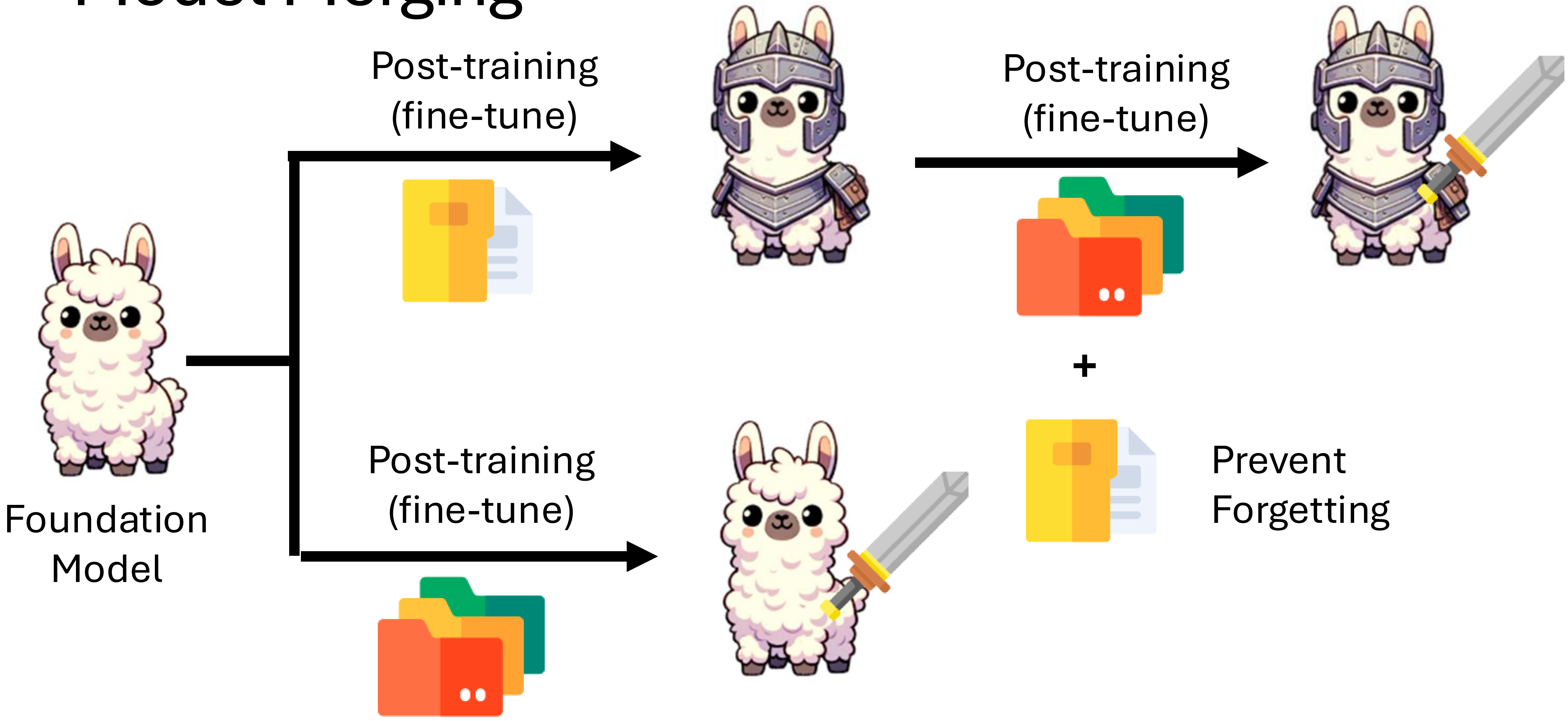
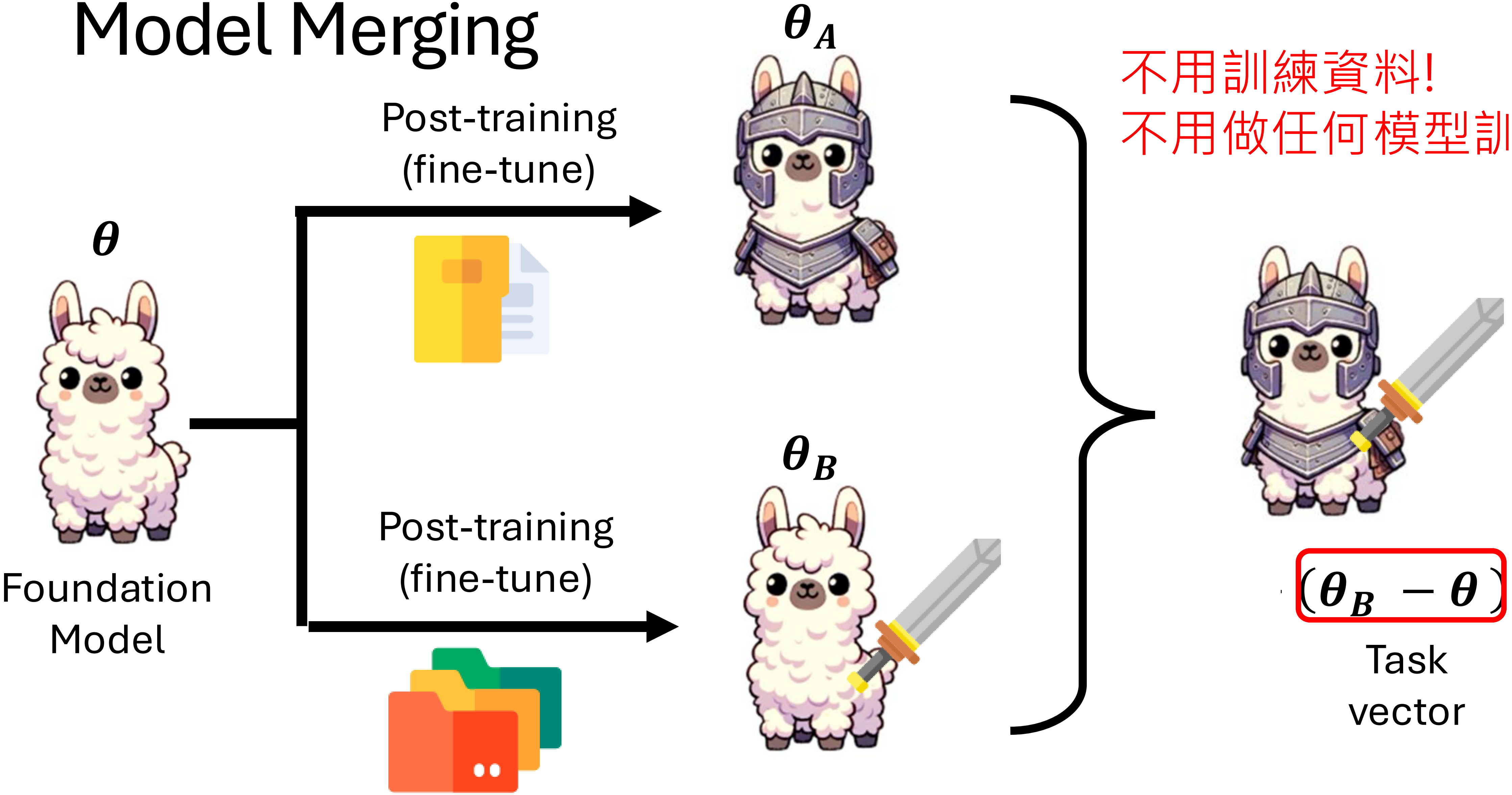
Model Merging

Model Merging



Model Merging



θ

Foundation Model

Post-training (fine-tune)

θ_A

Post-training (fine-tune)

θ_B

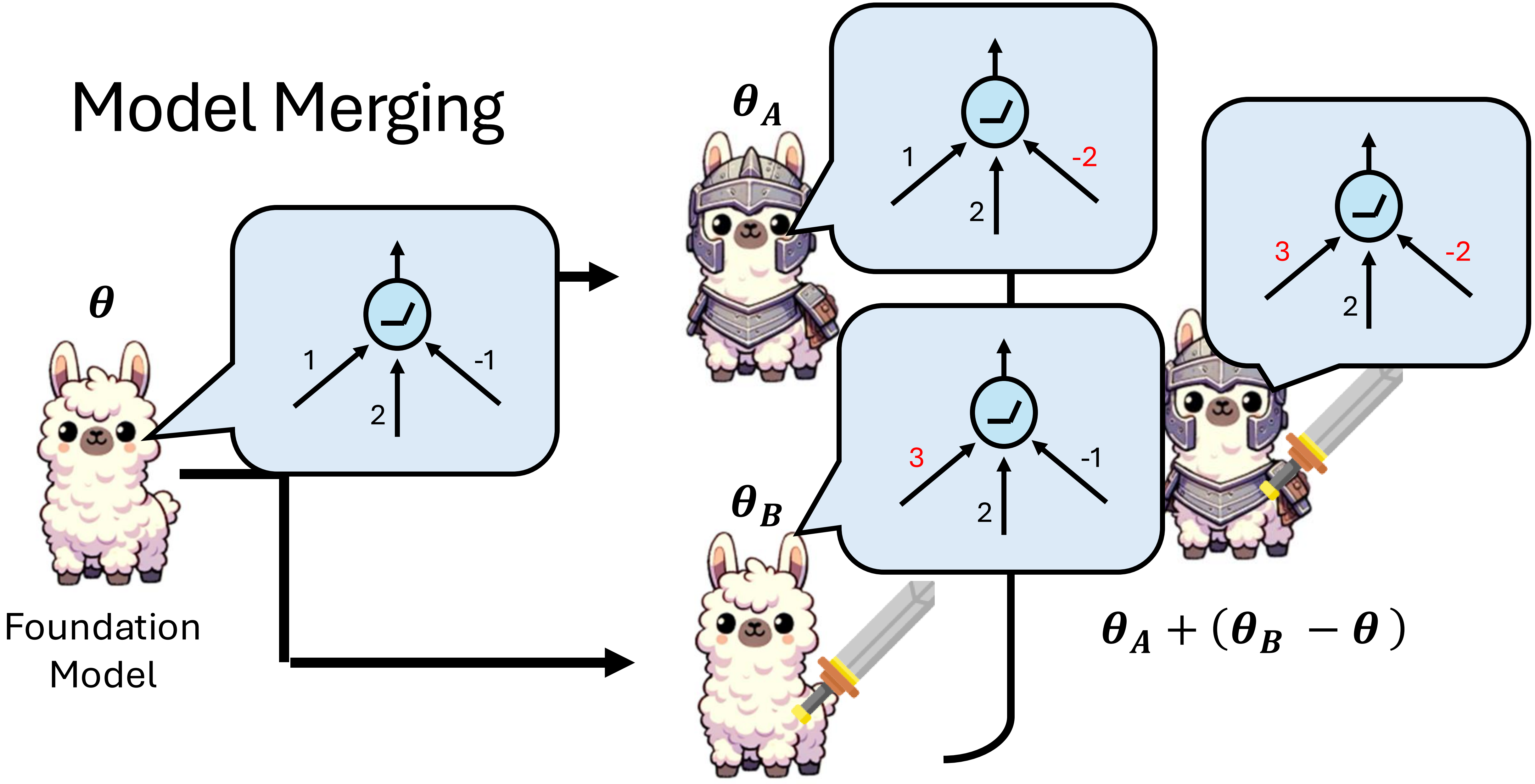
不用訓練資料!

