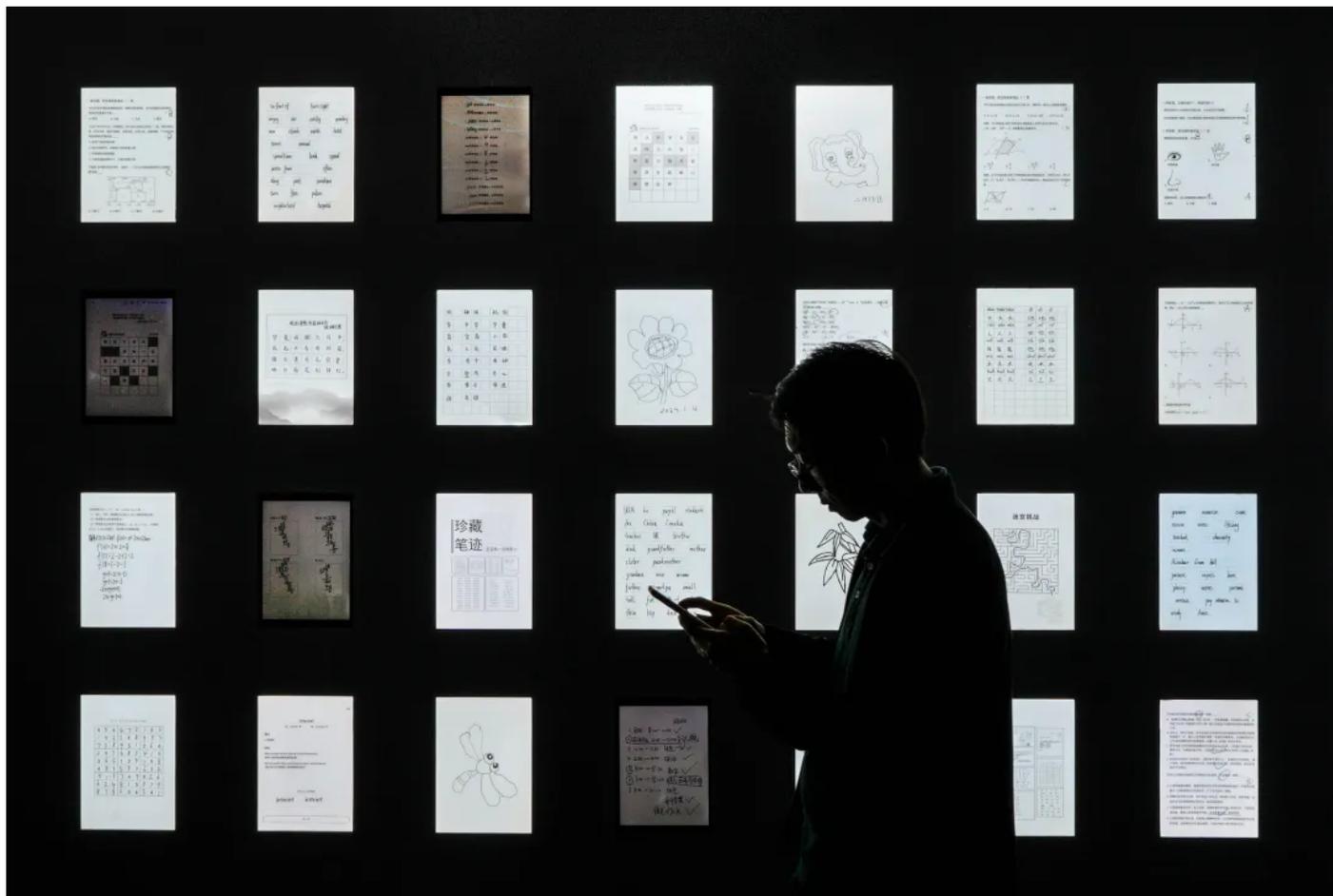


時薪20美元，人工智能背後的「零工做題家」

餵養機器像個打遊戲的過程，當機器夠聰明了，後面的做題家就可以自我消滅。



Andy Wong/AP/

近年我們迎來了一波科技大爆發：人工智能、大數據和機械人似乎從科幻般的遙遠未來被拉到我們面前，眼看就要對我們的生活造成巨大影響。都說人類站到食物鏈頂層是因為懂得使用工具，但這些工具會如何倒過來模塑人類社會？「工具人」欄目不定期刊出探討科技、社會與人文的深度報道，請點擊閱讀。

