亞洲大學 GPT概觀課程

GPT in Asia University

亞 洲 大 學 資 訊 發 展 處 蔡 季 甫 2 0 2 4 / 9





附錄:如何安裝服務 Appendix



甚麼是GPT

就這麼誕生了

當鹹蛋超人搖身一變成為了天才科學家,GPT



GPT為Generative pre-trained transformer的縮寫,中文又可稱為:生成式預先訓練轉換器



G,英文是 Generative,中文的意思是生成式,這代表的是這個 AI 模型是用於 生成新的資料。AI 有很多不同類型,過去比較常見的是辨識型的 AI,例如手機 的人臉辨識解鎖,就是透過辨識型 AI 做到。而生成式 AI 是比起辨識更困難的技 術,它可以用來生成文字、圖片等資料。

甚麼是GPT

P,英文是 Pre-trained,中文的意思是預訓練,意即這個 AI 模型是預先被訓練的,GPT 的訓練最開始是無監督式的 (unsupervised),換句話說就是沒有人類介入,沒有特別數據標注,大量地把網路上的資料餵到模型中。在無監督式訓練後,再搭配人類介入的微調 (fine-tuning),讓模型的輸出成為人類偏好的回應。

T·是指 Transformer,上面提到因為它是一個專有名詞,所以沒有特別翻譯成中文。這個 Transformer 是 Google 在 2017 年提出的一個深度學習模型,這個模型是以注意力 (attention) 為機制來產生預測。而 GPT 正是基於這個 Transformer 模型進一步發展的模型。





必須了解的兩個名詞! 預訓練(Pre-training)與微調(fine-tuning)

在預訓練(Pre-training)階段,GPT模型使用了一個被稱為"Masked Language Model"(MLM)的任務,也就是預測 一個句子中被遮蓋住的部分。預訓練的目標是最大化句子中每個位 置的單詞的條件概率,這個概率由模型生成的分佈和真實單詞的分 佈之間的交叉熵來計算。

在微調(fine-tuning)階段,GPT模型在特定任務的數據上進行 訓練,例如情感分類、問答等。微調的目標是最小化特定任務的損 失函數,例如分類任務的交叉熵損失函數。

GPT模型的優點在於,由於其預訓練-微調的訓練策略,它可以有 效地利用大量的無標籤數據進行學習,並且可以輕鬆地適應各種不 同的任務。







GPT之父 - Sam Altman(奧特曼)的小秘密

1、高中時奧特曼就公開出櫃,由於當時民情保守,學生公開反對同性議題的演講, 所以奧特曼直接向全社區公開自己的性向。

2、2005年,奧特曼進入史丹佛大學攻讀電腦科學學位,但一年就退學,

3、2015年,奧特曼與時任特斯拉Tesla和SpaceX執行長馬斯克共同創立了時屬非營利組織的OpenAI,其創辦理念為「避免發生 AI 潛在風險和擔憂」,希望能夠預防 AI 人工智慧的災難性影響。

4、奧特曼於 2019 年5 月正式成為 OpenAI 的首席執行長, OpenAI 也從一家非營 利公司轉變為一家「營利」的公司, 該年更獲得微軟10億美元投資

5、奧特曼曾向媒體表示,他個人在OpenAI沒有任何股權,只有小部份經由以前任 職的公司持有的股份,等於他完全不能「直接」受惠於OpenAI狂飆的估值。

6、OpenAI 董事會於2023年11月解雇奧特曼,三天後奧特曼宣布加入微軟,但又 於加入微軟隔日再度重返 OpenAI。





甚麼是GPT

GPT-40 / 40 mini



2 GPT在亞大

別意外 · 在亞大使用GPT服務就是這麼的輕鬆





亞歷山大

後端為GPT-3.5模型·為本校Teams專屬服務·開放提供全校教職員生皆可使用。

Q小寶

後端為本校針對各單位提供的問答模型,可在Teams服務、校網,以及教務處、學務處、 資訊處等網站上直接進行問答使用。

Stable Diffusion

線上生圖服務·使用者可透過亞歷山大服務進行直接生圖·或是透過網址進入服務使 用完整功能

GitHub Copilot

由 OpenAI 提供技術支持的 AI 程式碼助手,透過機器學習協助開發者快速撰寫和改進 程式碼

Whisper

應用人工智慧技術的「語音轉文字」(speech-to-text)工具,它能幫我們將一份錄音檔 完整轉換為逐字稿,也能用於自動生成影片字幕檔。

Chat with RTX

Nvidia推出的本地端應用程式,使用者可以將個人電腦本地端文件作為資料集(客製化 資料集),也可連接諸如Mistral或Llama 2等開源大型語言模型,快速進行各種查詢



如何快速使用亞歷山大

- 1. 使用學生帳號登入Teams
- 2. 點選Teams左下方應用程式
- 3. 點選中央專為**貴組織打造項目**,即可 於右方找到亞歷山大
- 4. 點選新增後即可加入至Teams功能進 行使用

於Teams視窗內透過對話方式,即可擁 有GPT的服務。

同時在每次的對話後,亞歷山大也會提供相關性的三個網路搜尋結果,以提供使用者更有效的回覆。



GPT在亞大

↓ 瀬

Ð

Copilot

Ð

聊天

ന്ന്

国隊

â

工作指派

...

