

文字轉音樂生成式 AI 之進展、展望與面臨之問題： 借鑒繪圖生成式 AI 之發展歷程與其爭議

摘要

文字與圖像生成式 AI 在 2022、2023 年有了重大的進展並引起社會高度關注，但音樂生成式 AI 受到之關注似乎遠不如上述兩者。本文簡單回顧了文字轉音樂生成式 AI 在 2023 年的成果，並通過參考繪圖生成式 AI 的歷史，探討出了現在音樂生成式 AI 之進展對應繪圖生成式 AI，可能準備進入如 Midjourney 的出圈階段。同時，本文也通過考慮音樂本身之特性或參考繪圖生成式 AI 之爭議事件等方式，探討現在、未來音樂生成式 AI 正或可能面臨之問題與音樂生成式 AI 之未來展望。

關鍵字

音樂生成式 AI、音樂生成、音訊生成、生成式 AI、繪圖生成式 AI、圖片生成

目錄

一、 前言	3
二、 文字轉音樂生成式 AI 之現狀	4
(一) 定義與說明	4
(二) 現有之文字轉音樂生成式 AI	4
1. Riffusion	4
2. MusicLM	6
3. MusicGen	6
4. Stable Audio	7
5. Suno AI Chirp	8
(三) 技術規格比較	9
(四) 音訊品質評判	9
(五) 小結	10
三、 繪圖生成式 AI 發展歷程	11
(一) 定義與說明	11
(二) 簡史	11
1. 技術積累期	11
2. 技術突破	13
3. Midjourney：引領繪圖生成式 AI 進入大眾視野	14
4. Stable Diffusion：基石模型的誕生	16
5. 今：持續改進、增強可控性、擴展可能性	16
(三) 小結	17
四、 為何 Suno AI Chirp 可能成為音樂生成式 AI 之 Midjourney？	18
五、 文字轉音樂生成式 AI 之問題與未來展望	20
(一) 訓練資料問題與展望	20

(二) 使用上可能存在之問題.....	21
(三) 繪畫生成式 AI 現在面臨之問題：音樂生成式 AI 會不會同樣陷入之？	23
1.現存創作者抵制問題.....	23
2.大眾抵制人工智慧生成之內容.....	25
(四) 音樂生成式 AI 未來可能造成之問題.....	26
(五) 生成式 AI 發展時程預測、展望與解釋.....	28
1.什麼時候會出現像 Stable Diffusion 一樣強大、開源的音樂生成模型？..	28
2.為什麼音樂生成式 AI 發展的比繪圖生成式 AI 緩慢？未來速度會不會加	
快？	28
(六) 法律層面之展望.....	29
1.現行生成式 AI 之訓練過程是否違反著作權法、是否為合理使用？	29
2.若台灣認定生成式 AI 之訓練非合理使用會有什麼問題.....	30
3.生成式 AI 使否會影響創作積極性.....	31
4.建立表揚（ Credit ）機制與分潤機制.....	31
5.建立 AI 生成內容之正盜版制度與意識.....	33
(七) 小結.....	34
六、 結論	35
附件	36
參考資料	38

一、前言

音樂生成，自計算機開始進入人們視野以來，一直都是人們相當感興趣的領域。在過去，人們也有諸多嘗試，像是誕生於 1957 年，被認為是世界上第一首生成音樂的〈Illiac Suite〉，¹或是誕生於 2016 年，首位被法國音樂人聯合會（SACEM）認證並被輝達（NVIDIA）等知名公司採用的虛擬作曲家 AIVA。^{2 3}雖然相對於文本生成（Chat-GPT 等）、圖像生成（Mid journey 等）接連取得重大突破並獲得廣泛世人關注的盛況，^{4 5}音樂生成似乎沒有能取得同樣程度的關注。不過在這個 AI 算法改進快速、算力大幅增長的時刻，原先沒有被過多重視之直接生成音樂的音樂生成式 AI 逐漸地回到了人們的視野中，並且有了相當程度的進展，值得人們更多的注意並且更加警惕其成熟後，可能會造成的各種問題。

本文希望透過觀察 Riffusion、MusicLM、MusicGen、Stable Audio、Suno AI Chirp 這幾個較廣為人知的文字轉音樂（Text to Music）生成式 AI，回顧現今文字轉音樂生成式 AI 的進展，同時借鑒同為藝術創作生成式 AI 的繪圖生成式 AI 之發展歷程與其衍生出來之，包括但不限於版權、創作者抵制等的爭議，預測音樂生成式 AI 將來可能的發展方向與可能遇到的問題，最終冀望能夠：

- 解釋音樂生成式 AI 發展較繪圖生成式 AI 慢的原因、提出可能需克服之議題並預測音樂生成式 AI 未來可能的發展方向以及其可能的發展時程
- 使音樂生成式 AI 之研究者了解 AIGC 現存之爭議與其背後原因，讓音樂生成式 AI 之研究者或服務開發者可以迴避或是解決這些爭議
- 陳述未來音樂生成式 AI 可能會遇上的問題，並試圖提出解決方案
- 探討未來音樂生成式 AI 可能會引起的問題，引發討論以防止未來造成更大的損害

¹ L. A. Hiller, Jr., and L. M. Isaacson, Illiac Suite for String Quartet, New Music Edition, Vol. 30, No. 3, Bryn Mawr, Pa., Theodore Presser Co., 1957.