今年三月，來自台灣的 Vicky，在求職網站上看到以下廣告：

「Outlier 正在尋找具備繁體中文流利程度的優秀作家，以協助訓練生成式人工智慧模型。這是一個遠端的自由工作機會，工時彈性，因此您可以在適合您的時間工作。」

這份兼職提供每小時22.5美元的「平均薪資」，「理想專業背景的示例」包含譯者、文案作者、記者、編輯，以及「人文學科或寫作相關領域」的畢業生。用中文撰寫的招聘廣告生硬且不自然，帶有明顯機器翻譯痕跡。Vicky 覺得，他們真的挺需要中文母語者，來訓練一下這個 AI。

Vicky 當時在英國攻讀第二個碩士，正值學期間空檔，加上對當前火熱的人工智能產業的好奇心，她決定投遞履歷。很快地，她通過篩選，進入公司平台，卻沒有任何工作可做。她求助 Reddit 等網路論壇，發現許多人都有相同經歷。整件事彷彿一場詐騙，但不清楚到底被騙了什麼。

九月，Vicky 拿到學位後回了台灣，求職卻屢屢碰壁。她想起這份兼職，重新打開 Outlier 平台，發現一系列訓練課程。求職不利、和父母一同住在南部鄉下老家的 Vicky，有大把的空閒時間。在完成幾個（不支薪）訓練後，她終於接到第一個任務，或套用公司內部的說法，可以開始「做題」了。至今 Vicky 仍未找到正職，卻在 Outlier 上累積了大約2000美元的收入。

這份工作反映了矽谷科技業嘗試將人工智能訓練「零工化」的產業趨勢。Outlier 的背後，是總部位於舊金山、估值約140億美元的人工智能公司 Scale AI。該公司提供數據標註服務，客戶包含 OpenAI、Meta、微軟、輝達等科技巨頭。公司網站上，Scale AI 聲稱能幫助客戶「使用最好的數據造出最好的模型」；這些「最好的數據」，實際上是透過層層轉包，由像 Vicky 這樣的「零工做題家」訓練和提供。

規模最大的 Outlier 外，類似的零工平台還有許多，如 DataAnnotation、SuperAnnotate、CrowdGen、Welocalize、Alignerr。如果當前人工智能發展的瓶頸，是缺乏高品質的訓練資料和人類反饋，那矽谷提出的解決方案，就是搭建跨國平台，將人工智能訓練拆解為一個個40分鐘左右的單項任務，以自由接案、按件計酬的方式，在全球範圍內招聘具高等教育水平的人才。

於是，追求自由移動的高學歷人才，和自由的全遠端工作一拍即合，在「流動性」上達成共識，成為零工經濟的最新型態。AI 訓練的零工化、普及化，也讓被排除在人工智能浪潮外的人文學科畢業生，有了一窺當前最火熱產業的機會。

當文學系學生碰上人工智能，試圖教會大型語言模型說人話，他們卻發現，擋在前頭的，不是技術限制，而是比 AI 還要機械的人、比語言模型還不透明的平台規則。

當數位遊牧遇上全球零工平台

「跟其他同等教育水平能找到的工作比，這份工作的薪水大概是兩到三倍，甚至四倍。」 Neal 告訴我。

Neal 是中國大陸北部一所大學的大三學生，主修英語語言文學，副修應用語言學。十月初，他無意間在社群媒體上滑到 DataAnnotation 的廣告，招募 AI 訓練員。「薪資非常誘人，最低有 20 美元一小時，我都覺得看起來像詐騙。」他說道。Neal 所在省分的法定最低時薪，同樣是 20 元左右，單位卻是人民幣。