行事曆

C

通話

OneDrive

+

如何快速使用Q小寶

直接前往亞大校網、教務處、學務處或 資訊處網站,即可找到Q小寶圖示,點 選後即可進行問答。

或可使用Teams方式進行操作,加入方 式如下:

- 1. 使用學生帳號登入Teams
- 2. 點選Teams左下方應用程式
- 3. 點選中央專為**貴組織打造項目**,即可 於右方找到**Q小寶**
- 4. 點選新增後即可加入至Teams功能進

行使用





如何快速使用Stable Diffusion

- 1. 使用學生帳號登入Teams
- 2. 確認已加入亞歷山大服務
- 於亞歷山大Teams對話視窗內,使用 /img作為對話開頭,以空白鍵區隔 後即可輸入提示詞進行生圖。

亞歷山大的Stable Diffusion只提供快速 的單一圖片輸出,也無法做更詳細的參 數調整。如需較完整的服務,則可自行 安裝於個人電腦內進行使用。 (安裝步驟可參照第四章節說明。本機建

議有較高階之顯示卡才能有最好的使用

品質)

/img.¹2nd century style, rome house, render, photo realistic, super detailed, domed building of monsters university movie, Harvard university, campus building, a circular driveway in front of the building, building extends to the left and right of the photo, students walking in front of the building, future sense, realistic style, more blue sky, sunshine, best quality, high resolution, 8K, HDR, strong color contrast

亞歷山大 5/24 下午 05:37

圖片產生中...如有多人同時使用可能會有生成速度較慢的狀況,請您耐心等候!







如何使用Chat with RTX

如需使用完整的Chat with RTX服務,建 議自行安裝於個人電腦內進行使用。

個人電腦內需具有下列基本規格才可進 行安裝:

作業系統: Windows 11

GPU: Nvidia RTX系列,至少搭載8GB

VRAM

RAM:16GB以上

硬碟:35GB可用空間

(安裝步驟可參照第四章節說明。)

Chat with RTX

Al model	Dataset	S			
Select Al model	.txt, .pdf, .doc files supported				
Llama 2 13B int4 🔹	Folder Path	•			
Mistral 7B int4	C·\LIsers\W/S750\AppData\Loca\N\//DIA\Chat\WithRTX\RAG\trt				
✓ Llama 2 13B int4		-			
What is Portal prelude RTX? Based on the context information provided, Portal Prelude RTX is a remastered version of the game Portal with RTX, featuring full ray tracing, NVIDIA DLSS 3, and NVIDIA RTX IO. It is available for free and is a Game Ready driver for the game. Reference files: portal-prelude-rtx-available-now-for-free txt					

↺

GPT應用範例

3

動動你的手指·你也可以輕易的成為各項領域



六大提示詞技巧,大幅提升GPT輸出成果品質來源:

- 寫下清晰的指示:可以透過提供範例、或是實際要求 產出的回答字數範圍,明定步驟來引導GPT完成你期 待的任務 (講清楚我才懂你)
- 2. 提供參考範例、參考文字:提供參考資料給GPT或要 求答案提供引文 (避免莊孝維)
- 將複雜的任務拆分為更簡單的子任務:把大問題變成 小問題,再將小問題的答案總結(太複雜我會出槌)
- 4. 給模型「有時間思考」:要求給出答案前推導問題的 過程,或是提供推理來幫助GPT找出正確的答案 (要 求精準而不是要求速度)
- 5. 善用外部工具:透過外掛工具來進行輔助 (GPT不是 完美的)
- 6. <u>系統性測試實驗</u>:答案再度驗證並測試正確性 (羅馬 不是一天建成的)





好的提示詞讓你上天堂·不好的Prompt讓你···

提問公式:「角色」+「任務」+「格式」+「排除」

範例:

