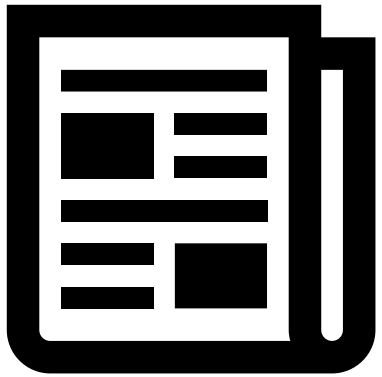
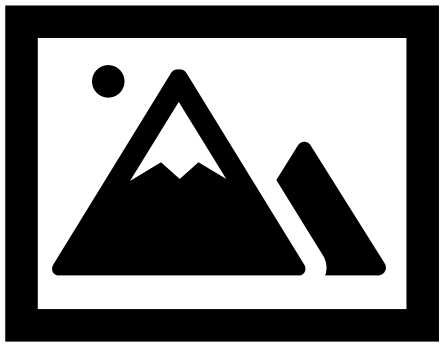

生成的策略

生成式人工智慧 (Generative AI)：機器產生複雜有結構的物件

盡乎無法窮舉　由有限的基本單位構成



文字



影像

聲音

文字由 Token 構成

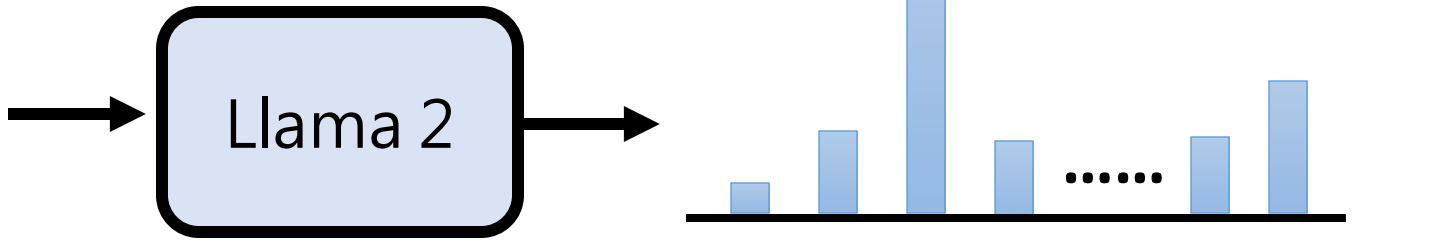
Tokens 65 Characters 373

A language model is a probabilistic model of a natural language. In 1980, the first significant statistical language model was proposed, and during the decade IBM performed ‘Shannon-style’ experiments, in which potential sources for language modeling improvement were identified by observing and analyzing the performance of human subjects in predicting or correcting text.

Text Token IDs

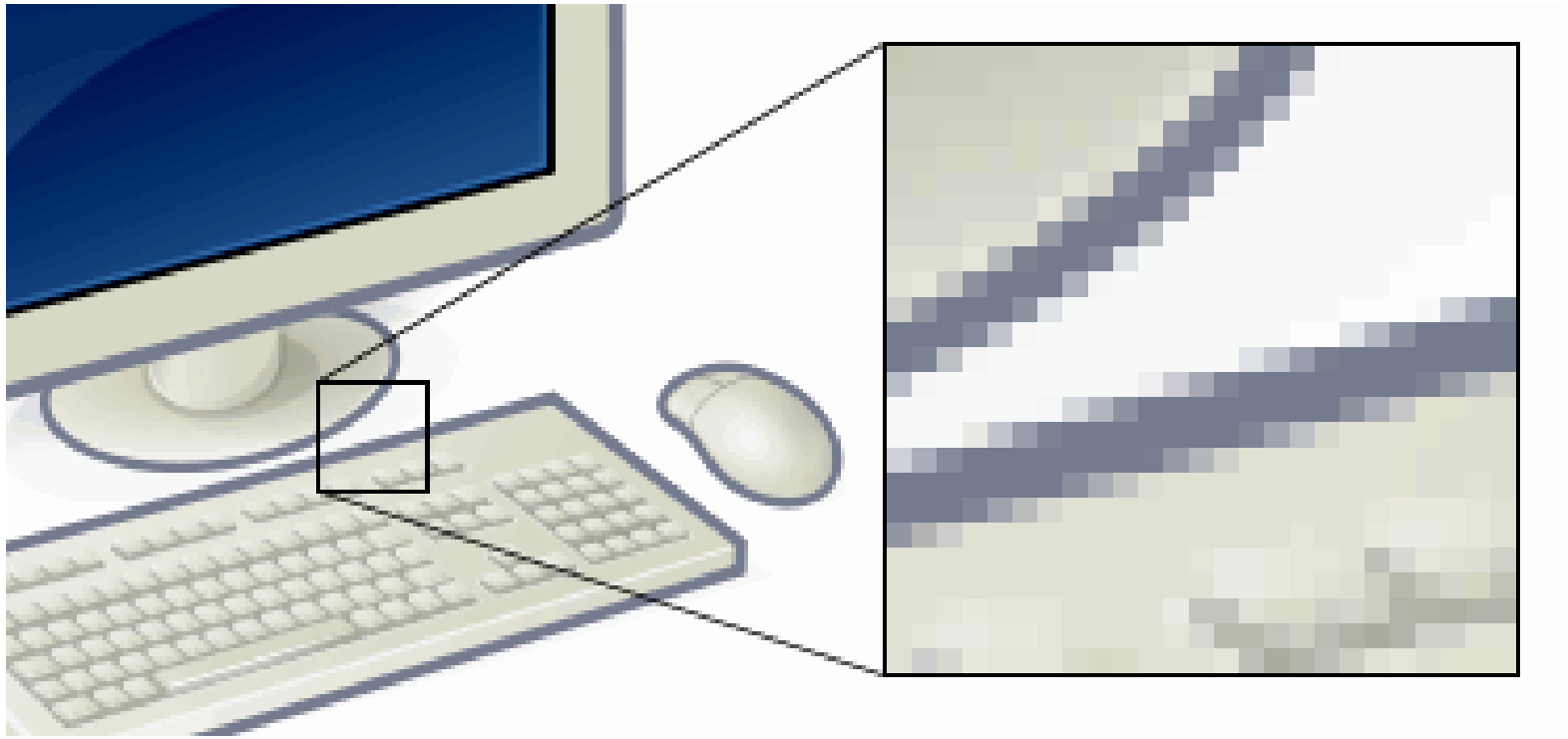
https://platform.openai.com/tokenizer

這門課是生成式 AI 導



32K 個不同的 Token

影像由像素(Pixel)所構成



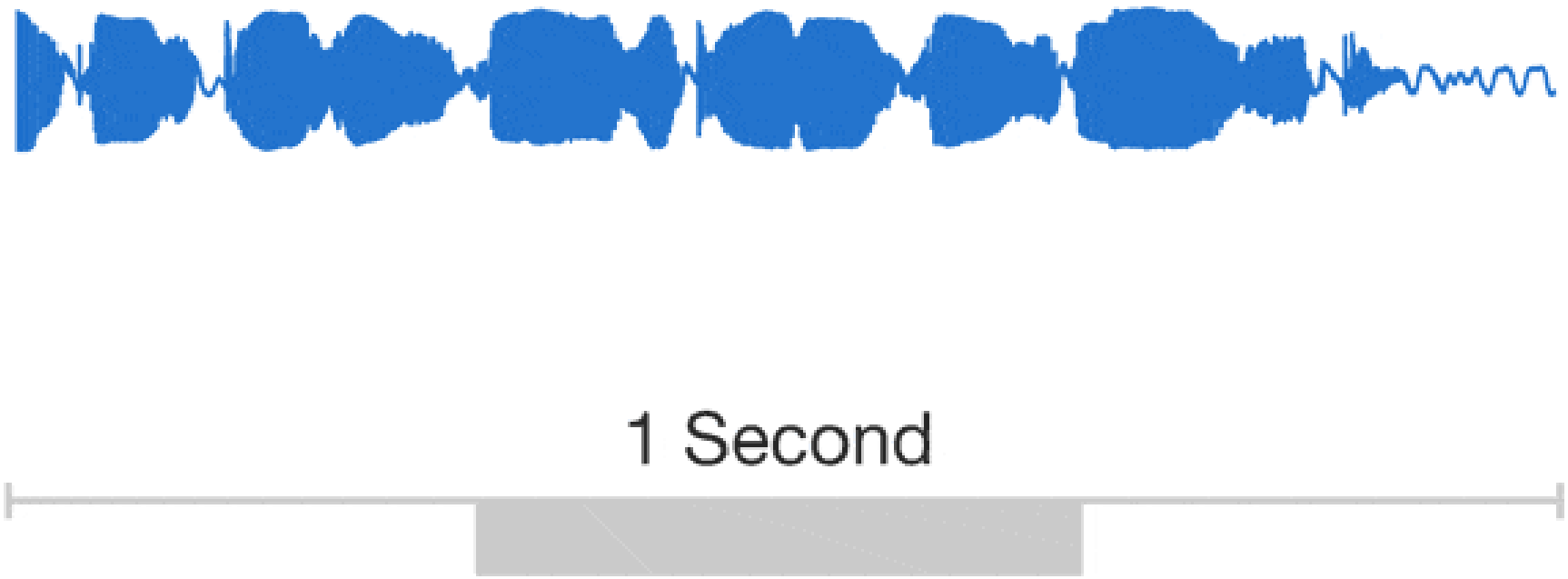
每一個像素可以有多少顏色取決於 BPP (Bit per Pixel)

8 BPP → 256 色
16 BPP → 65536 色
24 BPP → 1670 萬色

https://zh.wikipedia.org/zh-tw/%E5%83%8F%E7%B4%A0#/media/File:Pixel-example.png

https://web.archive.org/web/20211027193750/https://deepmind.com/blog/article/wavenet-generative-model-raw-audio

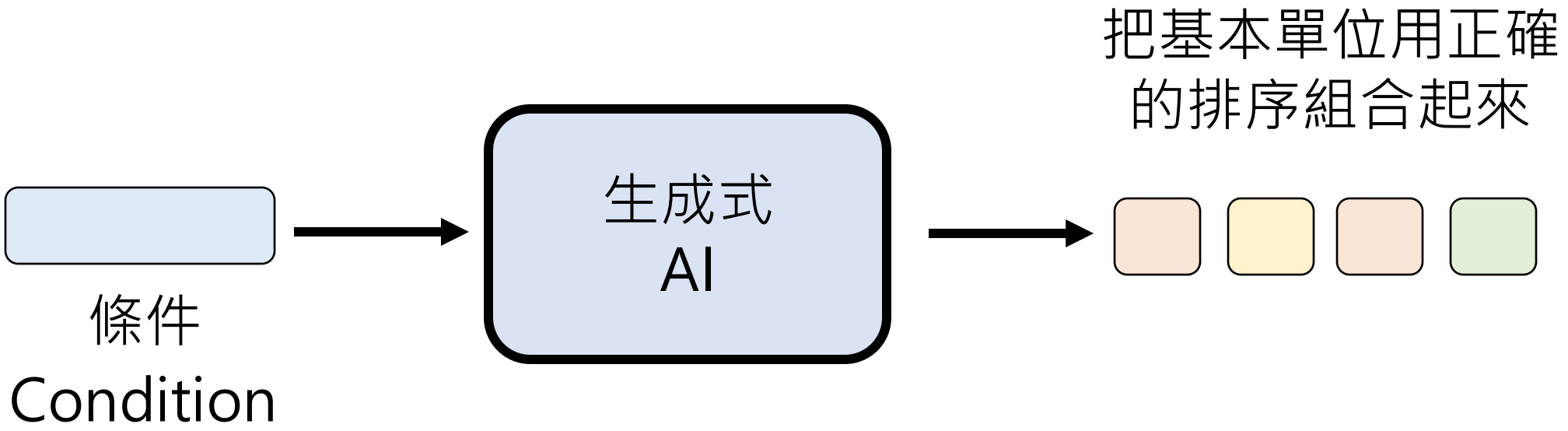
聲音由取樣點(Sample)所構成



取樣率 (Sampling Rate) 16KHz：每一秒有 16,000 個點

取樣解析度
(Bit Resolution)

