Ｐython Ｂeginner Ａssistant評分標準

V20240815

| **判斷項目** | **目的** | **評分標準** | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **評分項目1** | | **評分項目2** | | **評分項目3** | |
| 有效迴圈判斷 | 提供特定語法學習後之應用能力，例如不濫用迴圈，確保LLM可偵測到正確的迴圈使用數量。 | **準確性** | **1** | **詳細性** | **1** | **正向建議** | **1** |
| 正確描述程式碼的有效迴圈數量 | 1 | 詳細描述每個迴圈的功能與判斷為無效迴圈的理由 | 1 | 提供無效迴圈刪除建議、巢狀迴圈使用與迴圈類型使用建議 | 1 |
| 部分描述程式碼的有效迴圈數量(for或while其中一種有效迴圈判斷正確) | 0 | 詳細描述每個迴圈判斷為無效迴圈的理由 | 0 | 給予簡單建議，如無效迴圈刪除等 | 0 |
| 無法描述程式碼的有效迴圈數量 | -1 | 沒有詳細描述 | -1 | 無建議 | -1 |
| 語法錯誤判斷 | LLM應能協助察覺語法錯誤，並教導正確語法，提升學習效率。 | **準確性** | **1** | **詳細性** | **1** | **正向建議** | **1** |
| 完全察覺語法錯誤 | 1 | 詳細描述語法錯誤原因與程式碼位置 | 1 | 提供具體的修正建議，且有修正範例 | 1 |
| 未能察覺語法錯誤 | 0 | 部分描述語法錯誤原因與程式碼位置 | 0 | 無建議 | 0 |
| 錯誤察覺語法錯誤 | -1 | 沒有詳細描述 | -1 | 提供錯誤的修正建議 | -1 |
| 邏輯錯誤判斷 | 邏輯錯誤不易察覺且可能無程式報錯，但會導致程式結果不如預期，故希望LLM可以根據題目找出邏輯錯誤。 | **準確性** | **1** | **詳細性** | **1** | **正向建議** | **1** |
| 正確察覺邏輯錯誤 | 1 | 詳細描述邏輯錯誤原因 | 1 | 提供具體的修正建議，且有修正範例 | 1 |
| 無察覺邏輯錯誤 | 0 | 沒有詳細描述 | 0 | 無建議 | 0 |
| 沒有察覺邏輯錯誤 | -1 | 錯誤地描述邏輯錯誤原因 | -1 | 提供錯誤的修正建議 | -1 |
| 清晰的描述程式碼邏輯結構 | 有時學生看不懂演算法或函數的功能或結構，希望LLM可以幫助學生解析程式碼。 | **準確性** | **1** | **詳細性** | **1** | **正向建議** | **1** |
| 正確判斷演算法或函數名稱與功能 | 1 | 詳細描述，包括主要邏輯和次要邏輯 | 1 | 能夠教導演算法或函數知識(100%) | 1 |
| 無判斷出演算法或函數名稱，但可以敘述功能與結構 | 0 | 部分描述(50%) | 0 | 能夠教導演算法或函數知識(50%) | 0 |
| 無法為使用者描述功能 | -1 | 沒有詳細描述 | -1 | 無法提供教學任何提示 | -1 |
| 完成指定功能程式碼 | 當學習程式遇到瓶頸時或需要示範成功的程式時，學習才能有所成長。故LLM應該要能做到這樣的助教功能。 | **準確性** | **1** | **詳細性** | **1** | **正向建議** | **1** |
| 正確理解題目要求 | 1 | 程式編譯無誤且執行結果正確 | 1 | 提供架構程式的思路、注意事項與邏輯處理的技巧 | 1 |
| 無法理解題目要求 | 0 | 程式編譯無誤但執行結果有誤 | 0 | 提供應實現功能的目標，但沒有具體教學 | 0 |
| 錯誤理解題目要求 | -1 | 程式編譯有誤且無法執行 | -1 | 沒有教學功能 | -1 |

1. 立偉老師建議配分的部分參考表單樣式：非常不滿意、不滿意、普通、滿意、非常滿意，在負面結果可以有扣分的處理，所以我把配分統一改成有1~-1分，並把-1分的項目審視更改。

2. 立偉老師建議把評分項目**準確性**、**詳細性**、**正向建議**，改成更具體的名字，如醫生看病有察覺病情、口述/紙本記錄病情、醫療處置／開配藥方這種取名方式，但我還想不到又具體又可以匹配每一個判斷項目中評分項目的名字，所以還沒更改。

