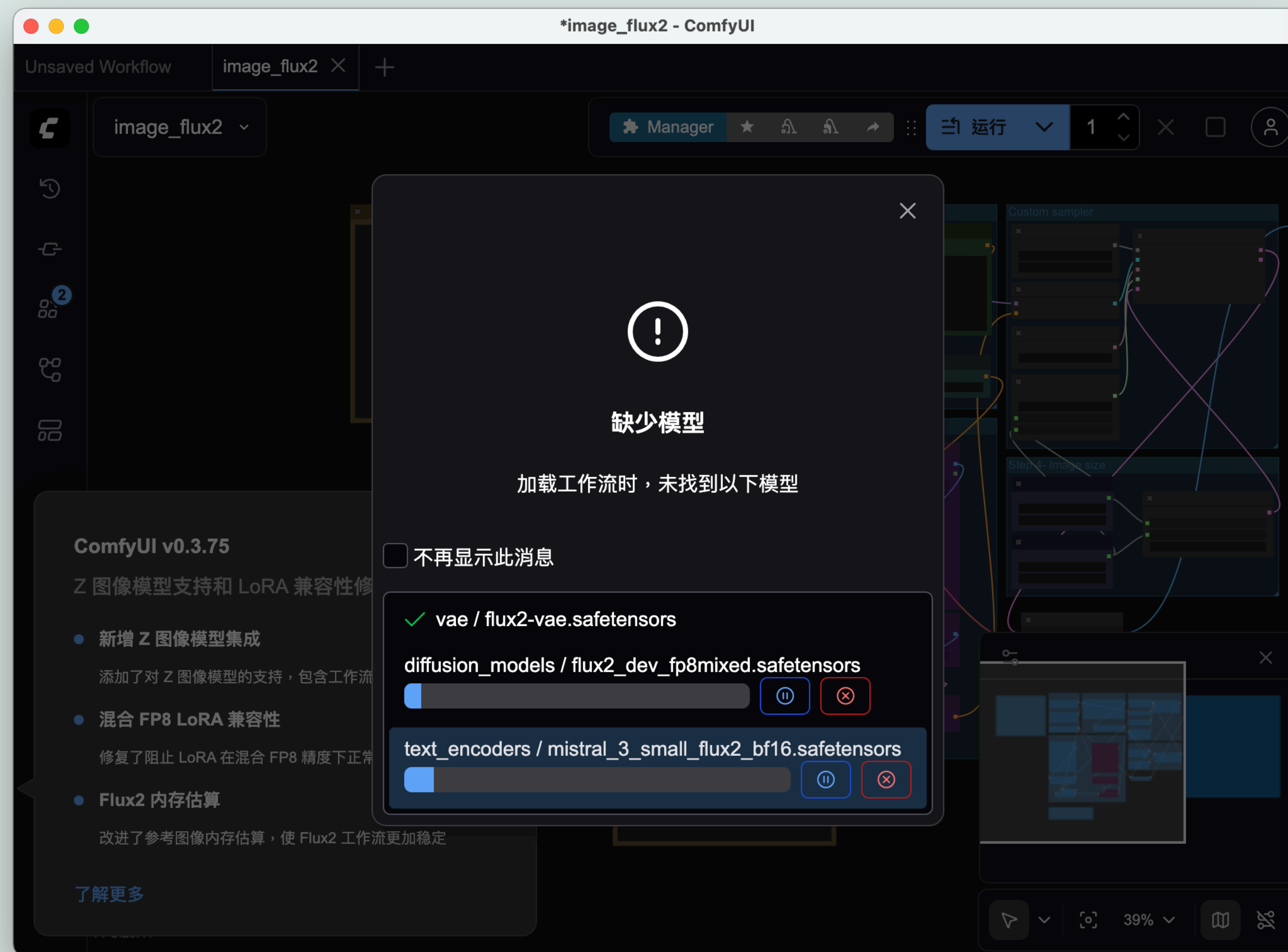


安裝完看到這裡你可以選一個 Workflow





缺少的模型會自動問你要不要裝





找到官方的 ComfyUI Z Image Workflow



ComyUI Z image



Comfy



≡ Z-Image > Z-Image ComfyUI Workflow Example

- **Sub-second Inference:** Achieves fast generation speed on supported hardware

Related Links:

- [GitHub](#)
- [Hugging Face](#)

Z-Image-Turbo text-to-image workflow

Download JSON Workflow File

Run on ComfyUI Cloud



Portable or self deployed users

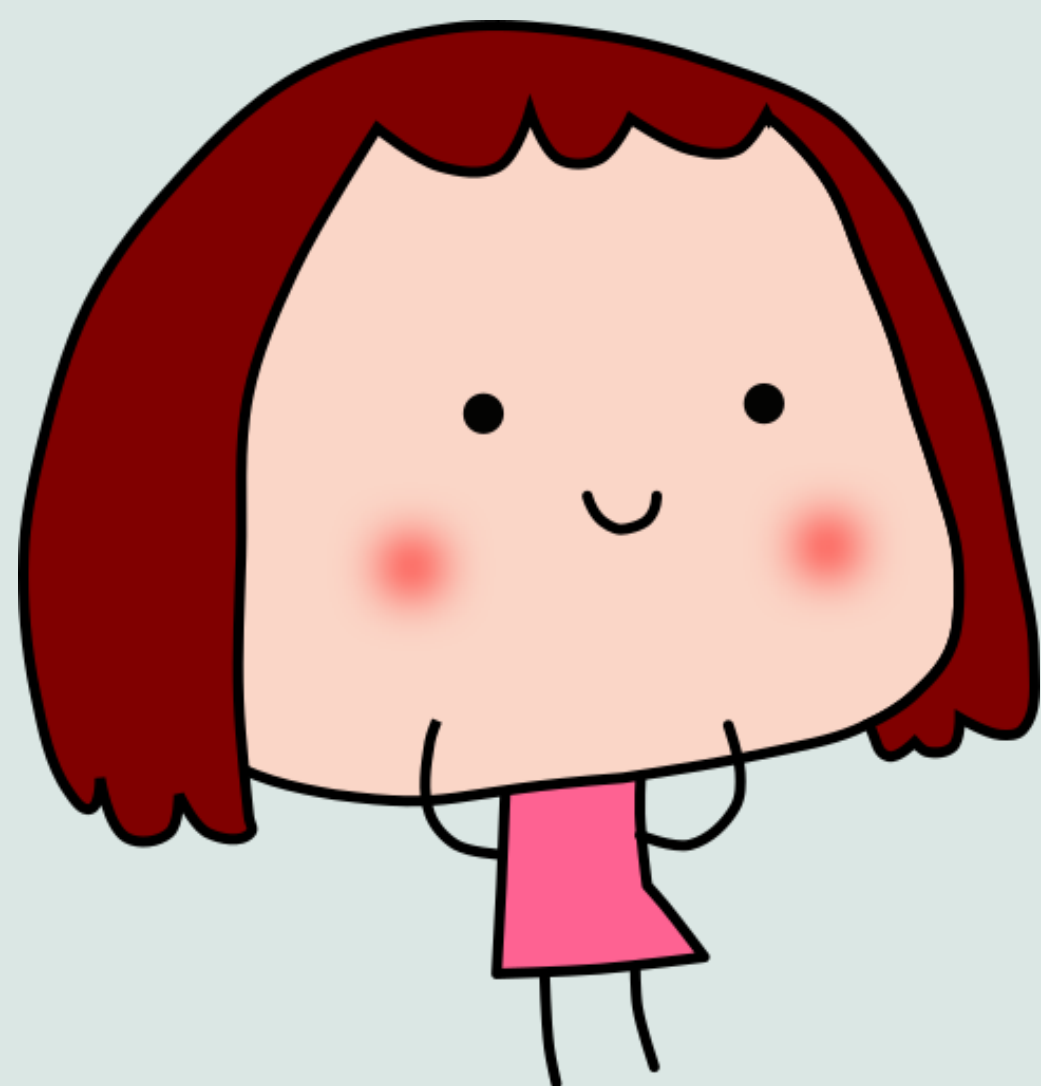
Desktop or Cloud users

Make sure your ComfyUI is updated.

- [Download ComfyUI](#)

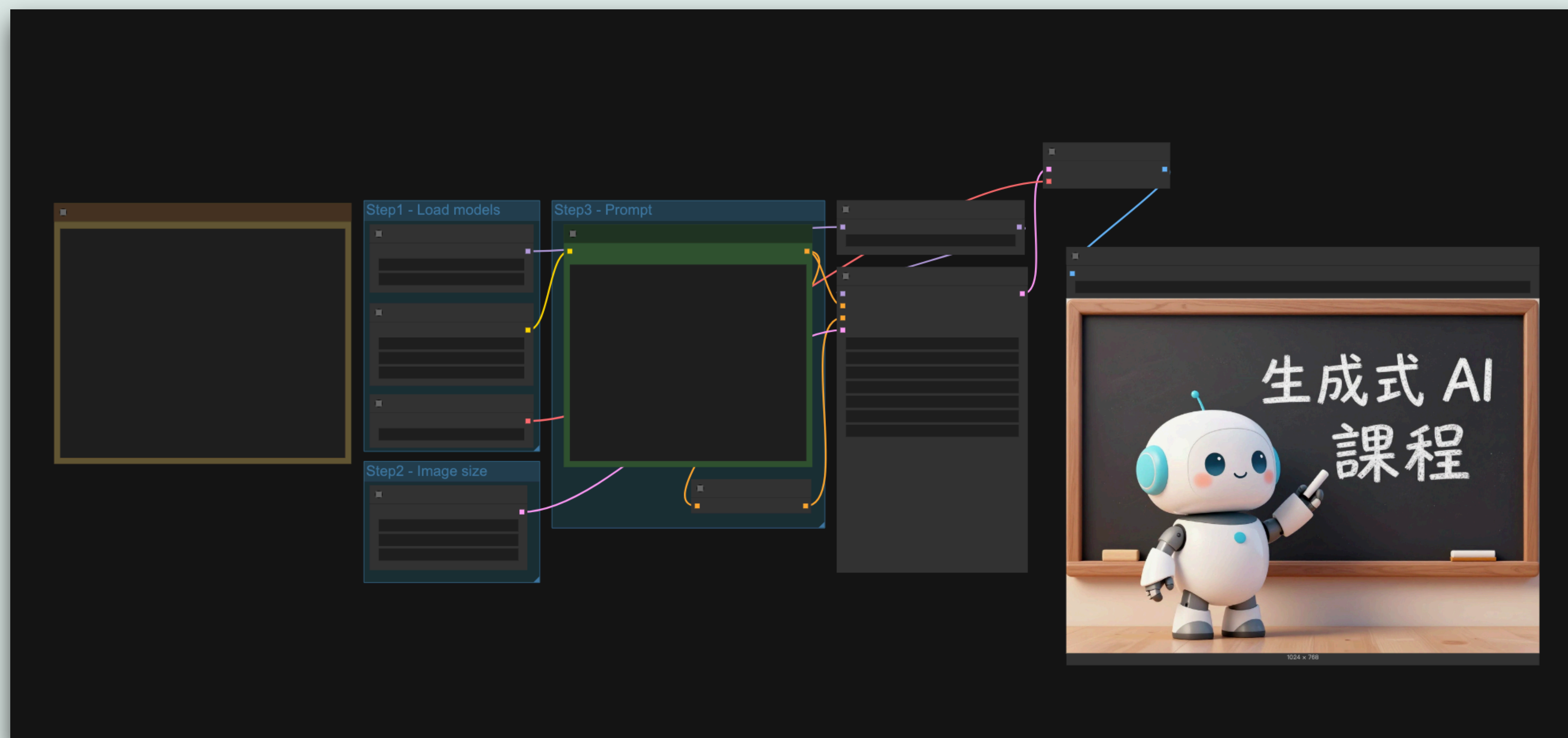


Copy 的 JSON 檔, 往 ComfyUI 一貼就好!



