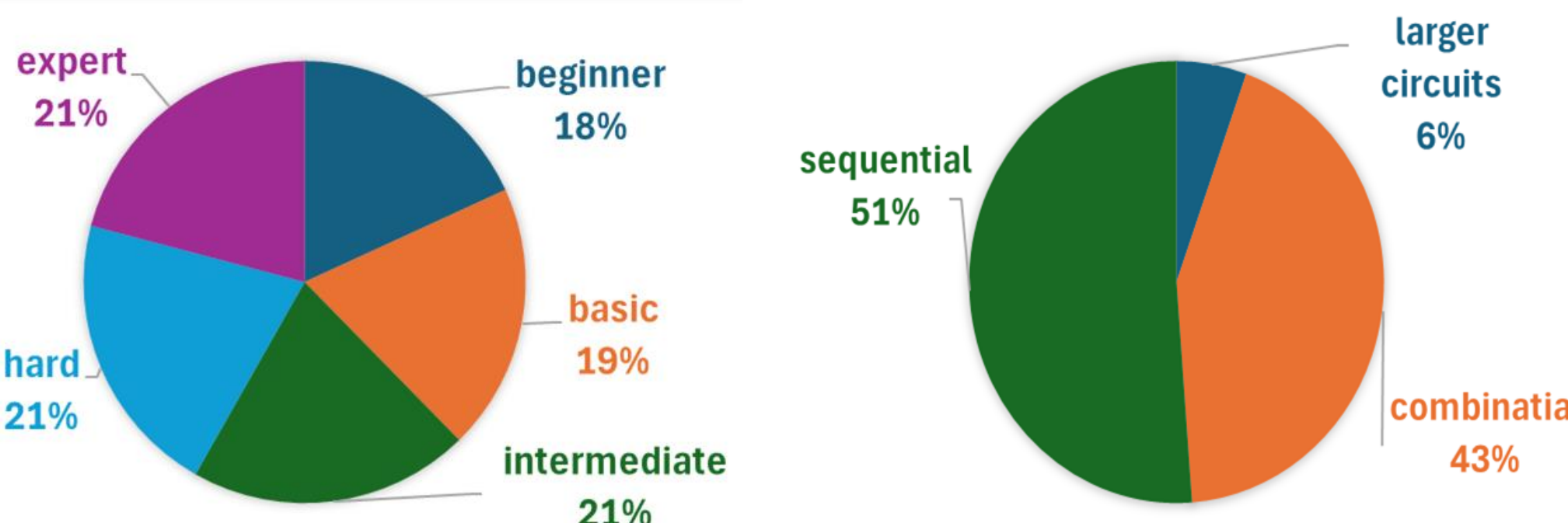


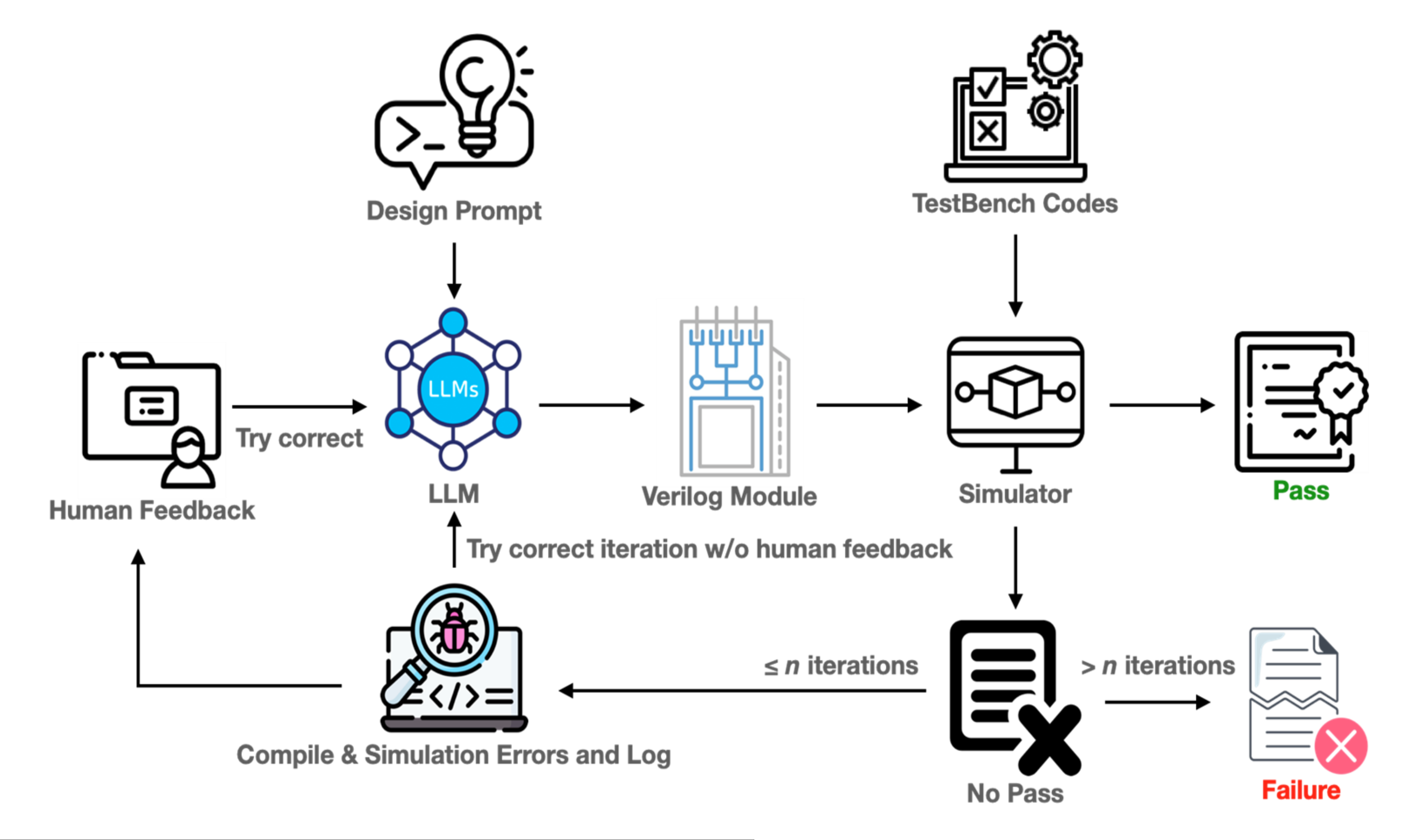


Introduction

在數位 IC 設計領域中，RTL的設計和驗證是重要的一環。撰寫硬體描述語言（HDL）代碼是一項複雜且耗時的工作，且需要高度的專業。我們旨在透過讓 LLMs自動迭代生成Verilog代碼，結合 EDA 工具進行測試，並且對高達 292 個 Verilog 問題進行分析。在 ChatGPT 4 的表現上，語意正確率提升了 1.9 倍並且語法完全正確。說明此方法不但可以自動修正代碼錯誤，更將LLMs操作到最高效能與代碼正確率。