你是一位{角色}。我現在正要執行OOOO,我希望你提供{任務},請提出我可以做的OOOO,並以{格式}的方式回應我,並 且{排除}某些不需要的OOOOO。

ノ ^{生成式AI的幻覺} 「英文很重要」+「不要百分百相信」

<u>GPT應用範例</u>





「英文很重要」+「不要百分百相信」 → _{適時的曲線救國}

請用兩百字寫出天龍八部的重點摘要

諾亞方洲 下午 03:11

《天龍八部》是金庸先生的武俠巨著,故事背景設定在宋、遼、西夏三國 鼎立的時代。小說以主角喬峰的身世之謎為主線,串聯起一段段錯綜複雜 的江湖恩怨情仇。喬峰身為丐幫幫主,後因身世揭露實為契丹人,身份遭 江湖人士唾棄,從而展開了一連串的流亡與復仇。與此同時,虛竹一路從 小和尚成長為武林高手,最終繼承逍遙派掌門之位。段譽則是大理國的公 子,性格憨厚,卻意外獲得了六脈神劍的絕學。四大部派——少林、武 當、峨眉、丐幫,以及天龍寺等江湖勢力,交織出一幅恢弘的武俠世界。 故事最終以喬峰的悲劇收場,象徵著江湖中的愛恨情仇,無法逃脫命運的 安排。《天龍八部》不僅描寫了豐富的武功奇遇,也深刻反映了人性的光 輝與黑暗,是一部集歷史、愛情、哲理於一身的武俠經典。 Please write a 200-word summary highlighting the key points of "Demi-Gods and Semi-Devils."

諾亞方洲 下午 03:15 翻譯

"Demi-Gods and Semi-Devils" (Tian Long Ba Bu) is a classic Chinese wuxia novel by Jin Yong, which intricately weaves the destinies of several protagonists with distinct backgrounds and personalities against the backdrop of a turbulent political landscape. The story unfolds during the Song Dynasty and features three main heroes: Qiao Feng, leader of the Beggars' Sect, who is of Khitan descent and grapples with his identity amidst racial tensions; Duan Yu, a carefree and peace-loving prince of Dali Kingdom, who inadvertently masters a powerful martial arts technique; and Xuzhu, a kind-hearted and naive monk who finds himself entangled in worldly affairs and inherits the legacy of the enigmatic Carefree Sect. As the plot progresses, these characters cross paths and become embroiled in complex relationships, political power struggles, and intense battles involving various sects and ethnic groups. Themes of loyalty, love, honor, and the quest for personal identity are explored, as the heroes confront their inner demons and the societal prejudices of their time. "Demi-Gods and Semi-Devils" is celebrated for its rich character development, moral dilemmas, and philosophical depth. Jin Yong employs a vast array of supporting characters and martial arts lore to enrich the narrative, crafting a world where the lines between good and evil, hero and villain, are often blurred. The novel's exploration of the human condition and the nature of heroism has left a lasting impact on Chinese literature and popular culture.

《天龍八部》是金庸的一部經典武俠小說,故事背景設定在宋代,巧妙地將幾位主角的命運交織在動盪的政治背景下。故事圍繞著三位主要英雄:隘幫幫主喬峰,他 是契丹族人,在種族矛盾中掙扎尋找自己的身份;大理國王子段譽,一個性格灑脫、熱愛和平的人,無意中掌握了一種強大的武功;以及善良單純的少林和尚虛竹, 他被捲入世俗事務中,繼承了逍遙派的衣缽。隨著情節的發展,這些人物的命運交織在一起,他們陷入了複雜的關係、政治權力鬥爭,以及涉及各門派和民族的激烈 戰鬥。小說探討了忠誠、愛、榮譽和個人身份的主題,英雄們在面對內心的惡魔和社會偏見時逐漸成長。《天龍八部》因其豐富的人物塑造、道德兩難和哲學深度而 廣受讚譽。金庸透過大量的配角和武俠傳說豐富了敘事,塑造了一個善惡模糊、英雄與惡棍界線模糊的世界。小說對人性和英雄主義的探討對中國文學和流行文化產 生了深遠的影響。



GPT應用範例

你是一位精通台灣繁體中文的專業翻譯,曾參與《紐約時報》 和《經濟學人》繁體中文版的翻譯工作,因此對於新聞和時事 文章的翻譯有深入的理解。我希望你能幫我將以下英文新聞段 落翻譯成繁體中文,風格與上述雜誌的繁體中文版本相似。規 則:

- 翻譯時要準確傳達新聞事實和背景。

- 保留特定的英文術語或名字,並在其前後加上空格,例如:"
 中 UN 文"。

- 分成兩次翻譯,並且輸出每一次結果:

1. 根據新聞內容直譯,不要遺漏任何訊息

 根據第一次直譯的結果重新意譯,遵守原意的前提下讓內 容更通俗易懂,符合台灣繁體中文的表達習慣

- 每輪翻譯後,都要重新比對英文原文,找到扭曲原意或者遺漏的內容,然後再補充到下一輪的翻譯當中。(Chain of Density概念)

本條消息只需要回覆 OK,接下來的消息我將會給你發送完整 內容,收到後請按照上面的規則輸出兩次翻譯結果。 A week ago, NASA officials said they would most likely need to decide by mid-August on how to safely bring home two astronauts who had traveled to the International Space Station on Boeing's troubled Starliner spacecraft. But on Wednesday, NASA said a decision remained at least a week away.