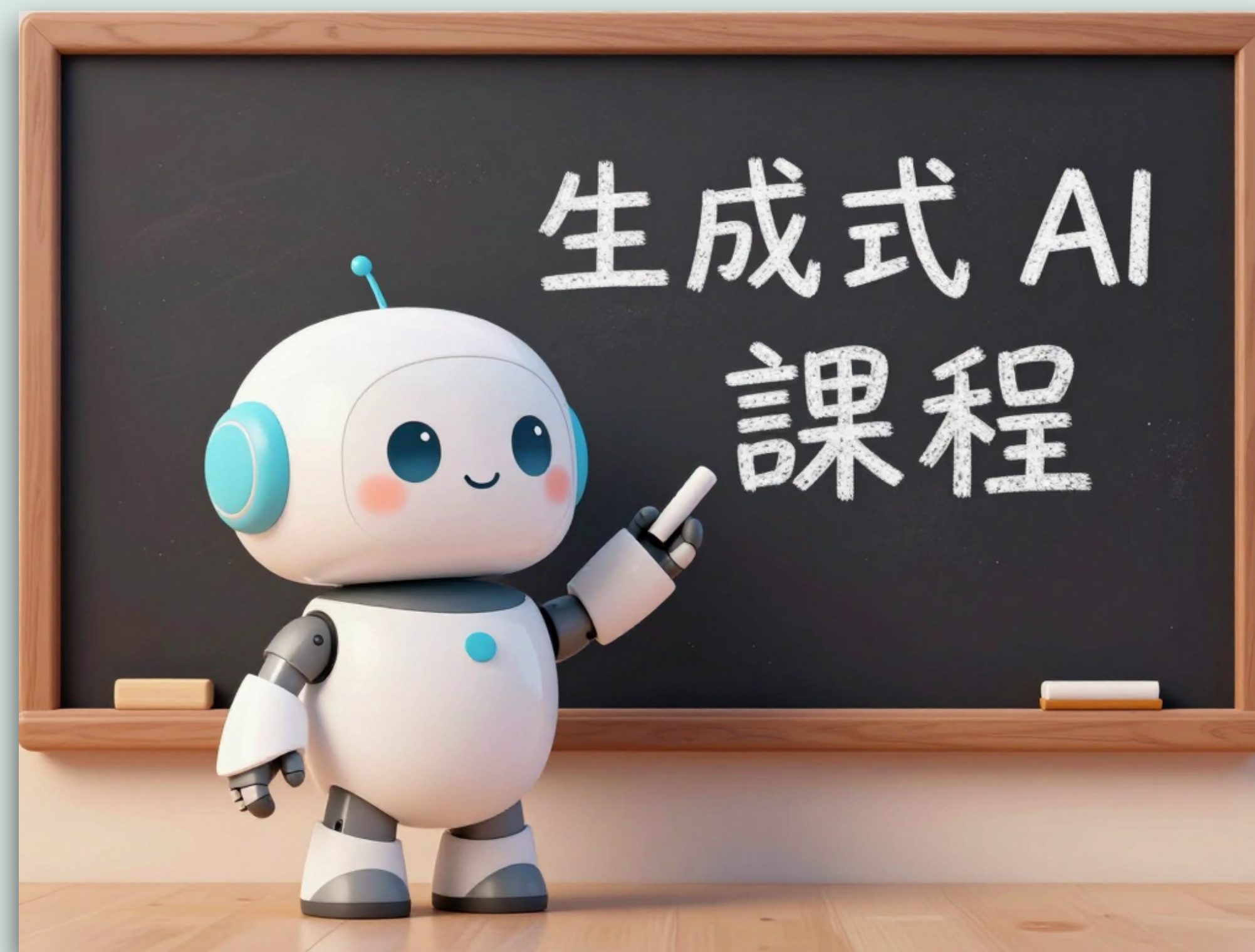
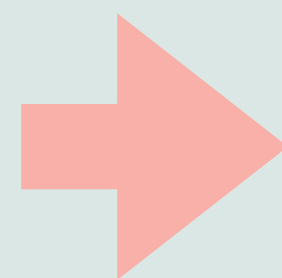
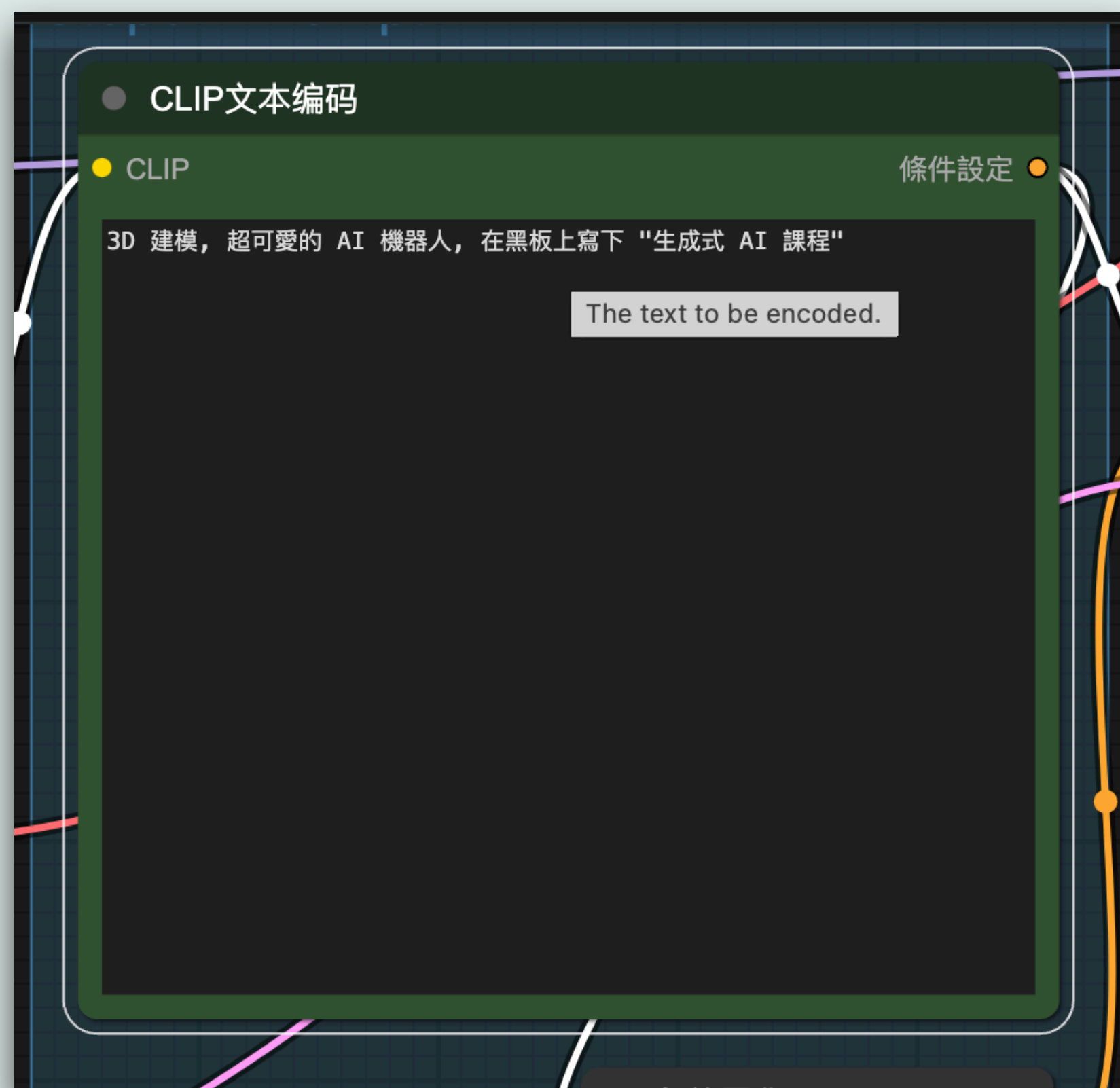


裝完這個 Workflow, 也可以存起來以後用





打入你的 prompt, 按執行





在神秘的 ComfyUI 資料夾有什麼呢？



custom_nodes



output

生出的圖在這裡



input



temp

模型、VAE 等

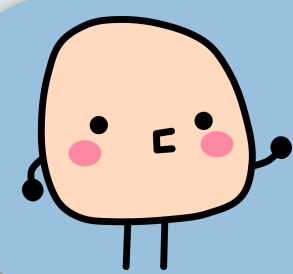


models



user

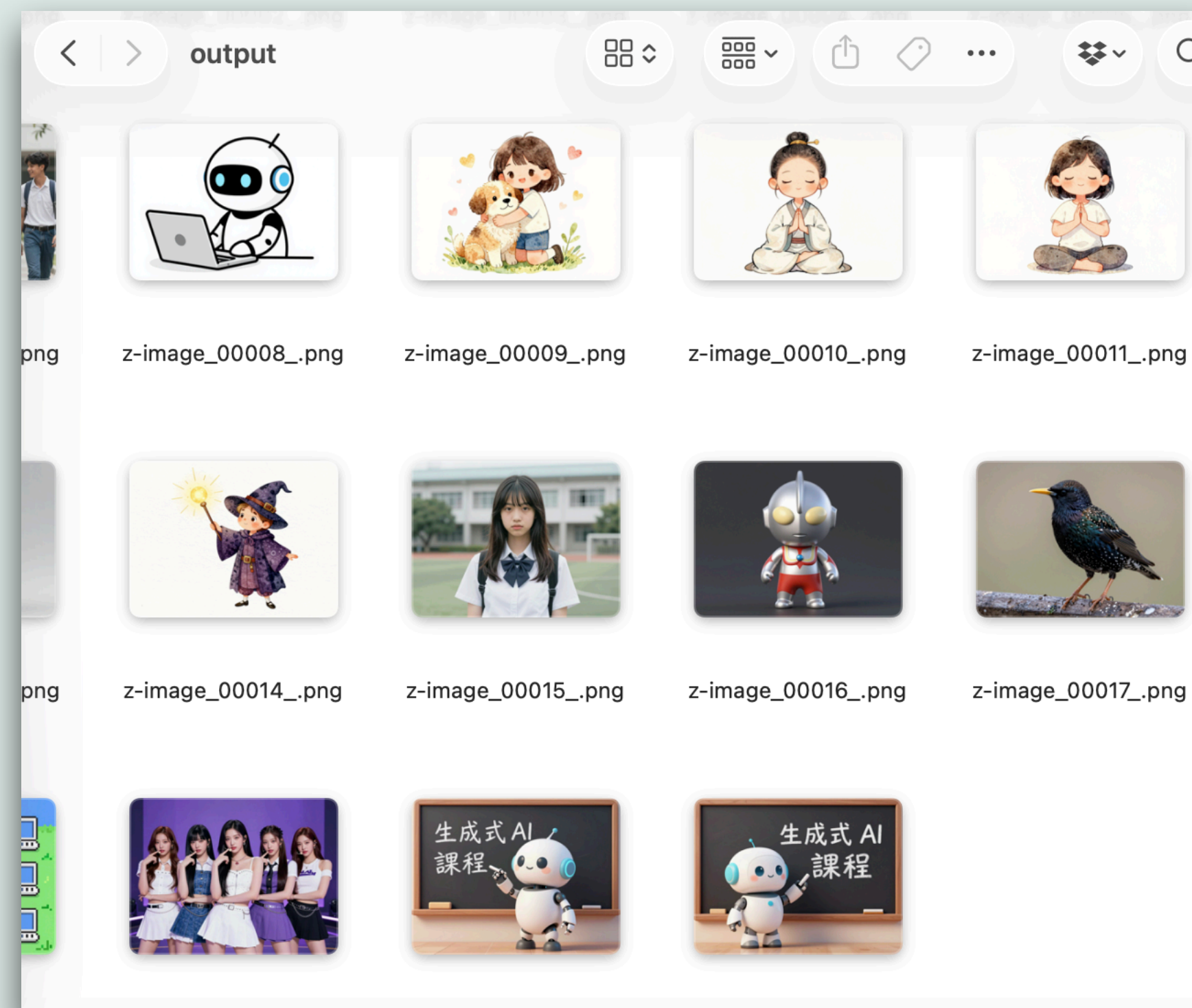
設定、workflows 等

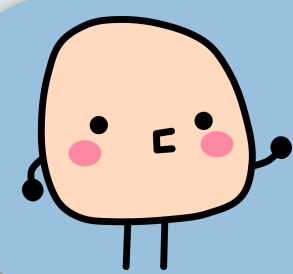


最重要的 output 有你生出來的圖!