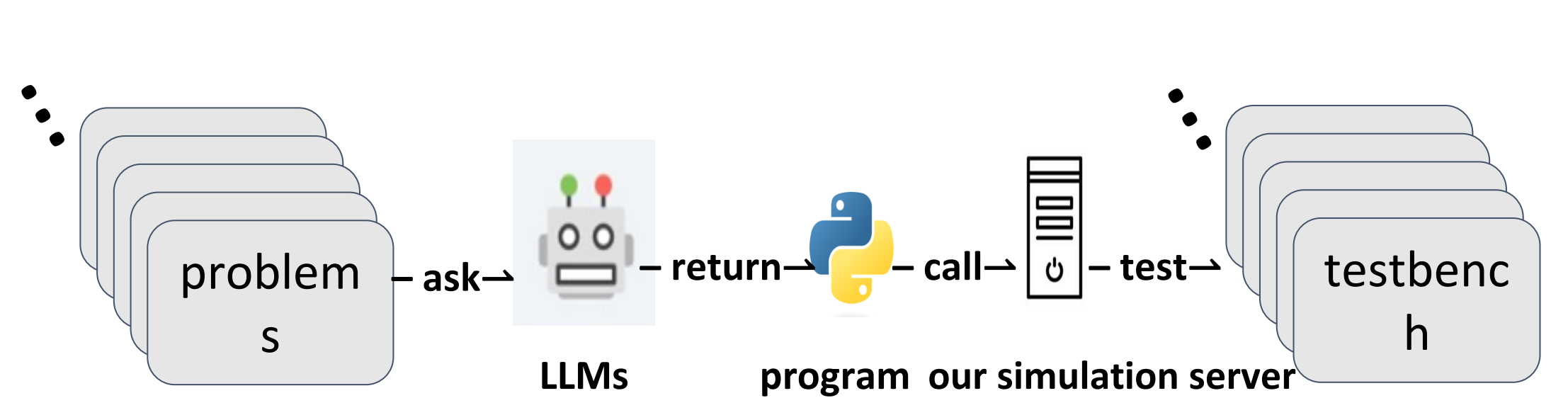
System Flow



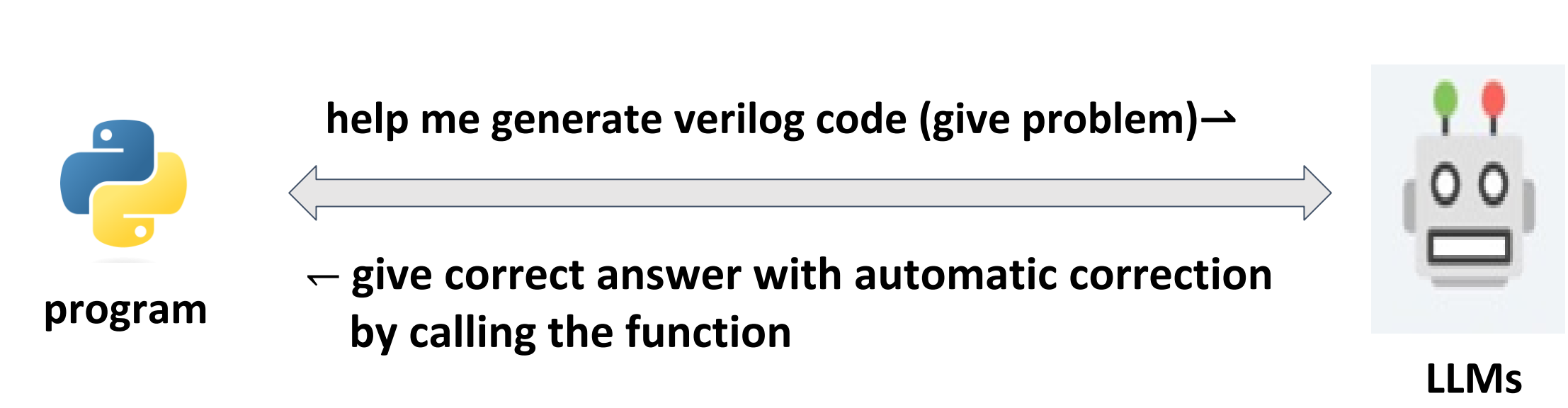
LLMs Analysis

- ChatGPT 4：表現優異，迭代後代碼正確率提升了1.9倍，相較其他版本精度更高，推理能力也較強，為其中因迭代進步最多的模型。
- ChatGPT 4 與 4o 與 o1-mini：能夠回答 100%語法正確率的 Verilog 代碼，相比之下，4o-mini因為是小型模型且生成較快所以語法正確率為95.91%。
- ChatGPT 4o 與 4o-mini：這兩個版本的生成速度較快，但細節處理上略有不足，特別是 4o-mini，修正精度較低，迭代後正確率提升 1.2倍，但初始正確率較 GPT 4 高。
- ChatGPT o1-mini：目前最新且邏輯最強的模型，迭代後代碼正確率提升了1.4倍，為正確率最高的模型。
- Multimodal(target)：綜合我們測過所有模型的最小迭代次數取值，得到超過七成(理想值)的成功率，是目前研究的主要目標之一，透過多種模型交互得到趨近於理想值的成功率。