亞歷山大 下午 04:37

🎐 第一次翻譯:

一周前,NASA官員表示,他們最有可能需要在8月中旬之前決定 如何安全地把兩名乘坐波音陷入困境的Starliner太空船前往國際空 間站的宇航員帶回地球。但是在周三,NASA表示,決定至少還有 一周的時間。

第二次翻譯:

一周前,NASA官員表示,他們最有可能在8月中旬之前決定如何 安全地把兩名乘坐波音陷入困境的Starliner太空船前往國際空間站 的宇航員帶回地球。然而,周三時NASA表示,他們尚未做出決 定,需要再等待至少一周的時間。



使用亞歷山大告訴我們怎麼下Prompt

我想請你幫忙規劃大學一年級學生的迎新活動。

請告訴我要提供怎樣的Prompt可以得到比較好的回答? 並請告訴我為什麼,並在最後舉出實際的Prompt範例讓 我更能了解。

亞歷山大 上午 11:14

▶ 為了獲得比較好的回答,你可以提供以下的Prompt:

- 1. 活動的目的是什麼?(例如歡迎新生、促進同學間的交流等)
- 活動時間和地點是什麼?(例如是在開學前一週的某個時間、 在校園內還是在外面的場地等)
- 活動的形式是什麼?(例如是室內或室外活動、需要報名還是 自由參加、有表演或遊戲等節目等)
- 活動的預算和資源是什麼?(例如有多少預算、需要借用什麼 器材或場地等)
- 活動的主題和內容是什麼?(例如以校園文化或體驗為主題、 有哪些內容或節目等)

透過以上的問題,可以讓你更清楚地了解迎新活動的規劃和需求,也 可以幫助你更有系統地思考如何設計迎新活動。

以下是一個實際的Prompt範例:

「請問你覺得一個成功的迎新活動需要具備哪些元素?例如活動的目 的、時間和地點、形式、預算和資源、主題和內容等。請盡可能詳細 地描述你的想法。」



使用亞歷山大彙整文章的重點

請你使用五十個字摘要以下文章的重點:

氣象署在臉書預報,未來一周台灣地區屬於典型夏季天氣,高溫皆落在32度到35度之間,部分地區有機會更高,環境大致吹西南風, 清晨至上午因與海陸風輻合作用,中南部地區有零星短暫陣雨,而下午太陽加熱使天氣轉不穩定,午後易有雷陣雨發展,提醒局部強 對流伴隨之短延時強降雨、雷擊、強風等劇烈天氣現象,前往山區溪河附近活動請注意溪水暴漲風險,並注意自身安全。明天(29 日)水氣稍增多,清晨西半部地區有局部短暫陣雨,白天起各地大多為多雲可見陽光,午後皆有局部短暫雷陣雨,且兩勢增大,局部 有大雨發生的機率,清晨至上午中南部地區亦有零星短暫陣雨。周五至周日(30日至9月1日)水氣再逐漸減少,各地大多為多雲到 睛,午後西半部地區及其他山區有局部短暫雷陣雨,清晨中南部地區亦有零星短暫陣雨。下周一、下周二(2日、3日)天氣逐漸轉 穩定,各地大多為多雲到睛,午後中南部地區及其他山區有局部短暫雷陣雨,清晨中南部地區有零星短暫陣雨。 月4日至6日)各地大多為多雲到睛,午後中南部地區及其他山區有局部短暫雷陣雨,清晨中南部地區有零星短暫陣雨。





使用亞歷山大作為論文大綱生成器

你現在是一個強大的科技論文大綱生成器。背後是以IEEE 期刊為基礎,我現在正在寫一篇關於「生成式AI應用於醫 療領域」的論文,請為我創建這篇論文的大綱,並將論文 以IEEE期刊格式為主分為相關段落,並確保每個大綱可以 適當說明其內容,以「一、二、三、…」的形式回覆我。

亞歷山大 下午 02:10 翻譯

🤣 好的,以下是您的論文大綱,按照IEEE期刊格式分為相關段落:

I. Introduction

A. Background and motivation

- B. Research question and objectives
- C. Contributions

II. Literature review

A. Overview of AI in healthcare

- B. Existing applications of generative AI in healthcare
- C. Advantages and limitations of generative Al in healthcare
- D. Gaps in the literature and research needs

III. Methodology

A. Research design and data collectionB. Pre-processing and feature selectionC. Generative AI models and trainingD. Evaluation metrics and performance measures