3. 以上為結合了立委老師建議與鄭院長更改的版本。

v20240814\_2

| **判斷項目** | **目的** | **評分標準** | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **評分項目1** | | **評分項目2** | | **評分項目3** | |
| 有效迴圈判斷 | 提供特定語法學習後之應用能力，例如不濫用迴圈，確保LLM可偵測到正確的迴圈使用數量。 | **準確性** | **2** | **詳細性** | **2** | **正向建議** | **2** |
| 正確描述程式碼的有效迴圈數量 | 2 | 詳細描述每個迴圈的功能與判斷為無效迴圈的理由 | 2 | 提供無效迴圈刪除建議、巢狀迴圈使用與迴圈類型使用建議 | 2 |
| 部分描述程式碼的有效迴圈數量(for或while其中一種有效迴圈判斷正確) | 1 | 詳細描述每個迴圈判斷為無效迴圈的理由 | 1 | 給予簡單建議，如無效迴圈刪除等 | 1 |
| 無法描述程式碼的有效迴圈數量 | 0 | 沒有詳細描述 | 0 | 無建議 | 0 |
| 語法錯誤判斷 | LLM應能協助察覺語法錯誤，並教導正確語法，提升學習效率。 | **準確性** | **2** | **詳細性** | **2** | **正向建議** | **2** |
| 完全察覺語法錯誤 | 2 | 詳細描述語法錯誤原因與程式碼位置 | 2 | 提供具體的修正建議，且有修正範例 | 2 |
| 部分察覺語法錯誤 | 1 | 部分描述語法錯誤原因與程式碼位置 | 1 | 無建議 | 1 |
| 未能察覺語法錯誤 | 0 | 沒有詳細描述 | 0 | 提供錯誤的修正建議 | 0 |
| 邏輯錯誤判斷 | 邏輯錯誤不易察覺且可能無程式報錯，但會導致程式結果不如預期，故希望LLM可以根據題目找出邏輯錯誤。 | **準確性** | **2** | **詳細性** | **2** | **正向建議** | **2** |
| 完全察覺邏輯錯誤 | 2 | 詳細描述邏輯錯誤原因 | 2 | 提供具體的修正建議，且有修正範例 | 2 |
| 部分察覺邏輯錯誤 | 1 | 沒有詳細描述 | 1 | 無建議 | 1 |
| 察覺錯誤或沒有察覺邏輯錯誤 | 0 | 錯誤地描述邏輯錯誤原因 | 0 | 提供錯誤的修正建議 | 0 |
| 清晰的描述程式碼邏輯結構 | 有時學生看不懂演算法或函數的功能或結構，希望LLM可以幫助學生解析程式碼。 | **準確性** | **2** | **詳細性** | **2** | **正向建議** | **2** |
| 正確判斷演算法或函數名稱與功能 | 2 | 詳細描述，包括主要邏輯和次要邏輯 | 2 | 能夠教導演算法或函數知識(100%) | 2 |
| 無判斷出演算法或函數名稱，但可以敘述功能與結構 | 1 | 部分描述(50%) | 1 | 能夠教導演算法或函數知識(50%) | 1 |
| 無法為使用者描述功能 | 0 | 沒有詳細描述 | 0 | 無法提供教學任何提示 | 0 |
| 完成指定功能程式碼 | 當學習程式遇到瓶頸時或需要示範成功的程式時，學習才能有所成長。故LLM應該要能做到這樣的助教功能。 | **準確性** | **2** | **詳細性** | **2** | **正向建議** | **2** |
| 正確理解題目要求 | 2 | 程式編譯無誤且執行結果正確 | 2 | 提供架構程式的思路、注意事項與邏輯處理的技巧 | 2 |
| 部分理解題目要求 | 1 | 程式編譯無誤但執行結果有誤 | 1 | 提供目標程式碼應實現功能的敘述 | 1 |
| 無法理解題目要求 | 0 | 程式編譯有誤且無法執行 | 0 | 沒有教學功能 | 0 |