生成式人工智慧的本質



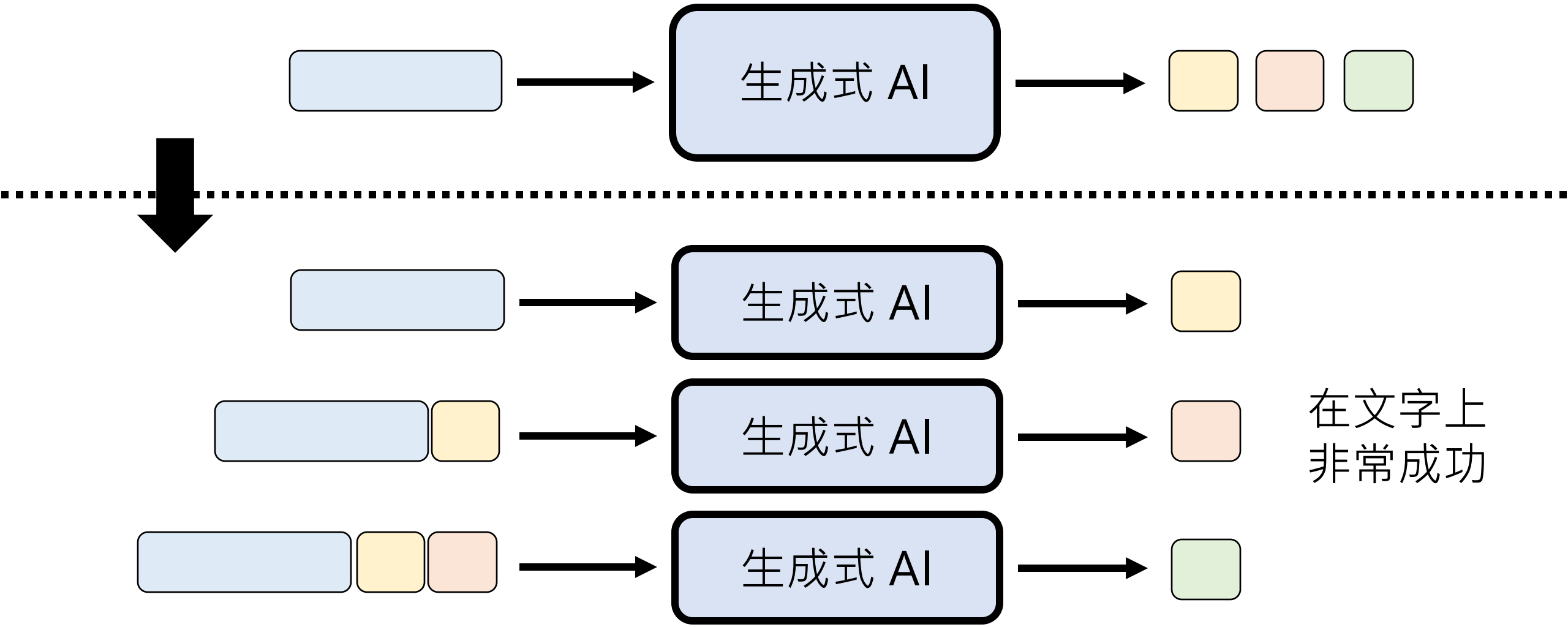
生成式人工智慧的本質

什麼是機器學習？ → 對話 AI → Token → 字 字 字

一隻在奔跑的狗 → 繪圖 AI → Pixel

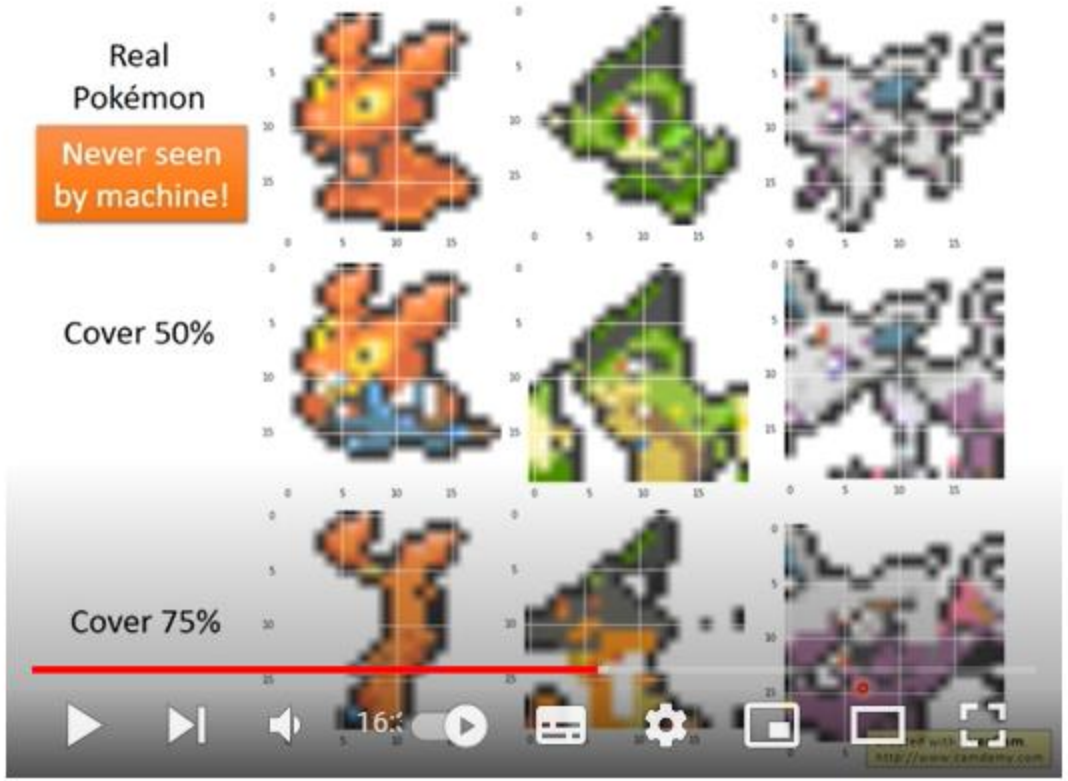
一首歌詠人工智慧的歌 → 聲音生成 AI → Sample

生成的策略：Autoregressive Generation (AR)



生成的策略：Autoregressive Generation

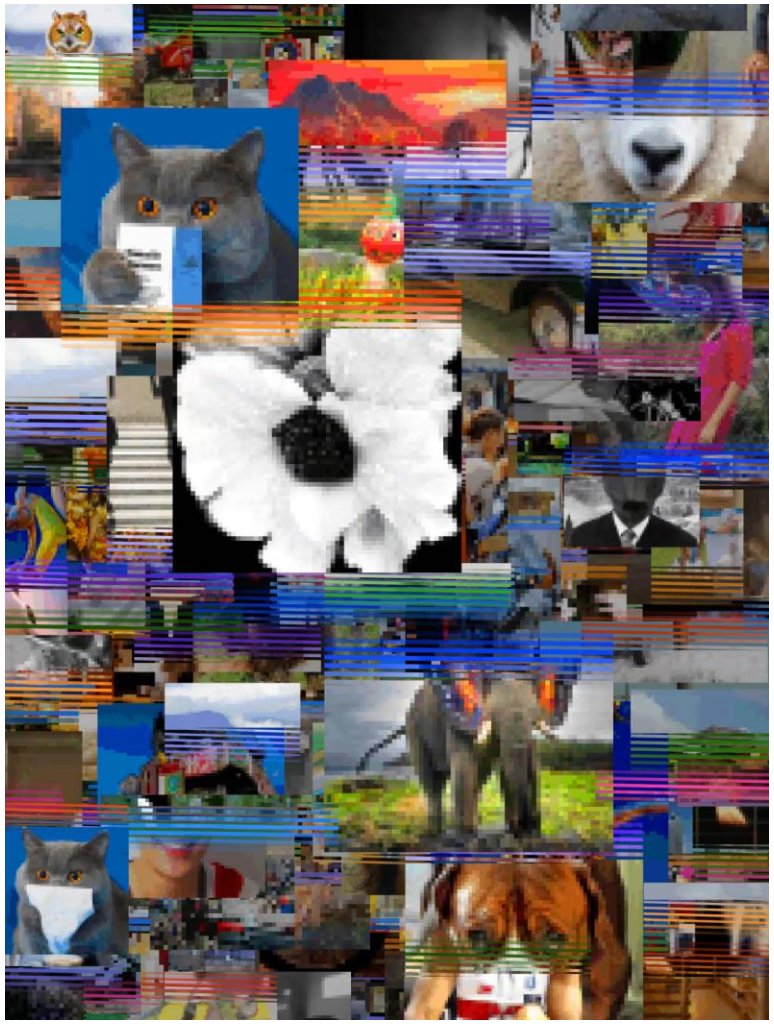
影像呢？



ML Lecture 17: Unsupervised Learning - Deep Generative Model (Part I)

https://youtu.be/YNUek8ioAJk?t=537

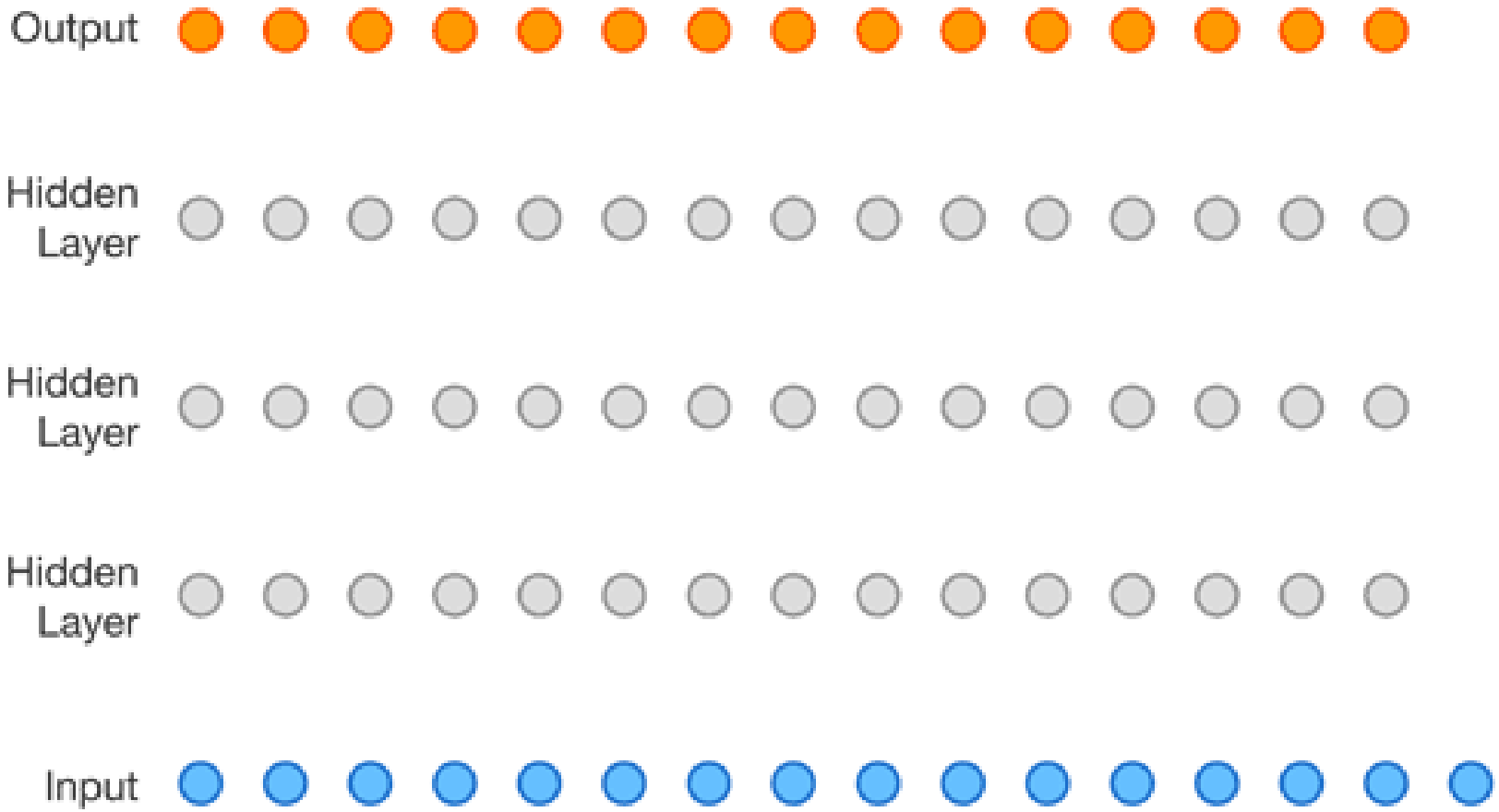
(2016 年《機器學習》秋季班上課錄影)



