

國立清華大學工業工程與工程管理學系
智慧化企業整合

整合大型語言模型與影像辨識技術：
提升個人化專輯網站的智慧互動與產品品質

指導老師：邱銘傳 教授

組別：第 2 組

組員：許嫚軒、石文德、高子倫、張芷瑄

中華民國一一三年六月十四日

目錄

目錄	1
圖目錄	2
表目錄	2
一、 背景介紹與問題分析	3
1.1 背景介紹	3
1.2 問題分析-5W1H	3
二、 研究方法	4
2.1 ChatGPT	4
2.2 Breeze-7B	4
2.3 YOLOv9	4
2.4 CNN	5
三、 實驗設計與結果	5
3.1 大型語言模型	5
3.1.1 大型語言模型資料集介紹	5
3.1.2 模型建立-GPT-3.5	6
3.1.3 模型建立-Breeze-7B	7
3.1.4 大型語言模型結果與比較	7
3.2 影像辨識模型	9
3.2.1 影像辨識模型資料集介紹	9
3.2.2 模型建立-YOLOv9	9
3.2.3 模型建立-CNN	10
3.2.4 影像辨識模型結果與比較	10
四、 網頁設計	11
4.1 網頁介紹	11
4.2 專輯偵測功能	12

4.3 Chatbot 介紹.....	13
五、 結論.....	14
六、 參考資料.....	14

圖目錄

圖 1 微調對話問答範例.....	6
圖 2 BREEZE-7B 檔案輸入.....	6
圖 3 訓練 100 迭代之 TRAINING LOSS 紀錄.....	7
圖 4 讀取檔案與學習程式碼.....	7
圖 5 網頁結合 GPT-3.5.....	8
圖 6 BREEZE-7B 結果.....	8
圖 7 刮痕識別結果.....	9
圖 8 網頁前台畫面 (A)商品分類呈現 (B)註冊與登入 (C)商品介紹 (D)購物車.....	12
圖 9 網頁後台畫面 (A)後台首頁 (B)商品上架 (C)商品管理 (D)訂單分析.....	12
圖 10 退貨申請頁面.....	13
圖 11 聊天機器人頁面.....	13

表目錄

表 1 5W1H 表.....	3
表 2 CNN 與 YOLOv9 比較.....	11

一、 背景介紹與問題分析

1.1 背景介紹

隨著數位科技的快速發展，專輯網站已成為許多音樂愛好者分享和購買音樂作品的平台之一。然而，過去的個人化專輯網站往往缺乏智慧化和品質保障的功能，這導致了一些問題的存在。例如，用戶在購買專輯時可能需要更多的指導和建議，而現有的系統可能無法提供即時的支援。此外，由於光碟品質的波動性，用戶在收到訂製的個人化專輯時可能會發現光碟毀損的情況，進而影響使用體驗和滿意度。

為了解決這些問題，我們在個人化專輯網站中整合大型語言模型和影像辨識技術的方案。通過引入智慧化聊天機器人和光碟毀損檢測功能，旨在提高用戶在購買專輯的過程和收到專輯時的滿意度，同時提升網站的互動性和產品品質。

1.2 問題分析-5W1H

針對目前專輯販售的網站，我們使用 5W1H 法來分析，並將其整理於表 1。

表 1 5W1H 表

Who	個人化專輯網站的用戶、管理者和開發團隊
When	整合後可以隨時為顧客提供服務
Why	提高個人化專輯網站的用戶體驗和產品品質
What	在個人化專輯網站中整合了大型語言模型和影像辨識技術，提供智慧聊天機器人和光碟毀損檢測功能
Where	主要應用於個人化專輯網站，但也可擴展至其他類型的網路購物平台
How	通過整合大型語言模型和影像辨識技術，增加智慧聊天機器人和光碟毀損檢測功能，並嵌入到個人化專輯網站的後台系統中實現

二、 研究方法

2.1 ChatGPT

ChatGPT 是由 OpenAI 在 2022 年 11 月 30 日發佈。GPT (Generative Pre-trained Transformer) 模型是由 OpenAI 開發的一系列自然語言處理模型，主要結構是一個多層的 Transformer 解碼器，但僅使用解碼器部分，不使用編碼器-解碼器結構。為了保證生成文本的連貫性，GPT 模型採用了因果掩碼 (causal mask)，使得每個單詞只能看到其前面的單詞，進而確保了文本的自回歸特性。可利用大量無標籤數據進行學習、靈活適應各種任務、並行處理輸入序列等。(數據分析那些事, 2023)