Andy Wong/AP/

受優渥薪資吸引，Neal 投遞了履歷，並花費四個小時完成了刁鑽的線上測試，試題包含評估 AI 生成的回應、續寫小說，以及英語朗讀。他成功通過測試，開始接到源源不絕的任務。兩個月來，Neal 每天工作四到五小時，假日有時間的話做得更久，至今已累積約5000美元的收入。

他也將這份兼職推薦給周遭朋友，卻少有人通過線上測試。Neal 反思：「我的教育背景，無論是語言學還是文學這一塊，還是比較相關的，能讓我想出一些比較高質量的 prompt（指令）。」

許多任務要求「挑戰 AI 智商的天花板」，這時候就必須「敦促」語言模型，下一些更專門的指令。Neal 舉例：「像我昨天就問它（語言模型）一篇 D.H. 勞倫斯的小說，要它總結內容。小說題目叫〈參孫和大利拉〉，不是這個作者非常有名的小說，然後 AI 就編造了一個故事，給出完全十萬八千里的回答。」

專業知識以外，語言能力也相當重要。多數時候，AI 訓練員使用中文和有待訓練的語言模型對話，模型以中文回應，做題家再對模型的回應進行評級，並以英文說明理由。測試和任務指令也都是英

文。Neal 評估，做這份工作需要母語者的中文水平，以及「雅思均分6.5」的英語水平。

透過 Outlier、DataAnnotation 等一系列零工平台，科技公司得以在全球範圍內，蒐羅具雙語能力、擁有人文學科學位的人才，如人在瑞典的 Sharon。

Sharon 同樣來自大陸，大學就讀中文系，擁有比較文學的碩士學位。在出版社做了幾年編輯後，她於去年遠赴瑞典，攻讀第二個碩士。Sharon 必須打工賺取生活費，但由於不會瑞典語，一直無法在當地找到工作。

今年七月，Sharon 在求職網站上發現 Outlier「AI 訓練員」招聘訊息，她馬上投了履歷，卻遲遲沒有回音。直到十月，她意外收到邀請入職的郵件，進入平台，做了好幾個不支薪的訓練課程後，她終於拿到任務，成為平台上幾千名簡體中文「做題家」之一。工作近兩個月來，Sharon 累積約 2000 美元的收入。

做題滿一定時數後，平台會給予部分工作者「內推」的名額，若內推進來的新人，也做滿 10 個小時，推薦者可獲得小額獎金。內推制度促使不少工作者將這份兼職分享至社群媒體，尋找有意加入的新人，賺取推薦獎金。某些貼文甚至收穫上百個贊，在小紅書上引發廣泛討論。

Sharon 推薦了大概十個人，其中只有一人成功讓她拿到獎金。「可能挺多人會卡在開頭吧，我個人的原因是瑞典這邊工作比較難找，我不介意前期要投入比較多的免費勞動，因為我的選擇很少。」她解釋：「如果一個人有很多選擇，可能這對他來說就不是一個好的工作。」

事實證明，這份全遠端兼職，對那些受過高等教育、卻不見得能找到匹配工作的人文學科畢業生來說，確實有著莫大的吸引力。如果暫時不想脫下孔乙己的長衫，那何不繼續當個做題家，還能賺外快？

一名人在紐西蘭的受訪者認為，這份工作非常適合留學生，一方面留學生具備雙語能力，另一方面彈性工時、居家辦公的自由度，能讓留學生兼顧課業和打工。Vicky 也是在留學英國期間，選擇向 Outlier 投履歷。雖然 Vicky 做的是繁體中文項目，與簡體中文相比，符合資格的做題家人數較少，她仍感受到非常多人在競爭這份工作。