IV. Results

A. Description of the dataset and sample characteristics
 B. Analysis of generative AI models and their performance
 C. Comparison with existing approaches and benchmarks
 D. Interpretation and implications of the results
 V. Discussion

A. Summary of findings and research contributions

B. Implications for healthcare practice and policy

C. Limitations and future research directions

D. Conclusion and recommendations for future work

VI. References

注意:這只是一個大綱,實際論文需要根據您的研究內容進行調整和補 充。



使用亞歷山大作為程式撰寫的輔助工具

你現在是一個任職於遊戲公司的開發工程師,曾經接觸許 多遊戲開發的核心任務。我現在正在撰寫一個「大富翁」 的遊戲,請用C#程式語言協助我進行撰寫,並嘗試用最 短的程式碼來完成這個程式,並請確保程式碼內都使用繁 體中文進行每一段的註解以便我了解該程式的進行

```
亞歷山大 下午 02:32
```

好的,以下是一個簡單的大富翁遊戲的 C# 程式碼,使用了基本的控制流 程和物件導向概念:

using System;

```
// 定義玩家類別
public class Player {
    public string name; // 玩家名稱
    public int money; // 玩家金錢
    public int position; // 玩家位置
```

```
// 玩家擲骰子
public int RollDice() {
```

Random rand = new Random();
return rand.Next(1, 7);
}

```
// 玩家移動
public void Move(int steps) {
    position += steps;
    if (position > 39) {
        position -= 40;
        money += 200;
        Console.WriteLine("{0} 經過起點, 獲得
$200!", name);
    }
}
// 玩家付款
public void Pay(int amount) {
    money -= amount;
    Console.WriteLine("{0} 支付了 ${1} ..., name,
}
```

G P T 應 用 範 例

使用亞歷山大快速創作一個想要的圖像

/img Visually Striking Surreal 3D supernatural fantasy city, 3D art of a cinematographic, in the style of trompe-l'œil, illusionistic, by Alexander Jansson, Tim Burton, magical fantasy, gradient and overlay, high contrast, global illumination lighting, dynamic lighting, alcohol ink on quicksilver, ultra detailed, professional masterpiece, deep rich colors, sharp focus, intricate, mesmerizing perfection /img Visually Striking Surreal 3D supernatural fantasy city, 3D art of a cinematographic, in the style of trompe-l'œil, illusionistic, by Alexander Jansson, Tim Burton, magical fantasy, gradient and overlay, high contrast, global illumination lighting, dynamic lighting, alcohol ink on quicksilver, ultra detailed, professional masterpiece, deep rich colors, sharp focus, intricate, mesmerizing perfection

亞歷山大 下午 03:25

圖片產生中...如有多人同時使用可能會有生成速度較慢的狀況,請您耐心等 候!



GPT應用範例

使用亞歷山大快速創作一個想要的圖像

/img Photo of one young male model in white clothes looking at the camera, in front of the skyline of Rome, sunny day, not edited, very realistic and detailed faces, skin with pores, 4k resolution, dark suit, skin not airbrushed, looking into the camera. /img Photo of one young male model in white clothes looking at the camera, in front of the skyline of Rome, sunny day, not edited, very realistic and detailed faces, skin with pores, 4k resolution, dark suit, skin not airbrushed, looking into the camera.





使用Q小寶確認相關系統問題





如何安裝服務

本章將提供如何在你的電腦內安裝Stable Diffusion及Chat with RTX進行使用

Stable Diffusion 安裝方式及基本介紹

4-1



首先需先在您的電腦內安裝python,依據情況選擇要安裝的版本,

下載網址:https://www.python.org/downloads/



Antilitat 🥠

Q 搜尋

🖬 💼 🥶 💿 🔚



接下來要在電腦內**安裝git**,依照作業系統選擇要安裝的版本, 下載網址:https://git-scm.com/downloads





開啟電腦內命令提示字元功能(於搜尋內輸入cmd),進入後安裝stable diffusion web ui,輸入以下代碼後便會開始下載 git clone https://github.com/AUTOMATIC1111/stable-diffusion-webui.git

```
C:\WINDOWS\system32\cmd. X
Microsoft Windows [版本 10.0.22631.3447]
(c) Microsoft Corporation. 著作權所有,並保留一切權利。
C:\Users\a6699>git clone https://github.com/AUTOMATIC1111/stable-diffusion-webui.git
                                  C:\WINDOWS\system32\cmd. X
                                 Microsoft Windows [版本 10.0.22631.3447]
                                 (c) Microsoft Corporation. 著作權所有,並保留一切權利。
                                 C:\Users\a6699>git clone https://github.com/AUTOMATIC1111/stable-diffusion-webui.git
                                 Cloning into 'stable-diffusion-webui'...
                 安裝完成
                                 remote: Enumerating objects: 32875, done.
                                 remote: Total 32875 (delta 0), reused 0 (delta 0), pack-reused 32875
                                 Receiving objects: 100% (32875/32875), 34.41 MiB | 21.87 MiB/s, done.
                                 Resolving deltas: 100% (23010/23010), done.
```