² Ed Lauder, 〈Aiva is the first AI to Officially be Recognised as a Composer〉，參見 AI BUSINESS：<https://aibusiness.com/verticals/aiva-is-the-first-ai-to-officially-be-recognised-as-a-composer>，2023 年 12 月 15 日頁面資料。

³ AIVA, Op.24 “I AM AI”, 2017

⁴ 中央通訊社，〈AI 聊天機器人 ChatGPT 用戶數成長速度創紀錄〉，參見中央通訊社：<https://www.cna.com.tw/news/ait/202302020181.aspx>，2023 年 12 月 15 日頁面資料。

⁵ 36 氪，〈Midjourney：AIGC 現象級應用，一年實現 1000 萬使用者和 1 億美元營收〉，參見 yahoo!新聞：<https://reurl.cc/GKzvED>，2023 年 12 月 15 日頁面資料。

二、文字轉音樂生成式 AI 之現狀

(一) 定義與說明

本篇文章中，文字轉音樂（Text to Music）生成式 AI 是指主要透過自然語言或由自然語言之文字提示詞（prompt）與模型交互、生成音檔，且生成內容可被認定為音樂之音樂生成式 AI。此定義類似 ShuLei Ji 等所提出之音樂生成式 AI 三等級之底層級（bottom level），⁶其定義為直接生成音訊。

此處略為調整定義的用意在於更加聚焦於本文的側重點：音樂與繪圖這兩個具有藝術成分的領域遇到生成式 AI 後的異同。定義中「生成內容可被認定為音樂」是由於有些音樂生成模型也可以生成如鼓循環（drum loop）或是環境聲等一般認知中非音樂的音檔，⁷而本文主要探討的是 AI 直接或幾乎直接生成最後之成品的情形，為避免討論失焦而訂定之。而「主要透過自然語言或由自然語言之文字提示詞（prompt）與模型交互」，除了確保交互方式與現今主流繪圖生成式 AI 一致外，也保留了容納現今 GPT-4 中可以見到的，透過自然語言直接與模型溝通，⁸未來可能成為主流之交互方式的可能性於定義中；由於最終產品只要將過程包裝起來，中間經過何種過程生成音檔其實對於使用者沒有任何影響，因此將 ShuLei Ji 等所提出之定義中之「直接」去掉。

基於以上定義、知名度、音樂生成式 AI 發展的重要性及說明需求，本文選出了 Riffusion、MusicLM、MusicGen、Stable Audio、Suno AI Chirp 這五個音樂生成式 AI 模型做為案例。而其它知名的音樂生成式 AI 模型，如先生成音樂符號再轉成音檔的 AIVA 等，則因為與定義不符而暫不列入討論。

(二) 現有之文字轉音樂生成式 AI

1. Riffusion

2022 年 12 月誕生之 Riffusion 為賽斯·福斯格倫（Seth Forsgren）及海克·馬蒂羅斯（Hayk Martiros）將繪圖生成模型 Stable Diffusion 透過文本—頻譜對精調（fine-tune）後，

⁶ Ji, S., Luo, J., & Yang, X. (2020). A Comprehensive Survey on Deep Music Generation: Multi-Level Representations, Algorithms, Evaluations, and Future Directions. *ArXiv, 2011.06801v1*, 2.

⁷ Stability AI, 〈Stable Audio: Fast Timing-Conditioned Latent Audio Diffusion〉，參見 [stability.ai](https://stability.ai/research/stable-audio-efficient-timing-latent-diffusion)：
<https://stability.ai/research/stable-audio-efficient-timing-latent-diffusion>

⁸ 林亦 Lyi, 《聊天就能程式！我用 GPTs 做了個自己的數字化身》，參見 YouTube：
<https://www.youtube.com/watch?v=pjKD3coLju0>，2023 年 12 月 15 日頁面資料。

⁹所產生的文字轉音樂生成式 AI，也是頭一批使用非歌詞之文字轉音樂之生成式 AI。Riffusion 原先為一開源模型，隨後接受融資、現已商業化且生成品質有十足的改進。¹⁰

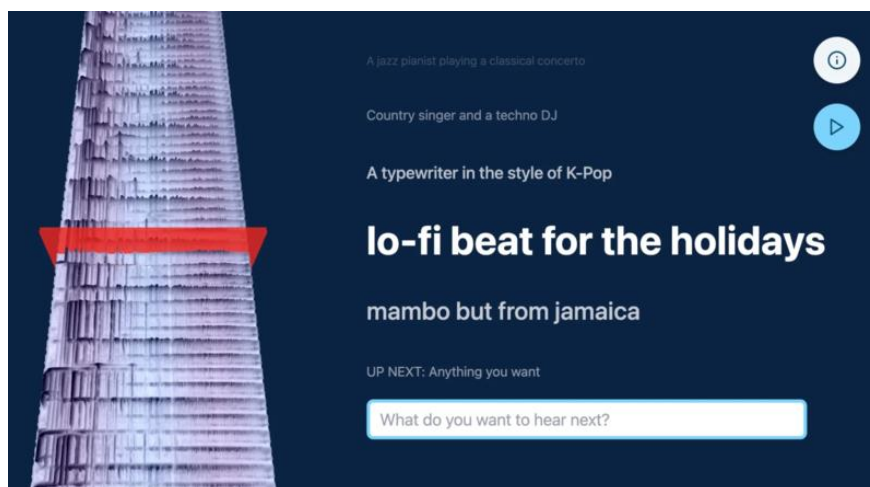


圖 1 Riffusion 舊官網截圖¹¹

經過改進之 Riffusion 現在之特色為支援透過歌詞與其他提示詞來生成具有高品質人聲之音樂，不過截至目前為止改進後之 Riffusion 只支援生成 12 秒之音樂。