「不論是想賺錢的人，想做兼職的人，夢想數位遊牧的人，或是跟我一樣好奇然後就進去了的人，很多人在搶很少的項目。」她說道。

不穩定的題目供給，不透明的平台規則

當全世界的工作者投入到這份兼職，但平台提供的任務數量有限，便導致僧多粥少，做題家常面臨無題可做的窘境。此狀態稱作「任務空檔」（empty queue），簡稱 EQ。

Reddit 上的「outlier_ai」看板，擁有 25000 名成員，EQ 是板友最常討論的問題之一。有些工作者因沒能通過測試而 EQ，有些因做題水平不佳而 EQ，但絕大多數的 EQ 來得莫名其妙，可能昨天做得好好的，今天就 EQ，或明明通過測試，卻自始自終都沒接到任務。怎樣才能有題可做，是一門玄學。

Sharon 注意到，某些簡中項目的工作組，擁有高達兩三千名做題家，這就註定不是人人都能接到工作。她猜測：「你更符合他們的要求和期待的話，你帳號的權重就會高一些，你獲得的任務量會更穩定一些。但整體來說，還是相當不透明，不知道到底有多少項目，項目有多少人在做，不知道評分規則到底是什麼。」

若做題品質長期保持在高水平，做題家能夠晉升為「審查員」（reviewer），批改其他人交上來的答案，並給予評分。做題家若長期被審查員打低分，可能喪失做題資格。哪怕對審查意見不服，也無法要求重審。Vicky 曾因不認同審查員的評分，對 Outlier 管理人員提出申訴，對方傳達的回應是：「你不要去反對他們的意見，如果你想繼續做這份工作，遵循他們的意見是最快的方式。」



Alexander Manzyuk/Reuters/

訪問前一週，Sharon 剛剛晉升為審查員。她體驗到審查員確實有更多權重：「不僅體現在任務更多，額外獎金更多，還包括優先被分配題目。一般獲得審查員資格後，就會進入該項目專屬於審查員的更小群組，我在的那個群組只有50人。有時候項目管理者會分享一些『內部資訊』，例如提前告知該項目有一波新題近期將上線，審查員能優先看到這些題目。」

即便如此，多位受訪者都表示，他們曾在 Reddit 等論壇上，看到有人明明已經做到審查員，帳號仍無緣無故被封禁。

Neal 對 DataAnnotation 內部人員的評價是「高冷」，覺得他們不太愛回答問題，不積極回覆郵件，開除人卻毫不手軟。「如果你一直符合他們的要求，工作項目上還是比較穩定的，能分到的工作也比較多。但如果他們覺得你可能哪方面做錯了，或不符合他們的要求，他們也會毫不留情地把你踢掉。」他說道：「具體不清楚評判標準是什麼，很多人一夜之間、也沒收到任何郵件，他們的工作就 gone 了。」

不透明的規則、不穩定的工作來源，導致工作者和平台，長期處在嚴重的權力不對等中。工作者在做題之外，往往還要付出大量的詮釋勞動，去揣測平台的想法、狀態。

三週前，由於多次從審查員那邊拿到低分，Vicky 喪失某一項目的做題資格。在訪問前一天，她卻意外收到 Outlier 的郵件，邀請她參加「重考」，若順利通過，能重新回到該項目組。她的理解是：「因為這個審查員的制度，他們篩掉太多潛在的貢獻者，合格的人數太少了，導致消耗任務的速度太慢，所以才又開了一個後門。」

當我問 Vicky 會不會參加重考，她回答：「我搞不好重考也不會過啊。」複習任務指令、重新參加測試需花費2到3個小時，沒有薪水可拿。

當缺乏線索、無從揣測的時候，工作者只能認命。Vicky 有次單日做了整整12個小時，「我不知道這個項目會持續多久，因此會想要在最短的時間內，把收入提高，它有可能兩三天就結束了，我不知道。」這樣一種「限量的急迫感」，究竟是僧多粥少的自然結果，還是平台方刻意塑造，同樣不得而知。