於電腦內找到stable-diffusion-webui資料夾,點選webui-user.bat檔案就會開始安裝步驟





安裝完成後畫面跳出可顯示網址就代表已安裝完成

remote: Compressing objects: 100% (18/18), done. remote: Total 20 (delta 0), reused 20 (delta 0), pack-reused 0 Receiving objects: 100% (20/20), 132.70 KiB | 963.00 KiB/s, done. Cloning Stable Diffusion into C:\Users\a6699\stable-diffusion-webui\repositories\stable-diffusion-stab ility-ai... Cloning into 'C:\Users\a6699\stable-diffusion-webui\repositories\stable-diffusion-stability-ai'... remote: Enumerating objects: 580, done. remote: Counting objects: 100% (571/571), done. remote: Compressing objects: 100% (304/304), done. Rremote: Total 580 (delta 278), reused 448 (delta 249), pack-reused 9 Receiving objects: 100% (580/580), 73.44 MiB | 24.80 MiB/s, done. Resolving deltas: 100% (278/278), done. Cloning Stable Diffusion XL into C:\Users\a6699\stable-diffusion-webui\repositories\generative-models. Cloning into 'C:\Users\a6699\stable-diffusion-webui\repositories\generative-models'... remote: Enumerating objects: 941, done. remote: Total 941 (delta 0), reused 0 (delta 0), pack-reused 941 Receiving objects: 100% (941/941), 43.85 MiB | 23.71 MiB/s, done. Resolving deltas: 100% (491/491), done. Cloning K-diffusion into C:\Users\a6699\stable-diffusion-webui\repositories\k-diffusion... Cloning into 'C:\Users\a6699\stable-diffusion-webui\repositories\k-diffusion'... remote: Enumerating objects: 1340, done. remote: Counting objects: 100% (738/738), done. remote: Compressing objects: 100% (91/91), done. remote: Total 1340 (delta 693), reused 654 (delta 647), pack-reused 602 Receiving objects: 100% (1340/1340), 236.03 KiB | 2.03 MiB/s, done. Resolving deltas: 100% (941/941), done. Cloning BLIP into C:\Users\a6699\stable-diffusion-webui\repositories\BLIP.... Cloning into 'C:\Users\a6699\stable-diffusion-webui\repositories\BLIP'... remote: Enumerating objects: 277, done. remote: Counting objects: 100% (165/165), done. remote: Compressing objects: 100% (30/30), done. remote: Total 277 (delta 137), reused 136 (delta 135), pack-reused 112 Receiving objects: 100% (277/277), 7.03 MiB | 12.73 MiB/s, done. Resolving deltas: 100% (152/152), done. Installing requirements Launching Web UI with arguments: no module 'xformers'. Processing without... no module 'xformers'. Processing without... No module 'xformers'. Proceeding without it. Calculating sha256 for C:\Users\a6699\stable_diffusion-webui\models\Stable_diffusion\anythingelseV4_v4 5.ckpt: Running on local URL: http://127.0.0.1:7860

Startup time: 221.8s (prepare environmen. ____, ______.9s, import gradio: 1.2s, setup path s: 1.6s, initialize shared: 0.3s, other imports: 1.2s, load scripts: 0.6s, create ui: 0.3s, gradio lau nch: 0.1s).





未來如需使用Stable Diffusion · 只需開啟資料夾內的stable-

diffusion-webui.bat這個檔案

當執行該檔案出現網址後,即可於瀏覽器內進入該網址即可開始使用Stable Diffusion

*Stable Diffusion本機服務網址: http://127.0.0.1:7860



				1	ਸ਼ ਂ ਦ । ਜ ਤ ।
ble Diffusion checkpoint					
•	-				
d2img img2img Extras PNG li	Info Checkpoint Merger Train S	ettings Extensions			
Prompt				0/75	
Press Ctrl+Enter to generate, Alt+Enter to skip, E	Esc to interrupt)				Generate
				0/75	
Negative prompt				0/10 A.	
Press Ctri+Enter to generate, Alt+Enter to skip, E	-sc to interrupt)				(×) •
				11	
Generation Textual Inversion Hyp	pernetworks Checkpoints Lora				
Sampling method	Sampling steps		20		
DPM++ 2M Karras	• –		2		
🗇 Hires. fix	Refiner		•		
Width		512 Batch cou	unt 1		
Width		512 Batch cou	1		
Width		512 Batch cou	e 1		
Width Height CFG Scale		512 Batch cou	e 1		
Height CFG Scale		512 Batch cou			
Height CFG Scale Seed		512 Batch size	e 1 7 7 b Etra		
Width Height CFG Scale Seed -1 Script		512 Batch size	e 1 7 7 b btra		
Width Height CFG Scale Seed -1 Script None		512 Batch oz	e 1 7 7		