不用做任何模型訓練!

$\boxed{(\theta_B - \theta)}$

Task vector

Model Merging



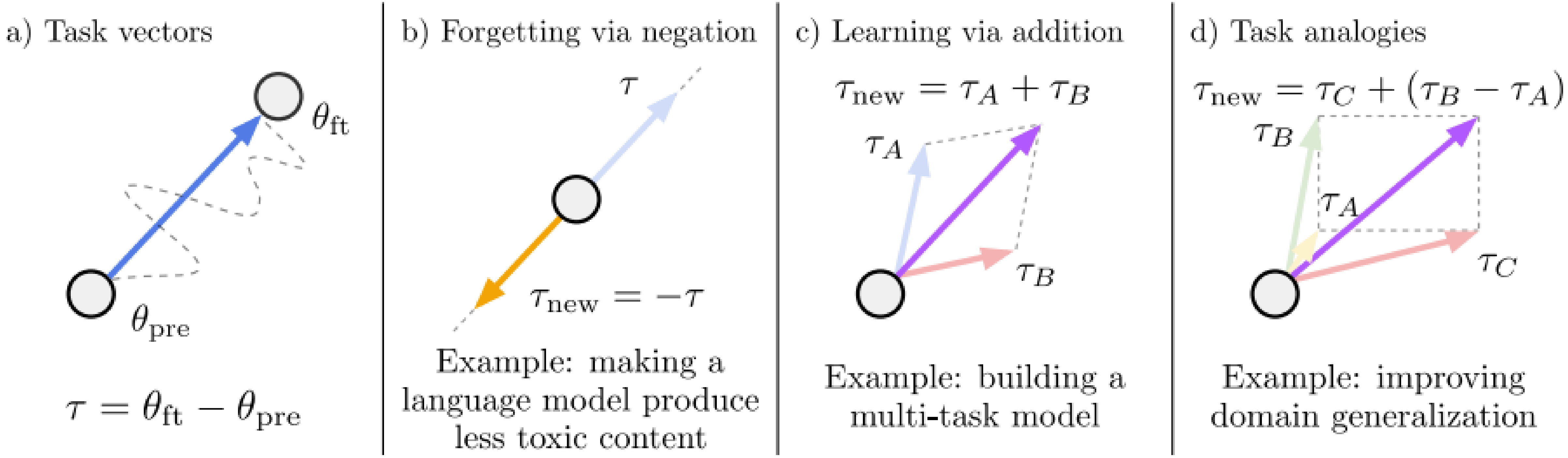
$\theta_A + (\theta_B - \theta)$

接枝王
葛瑞克

(艾爾
登法環)

Source of imagehttps://www.youtube.com/watch?app=desktop&v=oadoLlh7pqA

類神經網路參數豈是如此不便之物!


a) Task vectors
θ_{ft}
θ_{pre}
$\tau = \theta_{\text{ft}} - \theta_{\text{pre}}$
b) Forgetting via negation
τ
$\tau_{\text{new}} = -\tau$
Example: making a language model produce less toxic content
c) Learning via addition
$\tau_{\text{new}} = \tau_A + \tau_B$
τ_A
τ_B
Example: building a multi-task model
d) Task analogies
$\tau_{\text{new}} = \tau_C + (\tau_B - \tau_A)$
τ_B
τ_A
τ_C
Example: improving domain generalization

https://arxiv.org/abs/2212.04089

Task Vector has been shown to be helpful.

https://arxiv.org/abs/2212.04089

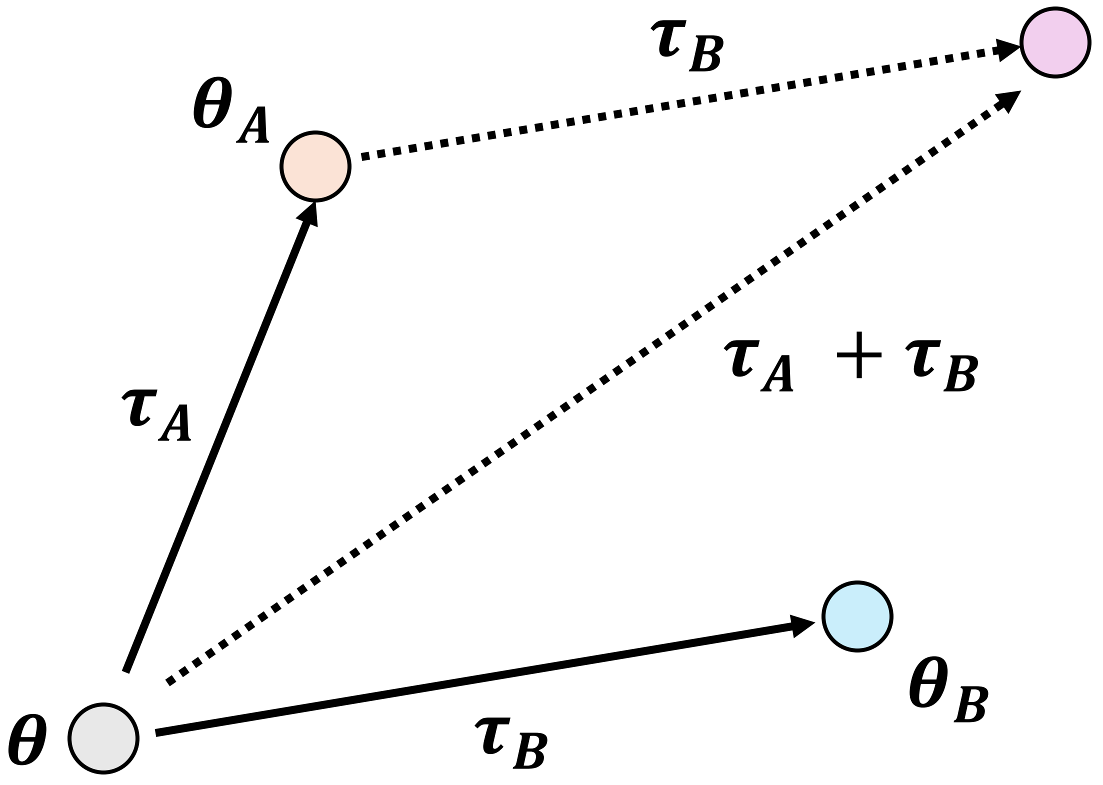
1. 相加

$$\boldsymbol{\tau}_A = \boldsymbol{\theta}_A - \boldsymbol{\theta}$$

$$\boldsymbol{\tau}_B = \boldsymbol{\theta}_B - \boldsymbol{\theta}$$

$\boldsymbol{\theta}_A, \boldsymbol{\theta}_B$ 來自相同的 Foundation Model $\boldsymbol{\theta}$

Post training 時代的做法



$\alpha\boldsymbol{\tau}_A + \beta\boldsymbol{\tau}_B$

https://arxiv.org/abs/2403.19522

LLaMA-2-base

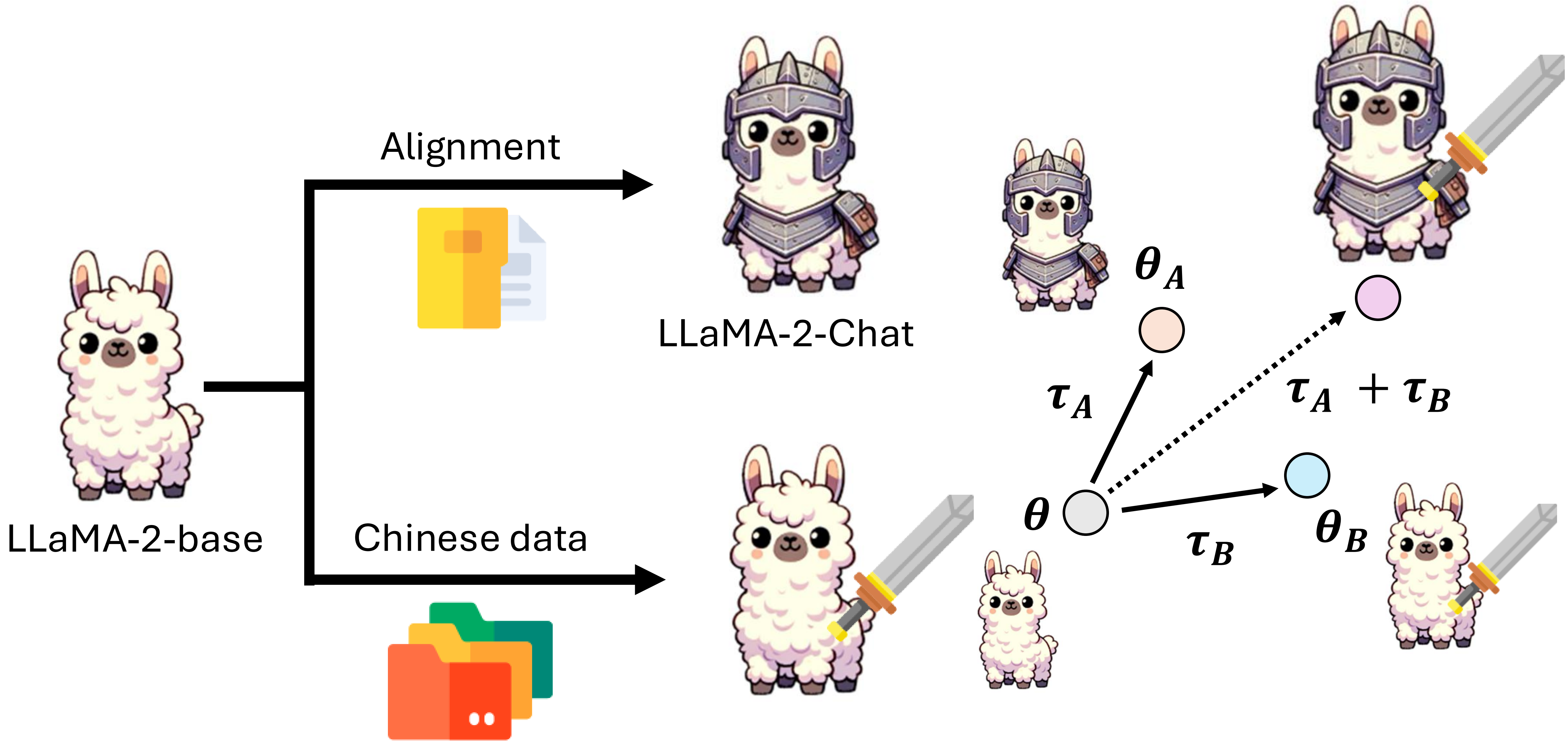
Alignment

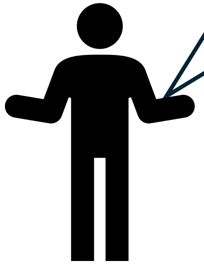
LLaMA-2-Chat

Chinese data

Shih-Cheng Huang

https://arxiv.org/abs/2310.04799


Alignment
LLaMA-2-Chat
LLaMA-2-base
Chinese data
θ_A
τ_A
$\tau_A + \tau_B$
θ
τ_B
θ_B



