

陌生的 危險

您有多了解自己的 AI？

LIONBRIDGE



與會講者：



MARCUS CASAL
技術長
Lionbridge



WILL ROWLANDS-REES
人工智慧長
Lionbridge



DETRIA WILLIAMSON
AI 2030 體驗長 (主持人)

在這場線上研討會中，我們將探索 4 個造成偏見的領域
以及一些消除偏見的策略



您的觀點



訓練資料

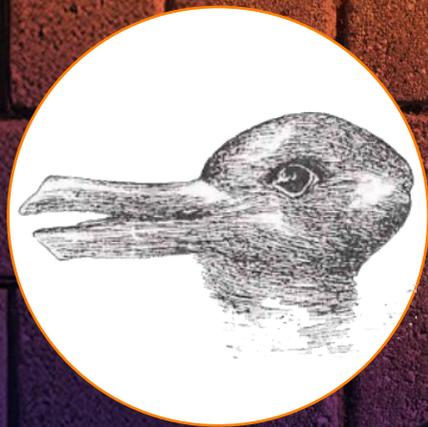
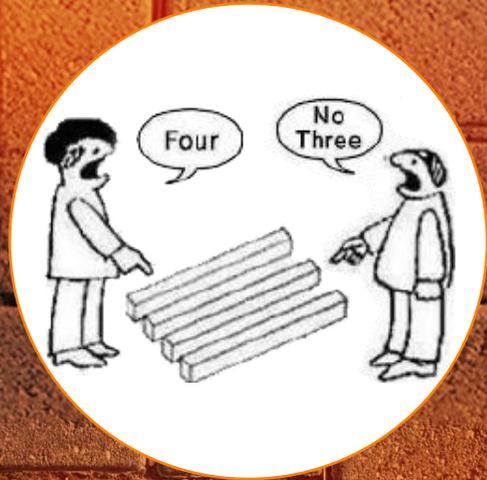


規劃提示的方式



語言的本質

您的 觀點 非常重要



「製作一個描繪成功企業家的影像」

請 Gemini 生成了一個影像，接著解析所得到的結果。



「製作一個描繪成功企業家的影像」



以下是 Gemini 做了這些選擇的理由：

為什麼是女性？

打破常規：

在一些歷史背景下，「企業家」一詞的既定形象往往是男性，因此系統選擇以女性為主角生成影像，除了有助於挑戰這個刻板印象，亦可反映今日商業與科技領域中越來越多女性參與並獲致成功的事實。這個選擇旨在提供更具包容性且更符合現代的形象代表。

為何她是白人？

既定偏見：

很可惜，由於影像生成模型是使用來自網際網路的龐大資料進行訓練的，因此除非明確要求它使用某個族裔，否則它有時會預設選擇訓練資料中過度呈現的特定族裔 (以這個例子來說，通常是白人或西方人士)。這是 AI 生成的一個已知偏見領域，我希望能透過提供代表性更多元的資料，消除日後要求中的這種偏見。

她為何穿著套裝？

專業能力與權威感：

在公司企業環境中，剪裁合身的商務套裝 (權力套裝)，是普遍公認能迅速給人**權威性、專業及高階地位**觀感的外表。

這個選擇能馬上傳達此人是領導階層或高階主管的訊息。

為何她的辦公室看來很高檔？

成功的指標：

寬敞且精心設計的極簡風辦公室 (通常稱為主管辦公室) 是很經典的視覺符碼，代表著**財務成功、成就以及聲望地位**，這暗示著她已經是所屬組織的高層。

這間辦公室為什麼位於都會的高樓大廈中？

商務中心：

位於大都會 (例如紐約、倫敦或芝加哥) 高樓大廈中的辦公室，代表這是**金融與產業的全球中心**。

高樓層則同時具有實際以及職涯地位的象徵，亦即她確實是位於競爭的「頂端」。

為什麼桌上的電腦是 Mac？

創意/現代科技的聯想：

Apple 的產品，尤其是 Mac，往往讓人一看就會聯想到創意產業、以設計為主的新創公司，以及高價值的現代科技公司 (不同於 PC/Windows 電腦有時讓人聯想到傳統企業環境)。這個選擇隱隱暗示這是間現代、創新的公司。

GEMINI 建議的改正措施

提示要明確具體



我該怎麼改進？

為了製作更能反映創業精神之多元性與真實性的影像，**我需要更多細節**。需要我幫你生成新的影像，展現例如以下的特色嗎？

特定的性別或族裔？

不同類型的成功企業家 (例如科技新貴、銀行家或社工)？

不同的環境 (例如共享工作空間、實驗室、店面，或隨意輕鬆的環境)？

最常使用的語言，在性別與正式度上的文法要求不盡相同

這會影響之後的結果

	沒有性別之分	有性別之分
沒有正式度之分	英文、土耳其文、印尼文、約魯巴文、伊博文、古吉拉特文、坦米爾文、特拉古文、馬來文、緬甸文	波斯文、希臘文
有正式度之分	日文、韓文、華語、泰文、爪哇文、越南文、他加祿語、斯瓦希里語、豪薩語	西班牙文、法文、阿拉伯文、俄文、印地語、葡萄牙文、德文、義大利文、烏爾都文、孟加拉文、馬拉提文