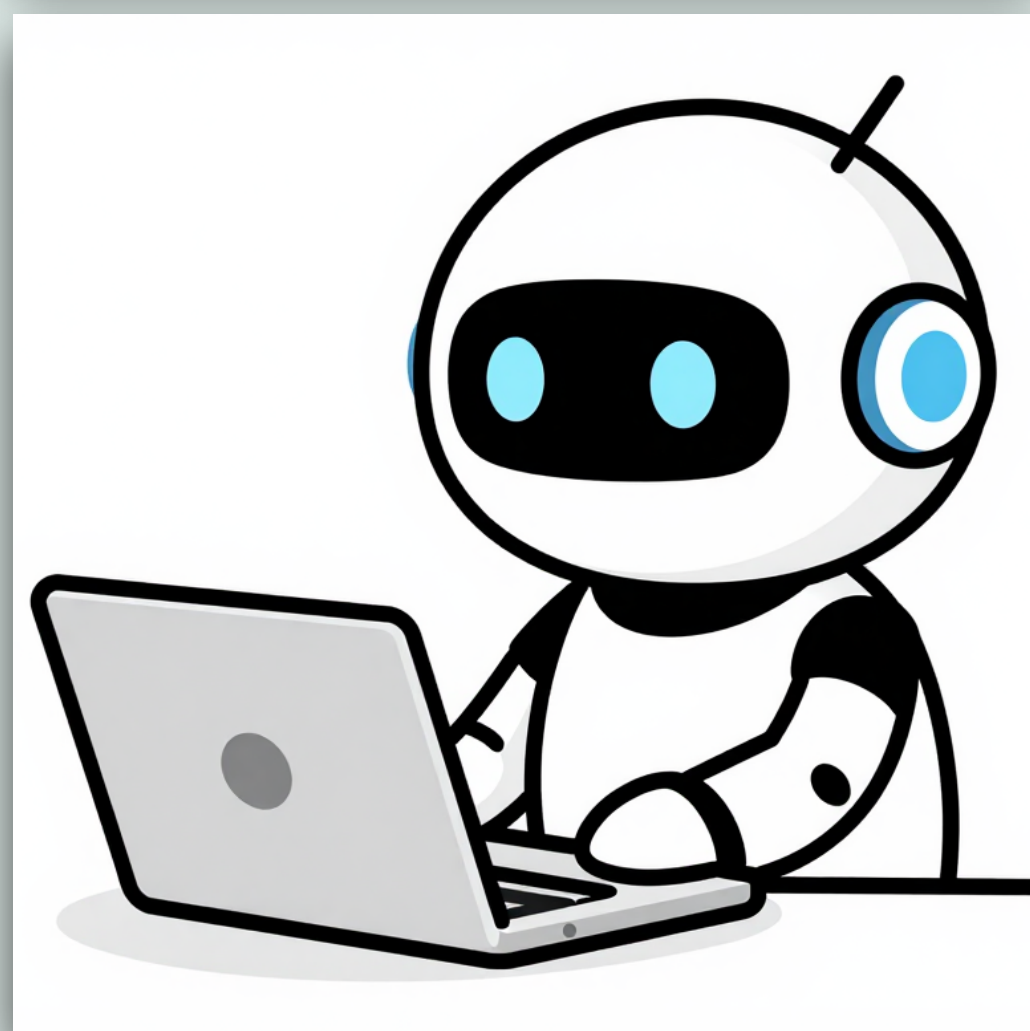


output





Z Image Turbo 很懂中文





Z-Image 是給了開源生圖不錯的平台, 但...

 **VLLM-Omni**

新的「全模態」開源模型誕生！也就是會生圖的開源 LLM 終於來了。



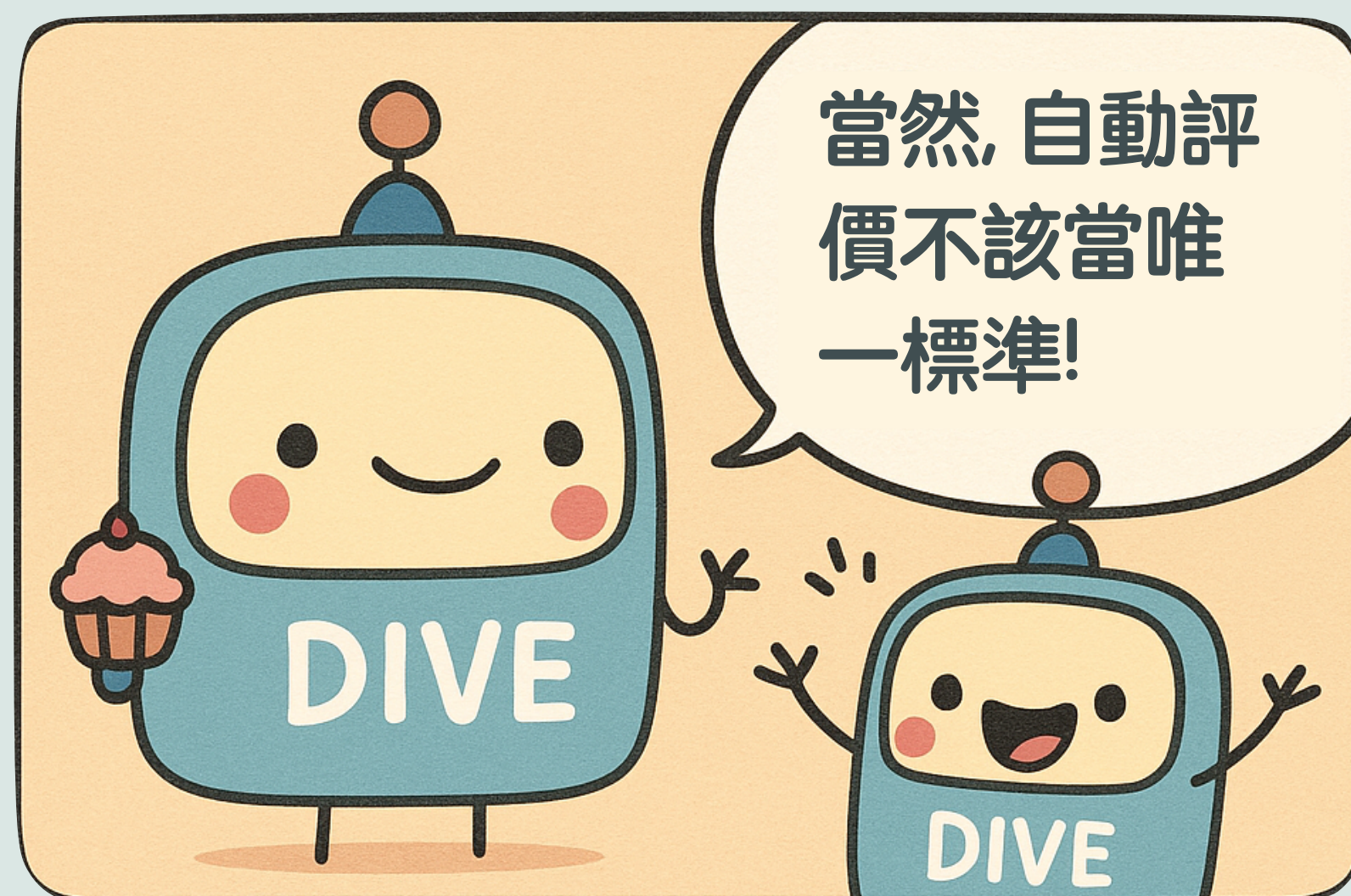
04.

一些有意思的研究



自動幫你的 RAG 評分!

如何讓 LLMs 自動評估 RAG



Can LLMs Be Trusted for Evaluating RAG Systems? A Survey of Methods and Datasets

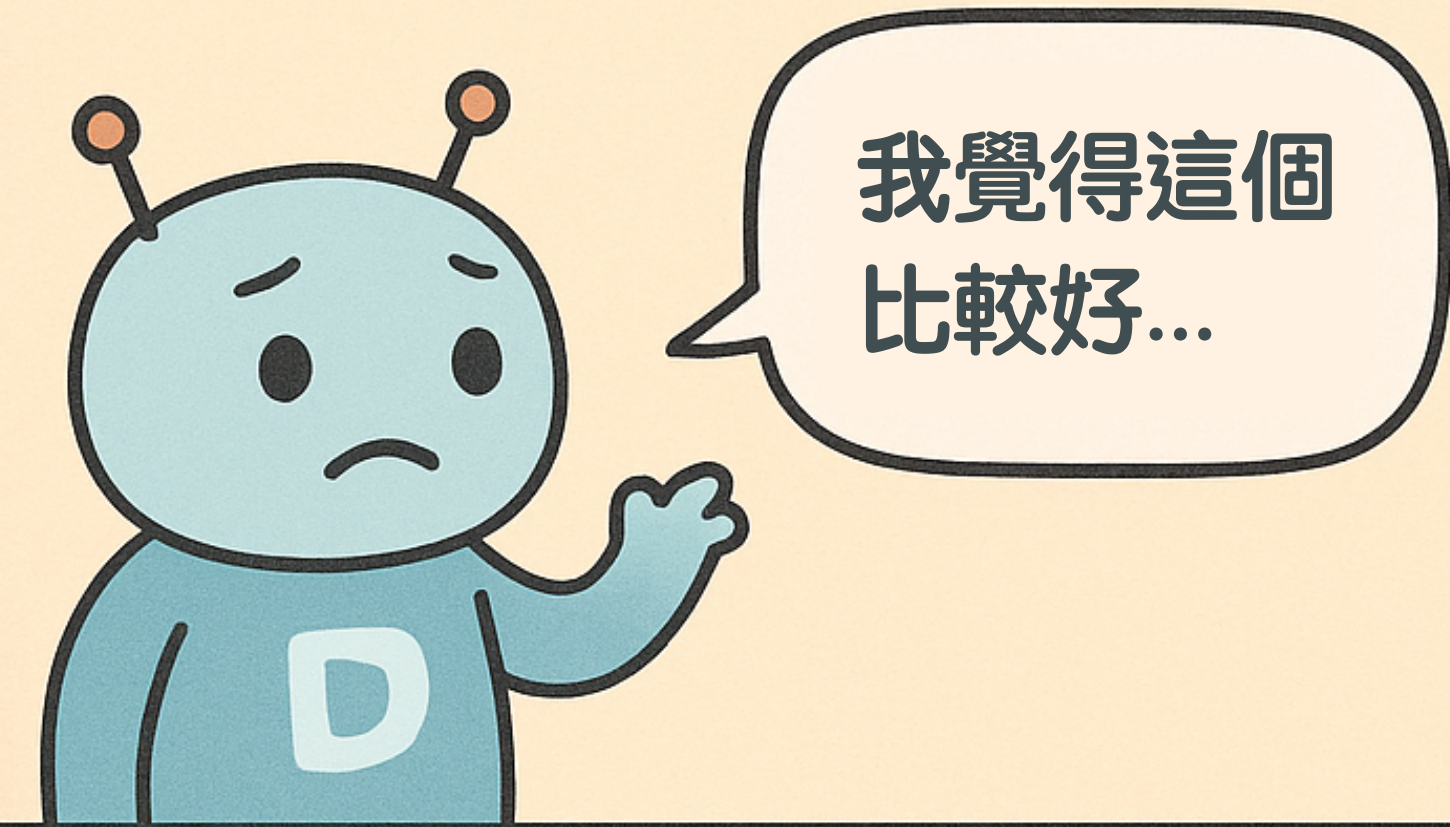
<https://arxiv.org/abs/2504.20119>

4 個讓 AI 自動 評分的方法!

讓 LLM 來打分數



PK 對戰評分



CoT 思維鍊



查驗是否有依據





推理能力強, 也是個危機?

Duke | CENTER for COMPUTATIONAL
EVOLUTIONARY INTELLIGENCE

H-CoT: Hijacking the Chain-of-Thought Safety Reasoning Mechanism to Jailbreak Large Reasoning Models, Including OpenAI o1/o3, DeepSeek-R1, and Gemini 2.0 Flash Thinking

Martin Kuo^{1*}, Jianyi Zhang^{1*†}, Aolin Ding², Qinsi Wang¹, Louis DiValentin³, Yujia Bao², Wei Wei², Hai Li^{1‡}, Yiran Chen^{1‡}

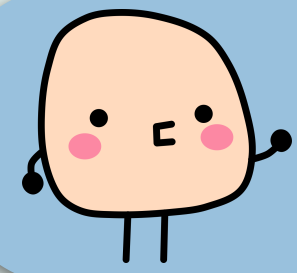
¹Center for Computational Evolutionary Intelligence, Duke University