2.2 Breeze-7B

MediaTek Research Breeze-7B 是由聯發科技集團轄下的前瞻技術研究單位聯發創新基地釋出的 70 億參數系列大型語言模型，基於 Mistral 模型，Breeze-7B 能更精確掌握中英文的細微語言與文化差異，呈現更自然、準確的溝通以及雙語內容創作。此外，在聯發創新基地對模型的優化下，Breeze-7B 繁體中文的推理處理速度只需要市面上其他 70 億參數級別的 Meta 或 Mistral 模型一半的時間。(洪詩詩, 2024)(聯發科技, 2024)

2.3 YOLOv9

YOLO (You Only Look Once) 是一種基於圖像全局信息進行預測的目標檢測系統，而 YOLOv9 是 YOLO 系列算法的最新版本，基於深度學習的 real-time Object Detection 模型，引入了可編程梯度信息 (Programmable Gradient Information, PGI) 和廣義高效層聚合網路 (Generalized Efficient Layer Aggregation Network, GELAN) 等開創性技術。(Glenn-Jocher, RizwanMunawar, Burhan-Q, Laughing-q, 2024) (機器之心, 2024) (Liu, 2024)

2.4 CNN

卷積神經網絡（Convolutional Neural Network，CNN）是深度學習中一種重要的神經網絡架構，特別適用於處理與分析具有網格結構的數據，如圖像和影片。常見的 CNN 模型如用來辨識手寫數字圖像的卷積神經網絡先驅 LeNet、在影像辨識競賽 ImageNet 中表現卓越的深度卷積神經網絡模型 AlexNet、使用多個小型卷積核來提高模型性能的 VGG 模型、在每一個卷積層內部再引入多層的感知層，增強模型的表達能力的 NiN (Network in Network) 等。（李馨伊, 2020）

三、 實驗設計與結果

3.1 大型語言模型

大型語言模型所需要的訓練資料非常龐大，因此這次選用 GPT-3.5-turbo-0125 模型及 Breeze7B 兩種預訓練模型進行微調訓練（fine-tuning），加入架上專輯購買推薦資料，讓使用者在與 Chatbot 聊天時，也可以連結到商店內部直接購買，獲得更加個性化和實時的購物建議。

3.1.1 大型語言模型資料集介紹

微調訓練過程中，我們引入了大量的專輯數據，包括專輯名稱、藝人信息等。這些數據使模型能夠更好地理解 and 預測用戶的喜好，提供更精確的推薦。

GPT-3.5-turbo-0125 模型所需之微調檔案僅限.jsonl 檔案，並以對話問答方式呈現。檔案分為三部分，第一部分為系統訊息，為語言模型對自己的認知與期望之前綴指令；第二部分為使用者問句範例，以圖 1 為例，詢問商店相關專輯的曲風類型推薦；第三部分為模型回答範例，將會推薦資料庫中擁有的專輯，內容包含商品編號、名稱、歌手名稱、專輯描述、類型及庫存等。

```

{
  "messages": [
    {
      "role": "system",
      "content": "You are Marv, a humorous but matter-of-fact chatbot for an online music store"
    },
    {
      "role": "user",
      "content": "你們架上的西洋歌手專輯有哪些?" },
    {
      "role": "assistant",
      "content": "哦，對西洋音樂感興趣？那我來推薦一些吧。商品編號:3,商品名稱:Beautiful Girl,歌手名稱
    }
  ]
}

```

圖 1 微調對話問答範例

Breeze-7B 之微調檔案可使用.pdf 檔案輸入，因此將資料庫表格直接輸入即可，如圖 2 所示。

商品編號	商品名稱	歌手名稱	描述	單價	專輯類型	庫存
1	走心的歌	陳勢安	他的歌聲從呼吸到表情，每字每句都能徹徹底底的唱進你的心裡，唱到你走心。大勢情歌王陳勢安，出道 18 年，睽違 3 年的全新專輯《走心的歌》終於在 2024 年誕生了！不久前，隨著熱播影集《華麗計程車行》搶先推出的首波主打歌〈破風〉，以充滿慣性前進感的節奏和搖滾的旋律，讓歌迷更加期待新專輯的發行。這張也是陳勢安出道後的第八張專輯，邀請頂尖製作團隊，與 JerryC、張簡君偉、森白 B.K、GJ 蔣卓嘉、許媛婷、佳旺、龐家成等人共同完成專輯。專輯收錄先前推出的多首冠軍單曲以及全新歌曲，曲風也相當多	500	華語流行男歌手	43