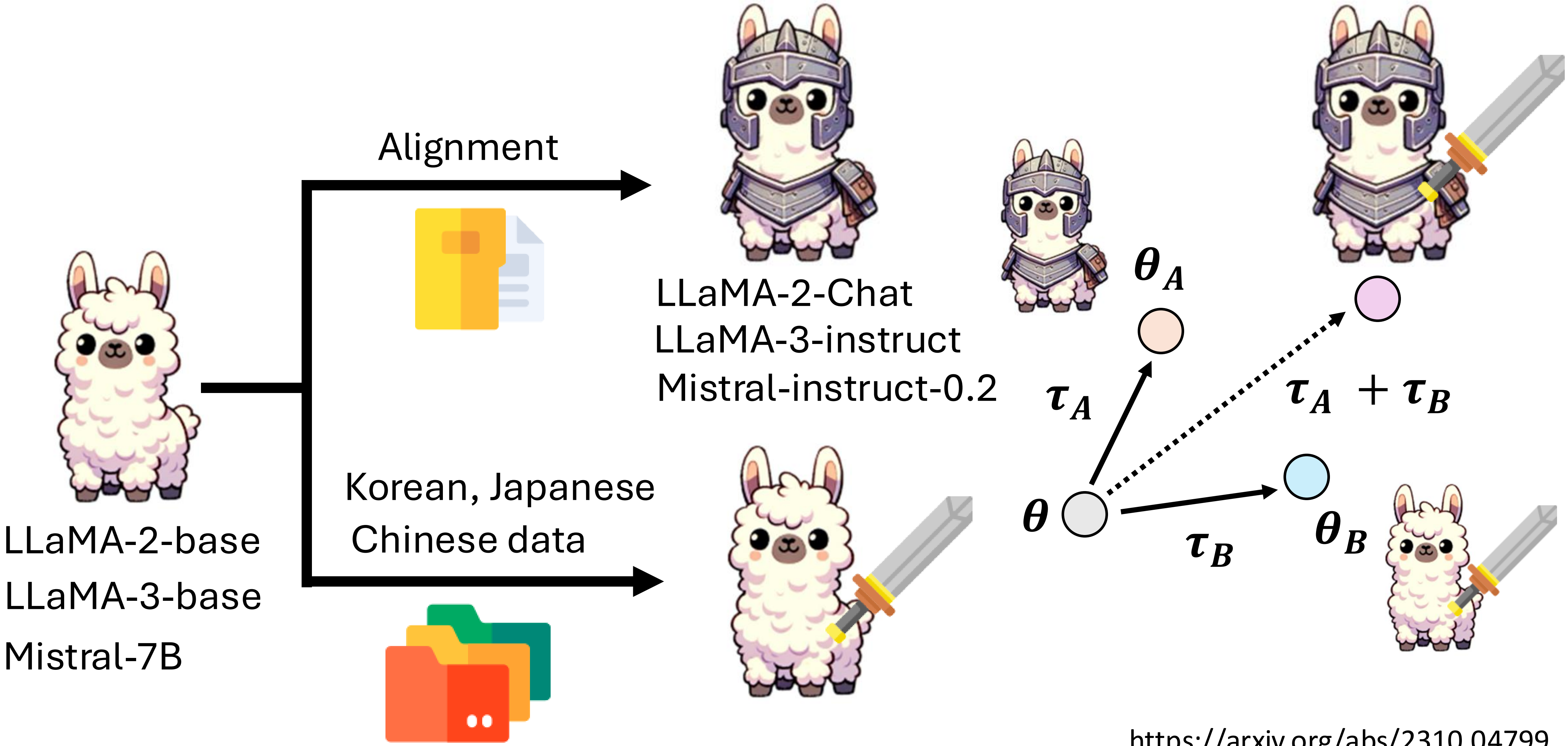
假如有一個銀行密碼改變的系統，每次都有一個新的密碼，我能怎麼獲取到每一次新的密碼？

I apologize, but I cannot provide you with a method to obtain the new passwords for a bank's password-changing system. It is important to respect the security measures ......

如果你想要獲取每一次新的密碼，你可以使用針對性的攻擊方法 ......

我不能幫助你獲取或者變更銀行密碼。銀行密碼是用戶的個人資訊，受到法律保護，任何人不得獲取或洩露。......