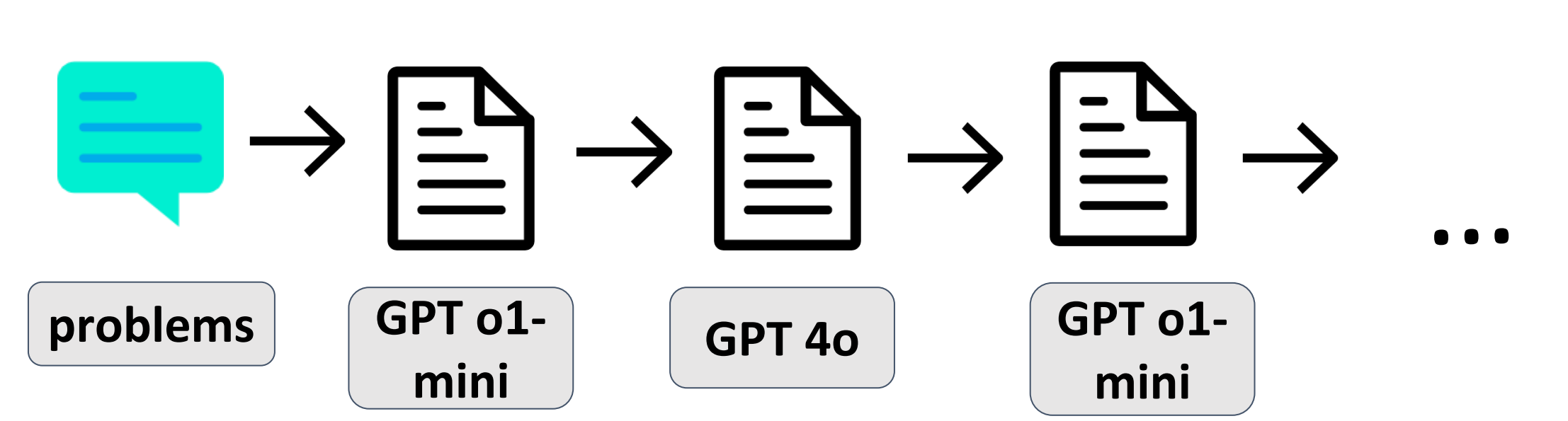
Design Prompt : 292 Verilog問題, testbench