Sharon 也有類似反思：「工作者非常被動，沒有選擇權，只能被動地等待任務來不來。作為一個被動方，你必須保持一個定期追蹤它們的習慣，每天花個幾分鐘或半個小時，這會變成你生活裏一個常規的事情。」

「如果一個人不想承受這種狀態，他就只能選擇離開。」

過渡性、遊戲化，人工智能背後的「幽靈工作」

平台上的零工做題家來來去去，水平參差不齊。Vicky 觀察：「他們（Outlier）還是不斷在僱用人，整個群體裏有像我一樣資深的人，同時也有一群滾動式、不斷進來的新人，搞不清楚狀況。」管理人員因此得一再回答同樣的問題，新人得一再經歷非常掙扎的過程，才能變成一個「可以用的工作者」。

「我一拿到正職工作，就不會出現在那裏了。」Vicky 說道。「我還是承認這份兼職其實非常不穩定，我畢業後一定會去找其他的工作。」Sharon 也說。

中國社科院青年學者孫萍在《過渡勞動：平台經濟下的外賣騎手》中，指出流動性、過渡性是中國外賣行業的基本樣態。當被問到為什麼送外賣，「過渡一下」是絕大多數外賣員的回答。孫萍團隊於2023年針對北京地區超過千名外賣員進行調查，僅13%受訪者表示願意一直送外賣，超過八成的人表示自己會在兩年內換工作。個別站點一年內的騎手流動率，介乎7成至9成之間。

孫萍將零工經濟下的「過渡勞動」，形容為一種「懸浮狀態」。零工工作者總想著要離開平台，進入下一個也許更穩定的生命階段。對平台而言，卻是「過渡成為常態」，儘管每天新人來、舊人去，勞動力的供給仍源源不斷。



Tingshu Wang/Reuters/

而一旦個體進入平台，勞動與生活的界線便逐漸模糊。孫萍寫道：「（外賣騎手）有訂單時工作，無訂單時休息，工作與生活的隨機轉換成為零工勞動者的日常。」正如做題家們必須在無題可做和快速刷題間來回迅速切換。

AI 訓練平台也透過「遊戲化」的激勵措施，來調節做題家的做題速率。Neal 告訴我，DataAnnotation 的項目多以7天為一個週期，當項目被放出來的第一天，費率會是最低的，到了第三、第四天，費率就往上加，若到了第五、第六天，還剩下很多任務，那說不好就給到30美元一小時。他做過最高的費率是37.5美元一小時。

Outlier 則是在一般任務之外，偶爾會出現限定時間的挑戰（mission），若完成挑戰，可獲得額外獎金。挑戰內容可能是：「假日完成至少1個項目」、「單日工作滿3小時」、「完成特定項目的20個任務」。若想最大化收入，工作者必須時時注意推出了哪些挑戰，儘可能配合平台方制定的工作節奏。哪怕這個節奏既不穩定、也無法預測，某些狀況下，卻帶來工作的快感。

「這個平台很有意思的是，你做完一個任務，它給的錢會立刻出現在你的帳戶，因此很有打遊戲的快感。」Sharon 說道：「我剛開始做的時候就有點上頭，我每做完一個，就要看一下我的帳戶，看到帳戶的錢有在動，就再回去做下一個。」

「玩遊戲我們累積的金幣不是都是虛擬金幣嗎，但做這個累積的是真實的金幣。」

絕大多數受訪者都同意，他們在做的 AI 訓練員，是一種零工工作。Neal 認為這份工作和外賣員、叫車司機類似，「自由度比較高，想做就做，不想做的時候就可以給自己放假」，「不同點就是 DataAnnotation 更偏向腦力活動，主要是智力上的輸出，而像滴滴、外賣這些更偏向體力活動」。