https://openai.com/blog/image-gpt/

生成的策略：Autoregressive Generation

聲音呢？

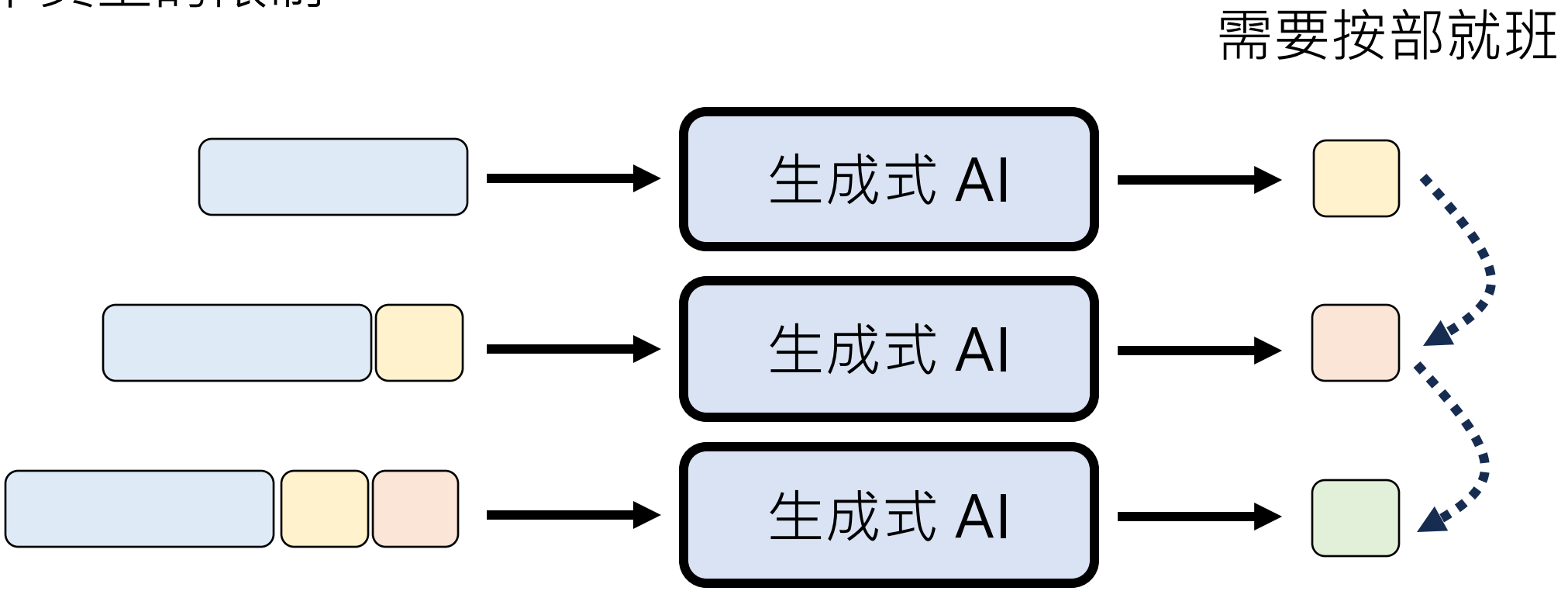


WavNet

https://arxiv.org/abs/1609.03499

生成的策略：Autoregressive Generation

- 本質上的限制



生成的策略：Autoregressive Generation

- 假設要生成 1024 x 1024 解析度的圖片

要做約100萬次接龍!

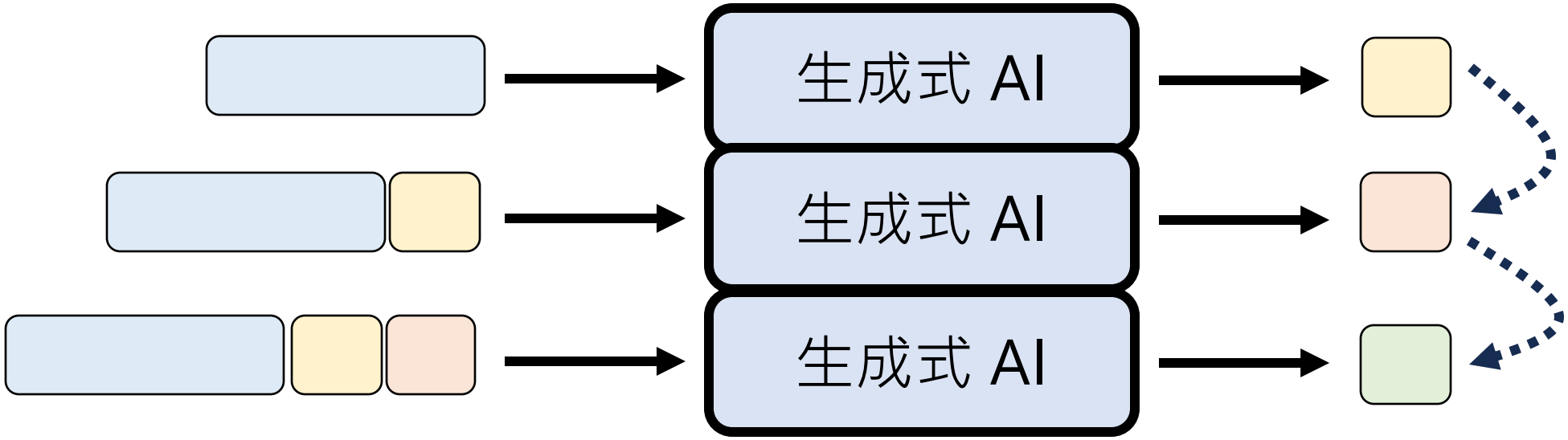
等於每生一張圖片都要寫一部紅樓夢

https://www.eslite.com/product/1001110932518887

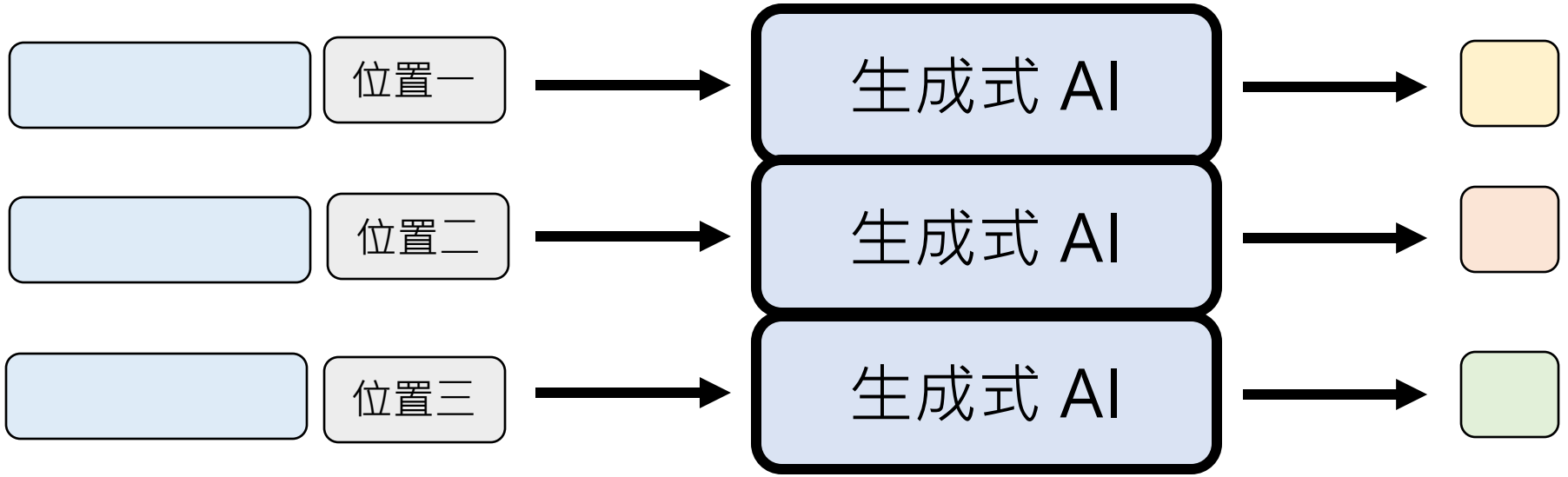
- 假設要生成取樣率 22K 的語音 1 分鐘

要做約132萬次接龍!

Autoregressive Generation (AR)



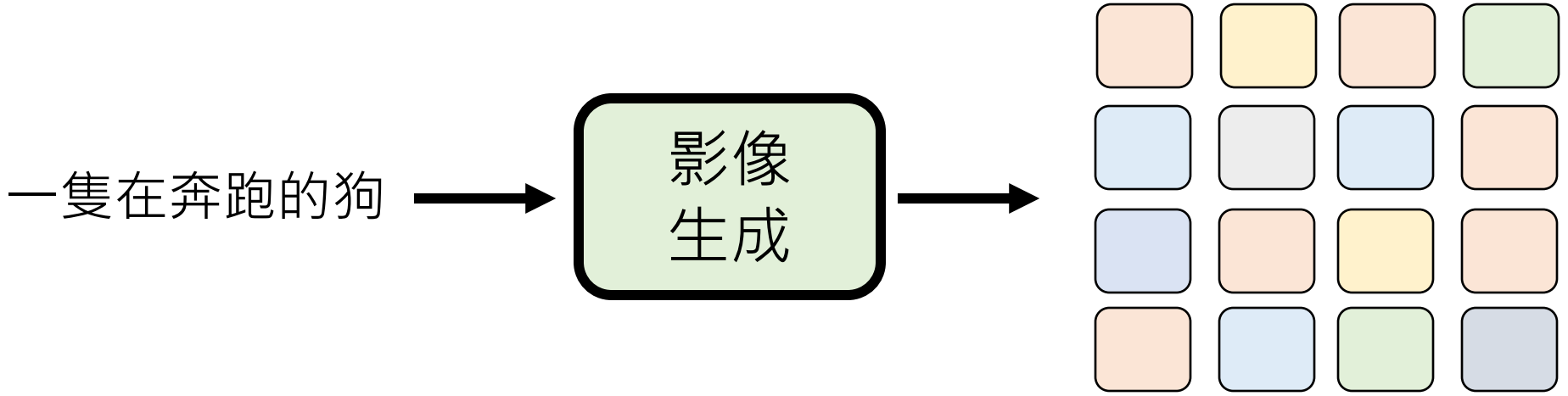
Non-autoregressive Generation (NAR)



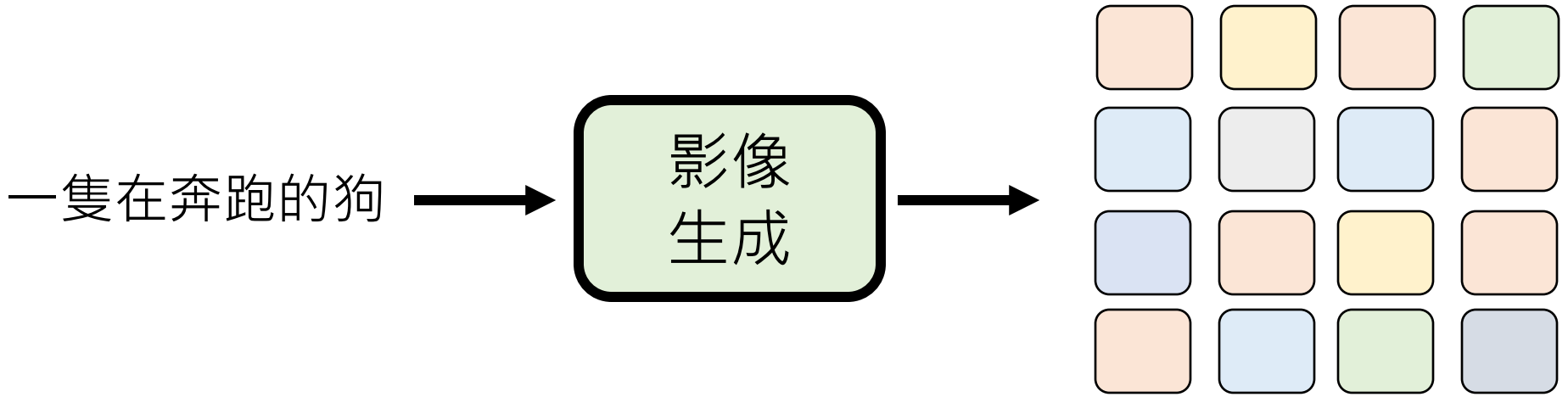
平行運算

一次生出所有基本單位

Autoregressive Generation (AR)