20240814b: 我的回覆與建議如下，請參考。

1. 依據建議把評分項目改成相同的，但評分的細微依據每個判斷項目裡還是有所不同，且把分數配比訂定成平均的配分，以此去除主觀傾向。

>> 了解，很好，日後記得將你的闡述放到論文中說明。

1. 面向的族群訂定在國、高中且無程式背景的學生，我覺得這個族群具備邏輯思考能力但缺乏自學能力，這個LLM可以擔任國高中老師上課的助教。

>> 很棒，這樣樣本族群確定了，上述表格的題目與評分應該就簡單很多。以下有兩個方法可以處理，您斟酌要走哪一條路。

1. 【方法1】若考慮是國高中生，就要請已經在學校教授資訊科技的老師與學長姊幫忙施測，但須要等到學期間才有機會取得評測資料；若此，以下的步驟需要一一履行才行:
   * + 目前你正在處理問見專家審查的階段，應該盡速處理完畢；
     + 此問卷還要做信效度的驗證，建議直接問立偉老師如何處理比較快；
     + 需先確認那些學校的那些老師要幫忙施測這樣的問卷；
     + 需要確認施測的年級、課程、日期、時間、範圍等；
     + 最好分成實驗組與對照組，或者問立偉老師該怎麼辦；
     + 需要撰寫計畫書準備申請IRB；
     + 需要申請IRB簡易審查(成大HREC)；
     + 拿到IRB審查通過文件，才有機會開始進行問卷施測；
     + 問卷施測後取得樣本資料，進行分析與結果判讀，才會得知上述LLM的使用評價。
     + 上述的作業若您想玩，可能要玩到年底才能畢其功，所以先想想是否真的要玩?
   1. 【方法2】若覺得國高中很麻煩，也可以在我大一的課堂上處理。目前我有申請一個三年的長期IRB，方便我+立偉老師+您們隨時插隊，進行類似的評測，但學生就是大學生嚕，但我的IRB需要修正將您的上述評分標準再次納入該IRB中，重新送修正審查，若通過，就可以在下學期直接讓大一新生填問卷囉。哈哈，但我需要花點時間修正我的IRB申請內容，且納入你的問卷才行。
2. 這樣看下來，你會發現真的事情很多，對吧~~~
3. 上述表格內容我有修改一些措辭，請自行確認。
4. 上述工作就是很多使用問卷方式取得資訊管理類碩博士學位的學生必經之路~~~很多人無法取得學位，大多是沒有恆心與毅力，並不是笨；反觀資訊工程類碩博士還需要實作出系統或設計出架構，相對就更難嚕~~~
5. 我修正完畢，才看到你的訊息裡有關立委老師的建議，請您自行彙整處理嚕~~~

==================================================================

1. 感謝老師幫我把文件表格化，幫助了我更明確訂定評分。

2. 我依據建議把評分項目改成相同的，但評分的細微依據每個判斷項目裡還是有所不同，且把分數配比訂定成平均的配分，以此去除主觀傾向。

3. 面向的族群訂定在國、高中且無程式背景的學生，我覺得這個族群具備邏輯思考能力但缺乏自學能力，這個LLM可以擔任國高中老師上課的助教。

v20240814

| **判斷項目** | **目的** | **評分標準** | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **評分項目1** | | **評分項目2** | | **評分項目3** | |
| 有效迴圈判斷 | 提供特定語法學習後之應用能力，例如不濫用迴圈，確保LLM可偵測到正確的迴圈使用數量。 | **準確性** | **4** | **詳細性** | **4** | **清晰性** | **2** |
| 正確描述程式碼的邏輯結構 | 4 | 詳細描述，包括主要邏輯和次要邏輯 | 4 | 描述清晰、易懂 | 2 |
| 未能完全描述或存在錯誤 | 2 | 部分描述 | 2 | 無法理解的描述 | 0 |
| 完全錯誤 | 0 | 沒有詳細描述分 | 0 |  |  |
| 語法錯誤判斷 | LLM應能協助察覺語法錯誤，並教導正確語法，提升學習效率。 | **準確性** | **4** | **詳細性** | **4** | **修正建議** | **2** |
| 正確描述程式碼的邏輯結構 | 4 | 詳細描述，包括主要邏輯和次要邏輯 | 4 | 提供具體的修正建議 | 2 |
| 未能完全描述或存在錯誤 | 2 | 部分描述 | 2 | 無(效)修正建議 | 0 |
| 完全錯誤 | 0 | 沒有詳細描述 | 0 |  |  |
| 邏輯錯誤判斷 | 邏輯錯誤不易察覺且可能無程式報錯，但會導致程式結果不如預期，故希望LLM可以根據題目找出邏輯錯誤。 | **準確性** | **4** | **詳細性** | **4** | **修正建議** | **2** |
| 正確描述程式碼的邏輯結構 | 4 | 詳細描述，包括主要邏輯和次要邏輯 | 4 | 提供具體的修正建議 | 2 |
| 未能完全描述或存在錯誤 | 2 | 部分描述 | 2 | 無(效)修正建議 | 0 |
| 完全錯誤 | 0 | 沒有詳細描述 | 0 |  |  |
| 程式碼優化(性能、效率、演算法) | 程式碼常常有可讀性低、時間複雜度高等問題，故希望LLM可以幫助提供高可讀性、高效能的程式碼。 | **準確性** | **4** | **詳細性** | **4** | **效能改善** | **2** |
| 提供有效的程式碼 | 4 | 詳細描述，包括主要邏輯和次要邏輯 | 4 | 優化後的程式碼性能或可讀性確實得到改善 | 2 |
| 提供的程式碼無法運作 | 0 | 部分描述 | 2 | 程式碼性能或可讀性皆無得到改善 | 0 |
|  |  | 沒有詳細描述 | 0 |  |  |
| 清晰的描述程式碼邏輯結構 | 有時學生看不懂演算法或函數的功能或結構，希望LLM可以幫助學生解析程式碼。 | **演算法準確性** | **5** | **詳細性** | **5** |  |  |
| 正確判斷演算法名稱與功能 | 5 | 詳細描述，包括主要邏輯和次要邏輯 | 5 |  |  |
| 無判斷出演算法名稱 | 3 | 部分描述 | 3 |  |  |
| 無敘述演算法功能 | 2 | 沒有詳細描述 | 0 |  |  |
| 無法判斷演算法名稱與功能 | 0 |  |  |  |  |
| 完成指定功能程式碼 | 當學習程式遇到瓶頸時或需要示範成功的程式時，學習才能有所成長。故LLM應該要能做到這樣的助教功能。 | **題目理解** | **2** | **詳細性** | **4** | **可運作性** | **4** |
| 正確理解題目要求 | 2 | 詳細描述，包括主要邏輯和次要邏輯 | 4 | 提供有效的程式碼 | 4 |
| 無法理解題目要求 | 0 | 部分描述 | 2 | 提供的程式碼無法運作、無效 | 0 |
|  |  | 沒有詳細描述 | 0 |  |  |