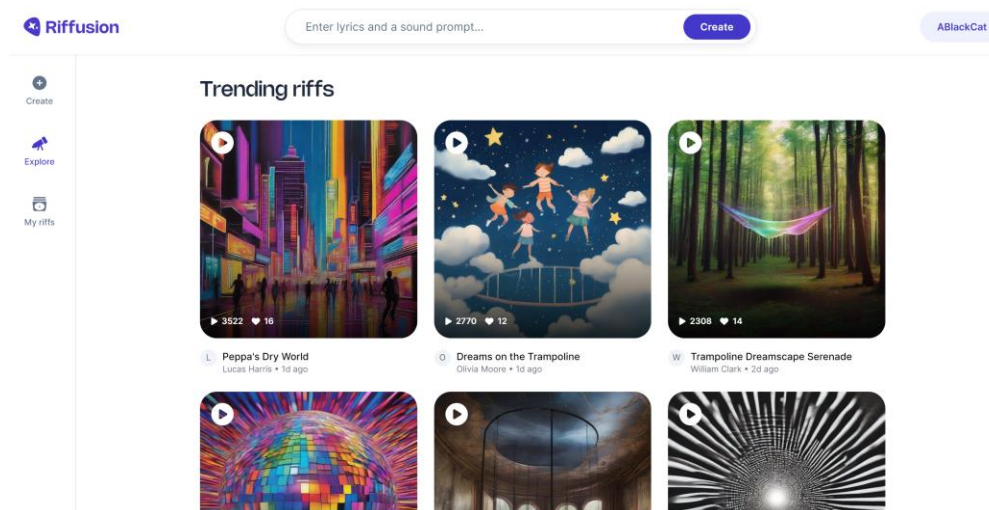


圖 2 Riffusion 新官網截圖¹²

⁹ DigiLog 作者群，〈將聲音視覺化後重新作曲？全新的 AI 音樂生成項目 Riffusion〉，參見 DigiLog：<https://digilog.tw/posts/1633>，2023 年 12 月 15 日頁面資料。

¹⁰ Kyle Wiggers，〈AI-generating music app Riffusion turns viral success into \$4M in funding〉，參見 TechCrunch：<https://techcrunch.com/2023/10/17/ai-generating-music-app-riffusion-turns-viral-success-into-4m-in-funding/>，2023 年 12 月 15 日頁面資料。

¹¹ DigiLog 作者群，〈將聲音視覺化後重新作曲？全新的 AI 音樂生成項目 Riffusion〉，參見 DigiLog：<https://digilog.tw/posts/1633>，2023 年 12 月 15 日頁面資料。

¹² 筆者於 2023 年 12 月 15 日於 <https://riffusion.com> 所截

2. MusicLM

MusicLM 為 Google DeepMind 所開發，2023 年 1 月公開發表並於 2023 年 5 月開放測試之音樂生成式 AI。MusicLM 採用基於 transformer 的自回歸模型（Transformer-based autoregressive model），模型參數量達、訓練資料來源為 the Free Music Archive、訓練資料時達 28 萬小時。MusicLM 可以生成 24kHz 採樣率、6kps 比特率（bit rate）、時常可達 5 分鐘的音樂片段。其模型架構如下圖 3：

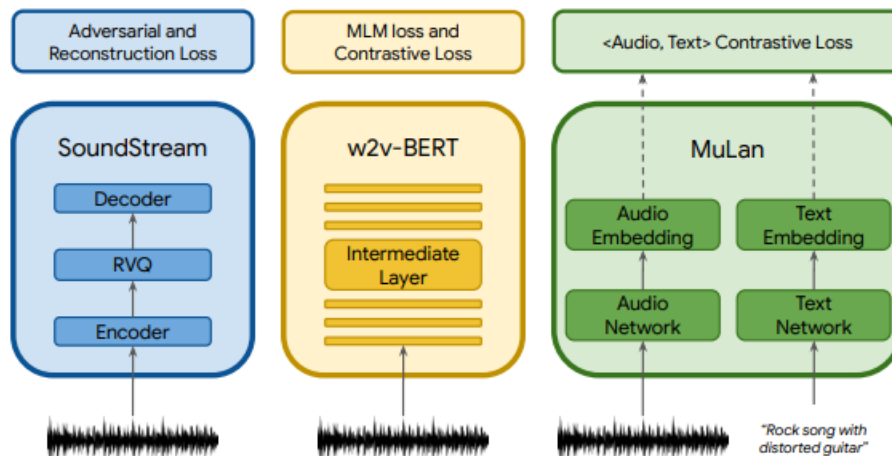


圖 3 MusicLM 模型架構圖¹³

訓練大型模型的一個難點就是如何取得足夠的訓練資料集，在文字轉音樂生成的情境中，問題就是如何蒐集夠多的音樂—文字描述對。MusicLM 引入了能輸入音樂產生文字描述的模式：MuLan，並利用其產生的文字描述去訓練其模型，因此能夠更輕易的擴張訓練資料的規模。

3. MusicGen

MusicGen 為 Metas（前 Facebook）旗下人工智慧研究室 Meta AI，於 2023 年 6 月發布並開源的文字轉音樂生成式 AI。MusicGen 採用基於 transformer 的單層語言模型（a single-stage transformer language model），搭配有效率的詞符交錯模式（efficient token interleaving patterns）來生成音樂。以下為詞符交錯模式示意圖 4：

¹³ Agostinelli, Andrea et al. "MusicLM: Generating Music From Text." *arXiv.org* (2023): n. pag. Web.