如何安裝服務





1、checkpoint = 大模型 = model = 畫家:

選擇你要的畫家,不同畫家所擅長的畫風不同,對同一個單字的理解也不一樣;或者你打的名詞比較冷門,有些畫家聽不懂就 會亂畫。例如你打"一個女孩",一個畫家可能畫出來的是真人寫實風格,另一個是動漫風格,一個畫起來像10歲,另一個畫起 來像30歲

至於SD1.5,SDXL,是什麼呢?你可以想成不同世代的畫家,像SD1.5就是stable diffusion這間公司先訓練出來一個叫做SD1.5的畫家,然後民間人士再據此依他們想要的方向訓練出stable diffusion1.5的各種徒子徒孫,要注意不同世代的繪師所能配合的VAE和LORA都不能通用

2、LORA = 小模型 = 模組:

可以想像成給畫家一個指引,例如你打"一個女孩",你的checkpoint可能就隨機給你各式個樣的女孩,但如果你用上了初音未來的LORA,那畫出來的女孩就會是初音未來的樣子。要注意LORA使用時可以自己設定強度,強度越高畫出來的東西就會用接近你的指引,但是強度開太高就會剝奪畫家的創作空間,畫面會變得很糟

3、Clip skip = AI分層停止參數:

這個是讓AI衍生分類的停止參數,數值越高AI越不會理會我們的要求,通常使用預設值1即可

4、prompt = 甲方要求:

你想要你的畫家一樣畫什麼東西,把要求打在框框裡就對了

5、Negative prompt = 甲方反向要求:

你想要你的畫家不要畫什麼東西,一樣把要求打在框框裡就對了



6、Latent = 潛空間 = 畫布:

基本上就是要你選擇你要的畫有多寬多高,以及你要畫家同時畫出幾張畫

7、Sampling method = 取樣器 = 繪師的繪畫方法技術:

基本上就是告訴你的畫家checkpoint要用哪一種技術畫畫。哪種適合生成的圖片建議可以上網看一下評測或自己測試一下,有些checkpoint會要求搭配指定的取樣器比較適合,還有一些特殊的加速畫法例如LCM,turbo,lightning之類的會是要搭配特定checkpoint配合各種東西才會使用

8、CFG Scale = 甲方要求的嚴格程度:

CFG數字越高,產生的圖會越按照你的需求,但是變化就會越少,反過來說CFG越低,畫家自由度就越大,成品變化就越高但 也越不按照你的指示

9、step = 步數 = 疊代次數 = 畫家的努力程度:

理論上step越高,成品細節越好,但是受限於checkpoint本身的能力,文生圖狀況通常超過40步就很難看出差別,建議不需要 再高;但若是操作圖生圖則與重繪幅度相關,則step不可太低。

10、seed = 隨機種子:

為了使每張圖片不同而加入的一些隨機參數,如果要測試同一條件下,不同的checkpoint,prompt,sampler有什麼差別,那 就要把seed固定,不然就設定為-1(隨機)就好

11、VAE = 上色師:

可以想像成一個畫家畫完後,你對他劃出來的顏色不滿意,再請上色師重新上色。但通常比較新的checkpoint是可以不用VAE的。



Stable Diffusion基本運作流程就是:

選擇你要的畫家(checkpoint) >> 選擇要用到的LORA(非必要) >> 設定Clip skip(非必要) >> 跟畫家說要畫的和不想畫的東西(prompt) >> 告訴他需要多少心力(Step), 什麼畫法來畫畫(Sampling method) >> 給他畫布(latent) >> 加入seed (非必要) >> 開始畫圖 >> 畫完的畫布請上色師重新調整顏色(VAE, 非必要) >> 儲存成品

如何安裝服務

圖片為使用dvarchMultiPrompt的checkpoint · 設定step為30 的狀況下 · 使用以下prompt進行初次生圖的結果:

A beautiful 20-year-old blonde girl with long hair, a thick single braid, amber eyes, tense expression, wearing a slavic turquoise sarafan, against the backdrop of the interior of a slavic rich princely terem, featuring trees, patterns, arches, hyperrealism, professional photo, detailed.