由於和平台方不存在僱傭關係，零工做題家不受任何法規的保護。薪資一般都轉帳到第三方平台如 PayPal 裏，工作者再自行提領。是否該報稅、該如何報稅的問題，也同樣由工作者承擔。

Sharon 便花了不少心力，試圖弄清瑞典的勞動和稅務法規。她告訴我：「在瑞典拿學生簽的話，兼職時間上限是每週20小時，具體操作起來有一定彈性。但需要考慮累積收入達到一定數額就得報稅。」不過，她對銀行要如何確認透過 PayPal 匯入的收入屬於「勞動所得」感到困惑。累積收入達5000美元的 Neal，還不太清楚他需不需要報稅。

從體力到腦力勞動，「零工化」的趨勢正在不斷擴張。而回顧歷史，會發現人工智能背後的人類勞動力，也正在經歷從血汗工廠到零工經濟的變化。

Outlier 的背後，是總部位於美國舊金山的人工智能公司 Scale AI。Outlier 的《隱私政策》有以下文字：「（本政策）凡提及『Outlier』、『我』、『我們』，皆指 Smart Ecosystem 及其附屬公司。」據《華爾街郵報》披露的登記文件，Smart Ecosystem 為 Scale AI 全資持有的子公司。

更早之前，Scale AI 於2017年創立 Remotasks 公司，將較低門檻的數據標註工作，外包到菲律賓、肯亞、奈及利亞和巴基斯坦等發展中國家，僱用低廉勞動力。在這些國家，Remotasks 存在拖欠工資、工資過低，以及無預警解僱員工等爭議。《華爾街郵報》形容，人工智能浪潮背後，是在「數位血汗工廠」裏工作的大量勞動者。

人類學家 Mary Gary 和計算機科學家 Siddharth Suri 則在《銷聲匿跡：數字化工作的真正未來》中，將驅動人工智能運作的隱形勞動，稱作「幽靈工作」（ghost work）。這些工作往往被刻意隱藏，以製造人工智能無所不能的假象。

如今，Scale AI 嘗試以零工經濟的模式，搜刮更高教育水平的勞動力，在全球僱用一群「零工做題家」。他們是以時間和知識餵養機器的新一代幽靈。做題家的工作具有自我消滅的特性，一旦語言模型達到母語使用者水平，就再也不需要人類的反饋和幫助。

但根據第一線 AI 訓練員的判斷，這天不會那麼快到來。

AI 能學會說台灣諧音梗嗎？

「如果 LLM（大型語言模型）都是這樣訓練出來的，我對這項技術好像也沒有什麼太大的信心了。」這是 Vicky 和 AI 密切互動三個月後的感想。

作為對語言高度敏感的繁體中文使用者，Vicky 相當在意 AI 回應中那些不自然、太像機器的表達，也非常認真地看待在地化的目標，努力讓 AI 講話更像台灣人。然而，阻擋她教會 AI 「說人話」的，不是技術限制，而是必須遵守的規則（guidelines），以及將規則強加於做題家的審查員和管理人員。

在其中一個項目裏，做題家必須從七個不同面向，對模型生成的回應給出重大問題、次要問題和沒有問題三種評級。七個面向為在地化、指令遵循、真實性、簡潔、寫作風格與腔調、無害，以及整體品質。關於如何區分這七個面向，訓練文件有近20頁的說明。為了評分標準化而定下的規則，卻往往有不足之處。

「有時候我會覺得某個回應以 guidelines 來說不錯，但是語氣上怪怪的，我就會很想要寫下來，」Vicky 說道，如某些明顯的冗詞贅字。她對語言的細膩想法，卻往往不被採納。「他們希望我們遵循 guidelines 去評分，有點在削弱我們對日常用語的主觀意見。」某些管理人員更指示，在寫評級理由時，不要使用第一人稱「我」，只能用第三人稱寫作，如「根據規則 X，回應 Y 有問題 Z」。