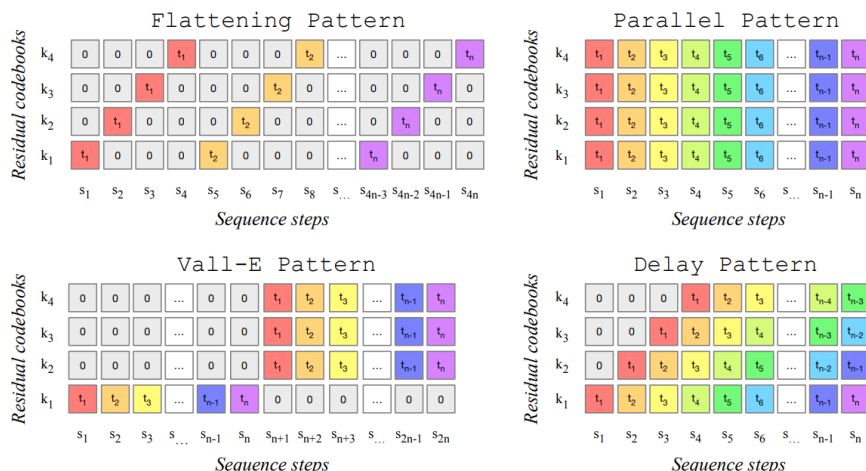


圖 4 MusicGen 詞符交錯模式示意圖¹⁴

MusicGen 之模型參數量最高可達 3.3 B，並且可以生成 32kHz 採樣率的音樂片段。

4. Stable Audio

Stable Audio 為開發 Stable Diffusion 之 AI 新創公司 Stability AI 於 2023 年 9 月所推出之文字轉音樂生成式 AI，採用與 Stable Diffusion 等類似的潛空間擴散模型（latent diffusion model）。此外，為了解決擴散模型（diffusion model）輸出長度固定與擴散模型訓練中將訓練用之音檔裁剪造成的問題，輸出結果為歌曲任意部份的問題，Stable Audio 引入了全卷積的編解碼器（fully-convolutional encoder and decoder），並且將裁剪後的訓練資料之開始時間與全曲時長作為條件提供給模型。其模型架構圖如下圖 5：

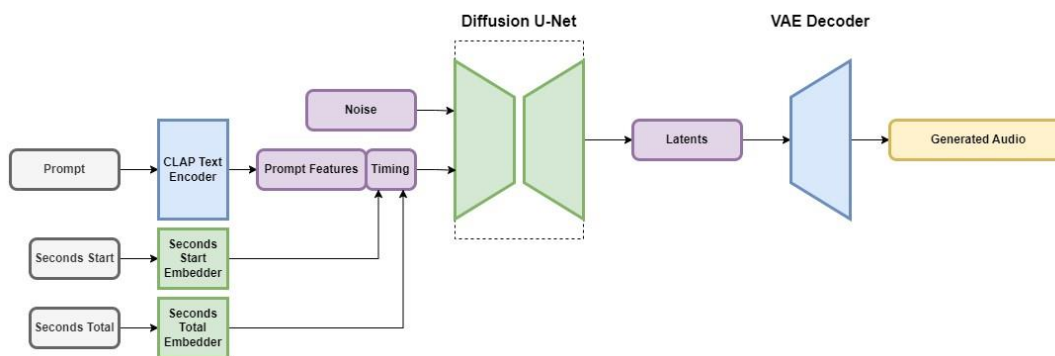


圖 5 Stable Audio 模型架構圖¹⁵

¹⁴ Copet, Jade et al. "Simple and Controllable Music Generation." *arXiv.org* (2023): n. pag. Web.

¹⁵ Stability AI, 〈Stable Audio: Fast Timing-Conditioned Latent Audio Diffusion〉, 參見 stability.ai : <https://stability.ai/research/stable-audio-efficient-timing-latent-diffusion>, 2023 年 12 月 16 日頁面資料。

Stable Audio 之模型參數量達 0.9B、訓練資料來源為音樂音效素材庫 AudioSparx、訓練資料超過 1.95 萬小時。Stable Audio 可以輸出 44.1kHz 採樣率、時長 90 秒或以上的音樂片段。

5. Suno AI Chirp

Suno AI Chirp 是 Suno AI 所開發之音樂生成式 AI 模型，雖然其未公開任何關於其模型之細節，不過由於其形式，包括支援輸入歌詞與提示詞生成具高品質人聲之音樂、伴隨音樂一起生成之圖片等，可以合理的猜測其採用之技術應該是與 Riffusion 相同，透過生成頻譜再轉換成音樂。Suno AI Chirp 相比 Riffusion 最大之差異在於它支援生成長達一分鐘之音樂片段。Suno AI 也是本文所提及之音樂生成式 AI 中，唯一一個預設有續寫功能的文字轉音樂生成式 AI。