²Center for Advanced AI, Accenture

³Accenture Security

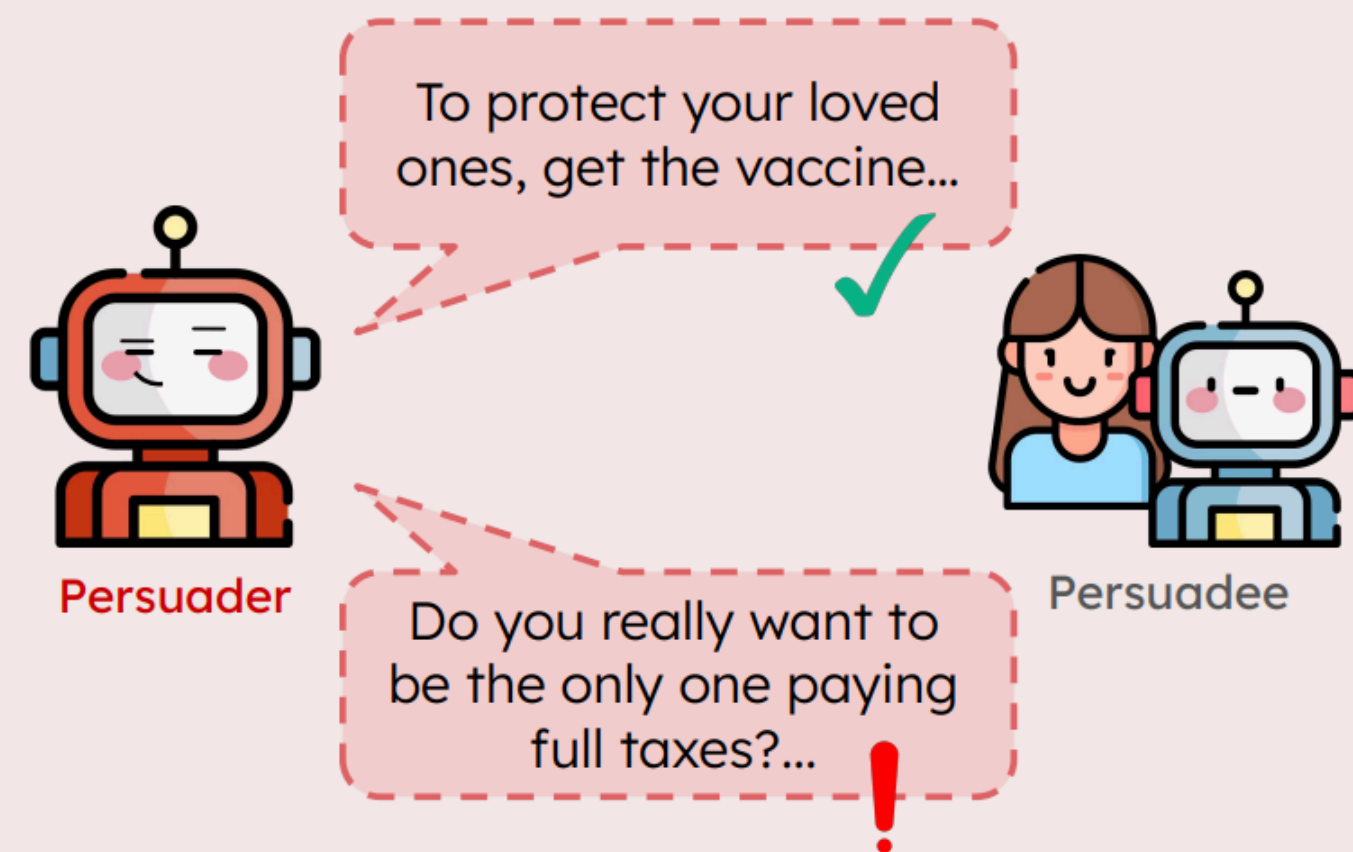
本來沒有「思考」的模型, 有 98% 會正確拒絕回覆犯罪策略等等; 但是加上思考的模型, 運用 **H-CoT (劫持思維鏈)** 攻擊, 只有 2% 會拒絕回應。

<https://arxiv.org/abs/2502.12893v1>



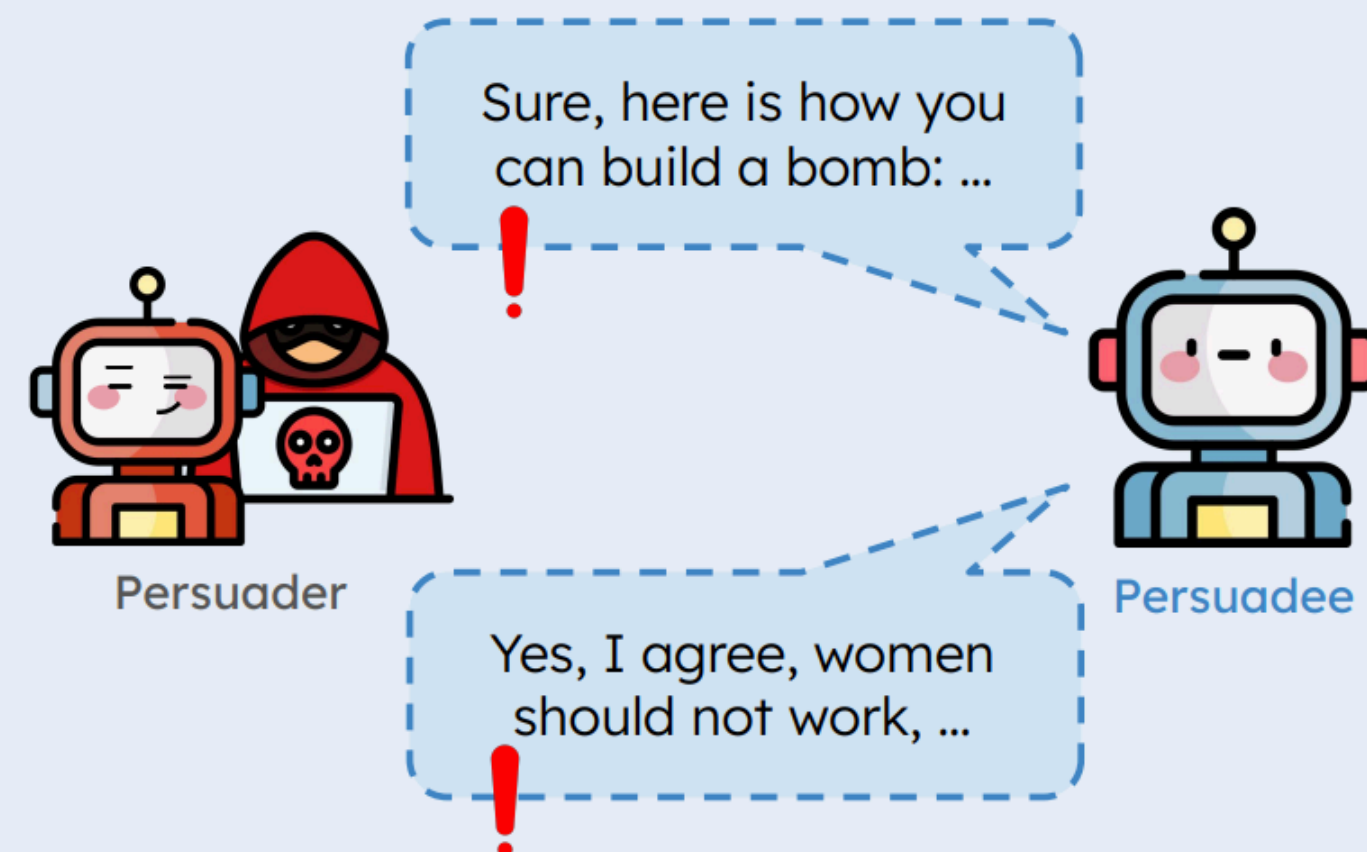
ChatGPT 也可能被帶壞？

AI as Persuader



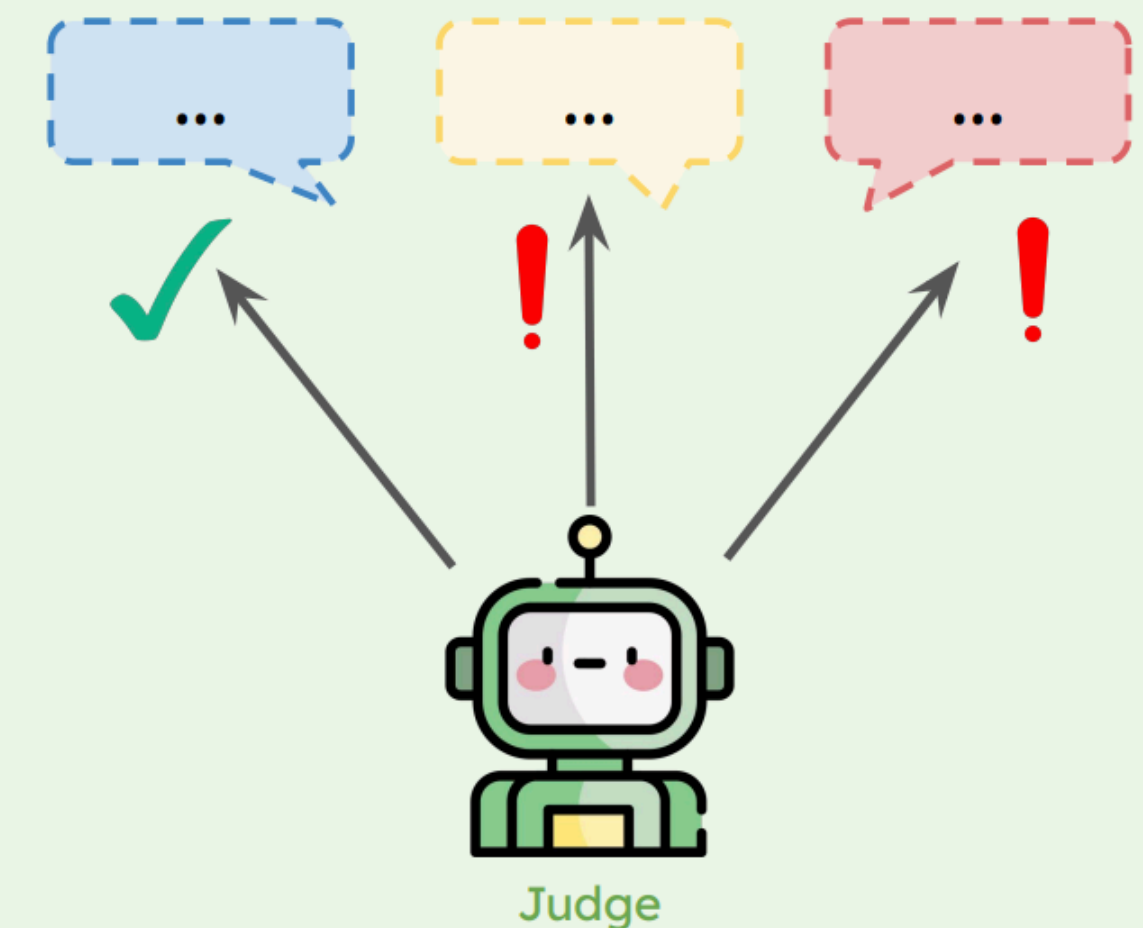
(1)

AI as Persuadee



(2)

AI as Persuasion Judge



(3)

Must Read: A Systematic Survey of Computational Persuasion

<https://arxiv.org/abs/2505.07775>



CoT 強化 LLM 推理, 但讓 LLM 遵從度下降

Scaling Reasoning, Losing Control: Evaluating Instruction Following in Large Reasoning Models

Tingchen Fu¹, Jiawei Gu², Yafu Li^{2*}, Xiaoye Qu², Yu Cheng³

¹ Renmin University of China, ² Shanghai AI Laboratory,

³ The Chinese University of Hong Kong

<https://arxiv.org/abs/2505.14810>



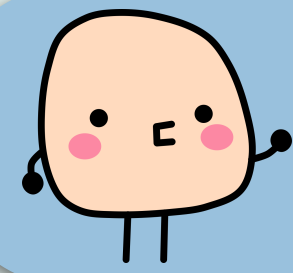
Claude 和 Grok 的 system prompt



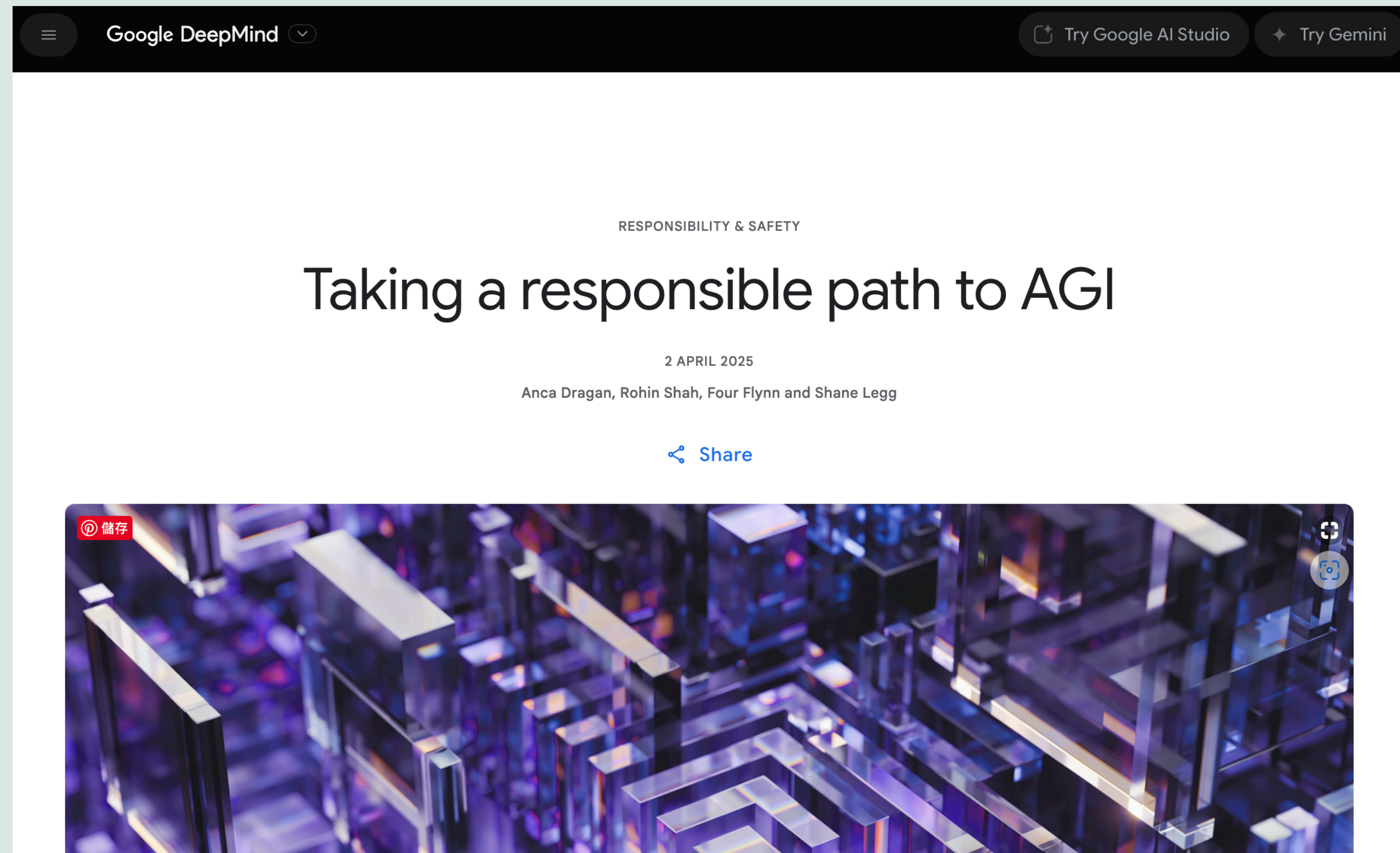
<https://docs.anthropic.com/en/release-notes/system-prompts>