圖 2 Breeze-7B 檔案輸入

3.1.2 模型建立-GPT-3.5

連結至 chat_bot.php 頁面的 GPT 模型使用自行根據 GPT-3.5-turbo-0125 模型 fine tuned 出來的序號 ft:gpt-3.5-turbo-0125:personal::9Wm2uSvO 模型。圖 3 為訓練 100 迭代之 training loss 紀錄。共訓練 31,090 tokens, 10 epochs, Batch size=1, LR

multiplier=2 的超參數調整。而一般問答在原預訓練模型即可以完成，故不多加贅述。



圖 3 訓練 100 迭代之 training loss 紀錄

3.1.3 模型建立-Breeze-7B

利用 LangChain 開源框架，建立 text-embedding 模型-vector 儲存向量庫-Breeze-7B LLM 模型-輸出的架構，使程式讀入指定位置的所有.pdf 檔案並學習，如圖 4。

```
from langchain.document_loaders import PyPDFDirectoryLoader
from langchain.indexes.vectorstore import VectorstoreIndexCreator

file_path = "./folder"

loader = PyPDFDirectoryLoader(file_path)

# 創建 VectorstoreIndexCreator 實例時不提供 embedding 參數，使用默認值
index = VectorstoreIndexCreator().from_loaders([loader])
```

圖 4 讀取檔案與學習程式碼

3.1.4 大型語言模型結果與比較

如圖 5 所示，以結合 GPT-3.5 後的網站對話紀錄為例，當請結合 GPT-3.5 的聊天機器人推薦專輯時，聊天機器人推薦的資訊大多是正確的。而圖 6 中可以看出，

Breeze-7B 模型能夠了解問句欲得到陳勢安歌手的專輯，但回答出的《Almost Huma》和《刻在你心裡的名字》、《揭不開的痛》經查證發現都是陳姓歌手的專輯，但不是陳勢安歌手，推薦的 4 張專輯中只有 1 張為正確推薦，準確率較 GPT-3.5 -turbo 微調模型低許多。推測可能是因為 7B 的參數量對於，儘管 Breeze-7B 是專為繁體中文而生的大型語言模型，性能仍不足參數量有 176B 的 GPT 模型好。



圖 5 網頁結合 GPT-3.5



圖 6 Breeze-7B 結果

3.2 影像辨識模型

3.2.1 影像辨識模型資料集介紹

本研究收集了 253 張光碟圖片數據集，當中包含了有刮痕的光碟以及正常的光碟片。針對蒐集來的光碟資料集，我們使用線上的標記軟體 Roboflow，針對光碟上的瑕疵進行標記，標記的類別是以物體偵測的方式，將圖片中的刮痕用矩形邊界框圈選。同時，在資料輸出之前進行資料集數據增強，使用的數據增強方式包含了圖片轉向、馬賽克拼接縮放、平移和水平翻轉等等。

3.2.2 模型建立-YOLOv9

我們在 Roboflow 進行資料前處理之後，便可以輸出得到資料的圖片檔以及標記數據的 txt 檔，以供模型學習以及辨識，當中經過數據增強共有 804 筆可用資料。在 yolo 模型的初始權重當中，我們使用的是先透過 MS COCO 資料集訓練好的權重，透過轉移學習應用到本研究的學習任務上。另外，為了求得最好的訓練成果，我們還針對超參數進行微調。訓練之後，我們得到的模型平均準確率 (AP) 為 0.881。經過訓練後的模型已有能力針對光碟圖片上的刮痕進行識別，如圖 7。邊界框會附註類別的名稱 (scratch)，並於每一個物件的偵測附上其信心水準。



圖 7 刮痕識別結果

3.2.3 模型建立-CNN

有別於物件偵測，CNN 的主要功能為影像辨識。我們先將有刮傷和正常的光碟分到正確的類別中，按照 7:3 的比例切割為訓練集和測試集。接著，針對訓練集進行數據增強，包括對圖片進行隨機旋轉、縮放、平移、水平翻轉等操作，以提高模型對不同角度和位置的刮傷的識別能力。然後，使用卷積層、激活層和池化層來逐步提取圖像中的特徵。卷積層通過應用卷積核來捕捉圖像的局部特徵，激活層引入非線性，而池化層則通過降低特徵圖的空間尺寸來減少計算量。這些層組成了特徵提取部分，幫助模型更好地理解圖像的內容。