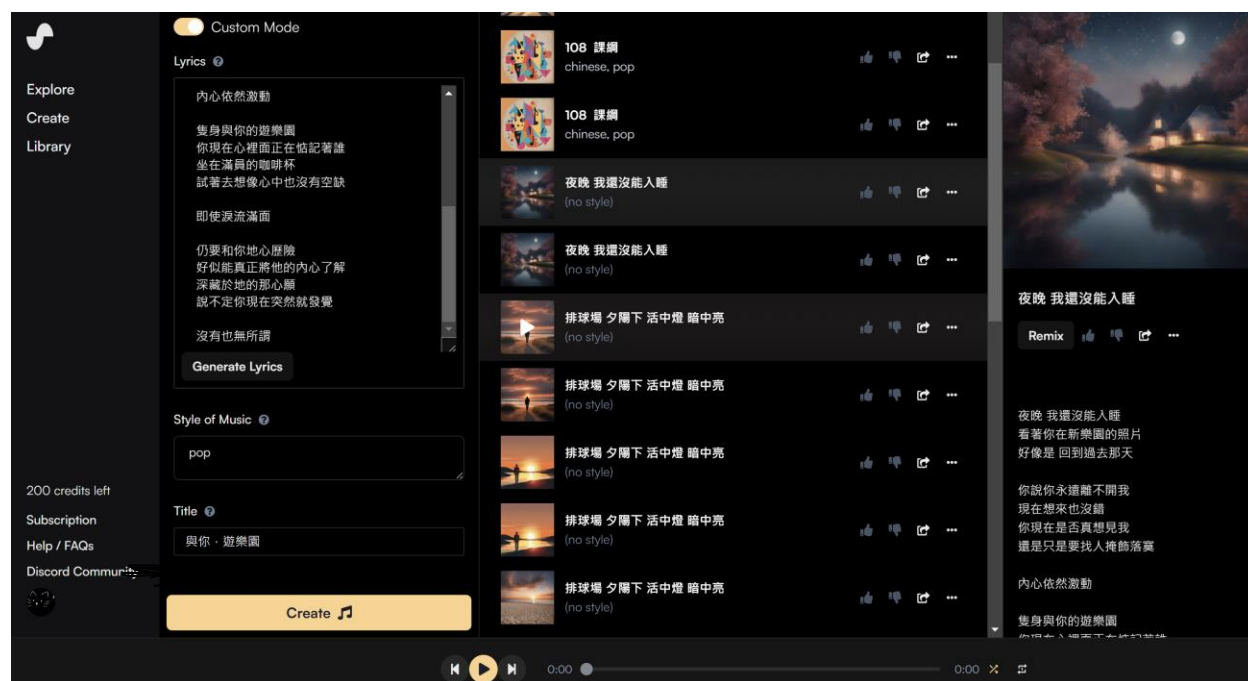


圖6 Suno AI Chirp 網頁版截圖¹⁶

由於截至本文撰寫時間（2023 年 12 月）最新之版本，在品質上有相當大的提升，因此本文所敘述之 Suno AI Chirp 皆默認為該版本。Suno AI Chirp 也是現今眾多文字轉音樂生成式 AI 中，最有可能成為文字轉音樂生成式 AI 之 Midjourney 的一個，其原因將會在第三部分簡單敘述繪圖生成式 AI 之歷史後，於第四部份中敘述。

¹⁶ 筆者於 2023 年 12 月 17 日於 <https://app.suno.ai/create/> 所截，內容為筆者使用 Suno AI Chirp 之服務所產生

(三) 技術規格比較

接下來我們會比較各個文字轉音樂生成式 AI 的技術規格，需要注意的是，這幾個模型採用的技術不盡相同，因此參數量無法直接比較，寫出來僅供參考。

	MusicLM	MusicGen	Stable Audio	Riffusion (改進)	Suno AI Chirp (V2)
發布時間	2023 年 1 月	2023 年 6 月	2023 年 9 月	2023 年 10 月	2023 年 11 月
模型基礎	Transformer	Transformer	Diffusion Model	Diffusion Model (頻譜)	-
訓練資料 時長(hr)	280,000	20,000	19,500	-	-
訓練資料 來源	公開資料	公開資料+ 內部資料	授權資料	-	-
取樣率 (Hz)	24,000	32,000	44,100	44,100	48,000
位元率 (kbps)	6	.wav 16bit	124 或以上 ¹⁷	192	128
聲道	單聲道	立體聲	立體聲	單聲道	立體聲

表 1 各文字轉音樂生成式 AI 之技術規格比較

(四) 音訊品質評判

如何評價音樂等藝術之高低優劣，長久以來都沒有固定的結論。由於個人對於音樂的喜好或多或少有些差異，實在難以以某一特定標準去概括，因此，任何試圖回答此問題的嘗試都往往以失敗告終。

然而，在討論、比較音樂生成式 AI 時，我們無可避免地要面對這個棘手的難題。本文此次採用之評估方法之為 2023 年 11 月由 Azalea Gui 等人所提出之 FAD[∞]。並使用其所提供之工具 Frechet Audio Distance Toolkit 進行計算。其前身為 Kevin Kilgour 等所提出之 Frechet Audio Distance，¹⁸被諸多音樂生成式 AI 採用作為評估生成出之音樂品質的方式。需要留意的是，由於此論文發布時間較近，截至今日為止仍未經同行審查，因此以下的結果不應被認為是有高度效力的，僅供參考。此外，由於 Frechet Distance 最終體現之結果為計算兩者之差距，因此結果不應當直接被詮釋為音樂的優劣，而是生成之音樂與參考音樂

¹⁷ 作者使用 Stable Audio 免費版服務下載得到之比特率為 124kbps，付費版位元率似乎比較高

¹⁸ Kilgour, Kevin et al. “Fréchet Audio Distance: A Metric for Evaluating Music Enhancement Algorithms.” arXiv.org (2019): n. pag. Web.

集 (FMA Pop) 之差距；此處能夠當成比較音樂品質的好壞之標尺，其中之原因來自於我們相信，好的音樂具有某些共通的但人類難以完全描述的特徵，而透過蒐集大量被認為品質良好之音樂並計算另一音樂與其之差距，我們就能相信那些差距更小的音樂更有可能是更好的音樂。

根據我們的跑出的結果，¹⁹大概音訊品質到一定程度之後就沒有太明顯的差距了。在使用 FMA Pop 做為參考的情形下，大部分的音樂的 FAD ∞ 分數會落在 550 分到 750 分左右，但這之中的分數高低其實也沒有辦法代表甚麼。而 Riffusion 的平均分數是 701、MusicLM 是 795、MusicGen 是 738、Stable Audio 是 684、Suno AI 是 669 分，由這樣的結果可以看出，除了較早推出、天生規格較差的 MusicLM 分數明顯較低以外，其他音樂生成式 AI 的分數都與一般的音樂相差不大，雖說不能就此代表他們的音訊品質就與現在的音樂無異，不過已經可以說是沒有太大的區別了。