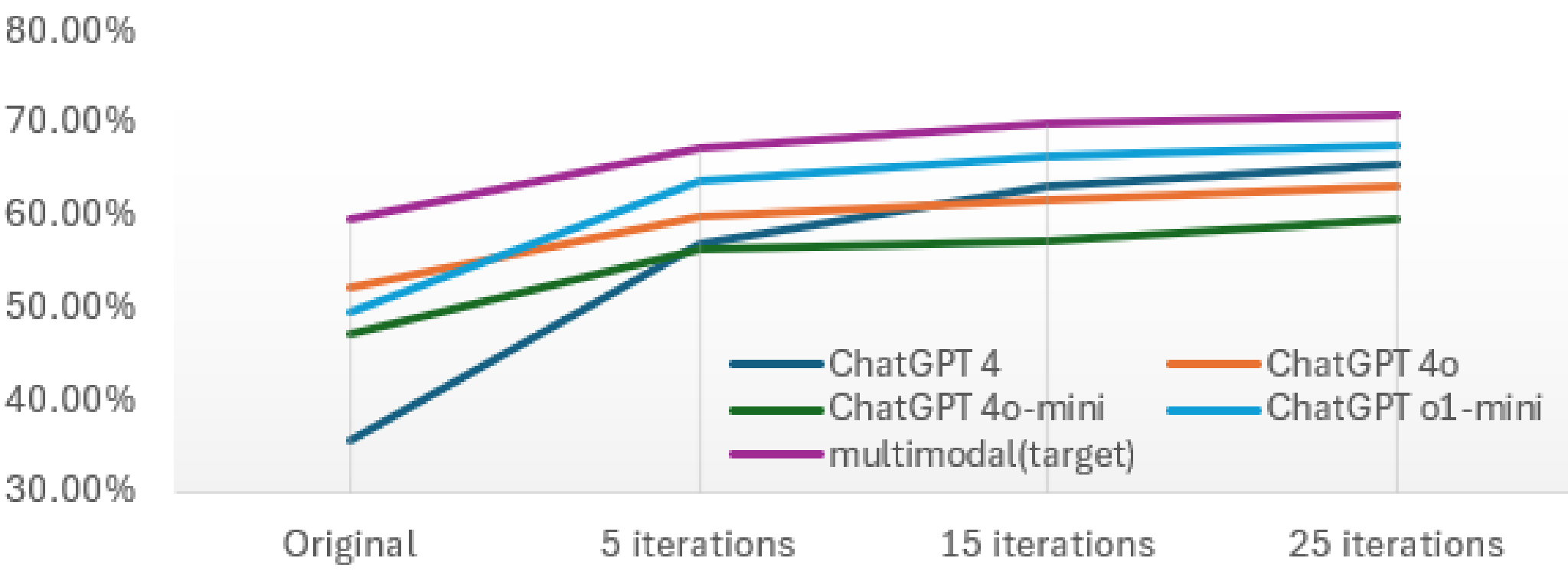
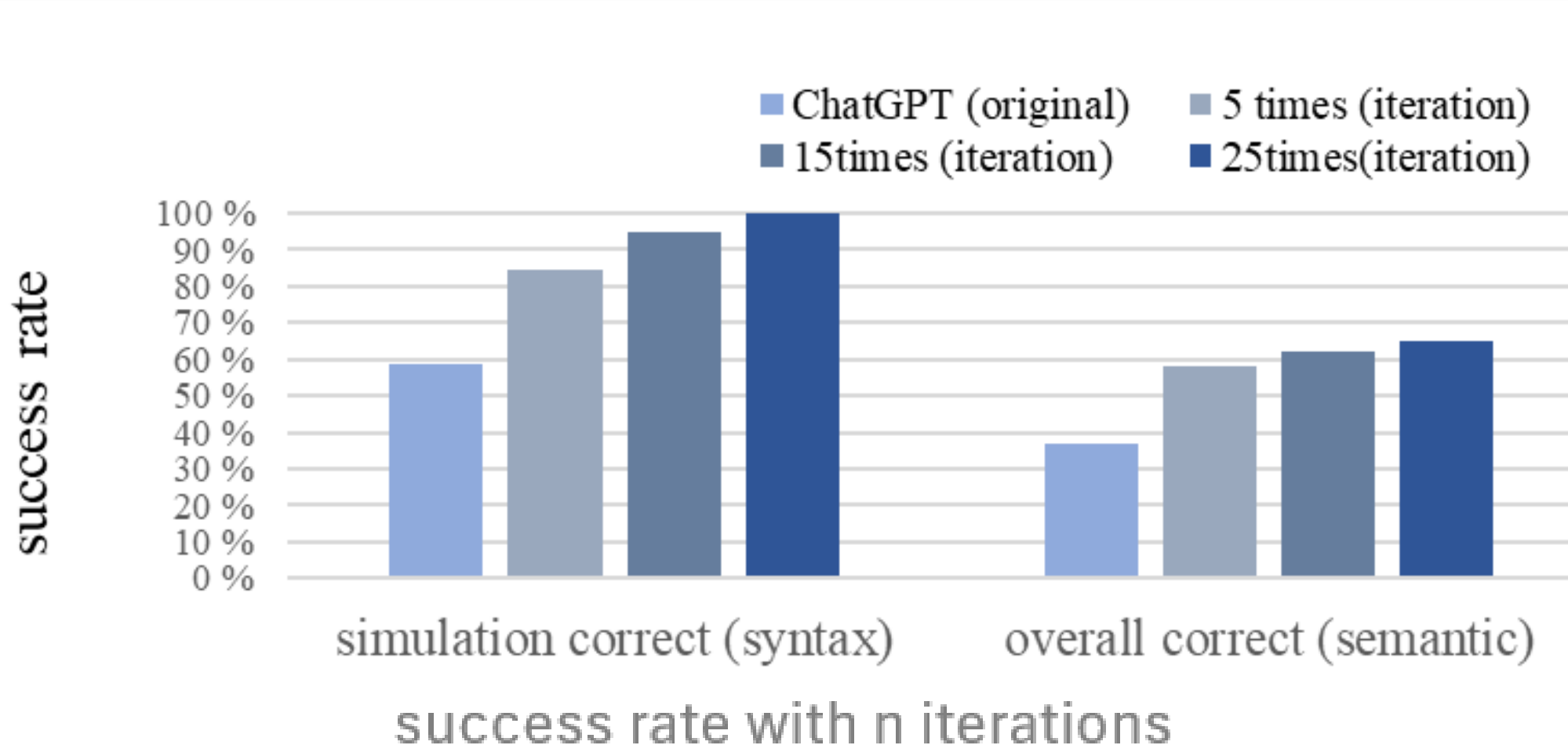
Function Calling : 讓 ChatGPT 主導程式