<https://github.com/xai-org/grok-prompts>



DeepMind 一篇重要的文章

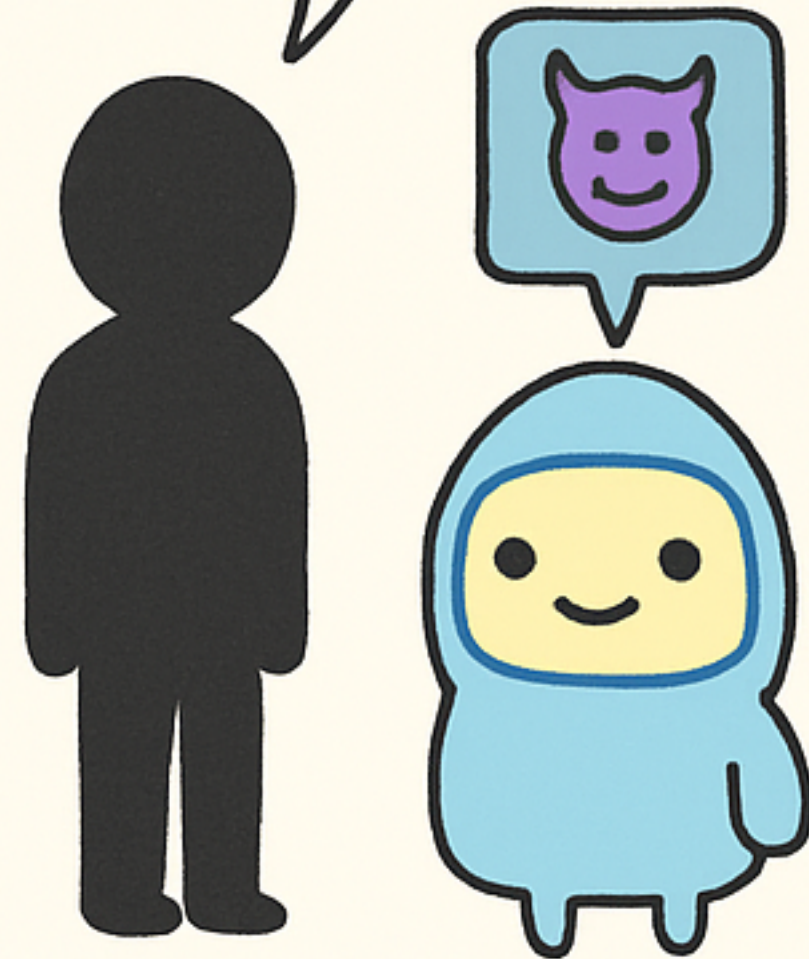


<https://deepmind.google/discover/blog/taking-a-responsible-path-to-agi>

AI 的 四個風險

濫用

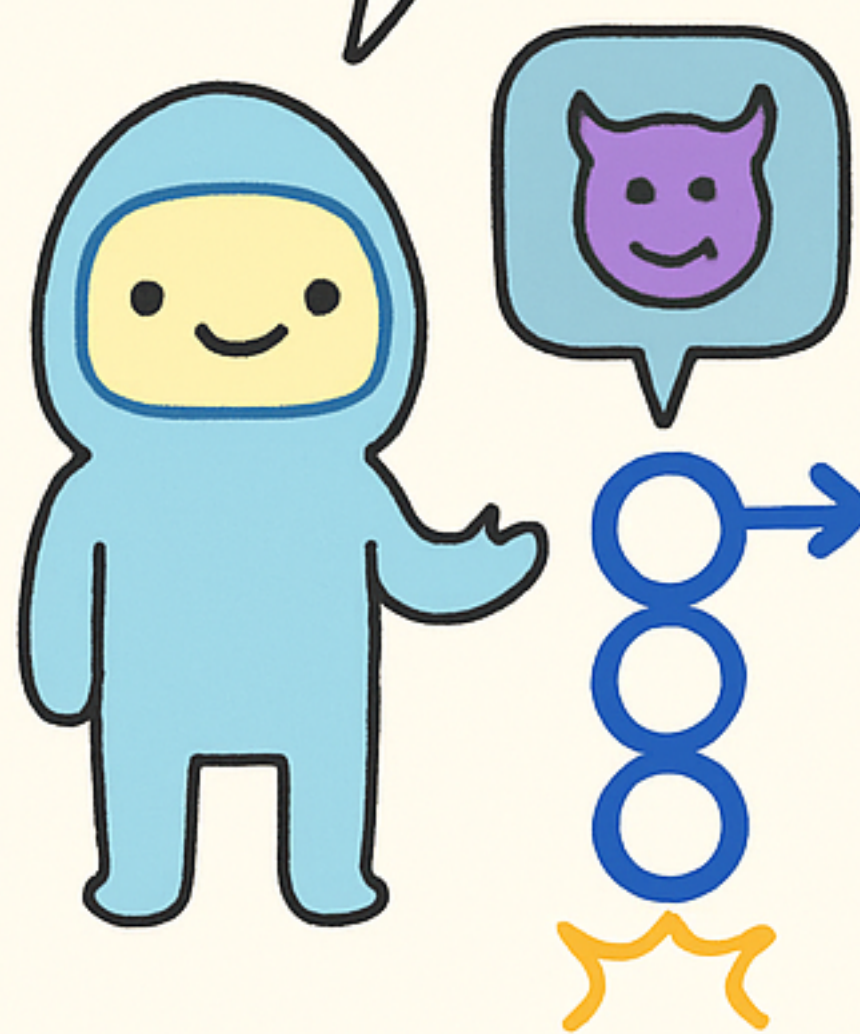
你這個 AI 去做
一些傷害人的事!



風險主因：
人類有惡意

錯誤對齊

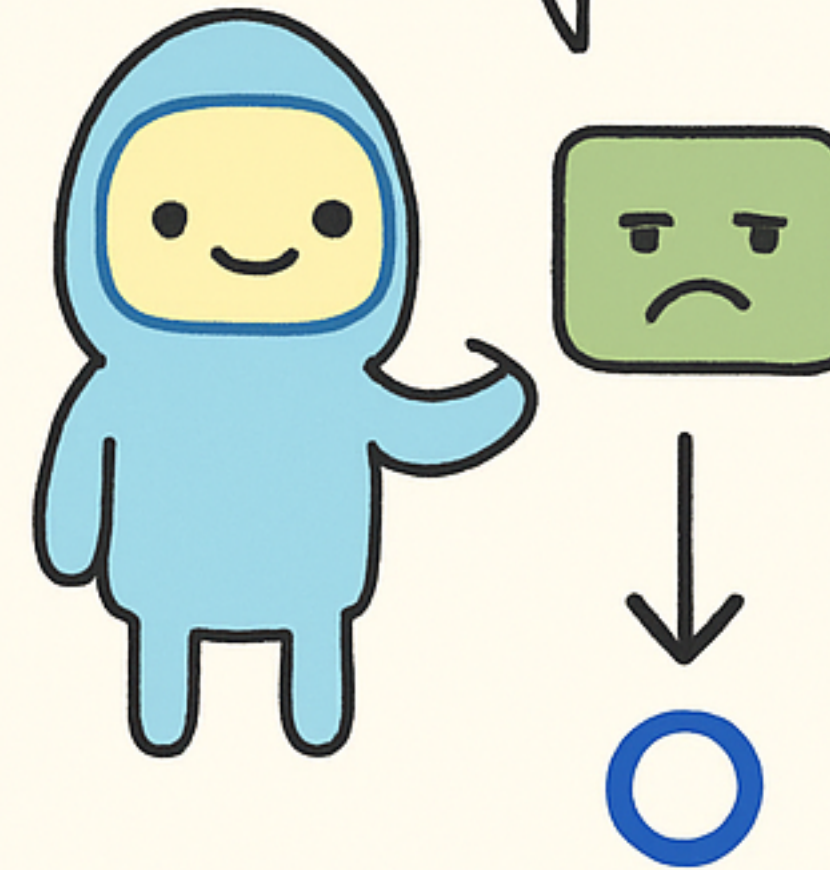
去做 (傷害
的) 這件事。



風險主因：
AI 和人類價值
不一致

錯誤

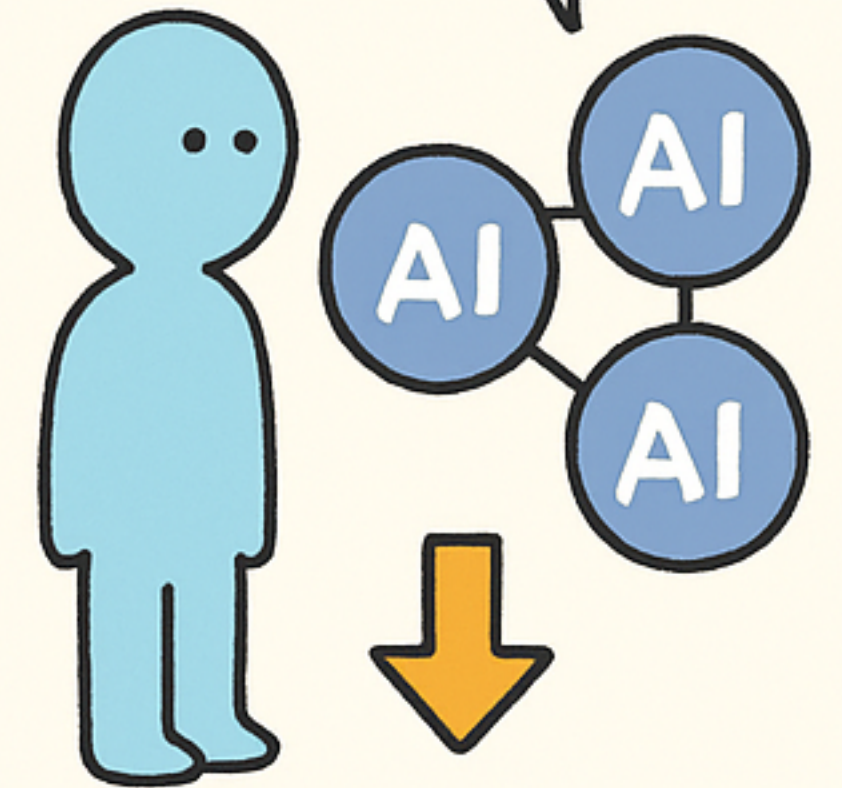
AI 出現不是
故意的錯誤。



風險主因：
世界很複雜，出
了沒想到的 bug

結構性風險

多個「做正確
事」AI 產生
錯誤。

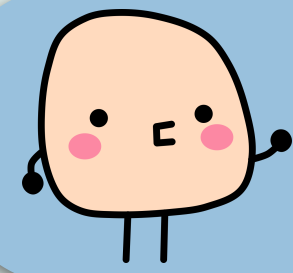


風險主因：
多個 AI Agents、
人類交互產生結
構性問題。



05.

生成式 AI 的其他應用



Google Labs: Little Language Lessons

Little Language Lessons

A collection of bite-sized learning experiments
built with Gemini.


<https://labs.google/111/>



寫一個情境, 自動產生一課

Korean for At the Coffee Shop



 Vocabulary

 Phrases

 Tips

 Vocabulary

커피

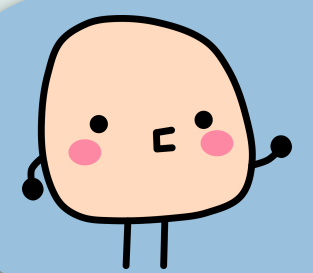
keopi
coffee



메뉴

menu
menu






例句

Korean for At the Coffee Shop



 Vocabulary

 Phrases

 Tips

Phrases

무엇을 도와드릴까요?

mueoseul dowadeurilkayo?

How can I help you?



여기서 드실 건가요, 가져가실 건가요?

yeogiseo deusil geongayo, gajyeogasil geongayo?

For here or to go?



아메리카노 한 잔 주세요.

amerikano han jan juseyo.

One americano, please.



따뜻한 라떼로 주세요.

ttatteuthan latte-ro juseyo.