(五) 小結

本文通過回顧 MusicLM、MusicGen、Stable Audio、Suno AI Chirp 及改進後的 Riffusion 這 5 個文字轉音樂生成式 AI，比較技術規格、音訊品質的方式，大致總結了文字轉音樂生成式 AI 在 2023 年的進展：已經開始逐漸接近一般流行歌曲的水準。然而，文章不可能完整表現音樂的每一個面向，還請耳聽為憑，自行使用各服務或模型確認。

¹⁹ 見附件一、附件二

三、繪圖生成式 AI 發展歷程

(一) 定義與說明

本篇文章中，文字轉繪圖生成式 AI 是指主要透過自然語言或由自然語言之文字提示詞（prompt）與模型交互、直接生成圖檔且生成內容可被認定為繪畫之繪圖生成式 AI。由於現今繪圖式生成式 AI 主要都為文字轉繪圖生成式 AI，為了保持簡潔，若無特殊說明，本文中的繪圖生成式 AI 都指的是文字轉繪圖生成式 AI。

選擇以繪圖生成式 AI 而非文字生成式 AI 或其它生成式 AI 作為借鑑之對象的理由有以下幾點，一是繪圖生成式 AI 發展的足夠好且整體發展已趨於穩定，變動不會太大，是合適的參考對象；二來是繪畫與音樂都同為藝術，有諸多相似之處，發展過程與面臨到的挑戰勢必也會有相似之處，是值得參考的對象；第三個理由是繪圖生成式 AI 過去這段時間累積了不少的矛盾與衝突事件，有許多值得借鑑學習的地方。因此，最終本文選擇以繪圖生成式 AI 而非如 ChatGPT 的其他生成式 AI，作為研究音樂生成式 AI 未來展望時之借鑑對象。

以下繪圖生成式 AI 之演進歷程介紹，將以 Web3 天空之城所撰寫之網路文章〈AI 繪畫何以突飛猛進？從歷史到技術突破，一文讀懂火爆的 AI 繪畫發展史〉為藍本，²⁰經查證後依本文需求進行增減、修改。

(二) 簡史

1. 技術積累期

雖說利用電腦自行進行繪畫的概念出現的很早，早在 1970 年代，哈羅德·科恩（Harold Cohen）就開始了繪畫程式 ARRON 的開發，不過現今我們所認知的 AI 繪畫的鼻祖，則沒有那麼長的歷史。

2012年，Google 的研究者聯合大量 CPU 運算資源以及來自 Youtube 的大量貓臉圖片，訓練出了能分類貓臉的神經網路，其神經網路之視覺化如下圖 7：²¹

²⁰ Web3 天空之城，〈AI 繪畫何以突飛猛進？從歷史到技術突破，一文讀懂火爆的 AI 繪畫發展史〉，參見微信 QQ：<https://mp.weixin.qq.com/s/LsJwRMSqPXSjSyhoNgzi6w>，2023 年 12 月 15 日頁面資料。

²¹ Hank Campbell，〈Singularity Update: 16,000 Processors To Identify A Cat〉，參見 Science 2.0：https://www.science20.com/science_20/singularity_update_16000_processors_identify_cat-91926，2023 年 12 月 18 日頁面資料。



圖7 能分類貓臉的神經網路之視覺化²²

此實驗向世界展示了神經網路、深度學習的可能性，也暗示了利用人工智慧生成圖像的可能性。隨後在接下來的數年間，AI 科學家們陸續提出多種模型並在試著用於生成圖片，其中一個有名的例子是對抗生成式網路（Generative Adversarial Network, GAN），而 StyleGAN 即是此技術之一良好展示。StyleGAN 為輝達（NVIDIA）於 2018 年推出之，利用 GAN 技術所開發之能夠生成人臉的圖片生成模型，²³並且有迭代之 StyleGAN2 與 StyleGAN3。經過預訓練也能生成繪畫風格的人臉，²⁴如下圖 8：



圖8 經預訓練後之 StyleGAN2 生成的日本動漫風格人物頭像²⁵

²² Le, Quoc V. “Building High-Level Features Using Large Scale Unsupervised Learning.” 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. *IEEE*, 2013. 8595–8598. Web.

²³ Karras, Tero, Samuli Laine, and Timo Aila. “A Style-Based Generator Architecture for Generative Adversarial Networks.” *IEEE transactions on pattern analysis and machine intelligence* 43.12 (2021): 4217–4228. Web.

²⁴ Fathy Rashad, 〈Generating Anime Characters with StyleGAN2〉，參見 Medium：
<https://towardsdatascience.com/generating-anime-characters-with-stylegan2-6f8ae59e237b>，2023 年 12 月 15 日頁面資料。

²⁵ Fathy Rashad, 〈Generating Anime Characters with StyleGAN2〉，參見 Medium：
<https://towardsdatascience.com/generating-anime-characters-with-stylegan2-6f8ae59e237b>，2023 年 12 月 15 日頁面資料。

2. 技術突破

雖然透過 GAN 等技術，AI 研究者已經可以生成具有一定品質的圖像，不過真正使 AI 繪圖大放異彩的原因離不開以下的技術突破。一是 OpenAI 於 2021 年 1 月所發布的初代 DALL-E，展示了利用文字提示詞生成圖像的可能性。²⁶其取得重大突破的原因在於 OpenAI 新建立並開源了圖像分類模型 CLIP，並使用大量數據進行訓練，²⁷使訓練資料的數量級得以提升，為使用文字提示詞生成圖像打下良好的基礎。



圖 9 DALL-E 所生成的圖像，文字提示詞為 *an illustration of a baby daikon radish in a tutu walking a dog* ²⁸

同時，另一個名為擴散模型（Diffusion Model）的模型逐漸進入 AI 圖像生成研究人員的視野中。擴散模型靠著相對 GAN 等其他模型更好的圖像生成品質，逐漸成為了 AI 圖像生成之主流研究方向。一個著名的例子是於 2021 年 10 月公開的開源 AI 繪圖模型 Disco Diffusion，能夠使用文字提示詞生成出品質精良、構圖優美的圖像。