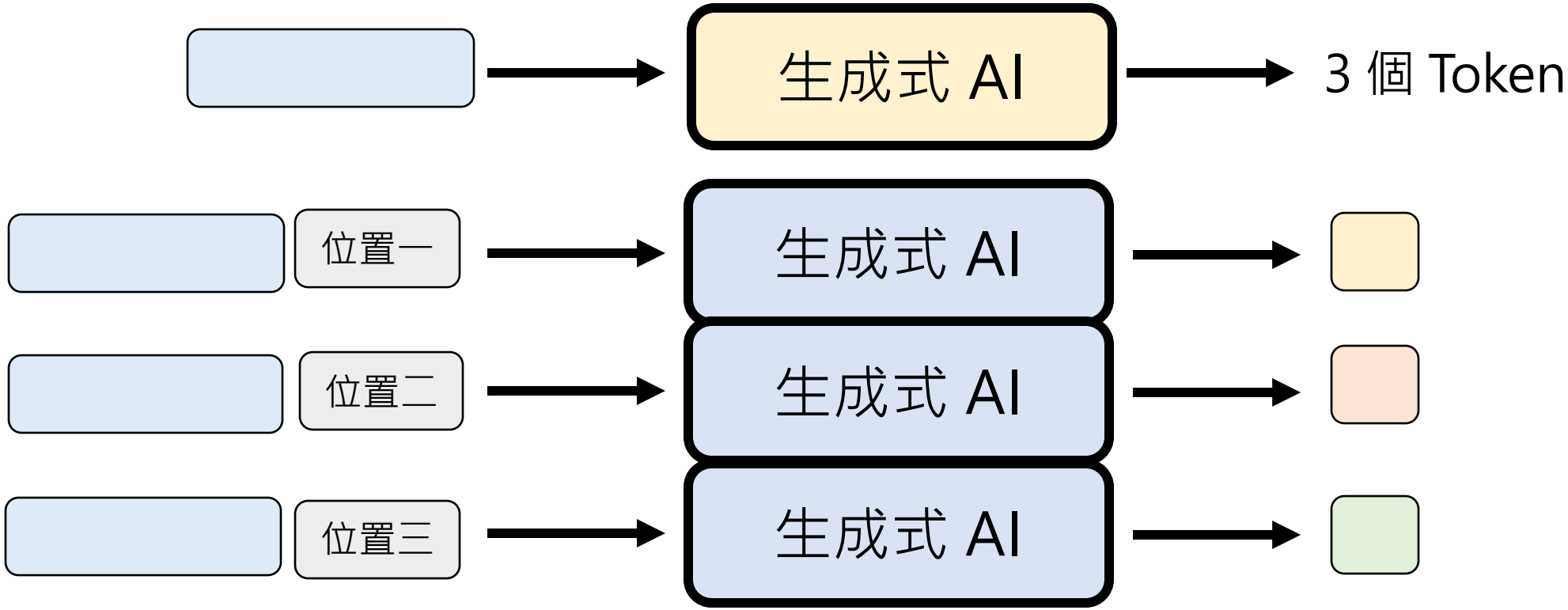


Non-autoregressive Generation (NAR)



Non-Autoregressive Generation

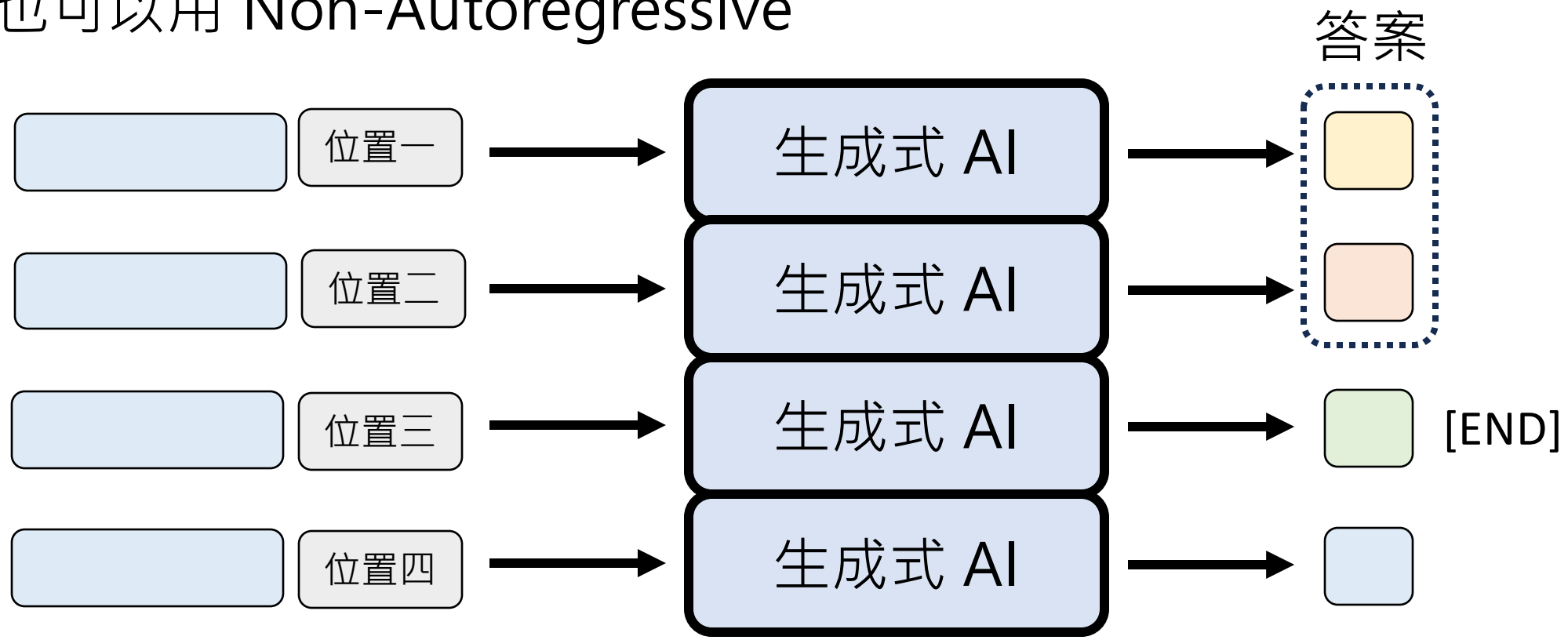
- 文字也可以用 Non-Autoregressive



Survey paper: https://arxiv.org/pdf/2204.09269

Non-Autoregressive Generation

- 文字也可以用 Non-Autoregressive

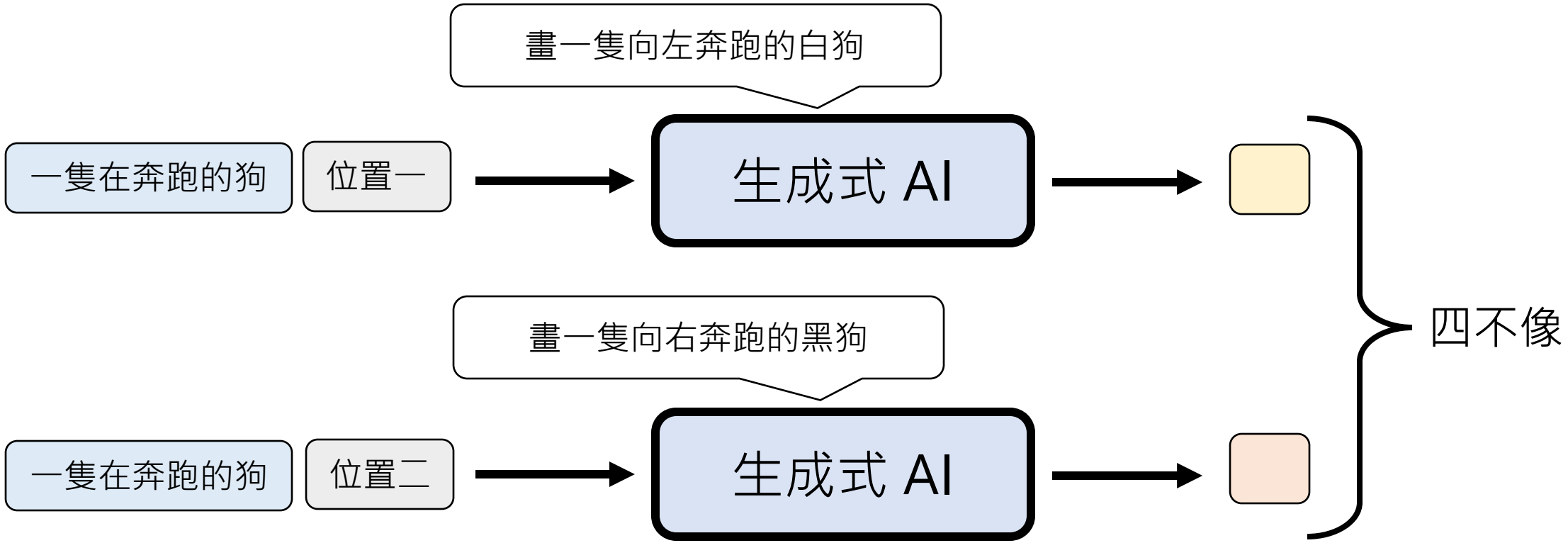


反正就是生成固定長度

Survey paper: https://arxiv.org/pdf/2204.09269

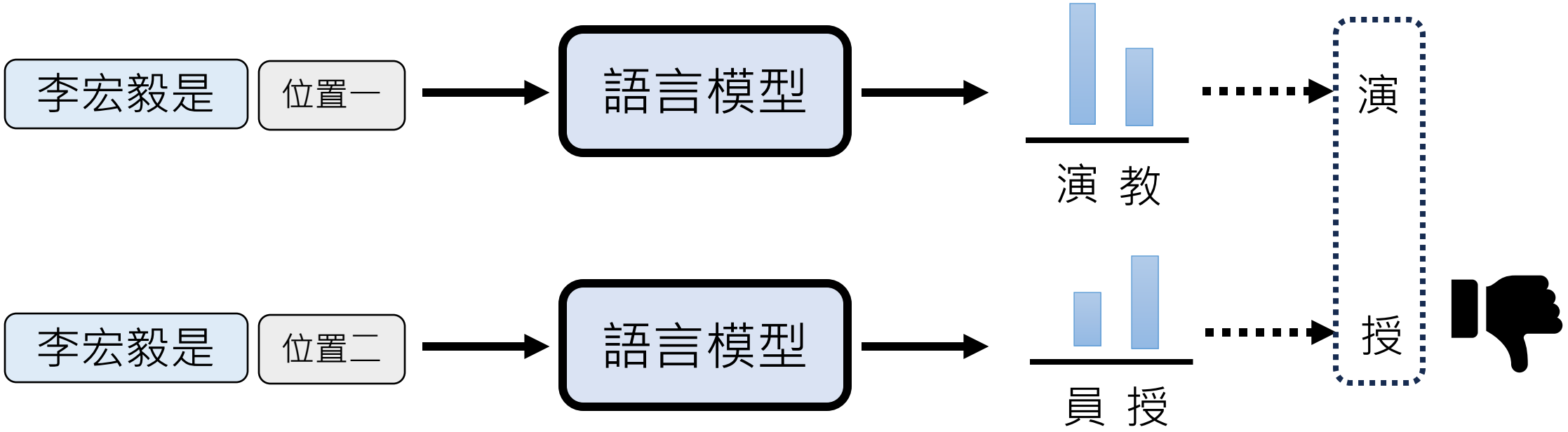
Non-Autoregressive Generation 的品質問題

- 生成往往需要AI自行腦補，給定條件仍有很多不同可能的輸出



“multi-modality problem”