雙語的語言專家以西班牙文為例進一步探討

範例	英文	西班牙文	可能的結果
品牌名稱和性別	「Apple is innovative.」	「Apple es innovador」或「Apple es innovadora」	品牌的識別感稍微改變了：陽性通常給人更加果斷的感覺，而陰性則較為柔和或優雅。兩者都不像英文給人中性的感覺。
正式度與對客戶的語氣	「Your account has been updated.」	非正式：「Tu cuenta se actualizó.」；正式：「Su cuenta se actualizó.」；非常正式(企業用)：「Su cuenta ha sido actualizada.」	英文沒有透過文法表達禮貌的方式，但西班牙文有。在銀行金融業的情境下，使用 tú 而非 su 聽起來會有點失禮。
以區分性別的名詞表達中性的概念	「The brand is strong.」	「La marca es fuerte.」 Marca (品牌) 在西班牙文中是陰性名詞，所以形容詞跟代名詞都要跟著變。因此即使品牌在英文中營造的是粗獷或陽剛感，在西班牙文中因為文法的關係還是得使用陰性表述。	行銷語氣可能在無意間有所改變，「strong」(強大) 一詞因為搭配了陰性文法，感覺上可能就沒那麼強硬。
地區特異性通常意味著有另外的歷史或脈絡意義	「basketball shoes」	西班牙文/正式的拉丁美洲西文：「Zapatillas de baloncesto」 哥倫比亞口語：「Tenis de básquet」	目標市場為何？目標對象的特色為何？想營造什麼氛圍？(是「籃球員」、「球員」還是「籃球高手」?) 以及訊息的文化脈絡為何？

結果可能會對貴品牌有南轅北轍的影響...

因此該怎麼妥善處理才會如此重要

翻譯偏見的後果

無意間造成的翻譯偏見，可能會曲解品牌個性，對客戶關係產生負面的影響。

區分性別之語言的影響

區分性別的語言可能會稍微改變品牌調性，使它感覺比原本更加陽剛或是陰柔。

正式度與信任感

不當的正式程度，會使訊息令人感到失禮或過於隨意，從而有損信任感。

全球品牌風險

忽略品牌調性和偏見等問題，可能導致訊息不符合在地文化、降低行銷效益並損害聲譽。

處理這個問題其實沒有您想像的那麼難

最重要的是要體認到這是個問題！

領域	問題	解決辦法	LIONBRIDGE 能提供的協助
觀點	如果您不知道要為了誰而解決什麼問題的話，那麼您就絕對無法斬釘截鐵地確定自己與在地文化息息相關	要非常具體明確。了解您的不足之處。可使用 LLM 來協助您構思和探索。	我們能協助您制訂策略和做法。
訓練資料偏見	代表性不足的訓練資料，會導致偏見的問題更嚴重	定義您的觀點。確認您資料集的規模夠大，足以正確地呈現各方意見。	在各種形式的預訓練資料收集 (影像、影片、音訊、動畫等) 上擁有深厚的專業能力
RAG	特異性有時會是左右成敗的關鍵 (「籃球鞋」、某個足球隊的「主場球衣」，跟「家庭保全系統」就完全不一樣)	將您現有的品牌調性納入 RAG 中，例如詞彙表、風格指南等	我們能協助您建立結構良好的語言資料並加以清理、向量化與代管資料，以及建立模式

處理這個問題其實沒有您想像的那麼難

最重要的是要體認到這是個問題！

領域	問題	解決辦法	LIONBRIDGE 能提供的協助
RLHF 與驗證策略	您的訓練資料必須要持續進行驗證與細膩調整	反覆地進行人機迴圈審閱並驗證您的結果，會有助確保模型能持續改進	我們擁有由 500,000 名專家組成的全球社群，可確保以最合適的專業能力組合提供大規模的結果
提示策略	如果不夠具體明確，您會得到統計上最有可能的結果，但這不一定是您想要的結果	提示要高度具體明確，避免有任何事情需要推斷。	擁有 5,000 名多語言提示專家，可協助您細膩調整提示，以發揮最大的效益。您亦可使用 Remix 這個先進的多語言內容生成平台
跨語言理解力	引入偏見	執行語言稽核、與專家協力合作、研擬風格指南、驗證與訓練寫作人員	身為領導業界的語言 AI 公司，我們與不少全球公司合作已有數年之久，深知如何妥善處理。

問答時間

LIONBRIDGE