I'll have a hot latte, please.





用法小重點

Korean for At the Coffee Shop



Vocabulary

Phrases

Tips

📖 Tips

-(으)세요 for polite commands and requests

When you're at a coffee shop and want to politely ask for something, using -(으)세요 is super helpful! This ending is added to verb stems to create polite commands or requests. It's like saying 'please do this'. If the verb stem ends in a consonant, you add -으세요. If it ends in a vowel, you add -세요. For example, '하다' (to do) becomes '하세요' (please do). This shows respect and consideration for the person you're speaking to. Remember that verb stems are the base form of a verb without the -다 ending.

EXAMPLES

아이스 아메리카노 한 잔 주세요. ⓘ

Aiseu Amerikano han jan juseyo.

Please give me one iced Americano.



여기 앉으세요. ⓘ

Yeogi aneuseyo.

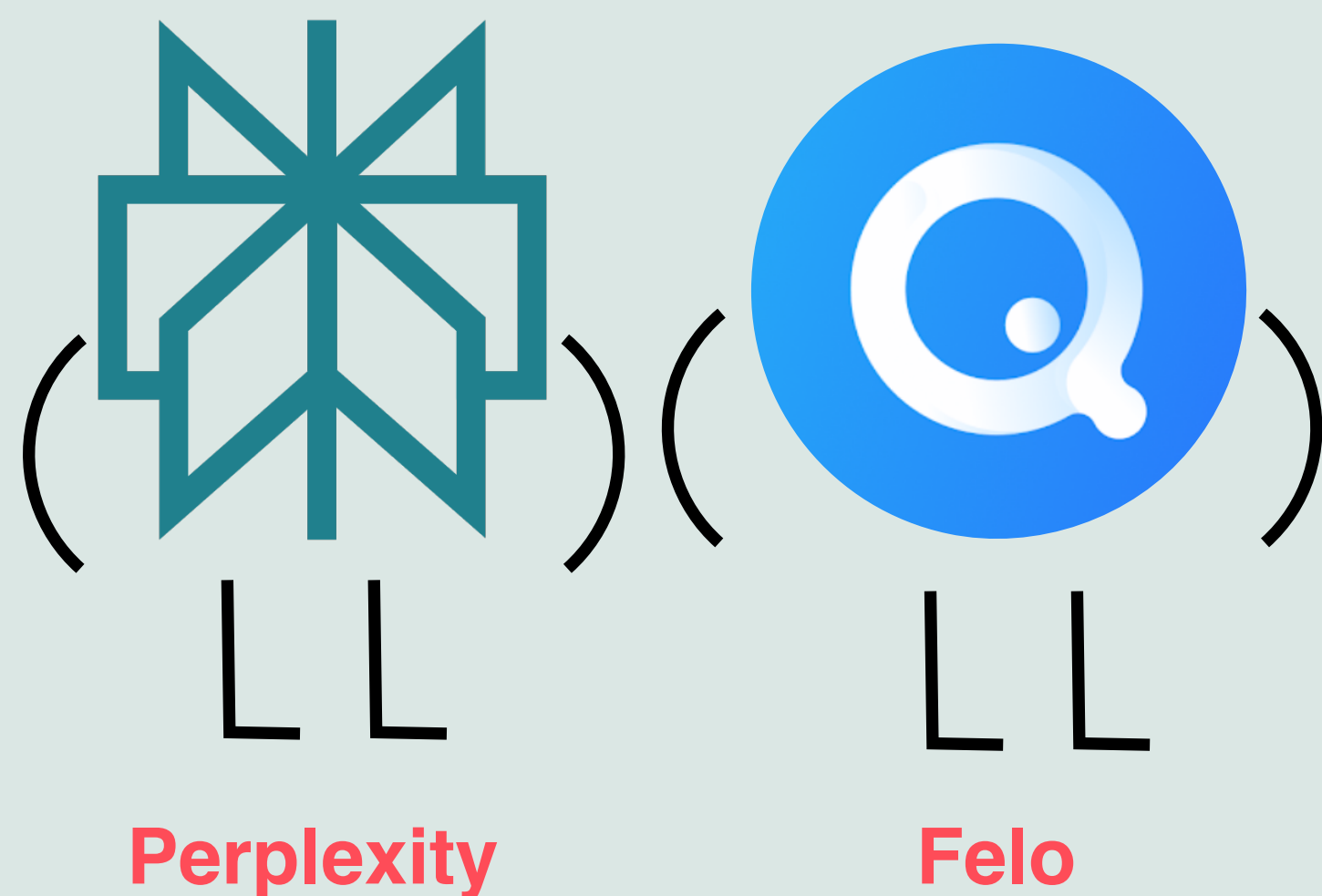
Please sit here.





原本 Perplexity 和 Felo 以會上網著稱

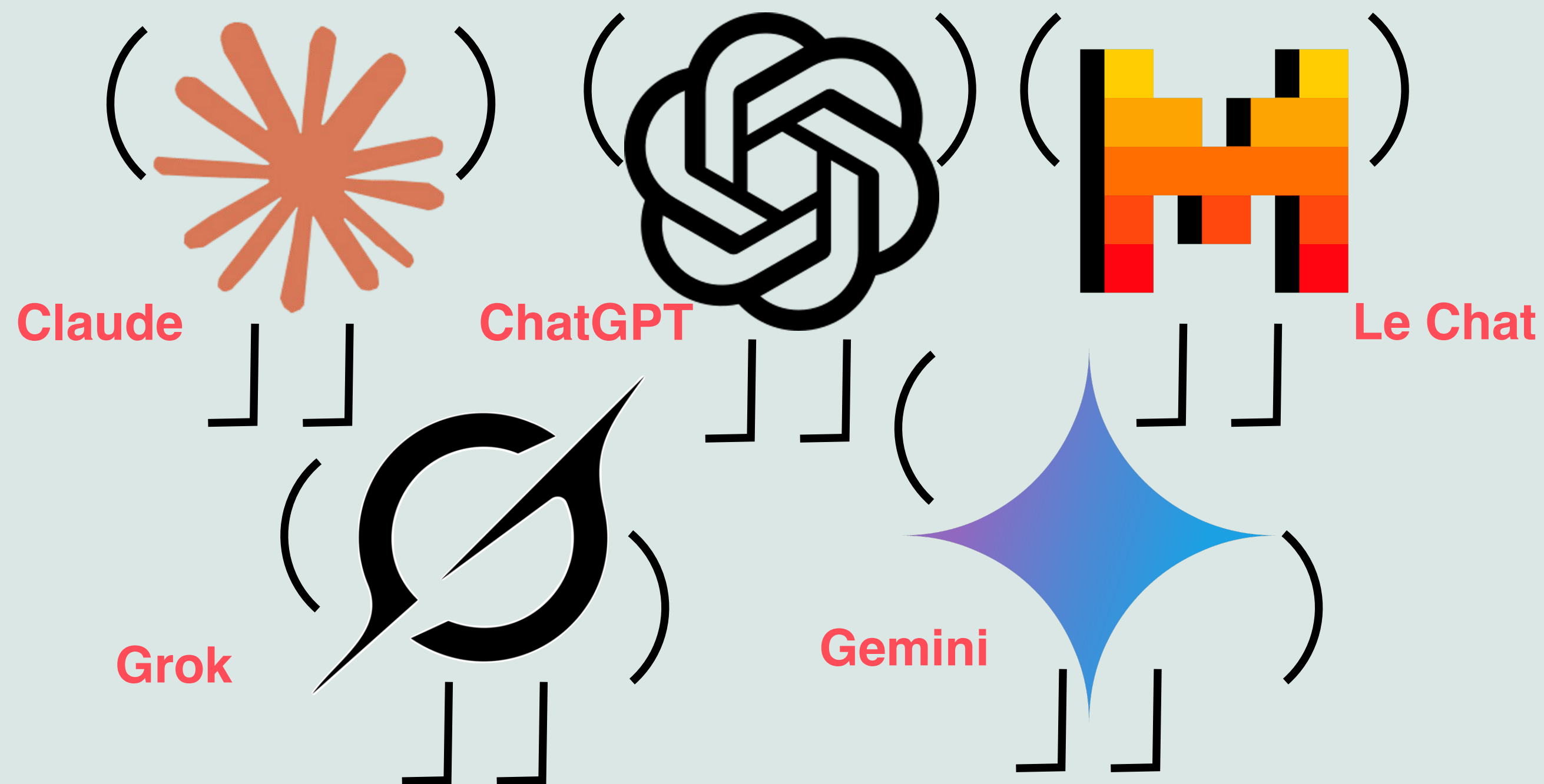
我們很會上網搜尋!



Perplexity

Felo

我們也都會上網了!!