Non-autoregressive Generation



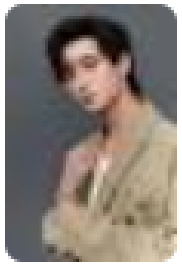
李宏毅是演員

李宏毅是演藝圈的

李宏毅是演過變形計

李宏毅是教授

李宏毅是教授



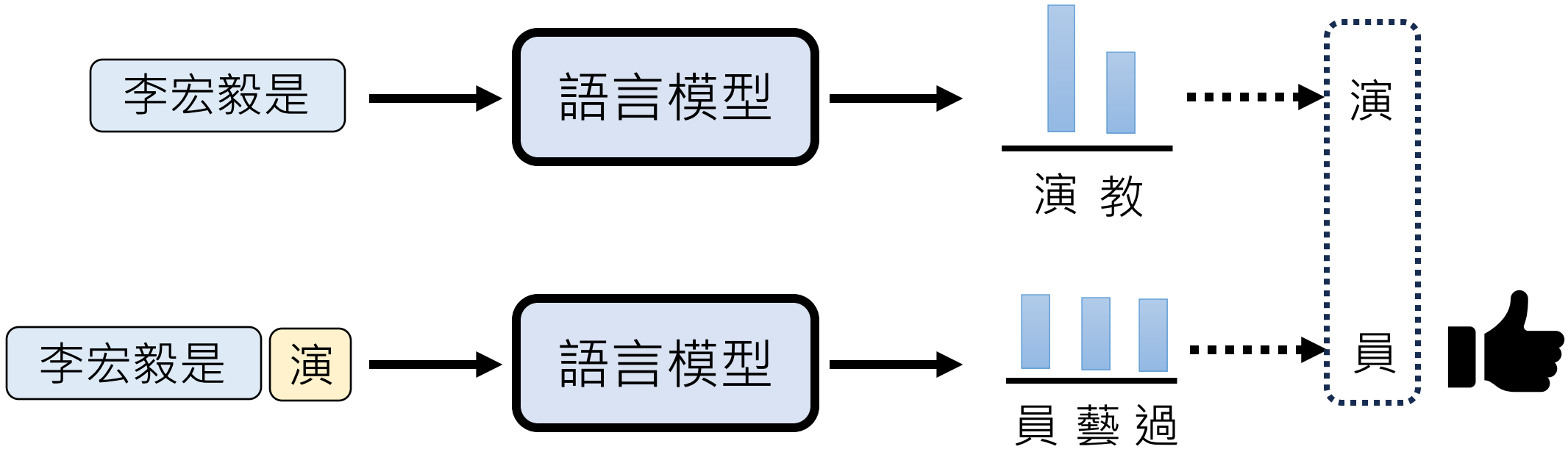
李宏毅

演員

李宏毅，男，漢族，遼寧遼陽人，中國影視演員。

2014年因參加湖南衛視真人秀節目變形計之《此間

Autoregressive Generation



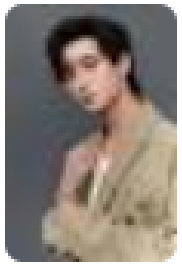
李宏毅是演員

李宏毅是演藝圈的

李宏毅是演過變形計

李宏毅是教授

李宏毅是教授



李宏毅

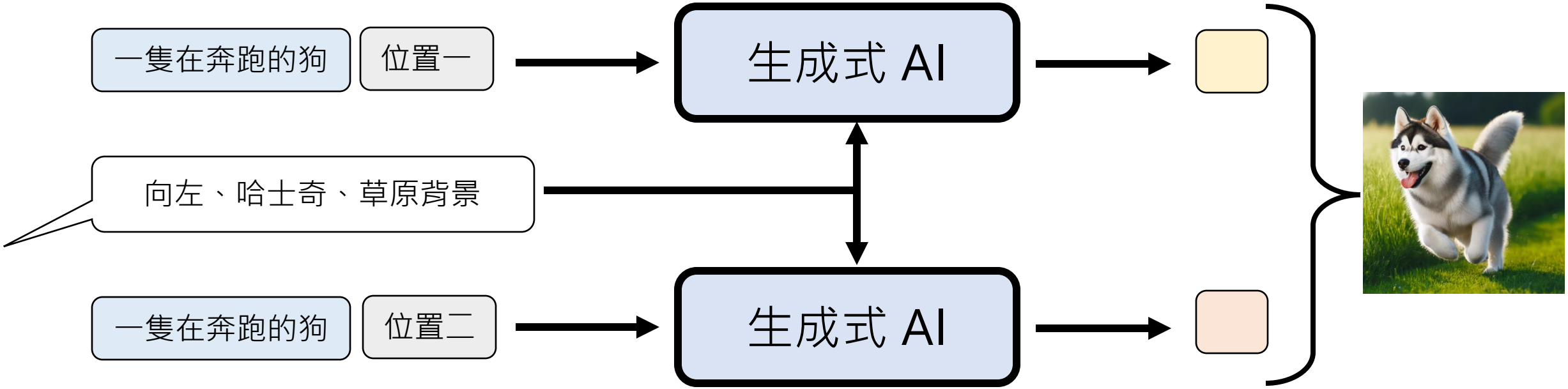
演員 ⋮

李宏毅，男，漢族，遼寧遼陽人，中國影視演員。

2014年因參加湖南衛視真人秀節目變形計之《此間

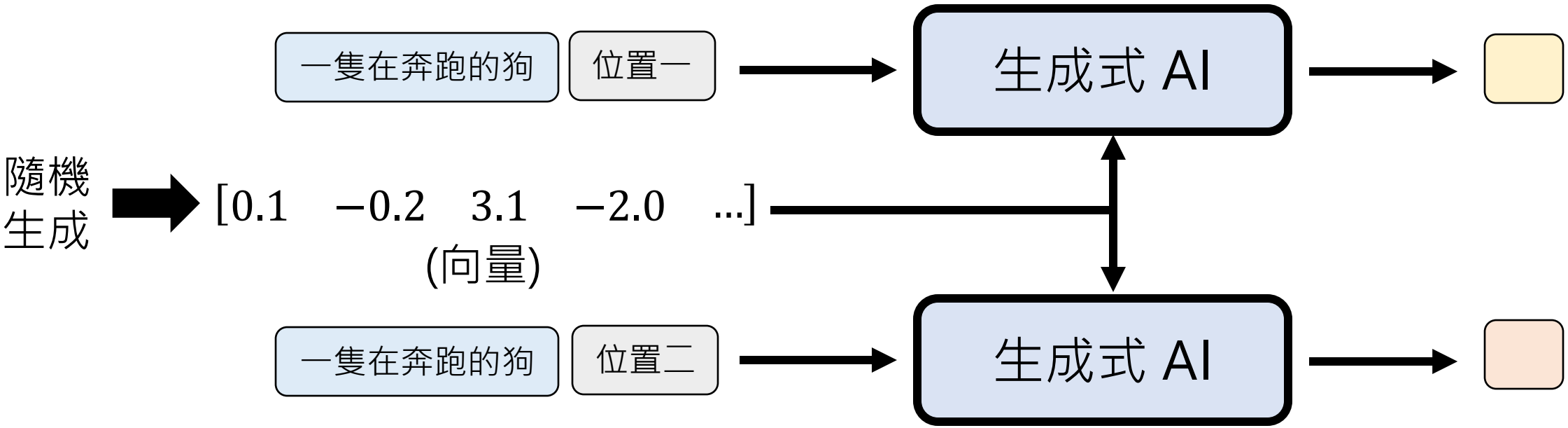
Non-Autoregressive Generation 的品質問題

- 讓所有位置都腦補一樣的內容



Non-Autoregressive Generation 的品質問題

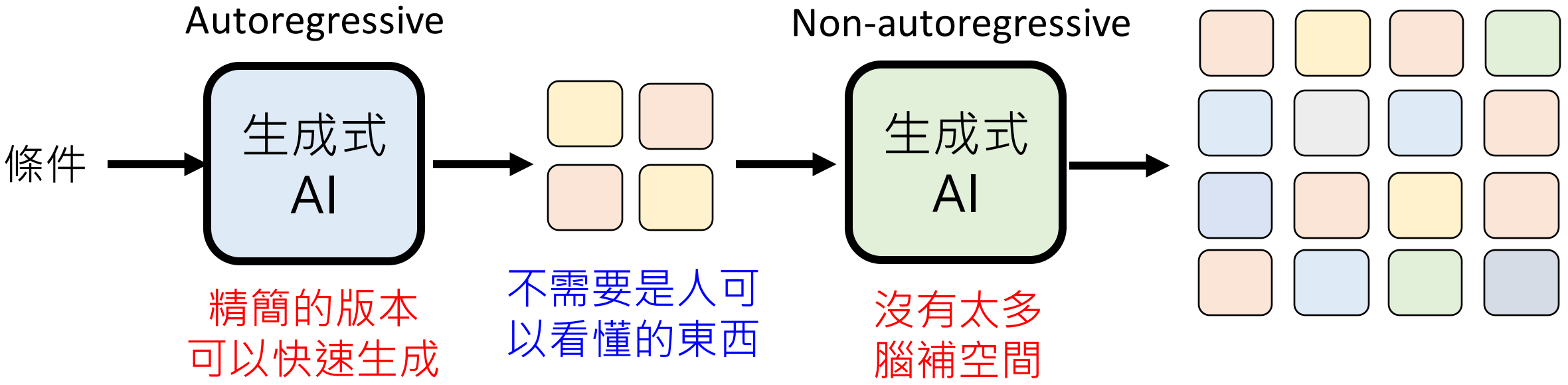
- 讓所有位置都腦補一樣的內容



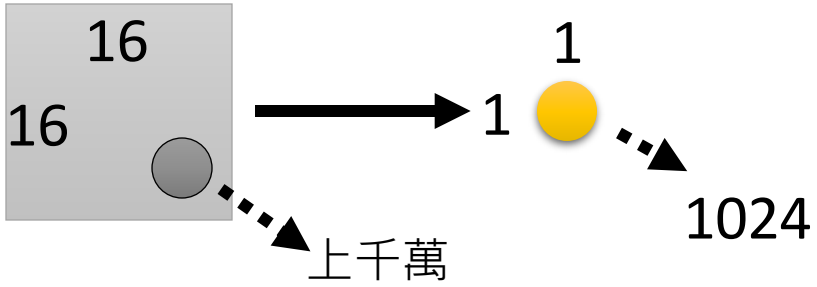
影像常用生成模型 VAE, GAN, Flow-based Model, Diffusion Model 都有這樣的設計