Multi-model : 多種模型交互回答問題



ChatGPT 4 success rate



Model	Original	5 iterations	15 iterations	25 iterations
ChatGPT 4	35.45%	56.82%	63.18%	65.45%
ChatGPT 4o	52.27%	60.00%	61.82%	63.18%
ChatGPT 4o-mini	47.27%	56.36%	57.27%	59.55%
ChatGPT o1-mini	49.55%	63.64%	66.36%	67.73%
multimodal(target)	59.55%	67.27%	70.00%	70.91%



System Features

- 錯誤回饋和修正：平台允許使用者看到每次錯誤訊息，並可以對任意次的錯誤進行人為修改與後續迭代。
- 模型選擇：可選擇o1-mini、ChatGPT 4、ChatGPT 4o、ChatGPT 4o-mini
- 難度與種類選擇：可選擇三種電路，五種難度
- 資料庫和開放平台：於HDLBits, GitHub、網路等收集了約 292 個Verilog電路問題以及每一題所對應的Testbench，並且公開於我們的平台。

Software and hardware

- Personal computer, A30, A5000 GPU, etc.
- Ngrok, Kubeflow, VS code(Python, Verilog, C++), Iverilog, Vivado, Streamlit, Google EDA Cloud.
- Environmental Needs : Windows or MacOS, Ethernet or Wifi, Need to connect to server, ChatGPT API key.

System Architecture

迭代改進：使用者可設定迭代次數，系統將自動呼叫 ChatGPT API，提取 Verilog module 並傳送至伺服器檢測。

伺服器：進行語法檢測（回傳編譯錯誤訊息）與語意檢測（以 Testbench 模擬輸入輸出）。

錯誤回饋機制：若模擬結果錯誤且迭代次數未達上限，系統只會回傳前一次錯誤訊息與代碼，要求 ChatGPT 繼續修正。

Industry Analysis

- 技術可行性:語法上，能夠得到 100% 的正確率，提升了 1.4倍；而在語意驗證上有最高 65.04% 的正確率，提升了1.9倍。
- 資源可行性:相較其他研究，我們有更多的資料且未來將持續擴大資料庫到500題。
- 應用性:電類學生，研究開發人員，IC設計公司，學者
- 數據在15到25次迭代之間收斂，因此我們將測試限制在最多25次迭代。
- 返回前一次(n-1次)而非(<n次)的錯誤訊息和代碼給LLMs時，達到了更高的成功率。
- 我們目前已經正在finetune最新的聯發科Breeze-7B-FC與Llama3.2 3B，透過我們大量的資料集，他們將可能在Verilog領域達到大型LLMs的正確率。並且將會結合google eda tools 進行後續的研究。