當做題家檢視 AI 是否忠於人類指令，審查員也在檢查做題家是否遵循規則。有時，做題家面臨兩難處境：究竟是遵循規則努力賺錢，還是貢獻自己對語言的「主觀意見」，即便那可能不是平台方想要的。



Ann Wang/Reuters/

做題家當然也有失手的時候。有一次，Vicky 要求模型將台灣22個縣市，依人口密度分成高、中、低三類。模型給出的兩個回應中，A 回應分錯三個縣市，B 回應僅分錯一個，但 B 回應包含一些語言方面的問題。在評價兩個回應何者為佳時，Vickey 選了 B，因為指令的核心要求是將縣市分類，而錯一個比錯三個好。

這個答案卻被審查員評為「非常糟糕」（very poor），只拿到1分（滿分為5分）。審查意見表示：「回應 A 有『指令遵循』方面的重大問題，回應 B 有『指令遵循』方面的重大問題和『寫作風格與腔調』方面的次要問題，一個重大問題優於一個重大問題加一個次要問題，所以 A 是更好的回應。」

哪怕不符直覺，Vicky 接受這個審查意見，她確實沒能遵循規則。但另一個例子涉及刻在台灣人基因裏的諧音梗，她咽不下這口氣。

該任務要求讓語言模型生成「詼諧幽默」且在地化的回應。Vicky 苦思良久後，決定叫 AI 想一系列手搖飲料店的諧音梗店名，如著名的八曜和茶（諧音「爸要喝茶」）。但 AI 給出的店名，在 Vicky 看來都不好笑，「不符合我們台灣人對諧音梗的認知」，她將這點寫在評級理由中交了上去。

審查員卻認為「不好笑」太過主觀，不構成一個重大問題，把做題結果打了低分。Vicky 非常不甘心：「AI 寫不出幽默的句子我完全可以理解，它本來就不夠在地化，所以才需要我們人的回饋嘛。但如果今天我們這樣的回饋都被審查員否定，那我實在不知道該怎麼處理這樣的任務。」

「那你覺得 AI 哪天能學會說台灣諧音梗嗎？」我問道。「我覺得很難，如果他們繼續給我們這麼多限制的話。」她解釋自己的困境：「如果你一直遵循規則，那 AI 生成的回答就會像是機器人，我們加入個人意見，就是希望 AI 往更『人性化』的方向發展，但審查員又把我們規則的方向拉。」

Sharon 也分享了類似經驗。她有時接到一些創作類的任務，要修改模型生成的小說甚至詩歌。

「不能說 AI 在語言表達上有什麼問題，但寫出來的故事非常幼稚、爛俗。」她因此需要對 AI 生成的小說進行全盤修改。更麻煩的是詩歌，「由於 AI 很難理解意象、隱喻之類的表達，它寫詩的能力是非常差的。」

Sharon 有時也和審查員的意見發生衝突。有一次，模型生成的回應中有「兩個麻雀」的說法，Sharon 認為沒有問題，審查員卻指出應該是「兩隻麻雀」，給了她最低的1分。

Sharon 並不服氣：「你只要去百度或 Google，就可以找到大量的例子寫『兩個麻雀』。在人與人對話的語境裏，中文的表達是非常靈活的，日常聊天裏經常會有不符合語法規範的說法，但這部分到底要不要去修改，就是一個問題。」

「那你有沒有想過，審查你的人有可能不是一個人，而是一個 AI？」我問道。Sharon 認為不是，因為 AI 建立在海量數據之上，而網路上有大量「兩個麻雀」的說法；反而人類由於從小到大接受的語文教育，才會堅持只能用「只」、不能用「個」去表達鳥類。

「我認為審查員肯定是人，因為 AI 沒有這麼機械，雖然這麼說很奇怪。」 Sharon 說完笑了。

(應受訪者要求，Vicky、Sharon 為化名)