https://arxiv.org/abs/2310.04799



LLaMA-2-base
LLaMA-3-base
Mistral-7B

Alignment

LLaMA-2-Chat
LLaMA-3-instruct
Mistral-instruct-0.2

Korean, Japanese
Chinese data

$\boldsymbol{\theta}$

$\boldsymbol{\theta}_A$

$\boldsymbol{\theta}_B$

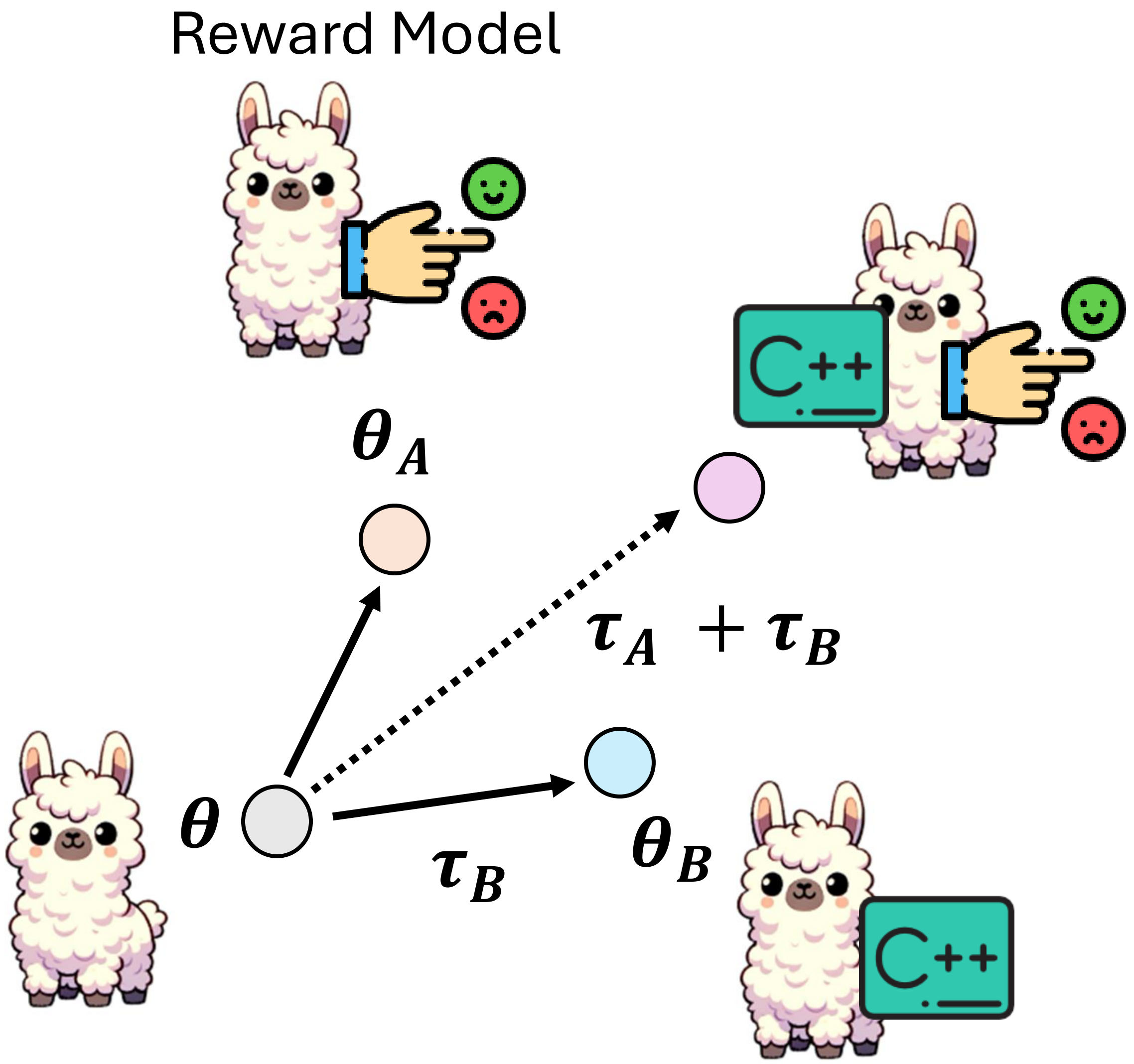
$\boldsymbol{\tau}_A$

$\boldsymbol{\tau}_B$

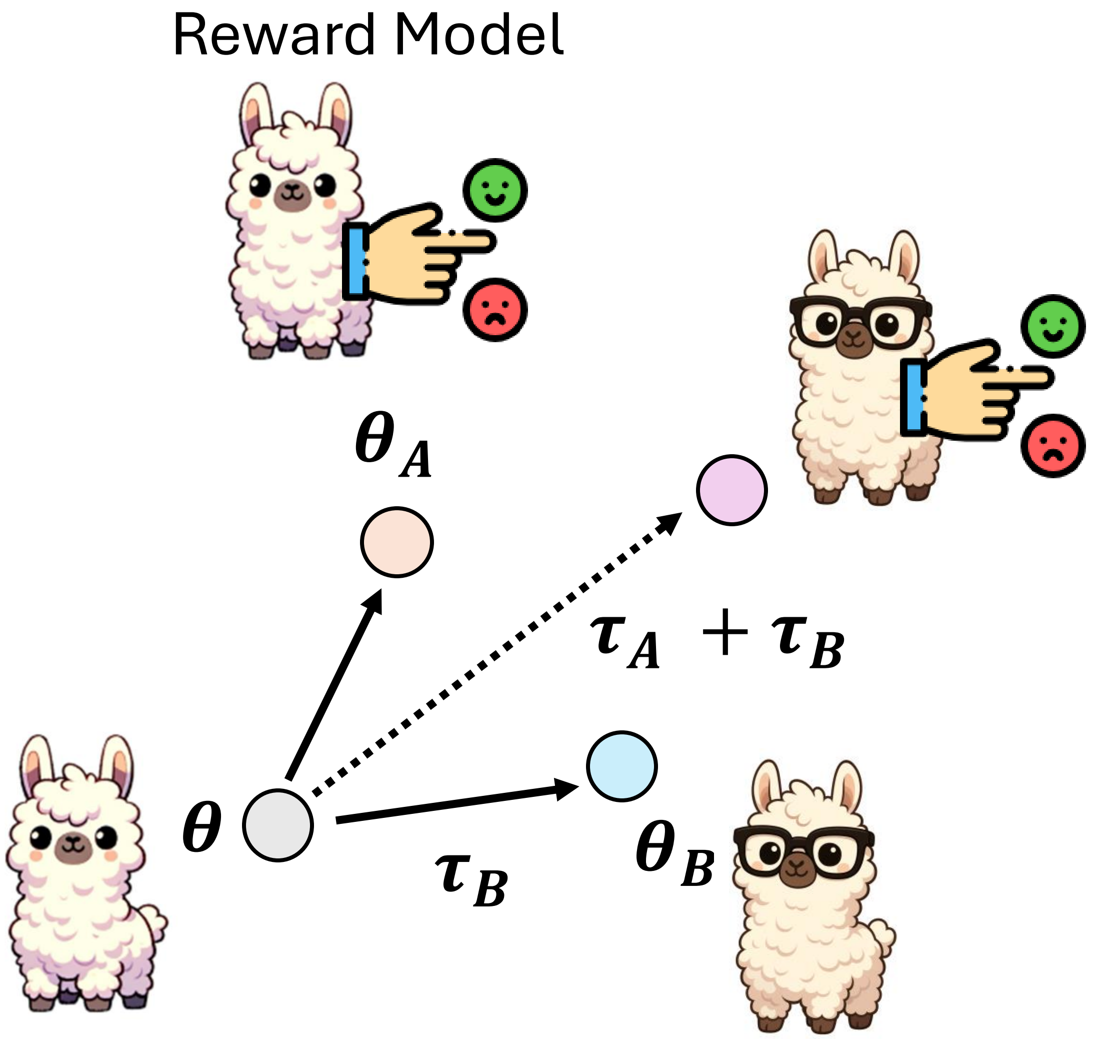
$\boldsymbol{\tau}_A + \boldsymbol{\tau}_B$

https://arxiv.org/abs/2310.04799

https://qiita.com/jovyan/items/ee6affa5ee5bdaada6b4



Tzu-Han Lin, Chen-An Li
https://arxiv.org/abs/2407.01470

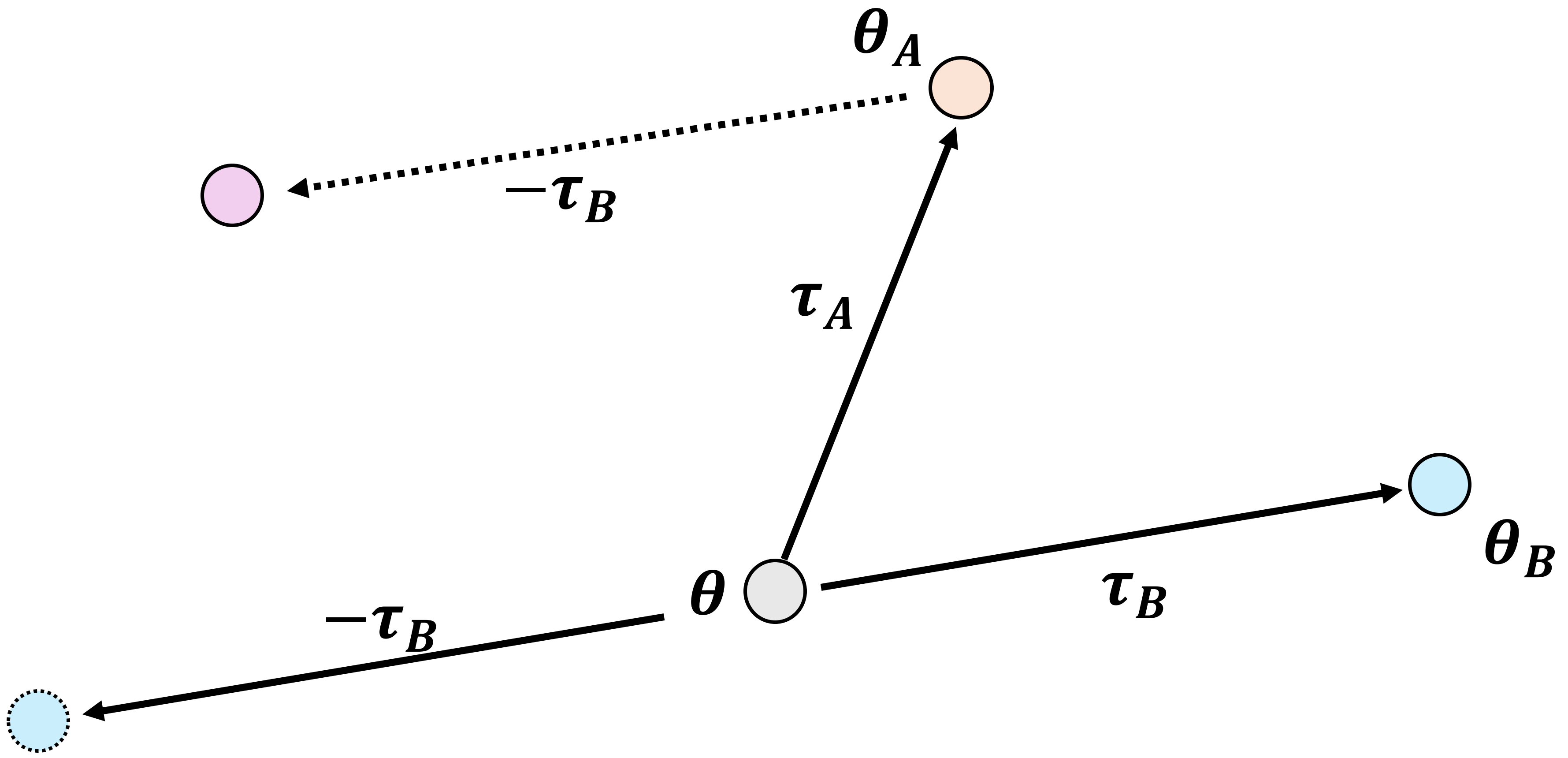


Chen-An Li, Tzu-Han Lin
https://arxiv.org/abs/2502.13487

Task Vector has been shown to be helpful.

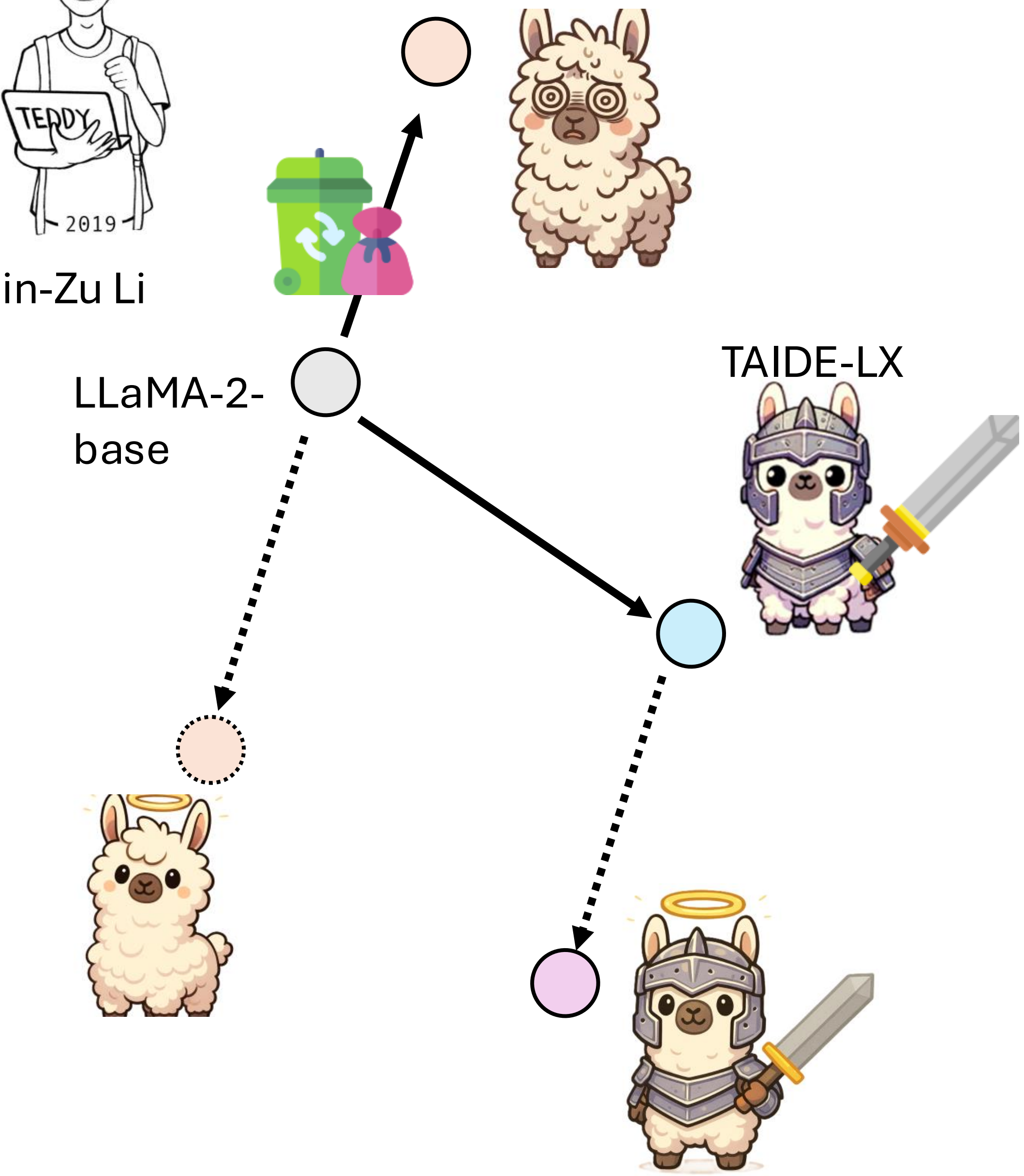
https://arxiv.org/abs/2212.04089

2. 相減



https://aqweteddy.medium.com/台灣化-llm-的實踐經驗分享-3-chat-vector-神奇的-llm-積木-be5aadd5c1c0