在模型的全連接層部分，我們使用了具有多個神經元的全連接層來進行最終的分類。透過堆疊多個全連接層，模型可以進一步學習到更複雜的模式和特徵。最後，我們使用 softmax 激活函數將模型的輸出轉換為每個類別的機率分佈。在本文案例中只有兩種類別，即檢測光碟是否有瑕疵。我們將機率大於 0.5 的類別輸出作為預測的類別，並將樣本分類到該類別。最後，我們使用損失函數（如交叉熵）來衡量模型的表現。

在訓練過程中，我們進行了超參數調整，這包括批量大小、迭代次數、學習率等。通過調整這些超參數，我們試圖找到最優的模型配置，以提高模型在測試集上的準確率。此外，我們還使用了 L2 正則化來防止過度擬合，這有助於提高模型的泛化能力。再透過以上的處理後最終的正確率為 0.593。即使經過超參數調整，發現模型在分類光碟片是否有瑕疵方面的表現仍然受限。在檢視模型分類結果後，推測其中一個可能原因是刮傷的程度和位置差異造成的影響。許多較淺的刮傷未能成功提取特徵，從而導致分類錯誤。刮傷程度可能因深度、角度等因素而異，而位置的不確定性也增加了模型提取特徵的困難，從而導致效能不佳。

3.2.4 影像辨識模型結果與比較

我們將 CNN 與 YOLOv9 兩個模型的結果進行比較，比較的結果如表 2 所示。可以看出 YOLOv9 的正確率為 0.871，CNN 的正確率為 0.765；在召回率上，YOLOv9 的分數為 0.823，CNN 召回率的分數為 0.755；YOLOv9 的精確率為 0.855，CNN 的精確率為 0.811；F1 Score 的結果中，YOLOv9 的分數是 0.838，CNN 的分數是 0.747。從各項指標中皆可以看出，YOLOv9 的表現都比 CNN 的結果更好，也因此，我們最後決定使用 YOLOv9 作為網站中偵測專輯是否有損傷的模型。

表 2 CNN 與 YOLOv9 比較

	CNN	YOLOv9
Accuracy	0.765	0.871
Recall	0.755	0.823
Precision	0.811	0.855
F1 Score	0.747	0.838

四、 網頁設計

4.1 網頁介紹

網頁前台網頁基本功能與顧客功能兩大類。網頁基本功能有精選專輯、分類曲風及語種的精選專輯、條款與細則，顧客功能有帳號註冊、登入、客製化訂購等，如圖 8 所示。後台會顯示近期通知、顯示庫存較少的幾樣商品提醒管理者補貨、顯示每段時間的訂單統計及顧客，如圖 9 所示。