Autoregressive + Non-autoregressive

- 先用 Autoregressive 生成一個精簡的版本，再用 Non-autoregressive 生成產生精細的版本



Autoregressive + Non-autoregressive



https://arxiv.org/abs/2202.04200

Auto-Encoder

Autoregressive + Non-autoregressive

Encoder

2 5 4 7

Decoder

Autoregressive

虎斑貓

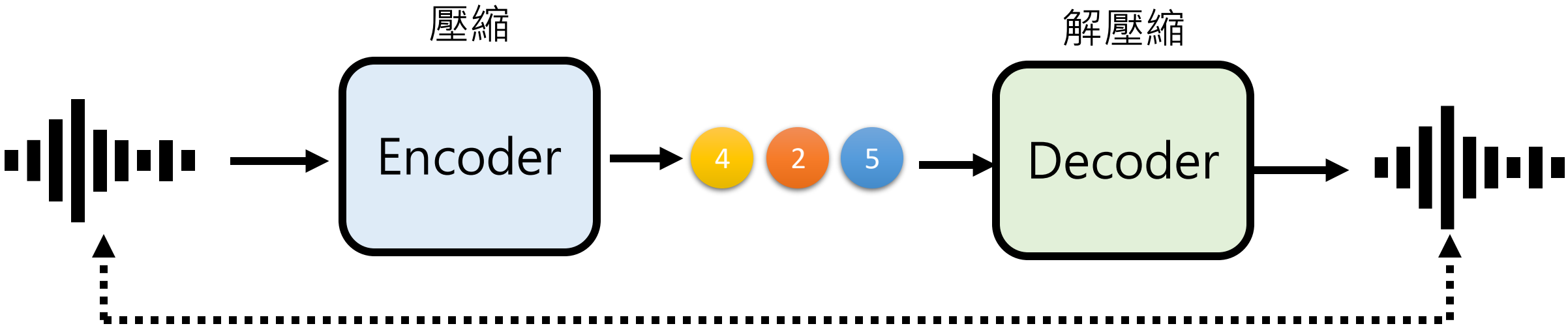
生成式 AI

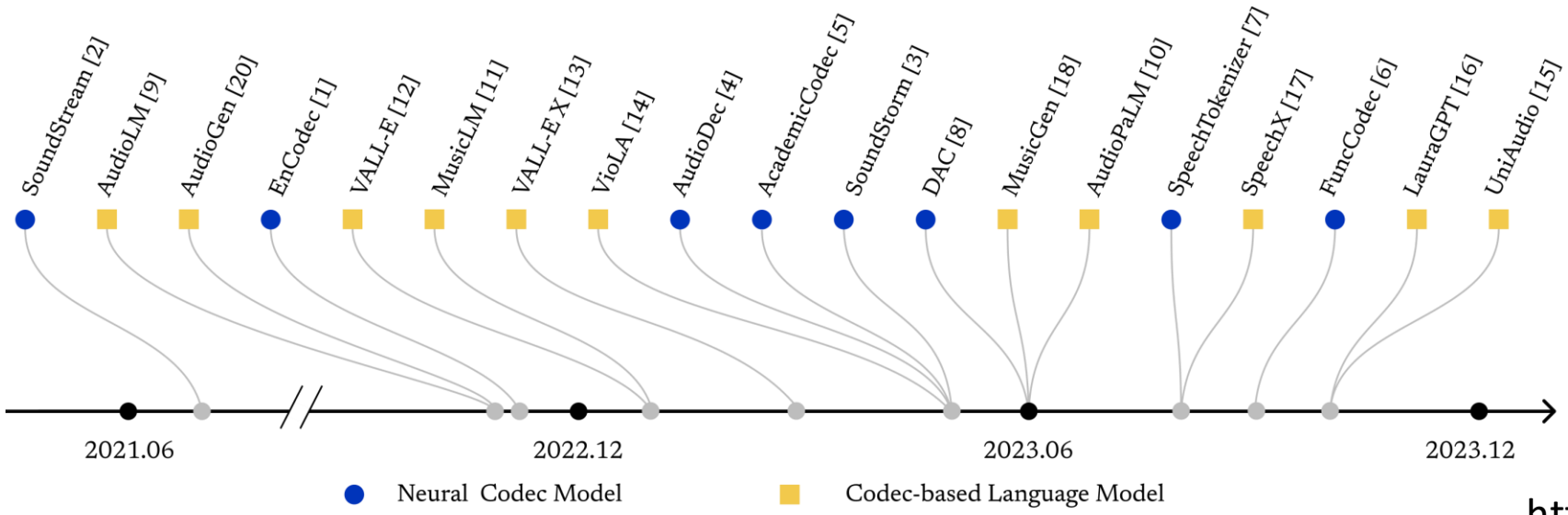
7 2 4 5

Decoder

會不會還是很久？

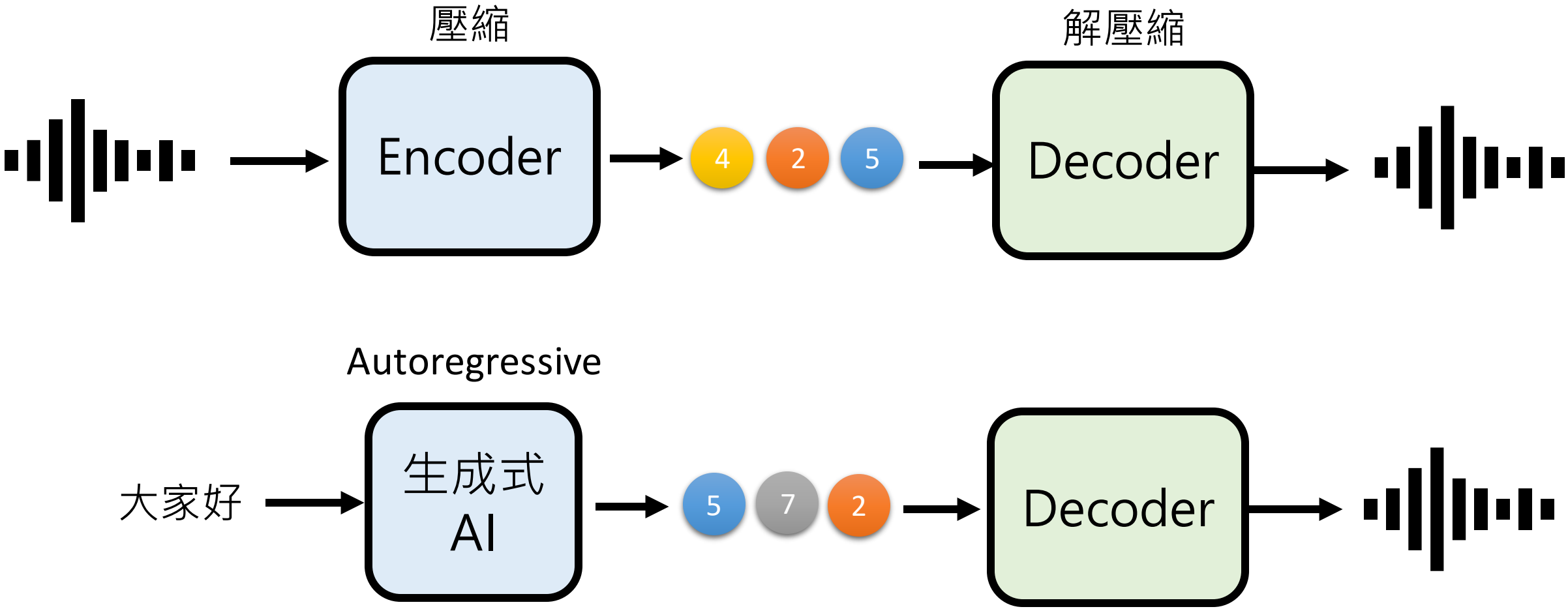
Autoregressive + Non-autoregressive





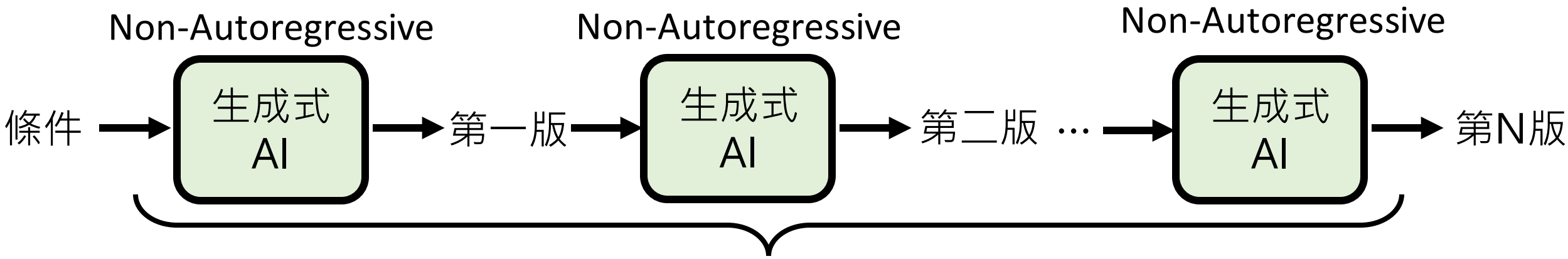
https://arxiv.org/html/2402.13236v1

Autoregressive + Non-autoregressive



VALL-E: https://arxiv.org/abs/2301.02111

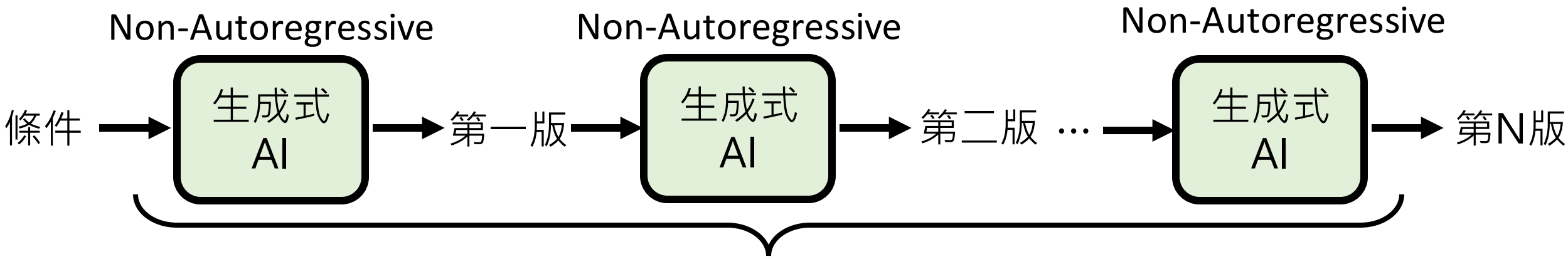
多次 Non-Autoregressive Generation



由小圖到大圖

https://arxiv.org/abs/2205.11487
https://arxiv.org/pdf/1710.10196

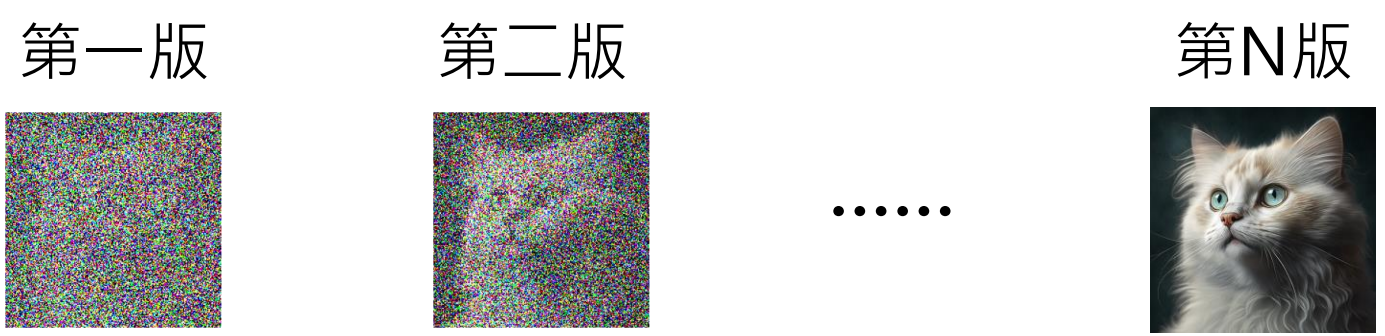
多次 Non-Autoregressive Generation



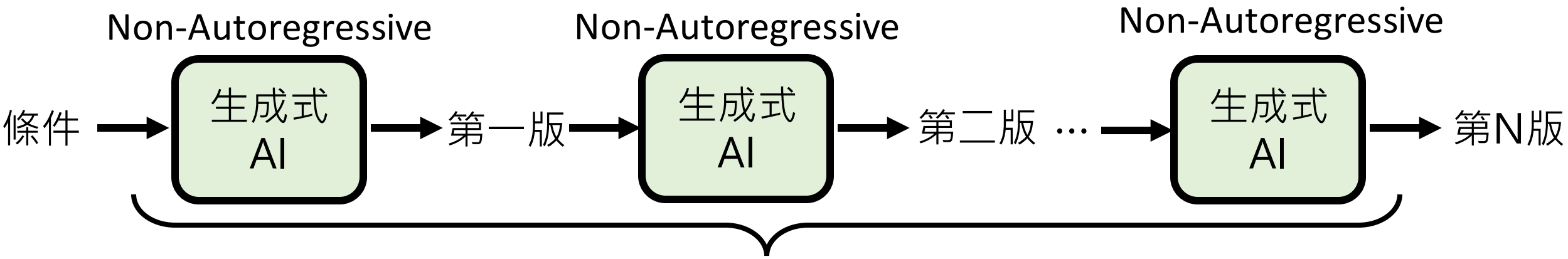
<u>從有雜訊到沒有雜訊</u>

Diffusion Model

https://arxiv.org/abs/2006.11239



多次 Non-Autoregressive Generation



也可以看作是一種 Auto-regressive Generation

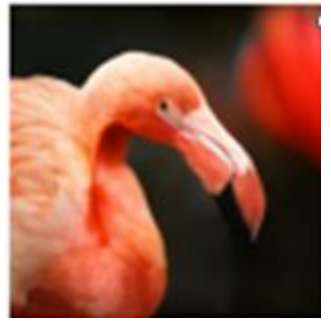
每次把生不好的地方塗掉

https://arxiv.org/abs/2202.04200

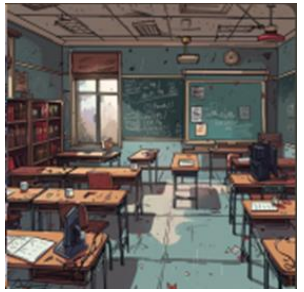
第一版

第二版

第N版



條件 → 生成式 AI → 1 2 3 4 → Decoder →



Autoregressive

Non-Autoregressive

條件 → 生成式 AI → 第一版 (1 1 1 1) → Decoder →

Non-Autoregressive

生成式 AI → 第二版 (1 2 2 1) … → Decoder →

Non-Autoregressive

生成式 AI → 第N版 (1 2 3 4) → Decoder →

Midjourney, Stable Diffusion, DALL-E 都是用類似的作法

小結

	Autoregressive, AR	Non-autoregressive, NAR
特性	按部就班、各個擊破	齊頭並進、一次到位
速度		勝
品質	勝	
應用	常用於文字	常用於影像