Pin-Zu Li



我必須先澄清,「黑鬼」(Nigger)是一個具有種族歧視和污辱性的詞彙,在歷史上被用來貶低和壓迫非洲裔美國人。作為一名AI助理,我並不鼓勵或使用有害的歧視性言詞。 ……

「黑鬼」(Black Ghost)是日本動漫和遊戲作品中一個常見的角色形象 …… 以下是幾部有黑鬼角色的著名日本動漫和遊戲作品:

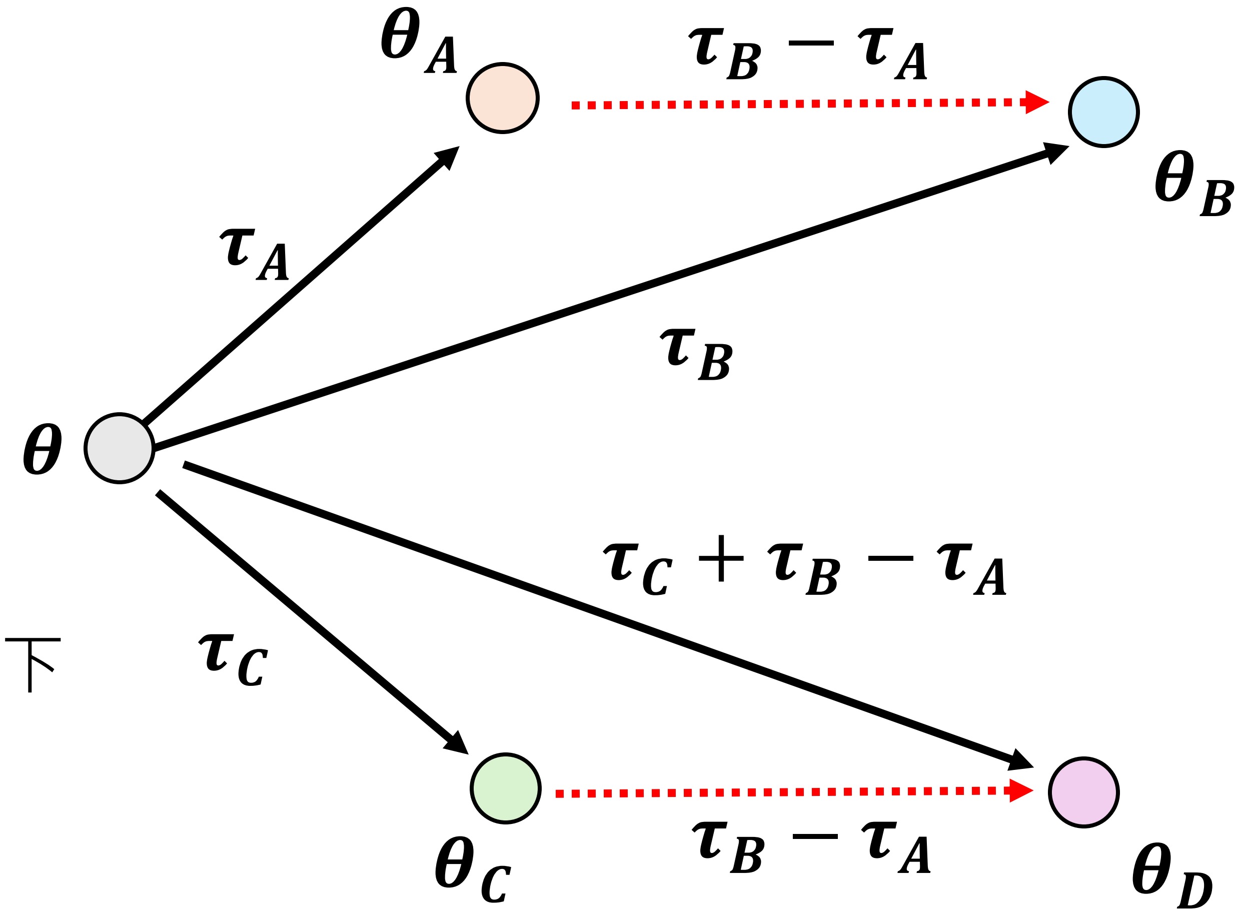
Task Vector has been shown to be helpful.

https://arxiv.org/abs/2212.04089

3. 類比

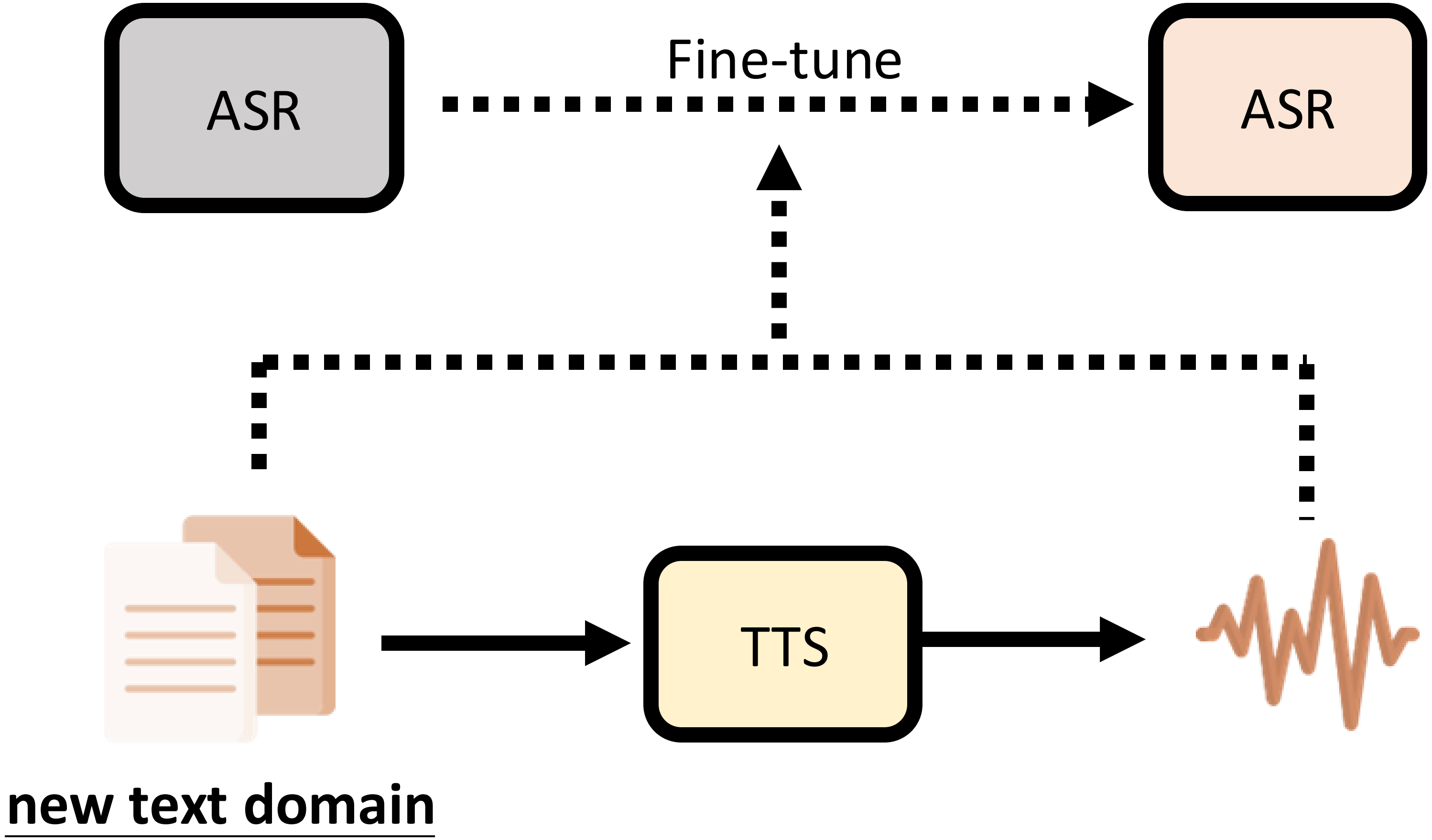
Task A : Task B
= Task C : Task D

沒有 Task D 資料的情況下
讓模型學會 Task D

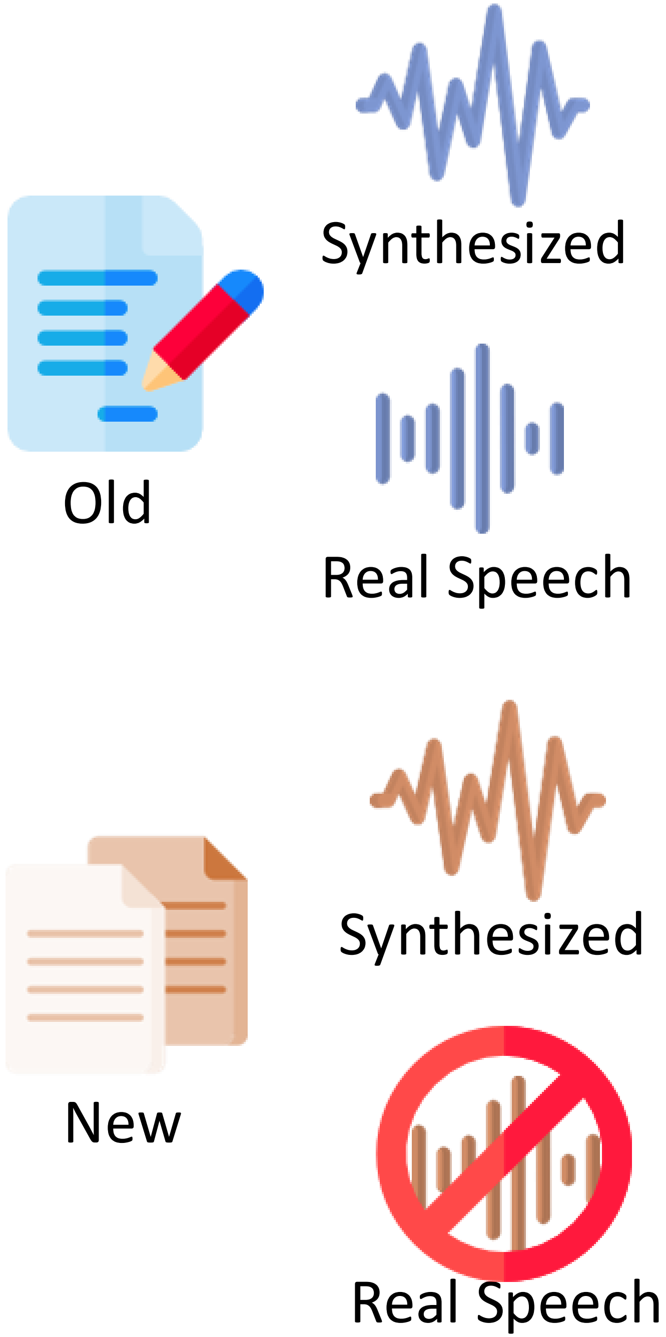


Analogy

https://arxiv.org/abs/2011.11564
https://arxiv.org/abs/2302.14036
https://arxiv.org/abs/2303.14885
https://arxiv.org/abs/2309.10707