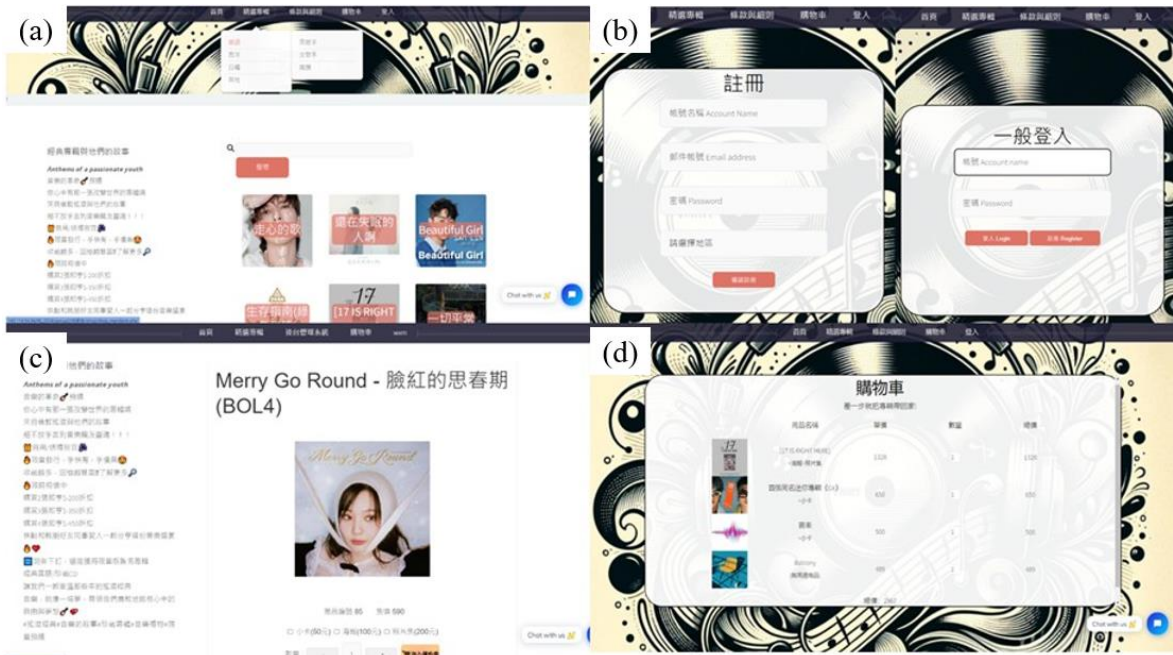


圖 8 網頁前台畫面 (a)商品分類呈現 (b)註冊與登入 (c)商品介紹 (d)購物車

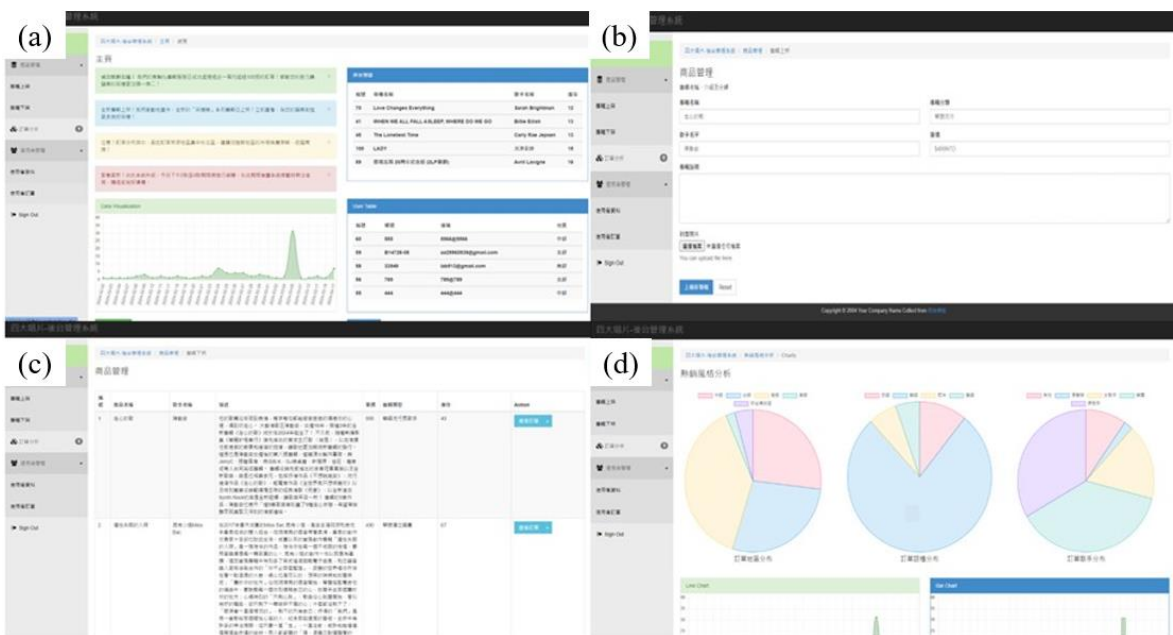


圖 9 網頁後台畫面 (a)後台首頁 (b)商品上架 (c)商品管理 (d)訂單分析

4.2 專輯偵測功能

我們新增了專輯偵測的功能。如 3.2 所述，在專輯退貨頁面上，顧客將專輯的

圖片上傳到網站，會連結到 YOLOv9 模型偵測光碟是否有損傷，便可以通知賣家進行後續的退換貨，如圖 10 所示。



圖 10 退貨申請頁面

4.3 Chatbot 介紹



圖 11 聊天機器人頁面

為了提升網站回應的人性化程度，我們整合了 GPT-3.5 模型與聊天機器人技術。當客戶提問時，系統會先檢查是否觸及到預設的關鍵問題；若是，則依照既定回答進行回應。反之，若客戶的問題未涵蓋預設範圍，系統將自動連接 GPT-3.5 模型進行更深入的互動，以確保提供準確而周到的客戶服務。如圖 11 所示。