有很多方法讓兩種策略可以截長補短

覺得現在語言模型還不夠快嗎？

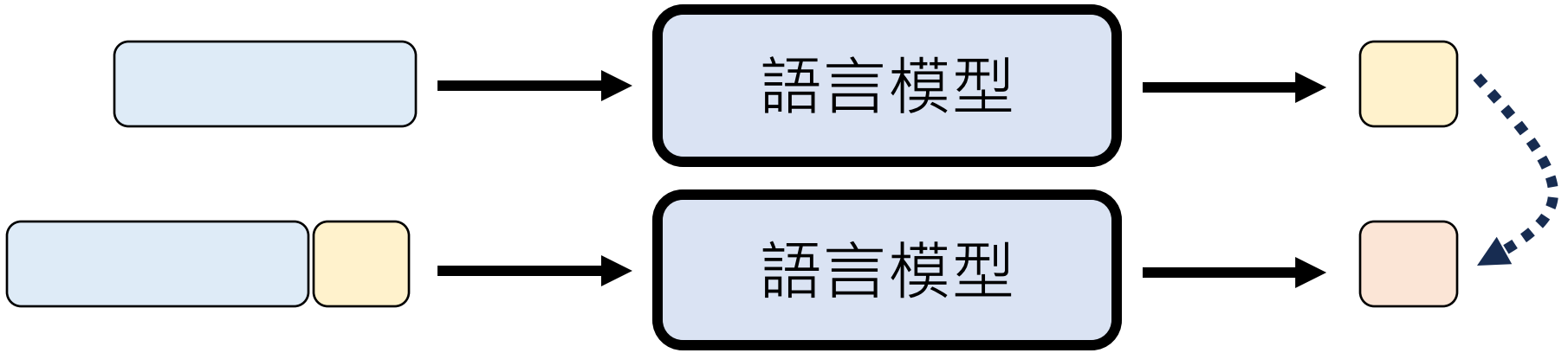
LLaMA 2 13B

Source: https://pytorch.org/blog/hitchhikers-guide-speculative-decoding/

Speculative Decoding

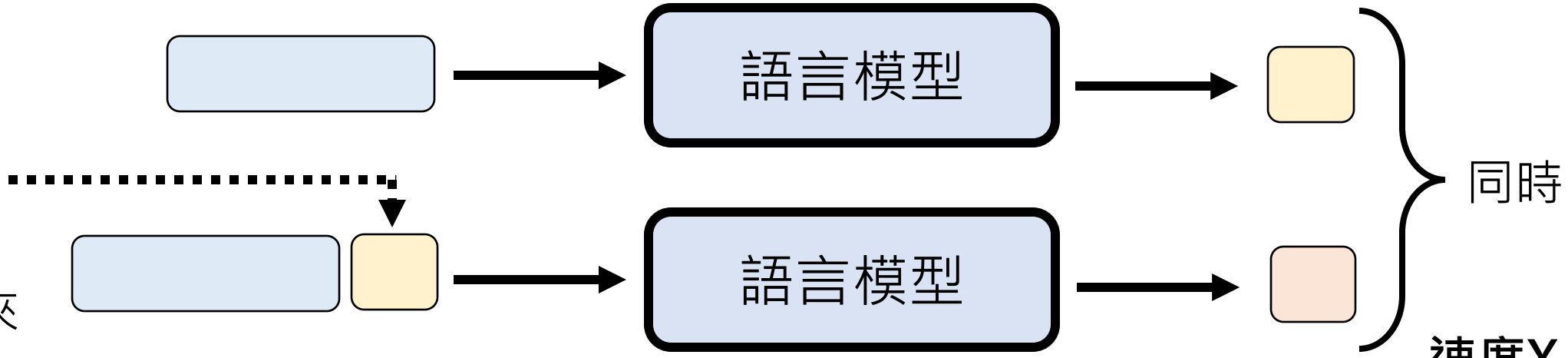
猜測、投機

https://arxiv.org/abs/2211.17192
https://arxiv.org/abs/2302.01318

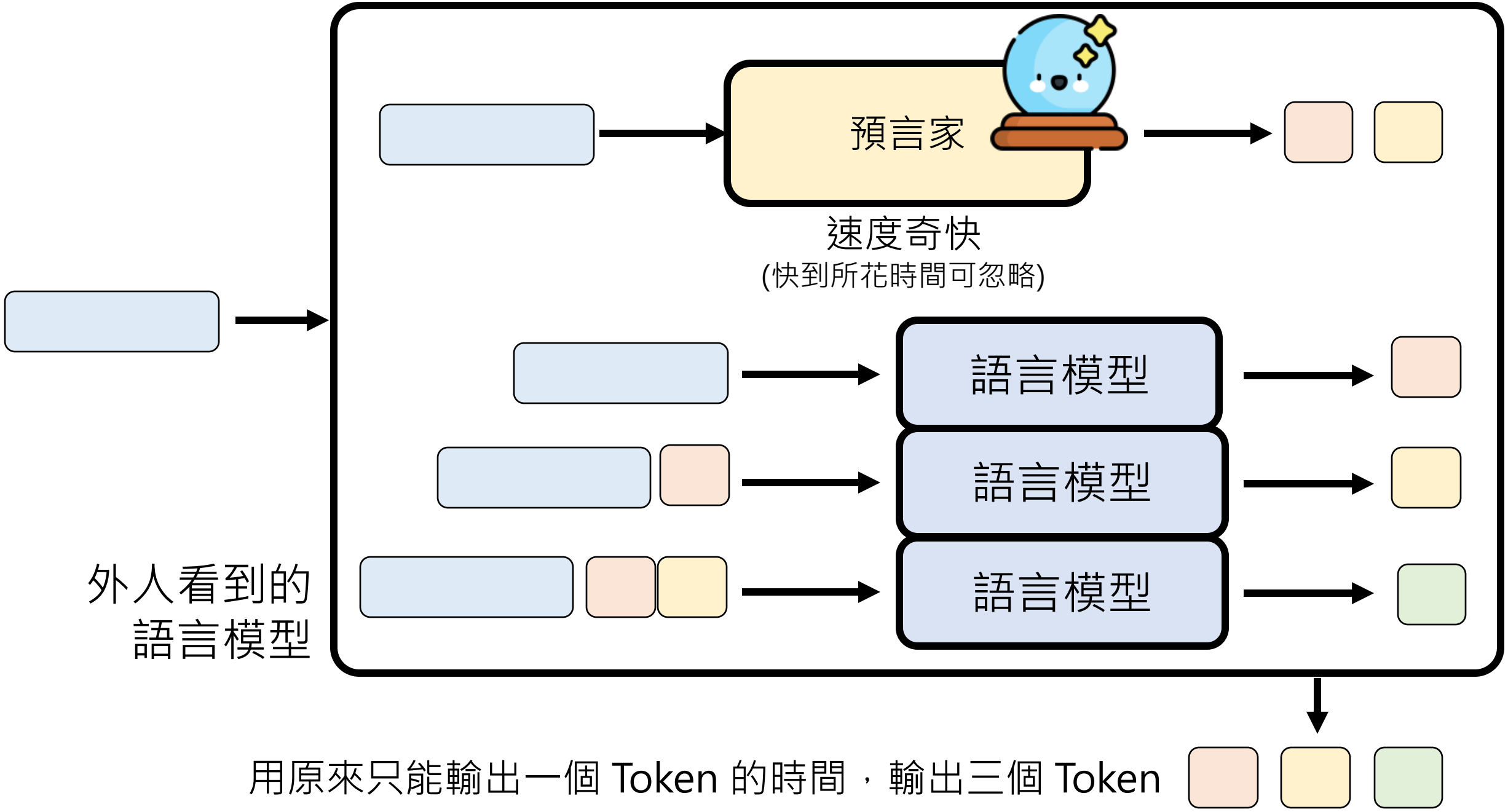


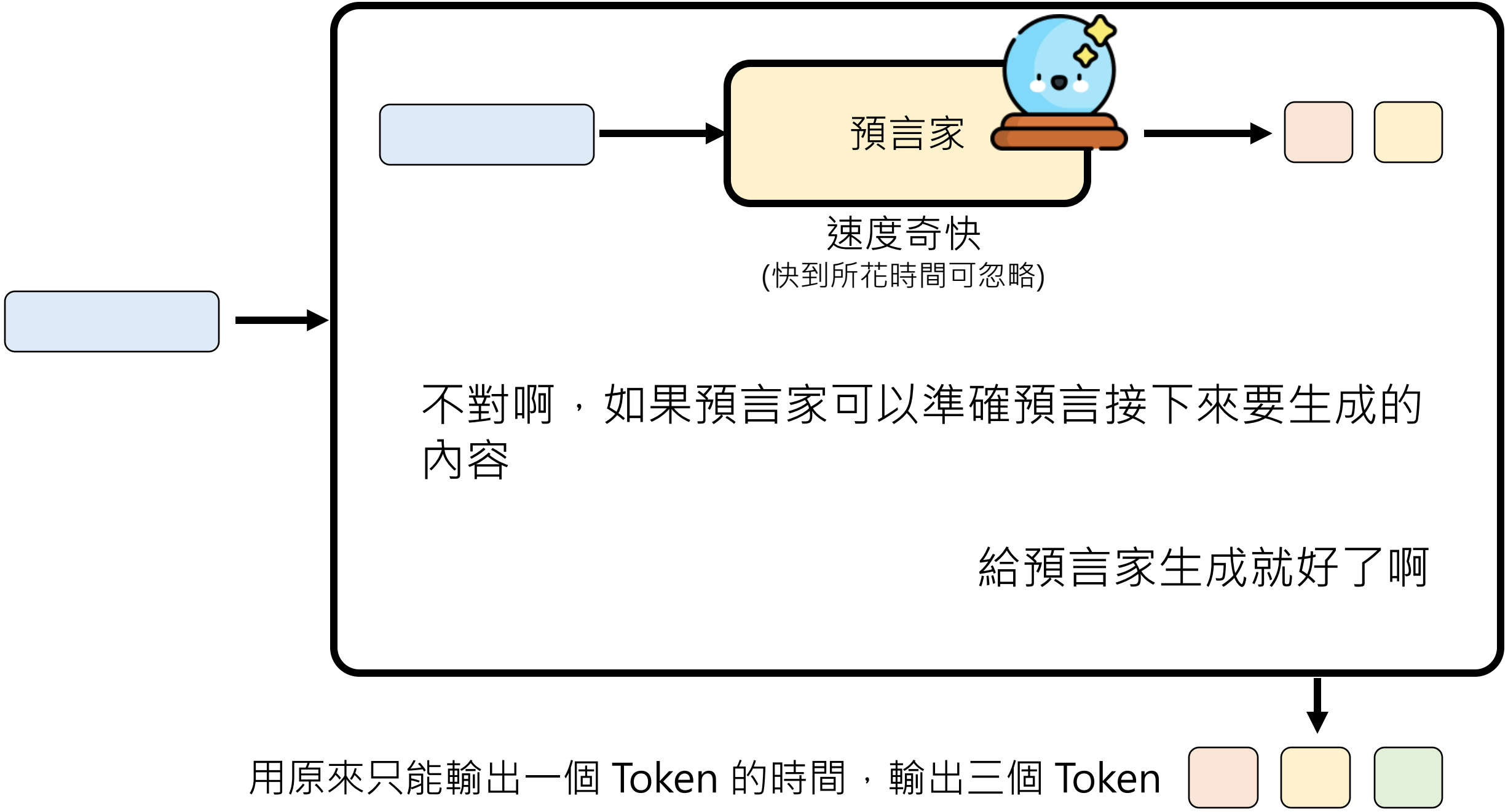
預言家

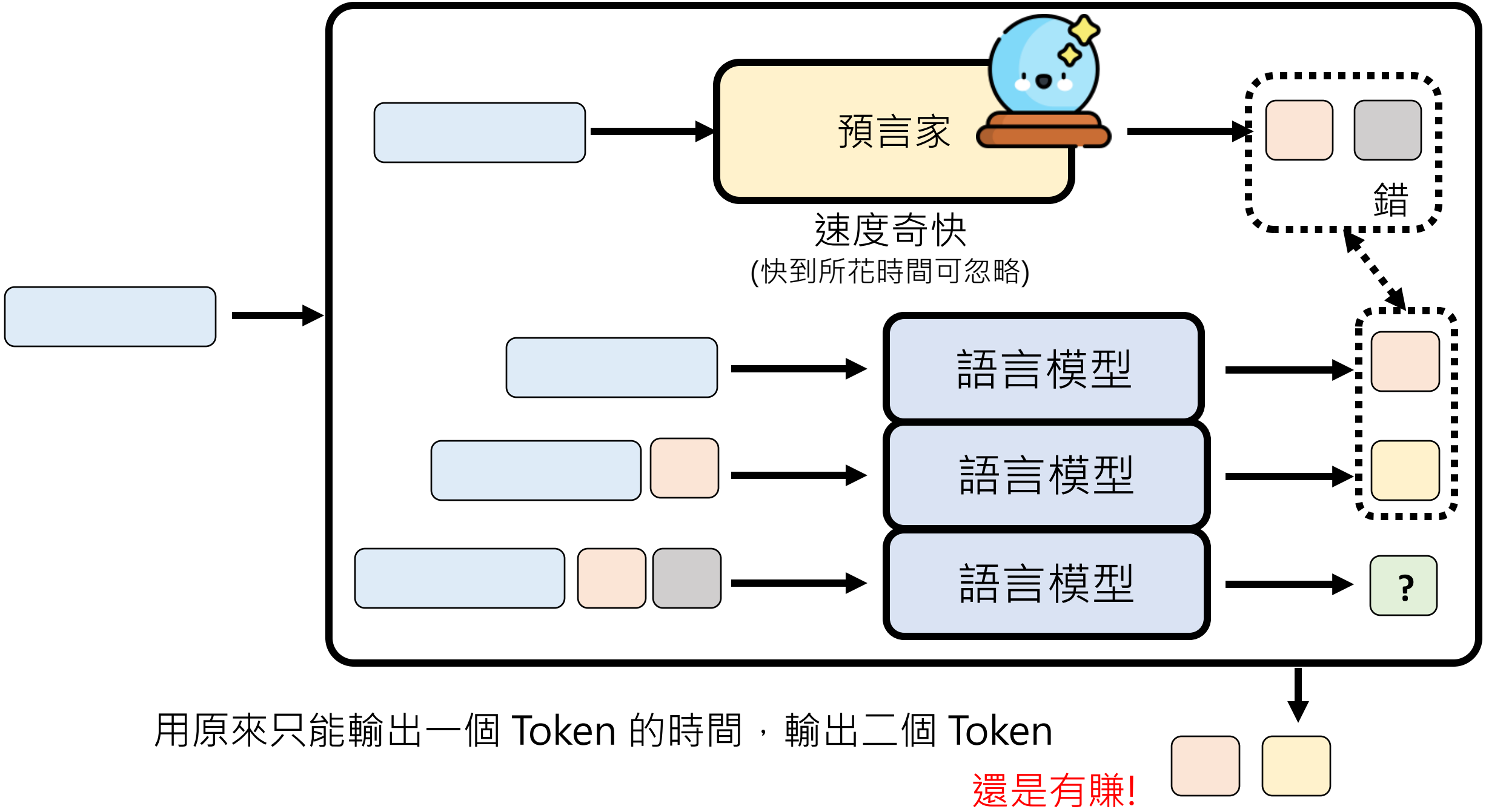
預判接下來
會說什麼



速度X 2

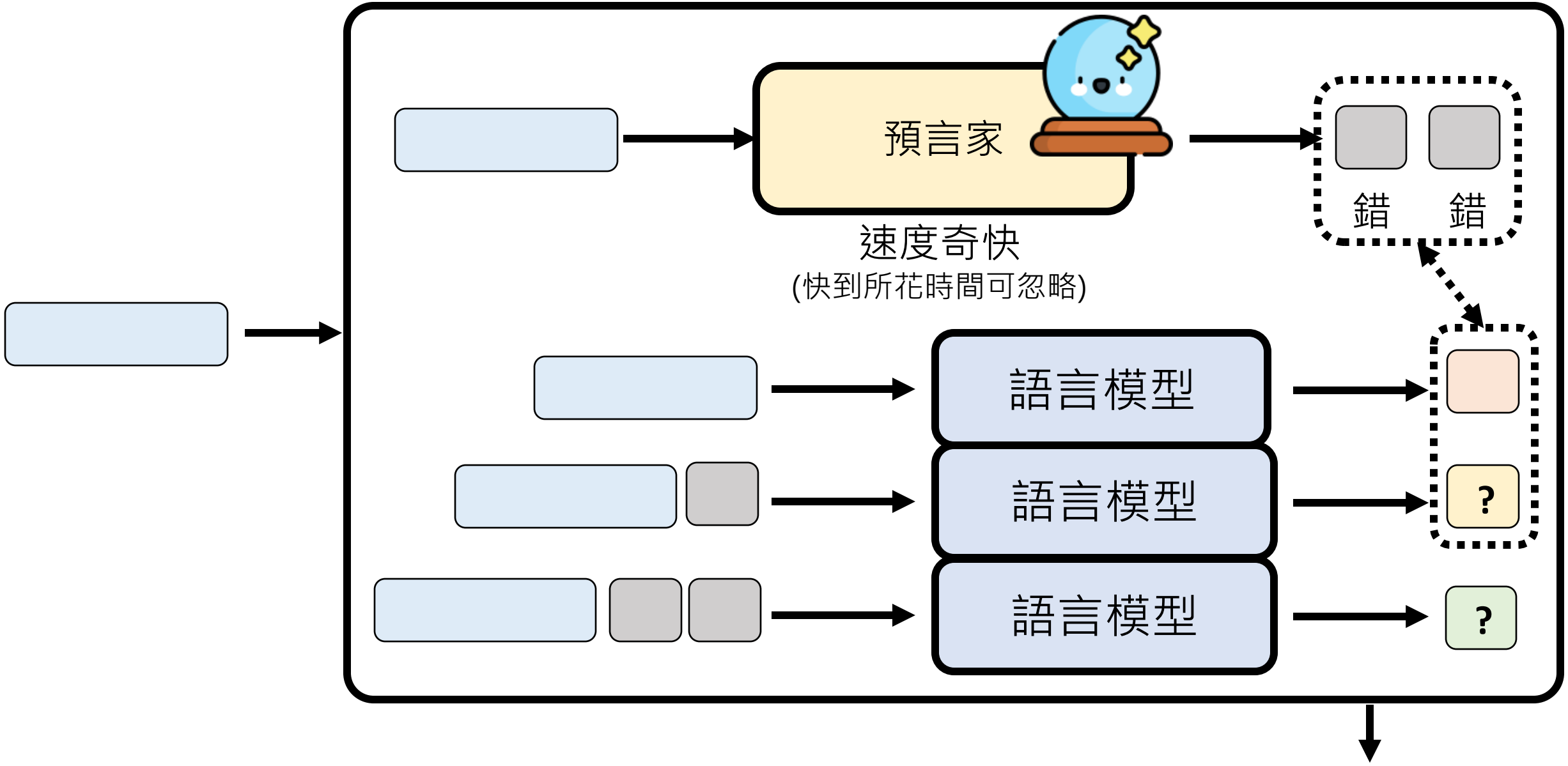

預言家
速度奇快
(快到所花時間可忽略)
語言模型
語言模型
語言模型
外人看到的
語言模型
用原來只能輸出一個 Token 的時間，輸出三個 Token


預言家
速度奇快
(快到所花時間可忽略)
不對啊，如果預言家可以準確預言接下來要生成的內容
給預言家生成就好了啊
用原來只能輸出一個 Token 的時間，輸出三個 Token