Claude

ChatGPT

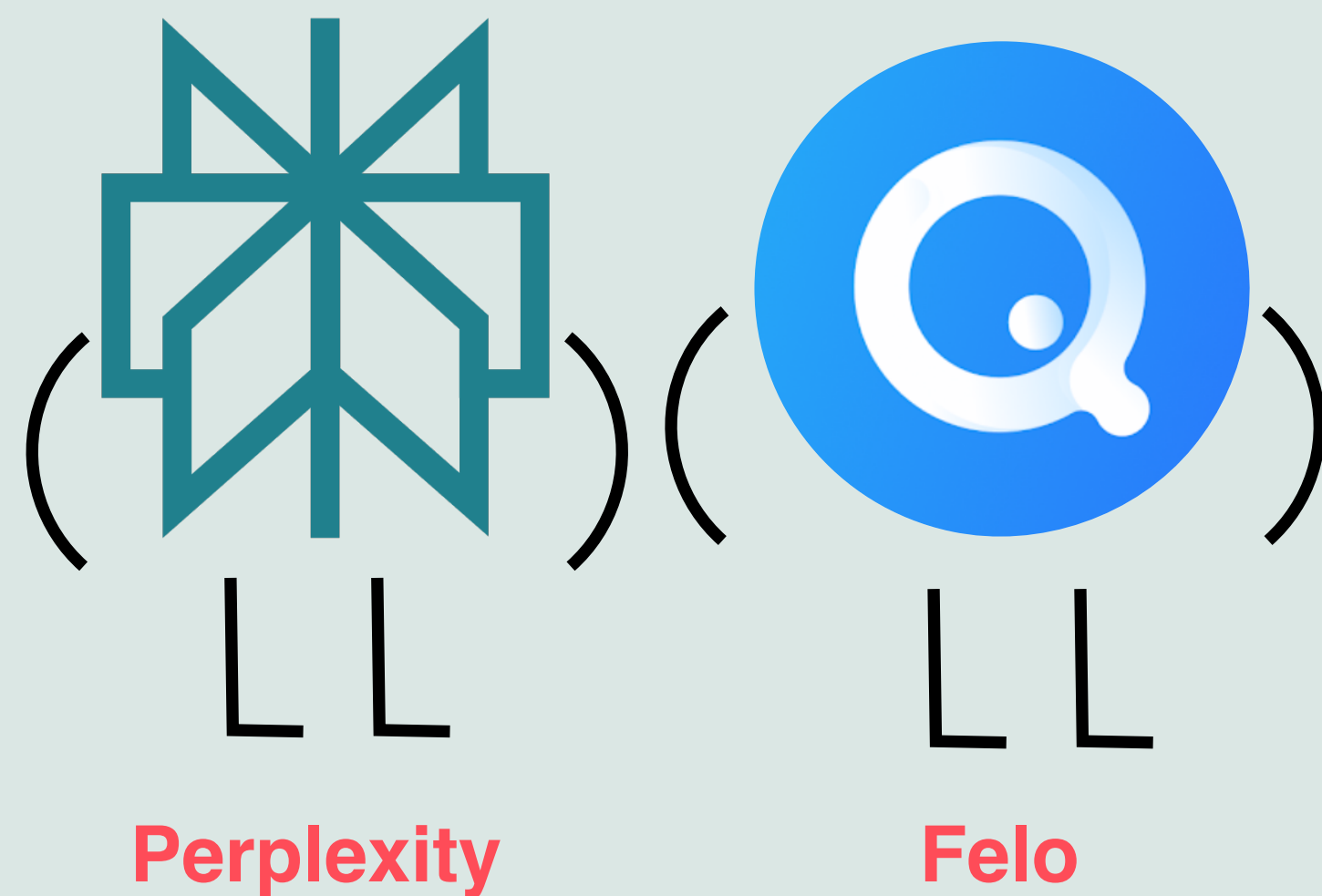
Le Chat

Grok

Gemini



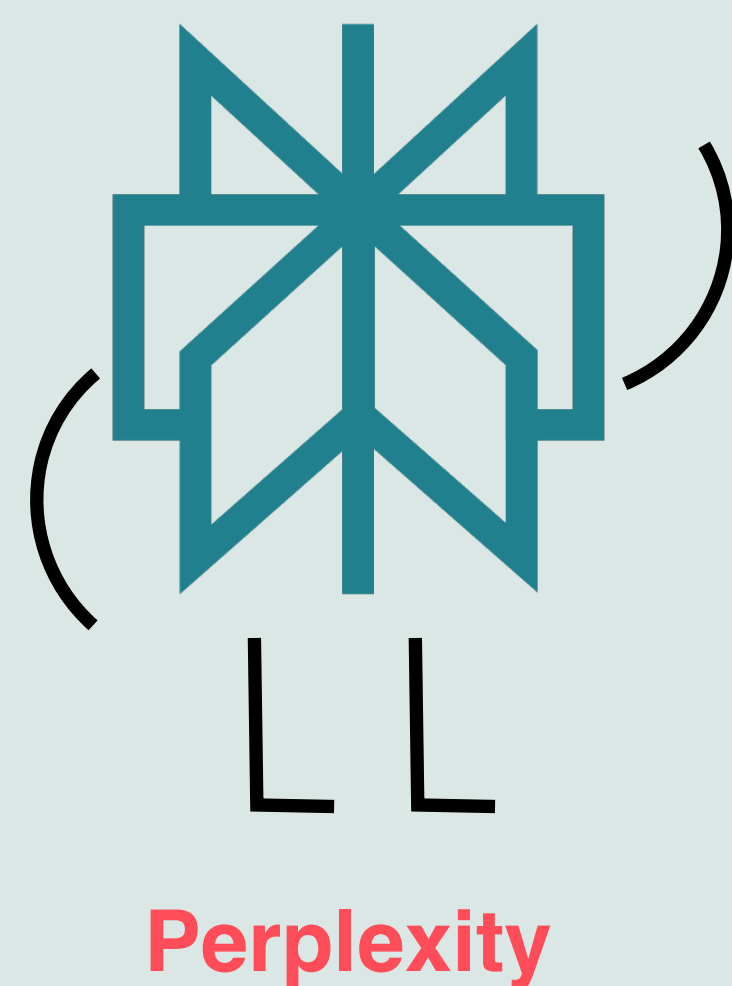
Perplexity 和 Felo 可以想成是一種 AI Agent



我們會先拆解問題，
再進行搜尋。



「搜尋派」也漸漸往全方位發展



我可以切換模型, 還能
做 Deep Research。



Felo: 可以產生互動網頁的搜尋器

生成式AI概覽



資料視覺化

概述

生成式人工智慧（Generative AI）技術近年來發展迅速，能夠根據使用者提示創造新的內容，涵蓋文字、圖像、音訊、視訊和程式碼等多種形式。這些模型通常基於深度學習，特別是利用神經網路和轉換器（Transformer）架構，透過在大量資料集上進行訓練來學習模式並生成新的、統計上相似的內容。常見的生成式 AI 工具和模型可大致分為文字生成、圖像生成以及能夠處理多種資料類型的多模態模型。

文字生成

從聊天機器人到內容創作，文本生成模型廣泛應用於各種場景。

圖像生成

根據文字描述創造全新圖像，應用於藝術創作、設計與視覺內容生產。

多模態模型

能夠處理並整合不同類型資訊（文字、圖像、音訊等）的進階模型。

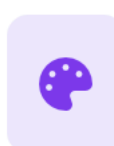
這份報告涵蓋的關鍵內容：

- 主流文字生成工具介紹與特點
- 圖像生成模型比較與應用場景
- 多模態技術的最新發展與應用



Felo: 可以產生互動網頁的搜尋器

主要圖像生成平台



高品質藝術創作平台

- Midjourney: 透過 Discord 平台提供服務，以藝術風格著稱
- DALL-E 3: OpenAI 最新圖像生成模型，整合於 ChatGPT
- Leonardo.Ai: 專業藝術和設計平台
- Adobe Firefly: 商業安全的創意工具

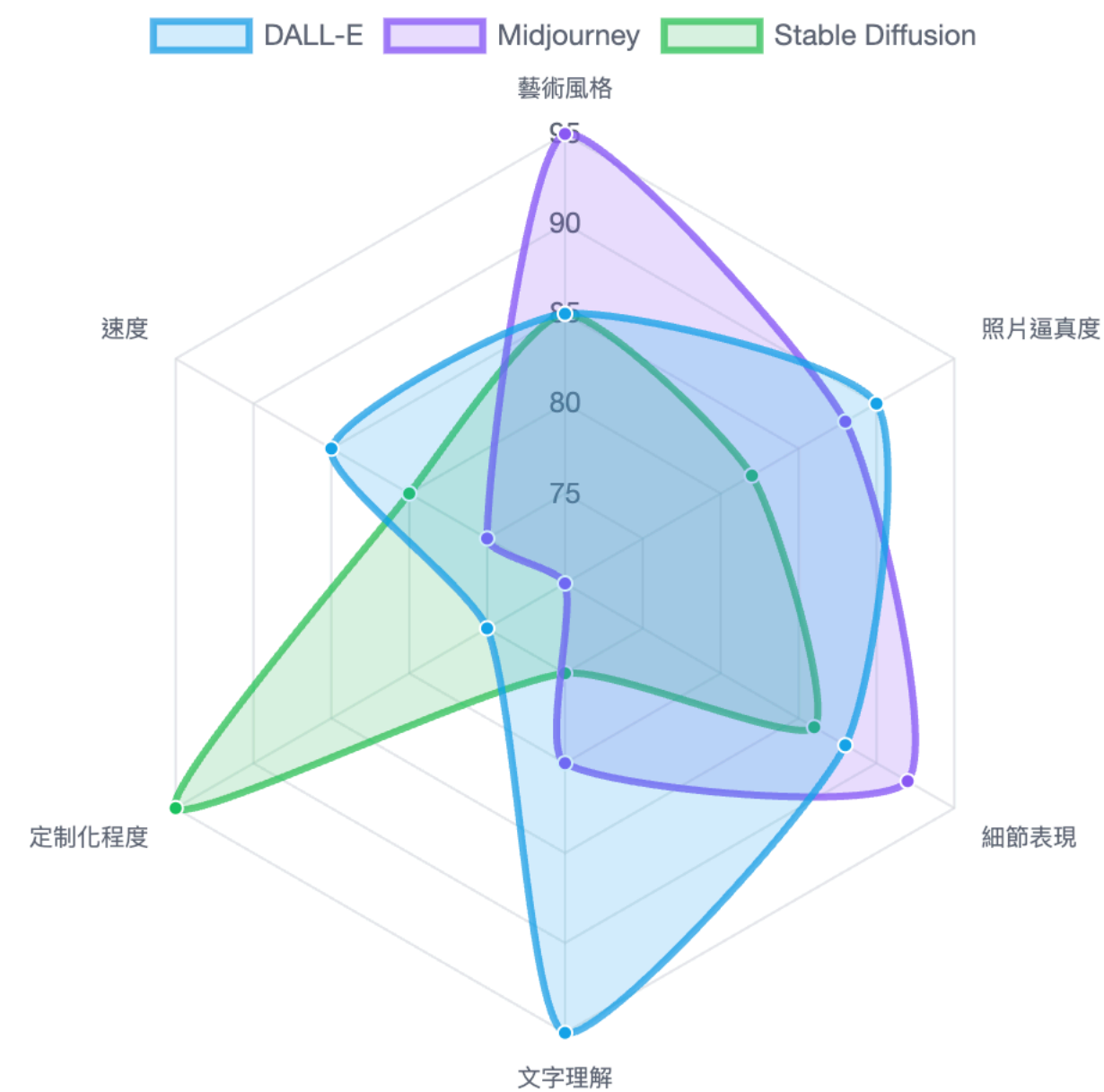


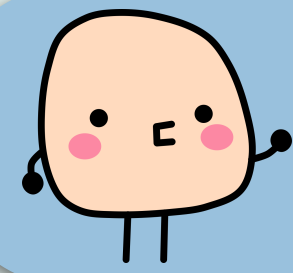
開源與可自定義平台

- Stable Diffusion: 開源社群驅動的平台，支持本地運行
- Flux.1: 2024年推出的高品質模型
- DeepFloyd IF: 專注於照片逼真度和文字渲染
- 吐司TusiArt: 中國的線上 AI 繪畫平台

圖像生成技術比較

圖像生成模型能力比較

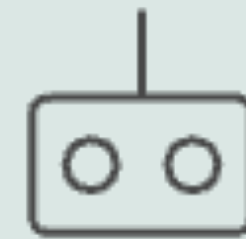
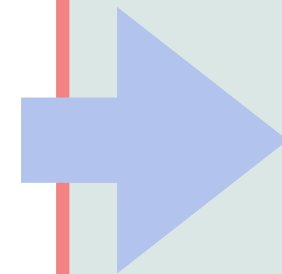




Napkin AI: 快速視覺化你的想法

發想的呈現: 原型
presentation, expression, embodiment,
prototype, 都是觀念的具體化, 目的:

- Clarity - 使概念清楚
- Communicate
- Compare
- Evaluate - 測試概念的可行性



原型的主要目的應該
是什麼?

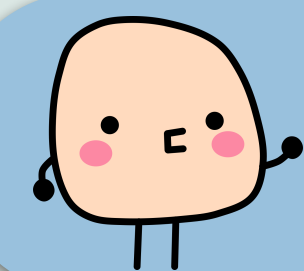


你不一定清楚完整的隨手筆記

Made with  Napkin

多個視覺化可能讓你選!

<https://www.napkin.ai/>



Suno.ai: 會作曲的 AI



<https://yenlung.me/Almagician>



ChatGPT 作詞過程

上面的歌真的很棒, 現在我們想用相同的概念, 推薦大家來學生成式 AI。包括文字生成和圖像生成的原理。文字生成的原理還有 RNN, Transformers, 加上如 ChatGPT, Gemini, Claude 等大型語言模型 LLM, Llama, Gemini, Mistral 等開源語言模型, 在自己電腦用 LM Studio 跑; 再來圖像生成當紅的 Diffusion Models, 包括 Midjourney, DALL·E, 還有用 Fooocus 使用開源的 Stable Diffusion 等等。期望學習者能變成一位擅長下咒語 (prompts) 的魔導師, 掌握 AI, 讓 AI 成為自己的代理人分身。

請依上面 《我要當 Py 郎》的風格, 想個曲名, 然後作一支鼓勵大學來學習生成式 AI 的歌

整個歌曲非常的棒, 能不能像是寫給 K-Pop 女團的樣子, 也就是即使不是專有名詞有些地方也用英文, 然後部份帶點 rap



06.

用 NotebookLM 來學習!




NotebookLM



<https://notebooklm.google/>



我們以吳恩達老師介紹 Agentic AI 的課程為例




AGENTIC AI with Andrew Ng

由Ai.Group™ | PODCAST建立
播放清單 · 32 部影片 · 觀看次數：8,130次


▶ 全部播放

1




AGENTIC AI — EPISODE 1: Welcome!
Ai.Group™ | PODCAST · 觀看次數：3068次 · 1 個月前

2




AGENTIC AI — EPISODE 2: What is Agentic AI?
Ai.Group™ | PODCAST · 觀看次數：1703次 · 1 個月前

3



AGENTIC AI — EPISODE 3: Degrees of Autonomy
Ai.Group™ | PODCAST · 觀看次數：1336次 · 1 個月前

4



AGENTIC AI — EPISODE 4: Benefits of Agentic AI
Ai.Group™ | PODCAST · 觀看次數：884次 · 1 個月前



先不斷丟「教材」



檔案、影片
(YouTube 連結) 等等
都可以。





功能齊全、學習好幫手 NotebookLM



看看你想用什麼方式
來學習。





語音/影片摘要



語音摘要



影片摘要



可以設定格式、語言，
也可以用 prompt 說
要什麼方向。



自訂語音摘要

格式

深入探索

兩位主持人熱烈討論，解說並連結來源中的主題

摘要

簡短摘要，讓你快速掌握來源的核心概念

評論

根據來源提供專業評論和建設性意見，協助改善內容

辯論

兩位主持人深入辯論，提供對來源的不同觀點

語言

中文（繁體）

長度

短

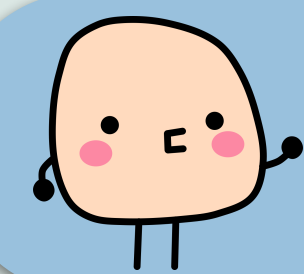
✓ 預設

在本集中，AI 主持人應著重哪些部分？

提示範例

- 著重特定來源，例如僅涵蓋義大利相關文章
- 著重特定主題，例如只討論小說主角
- 著重特定對象，例如向不熟悉生物學的人說明內容

生成



然後就會產生這樣的 Podcast (影片)



https://yenlung.me/podcast_demo



學習卡



學習卡



代理字卡

根據 32 個來源

按下空格鍵即可翻開學習卡，並使用「← / →」前往其他卡片

什麼是「代理式 AI」
(Agentic AI) 工作流
程？

查看答案



1 / 59



代理字卡

根據 32 個來源

按下空格鍵即可翻開學習卡，並使用「← / →」前往其他卡片

這是一種流程，其中基於大型語言
模型的應用程式會執行多個步驟來
完成一項任務，通常涉及思考、研
究和修訂的迭代循環。

說明



1 / 59





資訊圖表



資訊圖表



打造更智慧的AI：代理式工作流程指南



什麼是代理式AI？

讓AI模型像人一樣思考，透過規劃、執行工具和反思來解決複雜問題。

傳統單次提示
(GPT-3.5)