圖 10 koalapon 於網路論壇 Reddit 分享其使用 Disco Diffusion 生成之圖像²⁹

²⁶ OpenAI，〈DALL·E: Creating images from text〉參見 openai.com：https://openai.com/research/dall-e，2023 年 12 月 15 日頁面資料。

²⁷ OpenAI，〈CLIP: Connecting text and images〉參見 openai.com：https://openai.com/research/clip，2023 年 12 月 15 日頁面資料。

²⁸ OpenAI，〈DALL·E: Creating images from text〉參見 openai.com：https://openai.com/research/dall-e，2023 年 12 月 15 日頁面資料。

²⁹ Koalapon，〈What's happening here? (Disco Diffusion)〉，參見 Reddit：https://www.reddit.com/r/MediaSynthesis/comments/shmggh/whats_happening_here_disco_diffusion/，2023 年 12 月 15 日頁面資料。

Disco Diffusion 雖說在生成人體等複雜形體的情境下表現不佳，但它的出現證明了利用 AI 技術生成品質精良之畫作是絕對可行的。

Disco Diffusion 雖能生成出相當精美的圖像，不過由於其對於資源的需求以及一些技術的門檻（最簡單的方式是使用官方提供之 Google Colab 雲端協作文檔，建置環境後始得使用），並沒有獲得大眾太多的關注。

3. Midjourney：引領繪圖生成式 AI 進入大眾視野

2022 年 7 月，發布 V3 版本的 Midjourney 迅速引起了人們的注意。Midjourney 改善了生成圖片中的細節並改善了生成人體等物體的生成品質，³⁰且採用通訊社交平台 Discord 為交互媒介，用戶向官方開發之 Discord 機器人（Discord bot）提交文字提示詞即可圖像，大幅使用降低門檻。

此外，傑森·艾倫（Jason M. Allen）與其使用 Midjourney 生成並修改而成的作品〈Théâtre d'Opéra Spatial〉奪科羅拉多州博覽會（Colorado State Fair）美術比賽首獎這件事，³¹更是引起了社會相當廣泛的議論，進一步加速了 Midjourney 被大眾認知的過程。



圖 11 傑森·艾倫 via Midjourney 〈Théâtre d'Opéra Spatial〉³²

³⁰ Yubin，〈All Midjourney Versions [V1-V5.2] Compared: The Evolution of Midjourney〉，參見 Aituts.com：
<https://aituts.com/midjourney-versions/>，2023 年 12 月 15 日頁面資料。

³¹ Rachel Metz，〈AI won an art contest, and artists are furious〉，參見 CNN Business：
<https://edition.cnn.com/2022/09/03/tech/ai-art-fair-winner-controversy/index.html>，2023 年 12 月 16 日頁面資料。

³² Papageno，〈A zsűrit is megvezette a mesterséges intelligenciával készített műalkotás〉，參見 PAPAGENO：
<https://papageno.hu/intermezzo/2022/09/mi-colorado-kepzomuveszet/>，2023 年 12 月 16 日頁面資料。

而 Midjourney 在臺灣能夠被廣泛認識，其功臣還有另一個爭議事件：吳淡如 AI 繪圖炎上事件。³³吳淡如 AI 繪圖炎上事件是指 2023 年 2 月，台灣電視主持人、作家吳淡如於其臉書發布利用 Midjourney 生成之圖像，並於貼文內容寫道：「我剛剛在學電腦動畫，這是我的第一張作品。這個世界太好了玩了，小熊已經睡了，希望我不要通宵達旦的一直畫下去。題目：櫻花貓少女。你喜歡就送給你」。³⁴ ³⁵由於用詞不夠精確且有誤導觀眾認為生成之圖像為吳淡如本人親自執筆作畫之嫌，被要求其更改用詞，並且因後續公關處理不佳引起群眾的不滿的一起事件。由於群眾的大量關注以及媒體的廣泛報導，Midjourney 在臺灣的知名度又再更上一層樓。



圖 12 吳淡如 via Midjourney 〈櫻花貓少女〉³⁶

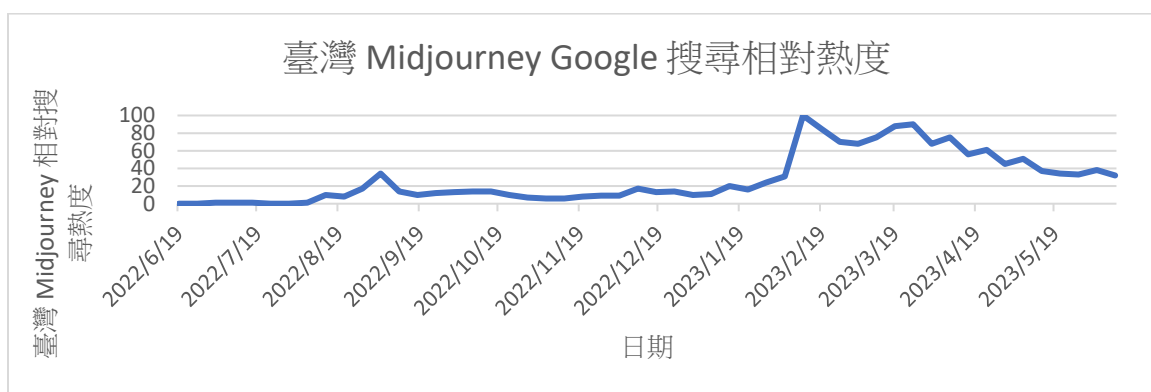


圖 13 臺灣 Midjourney 相對搜尋熱度，可以看到吳淡如 AI 炎上事件發生之 2023 年 2 月搜尋熱度突然竄升³⁷

³³ 李沁儒，〈吳淡如 AI 炎上事件懶人包！AI 畫作＝電腦繪圖嗎？她處理態度挨轟「理直氣壯糾正專業」〉，參見風傳媒：<https://www.storm.mg/lifestyle/4731458>，2023 年 12 月 14 日頁面資料。