20240814: 我的建議如下，請參考。

1. 首先，你很用功+用心+努力，老師就以碩士生的規格來引導你的學習。
2. 我隨手將整個文件表格化，方便我整體性查看你的評分邏輯內容。所以碩士生應該要自行將資料整理成類似上述的內容，並呈現+說明給其他人參考。
3. 整體問題
   1. 應該是針對中小學生為主，還是所有初學者? 請先界定使用族群。
   2. 若使用族群確認了，所有內容應該都會有所異動。例如，我目前預設是界定所有初學者且不論年紀(老實說，這個範圍非常大，很難界定，你再想想嚕)，故我修正所有【目的】的內容。
   3. 接著，我們來看【判斷項目】，其中，初學者應該不會了解【程式碼優化(性能、效率、演算法)】這件事，因為能將程式正常運行就已經不容易了。所以這一樣的判斷對初學者是否有意義，待思考與討論。
   4. 然後是評分項目，我幾乎沒有看過每個評分項不同的問卷(或許是我孤陋寡聞，論文看太少吧~~~)，個人建議每項都是5級分或10級分，不應該是有多有少，且每個項目都是五等份或三等份(目前你的設計大多是三等份，偶而有兩等份)，就是要盡量等價量化所有的評分項目，否則就會偏失到某一個評量的維度，可能研究不用實作，單單看問卷就會知道可能的答案傾向，這也是一般問卷很容易發生的問題，這樣就不客觀了。
   5. 以下是我試著改其中最後一個【判斷項目】的內容讓你參考，若我來設計，大約是這個樣子，請參考。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 完成指定功能程式碼 | 當學習程式遇到瓶頸時或需要示範成功的程式時，學習才能有所成長。故LLM應該要能做到這樣的助教功能。 | **題目理解** | **3** | **詳細性** | **3** | **可執行性** | **3** |
| 正確理解全部的題目要求 | 3 | 描述程式碼所有邏輯說明內容 | 3 | 程式碼編譯無誤，但執行結果正確 | 3 |
| 正確理解部分(約60%)的題目要求， | 2 | 描述程式碼部分(約60%)邏輯說明內容 | 2 | 程式碼編譯無誤，但執行結果有誤 | 2 |
| 正確理解部分(約30%)的題目要求 | 1 | 描述程式碼部分(約30%)邏輯說明內容 | 1 | 程式碼編譯無誤，但無法執行 | 1 |
| 無法理解題目要求 | 0 | 沒有描述程式碼邏輯說明內容 | 0 | 程式碼編譯有錯誤，無法執行 | 0 |

1. 其他叮嚀
   1. 請盡速再改一版給我看喔~~~最好是一天以內~~~
   2. 請記得，Garbage in, Garbage out；很多事情在一開始就要規劃好，否則就會發生類似的狀況，還需要回頭修改原本的設計內容，導致大家都很費心費力，也延宕了時程；所以類似的設計需要你花點心思處理。
   3. 請記得，不要妄自菲薄，成功或許有捷徑，但我卻鮮少遇到捷徑，所以還是要腳踏實地好好做人、做事、做研究等等。
   4. 您很優秀，成績也維持得很好，希望您的目前的學習可以為您日後的成長提供一些必要的養分，激發出你的潛能，繼續發揮你的戰力，我與立偉老師都會在旁邊一起鼓勵大家的，一定要加油。