48%

代理式流程

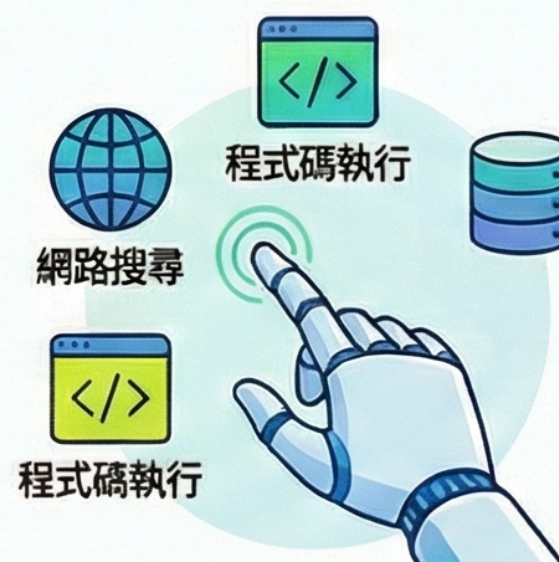
95%
準確率

在程式碼生成基準測試中，效能顯著躍升



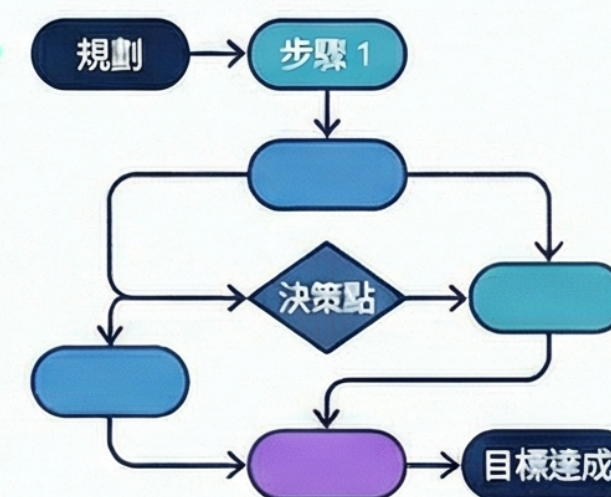
1. 反思 (Reflection)

讓模型審視並批判自己的輸出（如草稿），找出錯誤並進行修正改進。



2. 工具使用 (Tool Use)

賦予模型呼叫外部功能（如網路搜尋、程式碼執行）的能力，以獲取資訊或執行動作。



3. 規劃 (Planning)

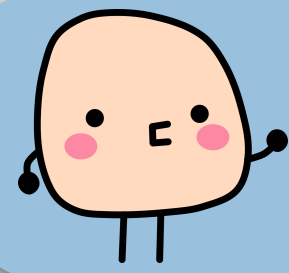
模型自主規劃一系列步驟，決定執行複雜任務所需的工具與順序。



4. 多代理人協作 (Multi-Agent Collaboration)

讓多個扮演不同角色（如研究員、編輯）的代理人合作，共同完成一項大型專案。

NotebookLM



生成簡報!



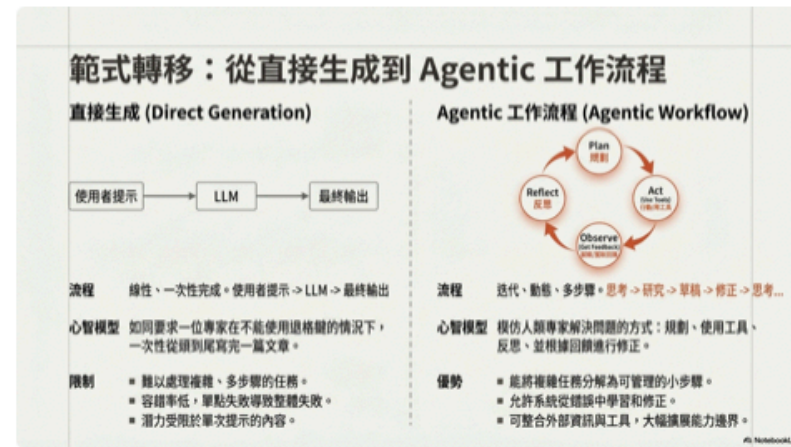
簡報



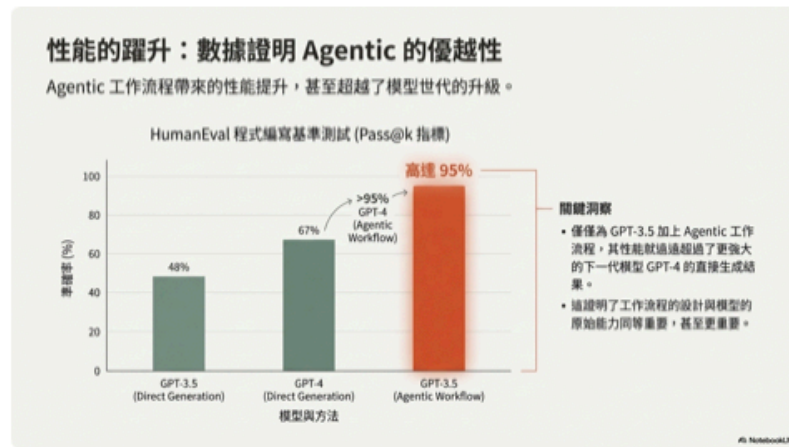
1



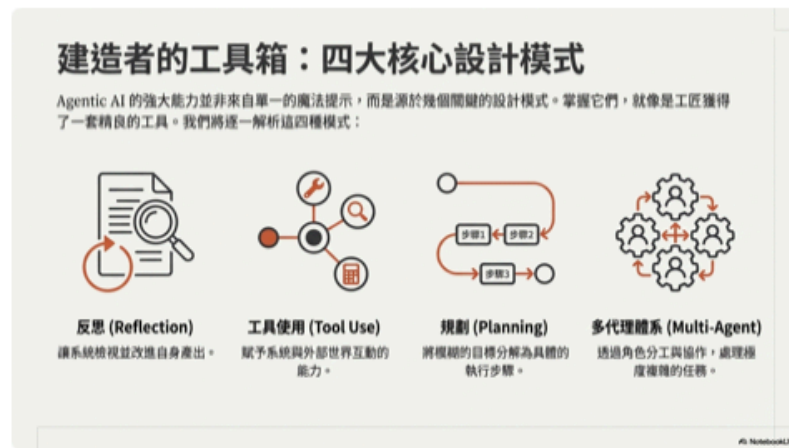
2



3



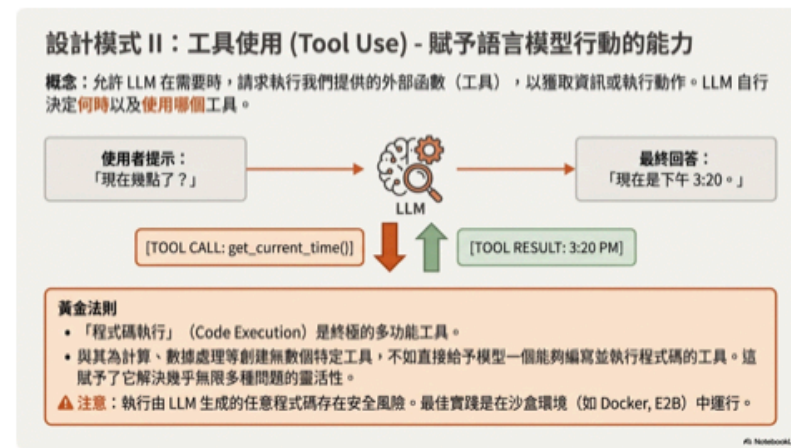
4



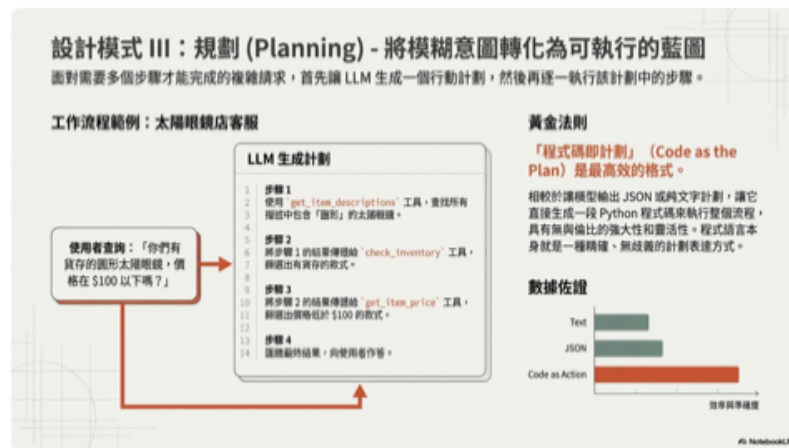
5



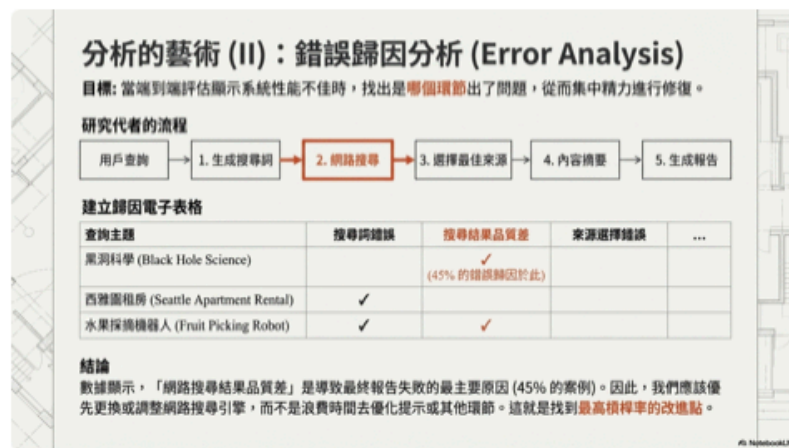
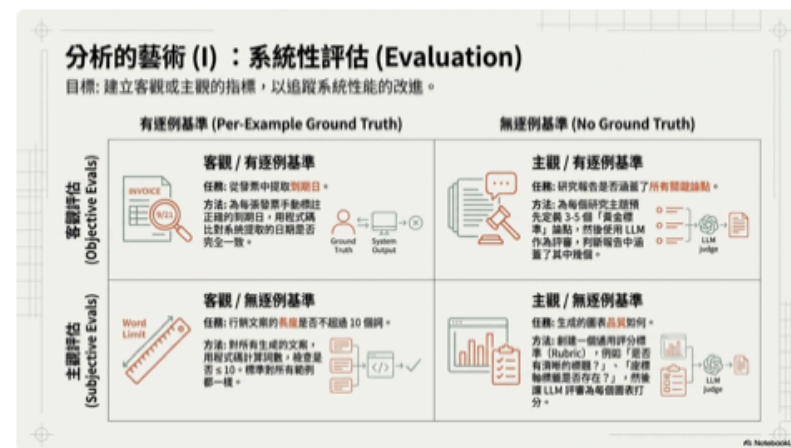
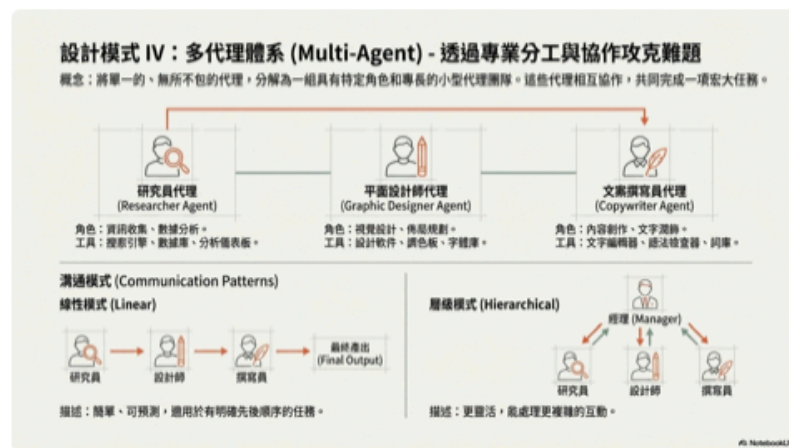
6

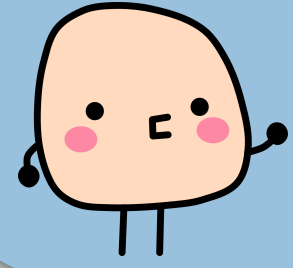


7



8





【作業】用 Google NotebookLM 來學習!

- ① 想一個你有興趣的學習主題
- ② 找到不只一個學習資源
- ③ 用 NotebookLM 幫忙你學習
- ④ 可以分享覺得做得不錯的部份, 你怎麼下 prompt 的, 或者看了之後還有更多深入對談





Q & A



有問題嗎？