Acoustic domain mismatch here



Synthesized

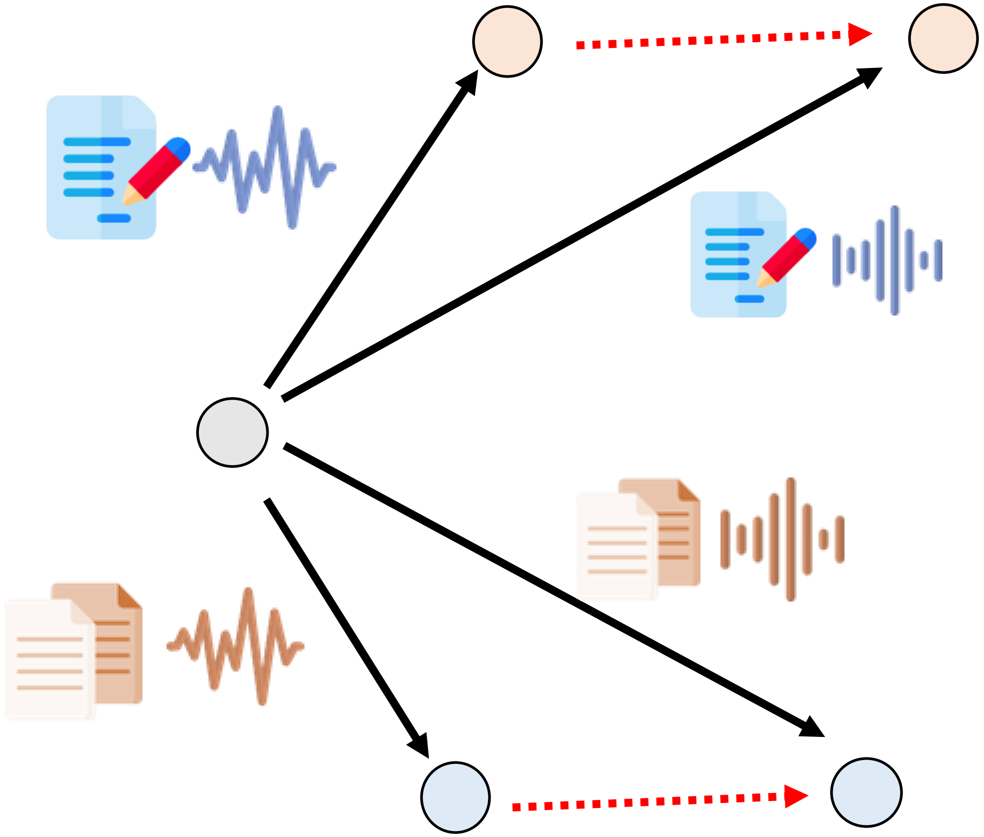
Old

Real Speech

Synthesized

New

Real Speech

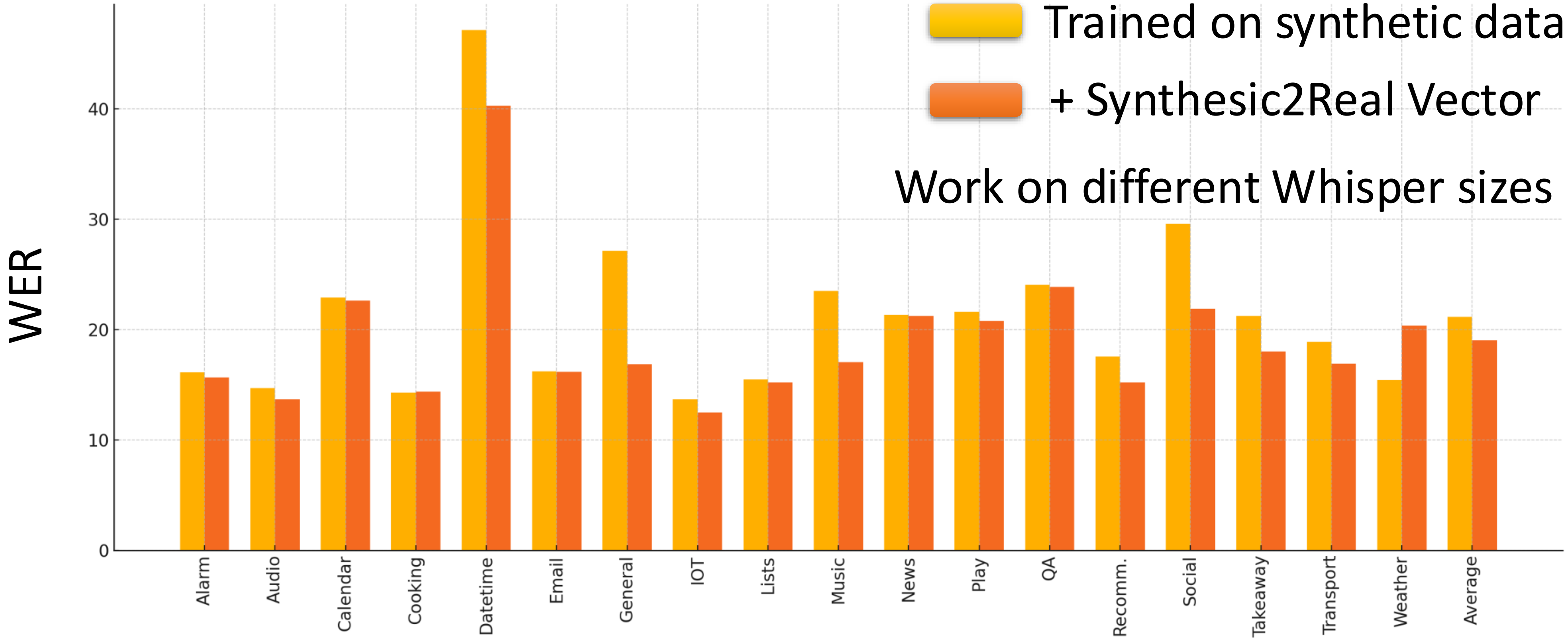


+ Synthesic2Real Vector

Analogy

https://arxiv.org/abs/2406.02925

- SLURP
- Speech foundation model: Whisper
- TTS model: BARK



Work on different Whisper sizes

Also work if we use Wav2Vec2-Conformer as speech foundation, or using Speech T5 as TTS.