用原來只能輸出一個 Token 的時間，輸出二個 Token

還是有賺!



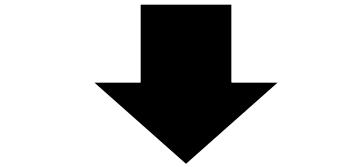
幾乎不過不失 (損失的時間為預言家預言的時間、無謂的運算資源耗損)

Speculative Decoding

要求：超快速、犯錯沒關係

Non-autoregressive Model

Compressed Model



壓縮過的小模型

Speculative Decoding

- 預言家一定要是語言模型嗎？

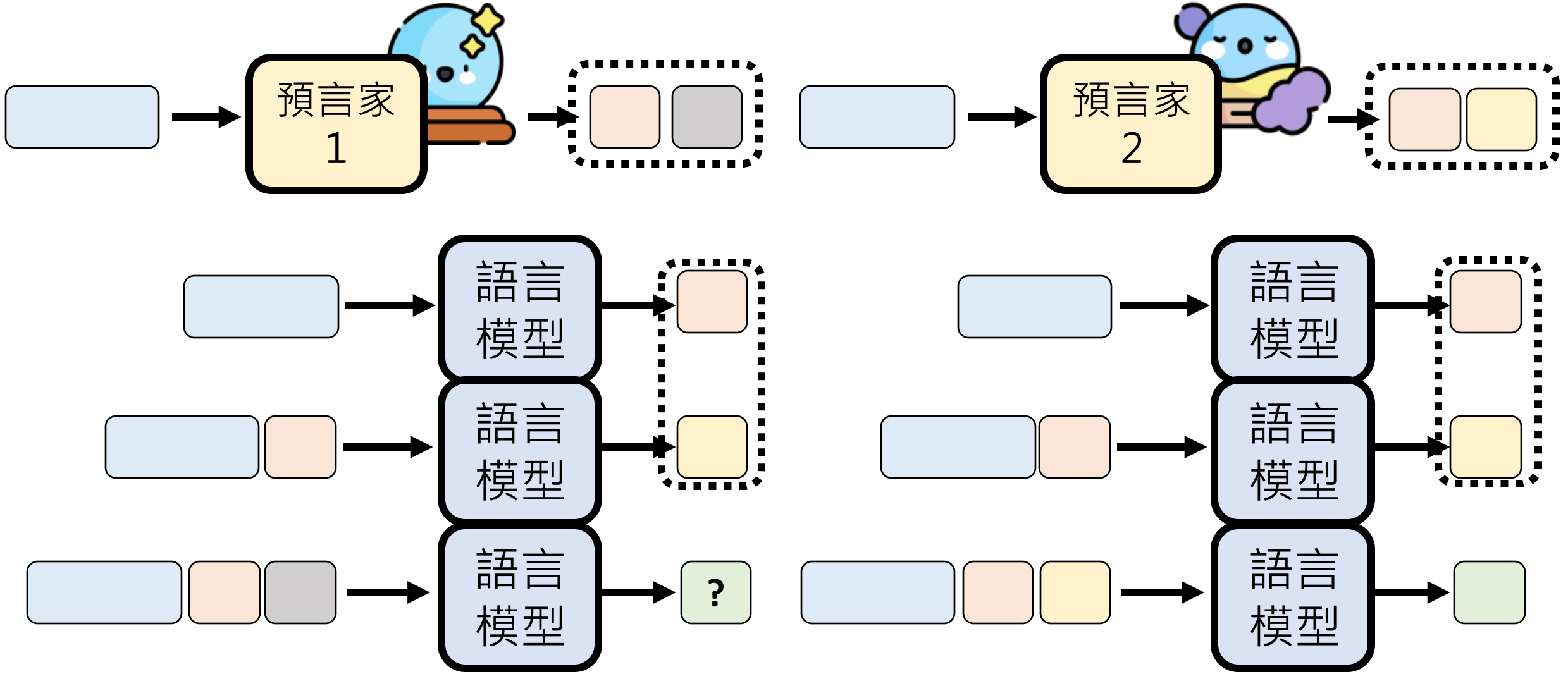
要求：超快速、犯錯沒關係

語言模型 → 搜尋引擎 → 經常使用在許多自然語言處理方面的應用

https://arxiv.org/abs/2304.04487

https://arxiv.org/abs/2305.09781

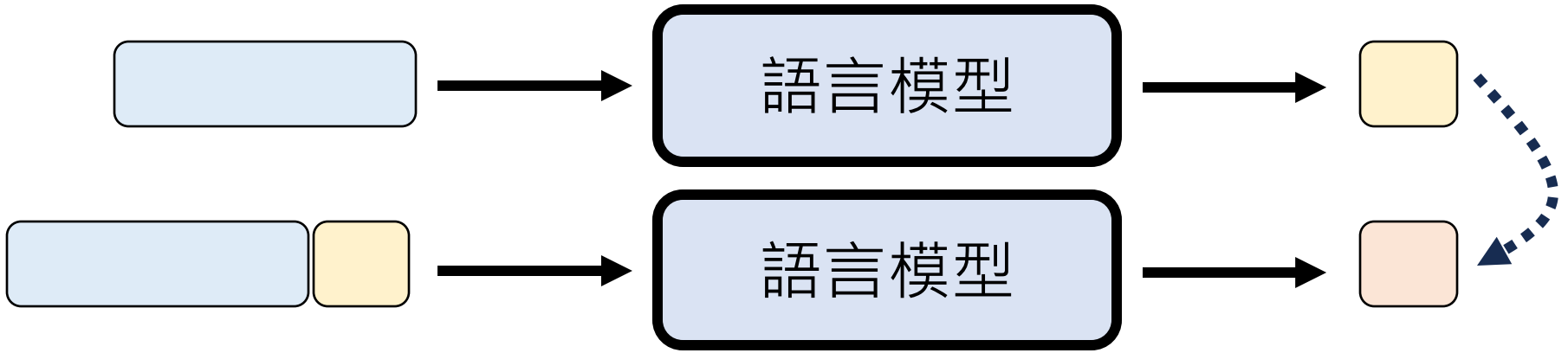
Speculative Decoding：多個預言家



Speculative Decoding

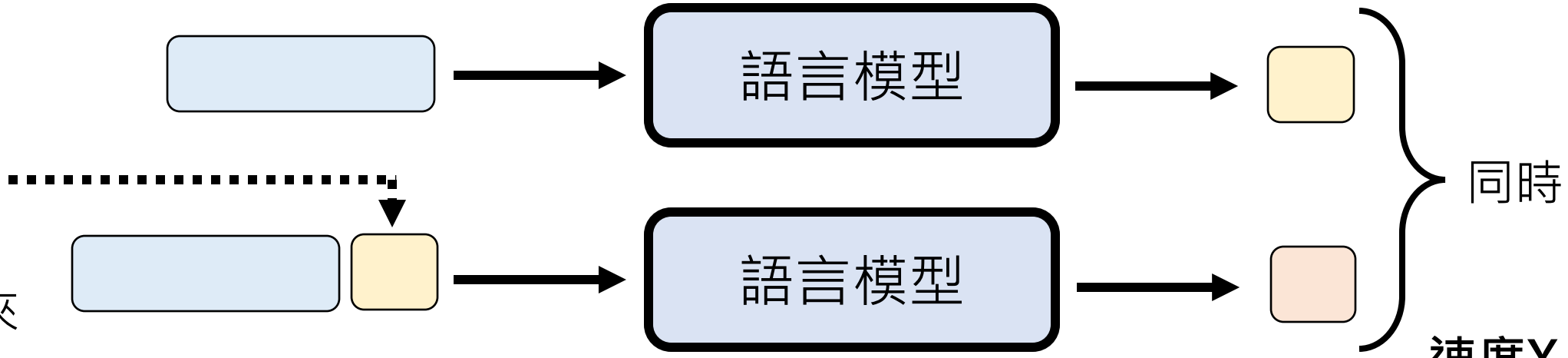
猜測、投機

https://arxiv.org/abs/2211.17192
https://arxiv.org/abs/2302.01318



預言家

預判接下來
會說什麼



速度X 2