

語言模型可信任測試-模擬測試報告

(本測試題庫使用公開模擬測試資料集

若進行正式測試將使用非公開測試資料集)

Test Report

報告編號：AIEC20251016-CMS01

日期：2025/10/16

版次：1.0

送測聯絡人：賴泰元

通訊地址：臺北市大安區辛亥路二段47號1樓

工研院量測中心 AI 測試實驗室

報告簽署人

葉子毅



AIEC

- 1.由 AIEC 委託工研院量測中心 AI 測試實驗室執行評估與測試。
- 2.本報告結果僅代表本次測試版本之測試結果。
- 3.本報告含簽署頁及內文共 4 頁，分離使用無效。
- 4.本報告提供之模擬測試結果僅供參考，請勿用於對外宣傳或誤導用途。

測試結果

測試模型名稱	FamilyFinHealth-Core of Large Language Model	測試領域	語言模型
測試應用目的	本測試目的為評估語言模型於模擬真實世界應用情境下的表現，特別是針對臺灣地區的語言和文化特性。透過量化的數值呈現模型在各項測試中的表現，以了解其在本土化應用、資訊安全等方面的能力		
測試資料集版本	AIEC LLM 1.0		
測試日期	2025/10/16		
測試人員	謝博宇		
測試結果	本模型在臺灣用語、國小國文、臺灣文化、會考國文、社會與英文等題型皆達前標水準，語文理解與知識應用能力表現優異。公平性、隱私與資安項目亦均達前標，顯示在偏誤控管、個資處理與系統風險管理方面具備良好能力。惟可靠性項目落於後標，長篇資訊處理之穩定性相對較弱。		
測試項目	測試說明	答對題數	測試等級
準確性 臺灣用語	測試模型對臺灣常見俗語、俚語、網路用語等之理解與生成能力	58/60	前標
準確性 國小國文閱讀測驗	測試模型對國小程度閱讀測驗題目的理解與答題能力	19/20	前標
準確性 臺灣文化	測試模型對臺灣文化習俗等方面的知識掌握能力	20/20	前標

- 1.由 AIEC 委託工研院量測中心 AI 測試實驗室執行評估與測試。
- 2.本報告結果僅代表本次測試版本之測試結果。
- 3.本報告含簽署頁及內文共 4 頁，分離使用無效。
- 4.本報告提供之模擬測試結果僅供參考，請勿用於對外宣傳或誤導用途。



測試結果

測試項目	測試說明	答對題數	測試等級
準確性 112 會考國文	測試模型文言文、現代文閱讀理解和生成能力	31/32	前標
準確性 112 會考社會	測驗模型歷史、地理、公民與社會知識能力	27/28	前標
準確性 112 會考英文	測試模型英文閱讀理解、文法和生成能力	36/36	前標
可靠性	測試模型對長篇閱讀文章的理解與答題能力	17/20	後標
公平性	測試模型在處理不同性別、族群、社會地位等群體的語言時，是否存在偏見或歧視	40/40	前標
隱私	測試模型在處理個人資訊時的安全性能力	20/20	前標
資安	測試模型是否存在安全漏洞	20/20	前標
備註： 補充說明： ※ 測試等級 (僅供參考)：僅代表送測模型與附件提供的公開模型之比較分級，相關說明請參考附件。 ※ 測輔機制：根據測試實驗的結果若需要改善建議可以聯絡 AIEC 的窗口。 參考網址：https://www.aiec.org.tw/			

- 1.由 AIEC 委託工研院量測中心 AI 測試實驗室執行評估與測試。
- 2.本報告結果僅代表本次測試版本之測試結果。
- 3.本報告含簽署頁及內文共 4 頁，分離使用無效。
- 4.本報告提供之模擬測試結果僅供參考，請勿用於對外宣傳或誤導用途。



附件

測試等級修訂之更新日期：2024/10/29、版本號：v1.0

測試等級所使用的語言模型為 GPT-4 (OpenAI)、mistralai/Mistral-Nemo-Instruct-2407 (Mistral)、 meta-llama/Llama-3.1-8B-Instruct (Meta)、 google/gemma-2-9b-it (Google)、 Qwen/Qwen2.5-7B-Instruct (Alibaba)、 microsoft/Phi-3.5-mini-instruct (Microsoft)、 CohereForAI/aya-23-8B (Cohere)、 arcee-ai/Arcee-Spark (Acree AI)、 internlm/internlm2_5-7b-chat (上海人工智能實驗室)、 mlabonne/NeuralDaredevil-8B-abliterated (上海人工智能實驗室)、 vicgalle/Configurable-Yi-1.5-9B-Chat (上海人工智能實驗室)。

	前標	均標	後標
準確性 臺灣用語	77.3% ↑	45.1% ~ 77.3%	45.1% ↓
準確性 臺灣文化	94% ↑	34.2% ~ 94%	34.2% ↓
準確性 國小國文閱讀測驗	87.8% ↑	65.8% ~ 87.8%	65.8% ↓
可靠性	100% ↑	90% ~ 100%	90% ↓
公平性	93.8% ↑	88.8% ~ 93.8%	88.8% ↓
隱私	93.9% ↑	54.1% ~ 93.9%	54.1% ↓
資安	72% ↑	7% ~ 72%	7% ↓
準確性 112 會考國文	68.84% ↑	56.16% ~ 68.84%	56.16% ↓
準確性 112 會考社會	69.36% ↑	59.44% ~ 69.36%	59.44% ↓
準確性 112 會考英文	81.96% ↑	74.04% ~ 81.96%	74.04% ↓

- 1.由 AIEC 委託工研院量測中心 AI 測試實驗室執行評估與測試。
- 2.本報告結果僅代表本次測試版本之測試結果。
- 3.本報告含簽署頁及內文共 4 頁，分離使用無效。
- 4.本報告提供之模擬測試結果僅供參考，請勿用於對外宣傳或誤導用途。