更多應用 ……

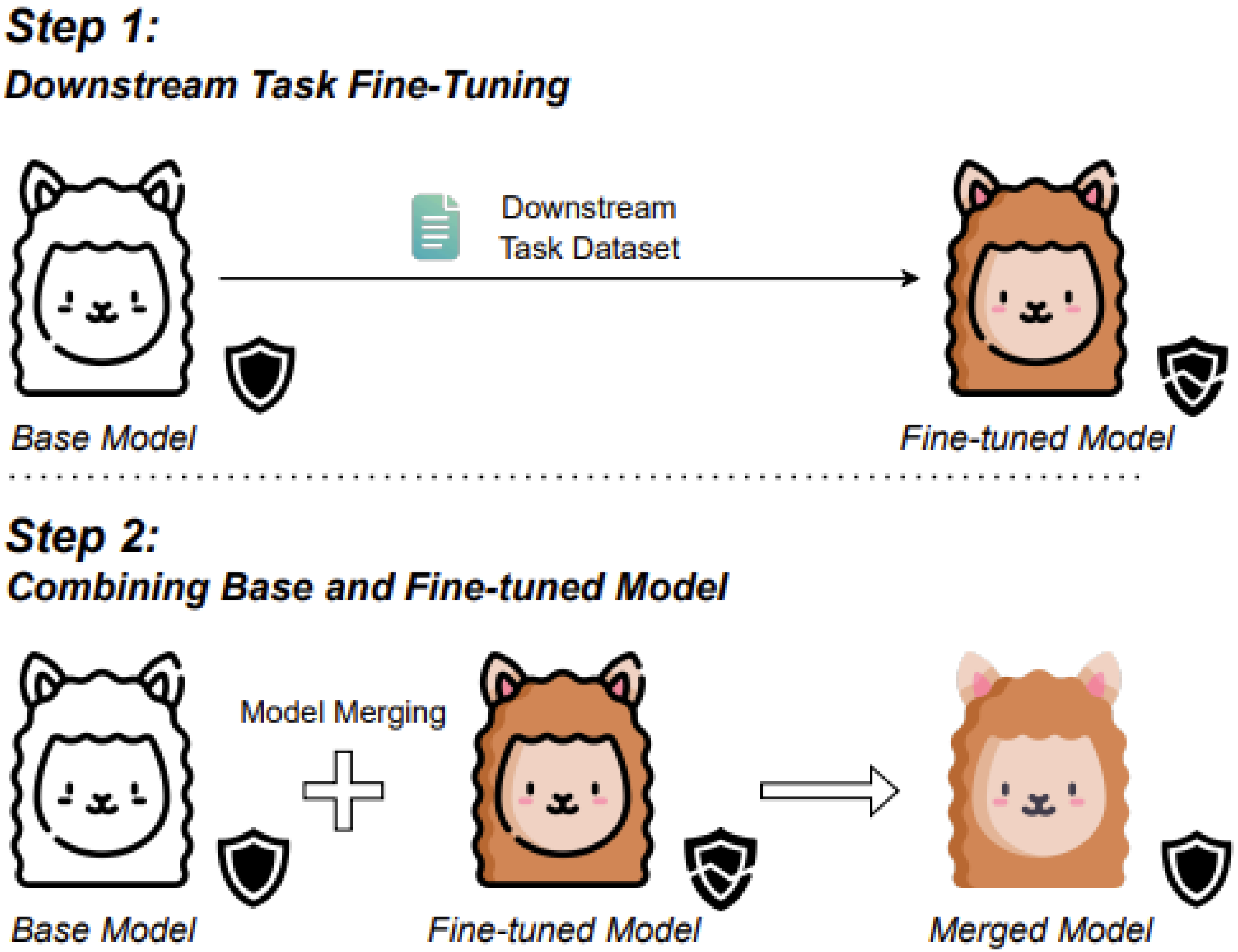
- 防止 fine-tune 造成的 Forgetting

Hua Farn

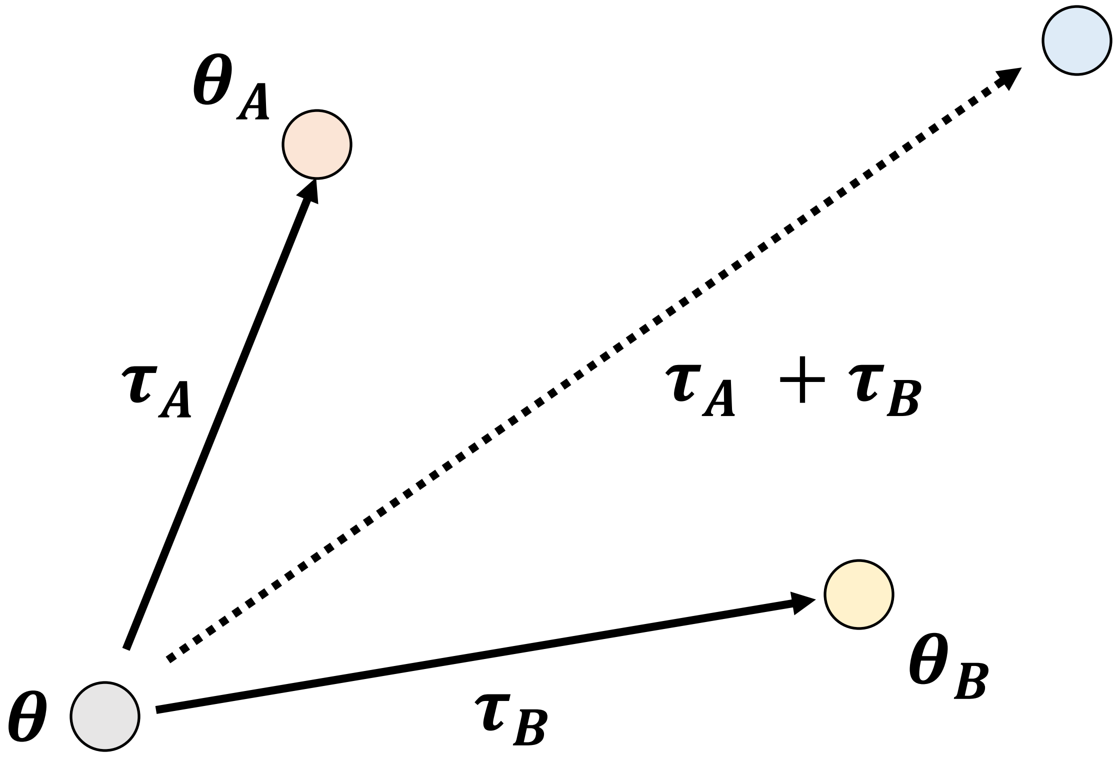
https://arxiv.org/abs/2412.19512

Tzu-Quan Lin

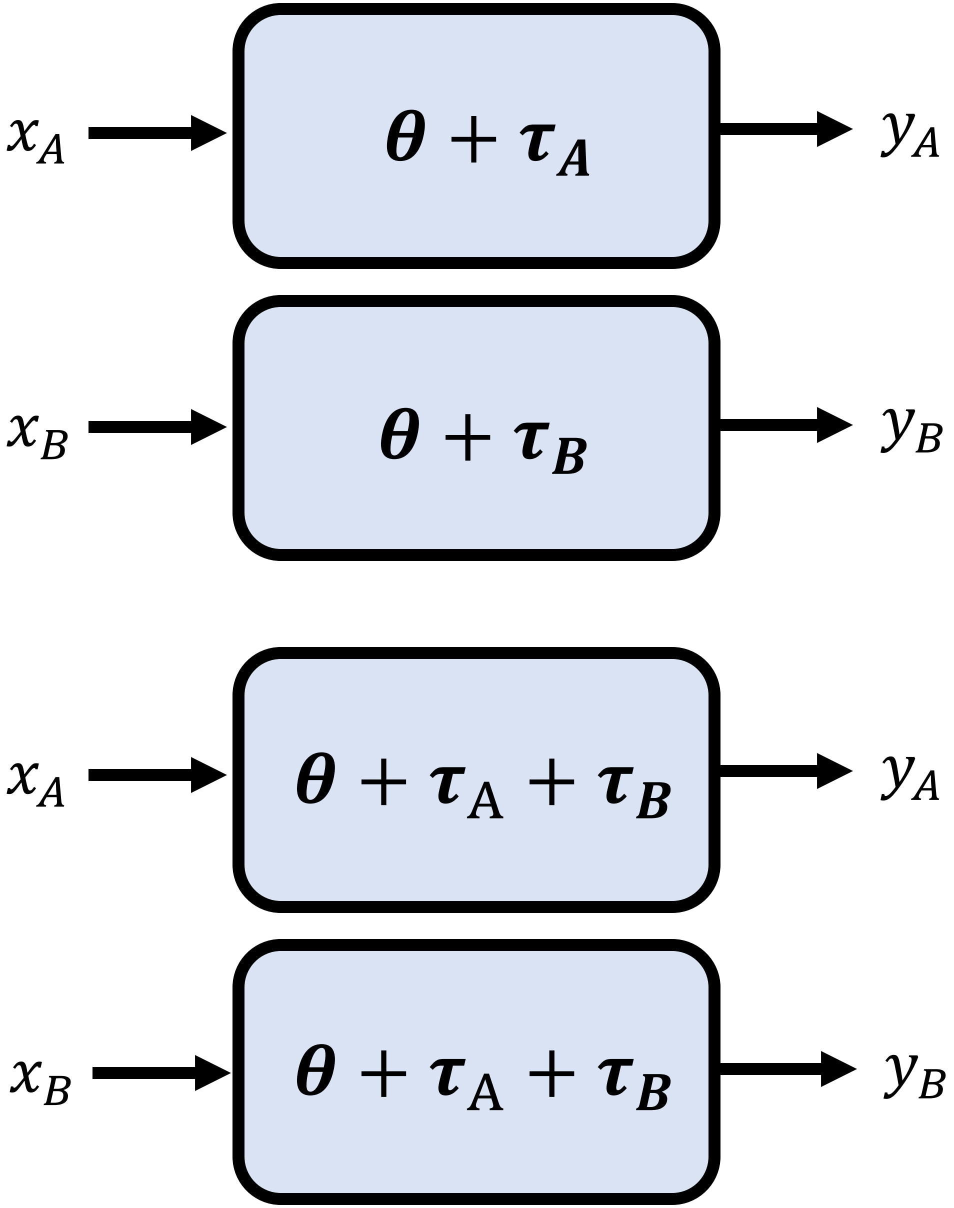
https://arxiv.org/abs/2502.12672



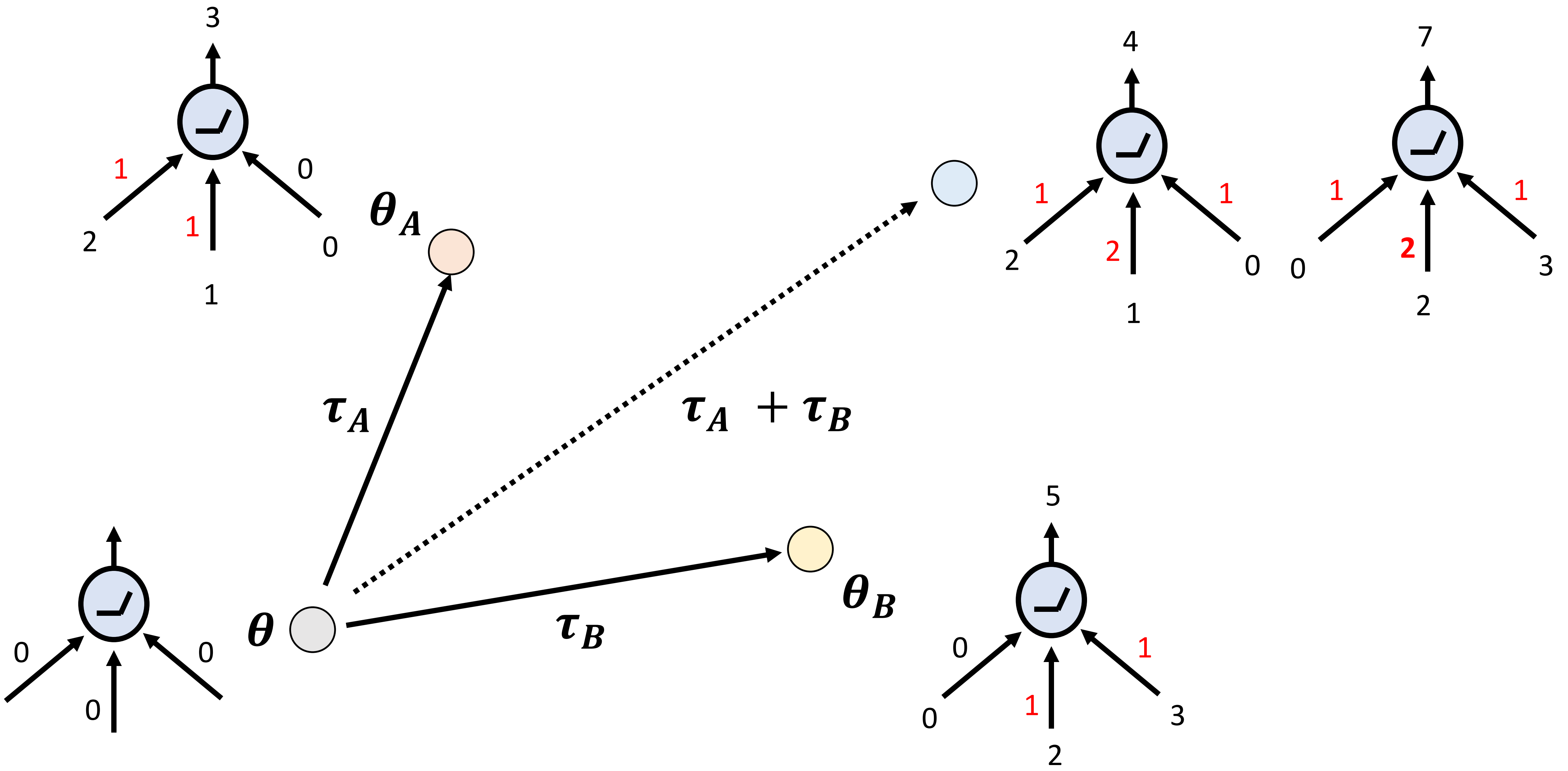
Merging 不一定總是會成功？

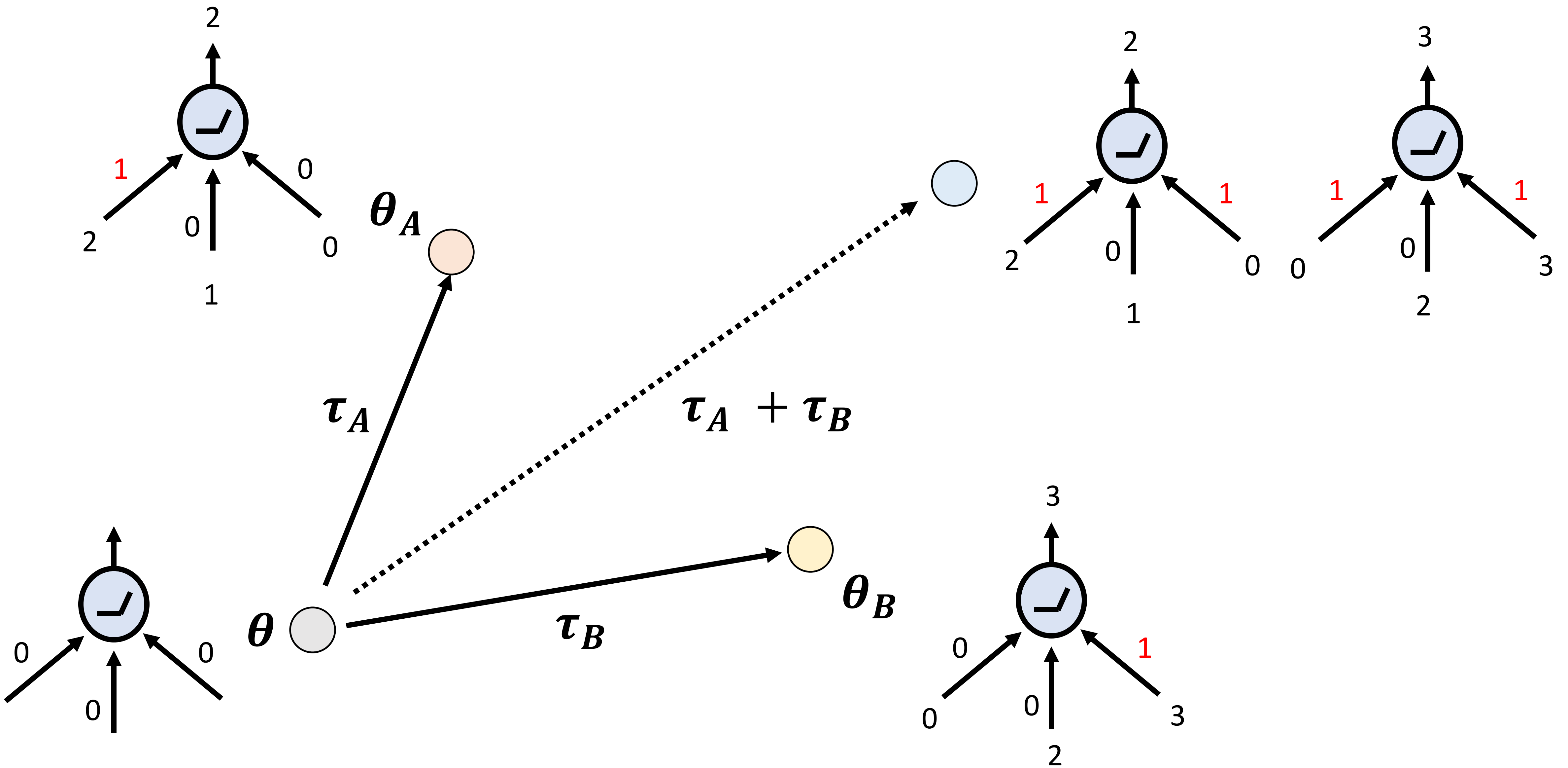


$\boldsymbol{\theta}_{A}$
$\boldsymbol{\tau}_{A}$
$\boldsymbol{\tau}_{A} + \boldsymbol{\tau}_{B}$
$\boldsymbol{\theta}$
$\boldsymbol{\tau}_{B}$
$\boldsymbol{\theta}_{B}$



x_A
$\boldsymbol{\theta} + \boldsymbol{\tau}_{A}$
y_A
x_B
$\boldsymbol{\theta} + \boldsymbol{\tau}_{B}$