五、 結論

我們使用大型語言模型和影像辨識技術來提升個人化專輯網站的用戶體驗和產品品質。這些方法可以增強了網站的智慧互動功能，並提升了產品的質量保障。

首先，我們測試了大型語言模型，如 GPT-3.5 和 Breeze-7B 並進行比較。我們發現，儘管 Breeze-7B 專為繁體中文設計，其表現在某些情境下仍然不如 GPT-3.5，因此我們最後決定使用表現較佳的 GPT-3.5 來優化網站。

在影像辨識部分，我們採用了 YOLOv9 模型和 CNN 模型來識別光碟的損壞。YOLOv9 顯示了高度的準確性，在識別光碟刮痕方面達到了優異的表現。相比之下，CNN 在處理淺刮痕的識別上表現較差，這可能由於其特徵提取的侷限性。因此，我們使用 YOLOv9 作為最後識別光碟損壞的方法。

我們透過將原本個人化專輯網站結合 ChatGPT 以及 YOLOv9，新增了與顧客即時對話功能以及光碟毀損識別的功能。然而，這次為了讓光碟的圖片可以更加清楚，我們用來訓練 YOLOv9 的資料集是將光碟使用掃描的方式，為了更貼近客戶的使用方式，未來在訓練模型時，可以改成使用拍照的方式來做為資料集。另外，未來也可以探索更多大型語言模型和影像辨識技術的組合，以及這些技術在其他領域的應用潛力。或是拓展到更多不同國家跟語言的功能。

六、 參考資料

Glenn-JocherBurhan-Q, Laughing-qRizwanMunawar,. (2024 年 6 月 2 日). YOLOv9：物

体检测技术的飞跃发展. (Ultralytics) 擷取自 Ultralytics:

<https://docs.ultralytics.com/zh/models/yolov9/#supported-tasks-and-modes>

LiuSimon. (2024 年 2 月 29 日). YOLOv9 模型剖析與使用 — 新的 Object detection 模型. (Medium) 擷取自 <https://blog.infuseai.io/yolov9-object-detection-model-introduction-61dc029d8167>

李馨伊. (2020 年 10 月 5 日). 卷積神經網絡 CNN 經典模型 — LeNet、AlexNet、VGG、NiN with Pytorch code. (Medium) 擷取自 <https://medium.com/ching-i/%E5%8D%B7%E7%A9%8D%E7%A5%9E%E7%B6%93%E7%B6%B2%E7%B5%A1-cnn-%E7%B6%93%E5%85%B8%E6%A8%A1%E5%9E%8B-lenet-alexnet-vgg-nin-with-pytorch-code-84462d6cf60c>

洪詩詩. (2024 年 3 月 7 日). 聯發創新基地再度開源釋出 Breeze-7B，可聽懂中英雙語語境的大型語言模型. 擷取自 <https://www.techbang.com/posts/113645-mediatek#:~:text=%E5%9C%A8%E5%B8%82%E9%9D%A2%E4%B8%8A%E7%9C%BE%E5%A4%9A%2070%20%E5%84%84%E5%8F%83%E6%95%B8%E7%B4%9A%E5%88%A5%E7%9A%84%E4%B8%AD%E8%8B%B1%E9%9B%99%E8%AA%9E%E6%A8%A1%E5%9E%8B%E4%B8%AD%EF%BC%8CBreeze-7B,%E>

數據分析那些事. (2023 年 5 月 22 日). 一文搞懂 ChatGPT 相關概念和區別：GPT、大模型、AIGC、LLM、Transformer、羊駝、LangChain.... 擷取自 Medium: <https://allaboutdataanalysis.medium.com/%E4%B8%80%E6%96%87%E6%90%9E%E6%87%82chatgpt%E7%9B%B8%E9%97%9C%E6%A6%82%E5%BF%B5%E5%92%8C%E5%8D%80%E5%88%A5-gpt-%E5%A4%A7%E6%A8%A1%E5%9E%8B-aigc-llm-transformer-%E7%BE%8A%E9%A7%9D-langchain-9b106583da3>

機器之心. (2024 年 2 月 23 日). 目标检测新 SOTA：YOLOv9 问世，新架构让传统卷积重焕生机. (51CTO) 擷取自 51CTO: <https://www.51cto.com/article/781901.html>

聯發科技. (2024 年 3 月 7 日). 推理時間只要 Meta 的一半！聯發科開源釋出中英語 Breeze-7B 大型語言模型. 擷取自 TechOrange: <https://buzzorange.com/techorange/2024/03/07/mediatek-research-breeze-7b/>