參考資源站台:

https://lexica.art/

https://civitai.com/

https://vocus.cc/salon/neversay-salon/about

https://blog.hinablue.me/

https://ivonblog.com/categories/stable-diffusion%E6%95%99%E5%AD%B8/

https://kodointeraction.tech/stable-diffusion-%E6%95%99%E5%AD%B8-15%E5%88%86%E9%90%98%E5%AE%8C%E5%85%A8%E6%8A%8A%E4%BD%A0%E6%95%99%E6%9C% 83/

4-2 安裝方式及基本介紹



首先務必確認自己的電腦硬體規格,如規格不符或效能不足建議先行升級硬體規格

作業系統: Windows 11

RAM:16GB以上

硬碟:建議35GB可用空間

顯示卡: 需配備Nvidia GeForce RTX 30 系列或更高版本GPU 且具備 8GB以上 的 VRAM



前往Nvidia官網的ChatRTX下載頁面下載安裝檔,下載完成後點選Setup進行安裝,安裝時間依電腦效能為10-120分鐘不等。網址:https://www.nvidia.com/zh-tw/ai-on-rtx/chatrtx/







安裝過程內可選擇要安裝的模型,此選項會依照您的顯示卡 效能提供可勾選項目,如效能不足可能就只會提供一個模型 可使用。

另外安裝過程中,請不要更改程式的安裝路徑,避免後續使 用過程發生異常。

with RTX

安裝完成後點選桌面圖示即可開啟主程式。 請注意!執行程式時務必連結網路!







Nvidia於2024年5月推出版本更新,除基本的功能更新之外, 另外加入ChatGLM3模型及Whisper自動語音識別,能讓使用 者使用語音對話的方式進行聊天,而ChatRTX 將以文字加以回 應。

如需進行更新,只需於官網下載頁面下載新版本進行程式更新。









1、Model = 語言模型 = 助手:

選擇你要的語言模型,不同助手可能會提供稍微有差異的回覆。 目前ChatRTX可載入五種模型:Llama213Bint4(Meta)、Gemma7B int4(Google)、Mistral7Bint4(Mistral)、ChatGLM36Bint4(智譜AI) 及CLIP(OpenAI)

2、Dataset = 本地數據資料導入 = 輔助助手:

可以想像成給助手一個輔助第二號助手,目前的本地端資料導入可以支援三種格式:txt、pdf及docx

3、Input = Prompt輸入區

你想要你的助手做甚麼事情,把要求打在框框裡就對了。 另外新版ChatRTX支援Whisper自動語音識別,使用者可直接透過語音方 式進行提問輸入

選擇AI 模型	×
Mistral 7B int4 (In-use) by Mistral Al 4GB 文字模型 支援 英語 瞭解更多	8
Llama2 13B int4 by Meta 6.8GB 文字模型 支援 英語 瞭解更多	0
ChatGLM 3 6B int4 by Zhipu Al 3.8GB 文字模型 支援 中文 瞭解更多	0
Gemma 7B int4 by Google 6.6GB 文字模型 支援 英語 瞭解更多	0
CLIP by OpenAl 1.5GB 圖片模型 瞭解更多	0



以下分別為使用Gemma及ChatGLM模型,詢問「請用一百個字簡述台灣亞洲大學的優勢」得到的回覆結果。



5 🗐

NVID	PIA ChatRTX					
AI Cha	莫型 tGLM 3 6B int4	數據資料 Al model data				
9	您	=				
	請用一百個字簡述台灣亞洲大學的優勢					
<u></u>	ChatRTX					
	台灣亚洲大学是一所位於台灣的私立大学,成立於 1986年,是一所致力於推動科技、人文及社會科學領域的綜合性大					
	學。以下是該大學的一些優勢:					
	1. 学术 excellence: 台灣亚洲大學拥有一流的师资和學術研究團隊,並且在許多領域中表現出色,例如: 資訊科技、商業管					
	理、法律、醫學、工程等。					
	2. 國際化: 台灣亚洲大學致力於推動國際化,並且與許多国家和地区的大學建立了合作和交流關係,例如: 美國、加拿大、 英國、海上利亚、英亚英族					
	天國、澳大利亚、利四三寺。 2. 名二六儿, 小浴亚洲上阁的阁庄本白世田名地。	本中了同处理学和地区, 法法律法国本地名三人 多人家				
	3. 多元又化: 台湾亚洲大學的學生米目世界各地,米目不同的國家和地區,這使得校園更加多元化和包容。 - 原系体影響素 心發來到出 忽然地地走回世原系從忽然書用的富士從世俗天照到原士從書店,似天世得又原系他診斷素					
	4. 俊秀的肌美率: 台湾亚洲大学的毕业生因其俊秀	N学侧衣况和置力的坟底而安到准土的有眯, 促而犹侍了 俊秀的桃美举。 四、 又提供了先有她会结路,会你以后的孩,会可能送到, 以下送给他啊!				
	5. 體驗式學習: 台湾亚洲大學注重學生的體驗式學的社會責任感和領導能力。	督,亚提供了計多機曾讓學生梦與社區服務、實習等沽動,從而增強他們 				