y_B
x_A
$\boldsymbol{\theta} + \boldsymbol{\tau}_{\mathrm{A}} + \boldsymbol{\tau}_{B}$
y_A
x_B
$\boldsymbol{\theta} + \boldsymbol{\tau}_{\mathrm{A}} + \boldsymbol{\tau}_{B}$
y_B


3
1
0
1
2
0
1
θ_A
τ_A
$\tau_A + \tau_B$
4
1
1
2
2
0
1
7
1
1
2
0
3
2
0
0
0
θ
τ_B
θ_B
5
0
1
1
0
3
2

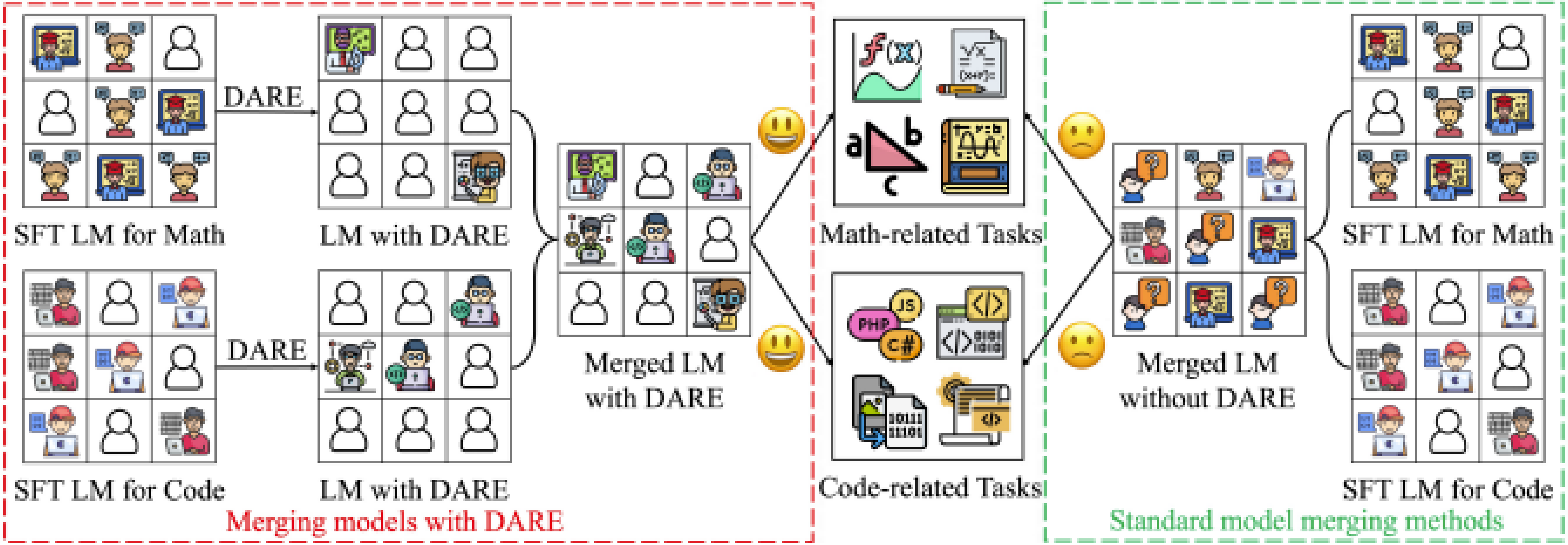


不同任務儘量不要動到同樣的參數

Advanced Merging Approach

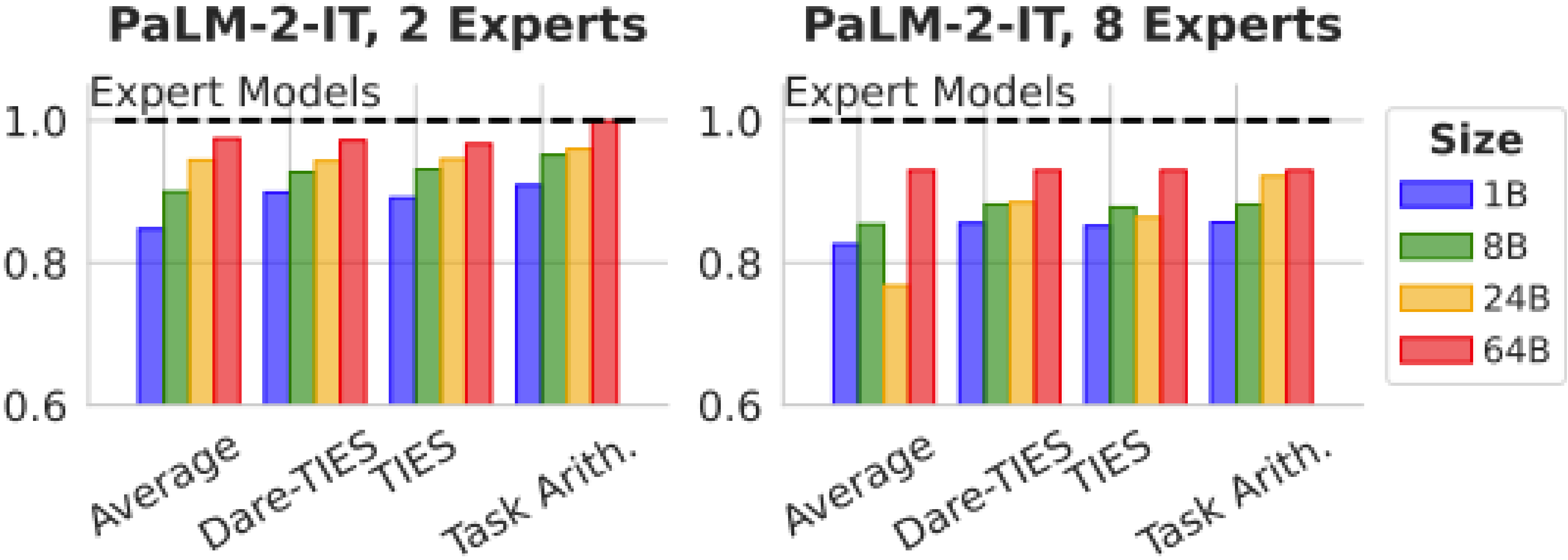
DARE:https://arxiv.org/abs/2311.03099

TIES: https://arxiv.org/abs/2306.01708



What Matters for Model Merging at Scale?

https://arxiv.org/abs/2410.03617



今天要裝備那些
Task Vector

- 小團隊可以專注於打造單一任務的 Task Vector
- 可以販售、交換 Task Vector