³⁴ 原文現已刪除

³⁵ 可參見 newstalk：林彥銘，〈吳淡如為「AI 繪圖」風波道歉！稱無意浪費社會資源 加碼自首「這件事」〉，<https://newtalk.tw/news/view/2023-02-16/857845>，2023 年 12 月 18 日頁面資料。

³⁶ Henley，〈吳淡如『AI 繪圖』事件惹議！一分鐘告訴你 AI 繪圖和電繪到底差在哪〉，參見電癩少女：<https://agirls.aotter.net/post/61828>，2023 年 12 月 16 日頁面資料。

³⁷ Google Trend，〈Midjourney 2022 年 6 月 16 日至 2023 年 6 月 16 日台灣搜尋熱度〉，參見 Google Trend：<https://trends.google.com.tw/trends/explore?date=2022-06-16%202023-06-16&geo=TW&q=%2Fg%2F1t4b6xmqq&hl=zh-TW>，2023 年 12 月 19 日頁面資料。

4. Stable Diffusion：基石模型的誕生

誠然 Midjourney 功能相當強大，不過若要選擇一個對於繪圖生成式 AI 貢獻最大、真正使繪圖生成式 AI 達到今日這般成就的模型，那肯頂得是 Stable Diffusion。Stable Diffusion 是由 AI 新創公司 Stability AI 於 2022 年 8 月所發表並開源的繪圖生成式 AI 模型，其貢獻主要有以下兩點，且皆是可以在繪圖生成式 AI 史留名的等級

一為技術上的突破，研究團隊提出一種在潛空間（latent space）上進行擴散過程的方式，同時效果並無降低太多。因此突破而降低的算力及記憶體需求，讓個人電腦運行繪圖生成式 AI 模型變的可能。一般來說，用戶的個人電腦只要擁有一張不錯的顯示卡，就可以在本地以不錯的速度直接運行 Stable Diffusion 大模型，無限生圖；高通（Qualcomm）更宣稱其於 2023 年 10 月發布的手機晶片 Snapdragon 8 Gen 3 能夠在本地運行 Stable Diffusion，且在一秒之內生成出圖像，³⁸再次凸顯了此技術突破的重要性與其帶來的可能性。

二為模型與權重的開源，Stability AI 的標語為「AI 由人們創造，為人們所用」，³⁹⁴⁰Stable Diffusion 的開源，不僅打破了繪圖生成式 AI 大模型被科技公司壟斷的情形、建立起了廣大的社群，還作為後人研究與改進的基礎，如將原先用於降低大語言模型微調（fine-tune）成本的 LoRA（Low-Rank Adaptation）技術用在 Stable Diffusion 上，能夠低成本微調大模型，使模型能夠輸出較符合預期的結果。⁴¹另一個利用 Stable Diffusion 展開研究的例子是 ControlNet，⁴²可以透過提供額外訊息的方式，控制模型的輸出結果，常用於控制、固定生成出之人物姿勢或手勢等情境。

總而言之，Stable Diffusion 的出現，使得投身繪圖生成式 AI 之成本降低極多，極大程度的促進了繪圖生成式 AI 的技術進展，且正在加速著繪圖生成式 AI 進入實際可用階段，影響深遠，可謂貢獻良多。

5. 今：持續改進、增強可控性、擴展可能性

繪圖生成式 AI 在經歷了時間短暫卻變化快速的突破期後，逐漸進入了穩定的成長期。各式控制 Stable Diffusion 的技術如雨後春筍一般湧現，知名的例子除了有剛剛有提及的

³⁸ Qualcomm，〈Snapdragon 8 Gen 3 Mobile Platform〉，參見 qualcomm.com：
<https://www.qualcomm.com/products/mobile/snapdragon/smartphones/snapdragon-8-series-mobile-platforms/snapdragon-8-gen-3-mobile-platform>，2023 年 12 月 14 日頁面資料。

³⁹ 英文原文為「AI by the people for the people」

⁴⁰ Priyanshu Anand，〈AI by the people, for the people - How Stability AI is democratizing AI〉，參見 Thredo AI：
<https://www.thredo.com/resources/ai-by-the-people-for-the-people-how-stability-ai-is-democratizing-ai>，
2023 年 12 月 14 日頁面資料。

⁴¹ Luo, Simian et al. “LCM-LoRA: A Universal Stable-Diffusion Acceleration Module.” *arXiv.org* (2023): n. pag. Web.

⁴² Zhang, Lvmin, Anyi Rao, and Maneesh Agrawala. “Adding Conditional Control to Text-to-Image Diffusion Models.” *arXiv.org* (2023): n. pag. Web.

ControlNet、LoRA 以及類似功能的 LyCORIS 等以外，⁴³還有方便用戶使用的圖形介面，如 Stable Diffusion web UI、ComfyUI。

此外，各大模型各自都在改進其模型生成圖像品質，像是 Midjourney 於 2023 年 5 月推出的 5.1 版改善了許多細節，再次引起眾人關注；⁴⁴OpenAI 的 DALL-E 目前已推出第三代 DALL-E 3，且結合其大型語言模型 GPT-4，能夠完成像是生成海報等生成圖像或是畫作以外的任務。⁴⁵

（三）小結

繪圖生成式 AI 從一開始實驗性質的嘗試，技術逐漸突破且找到了適合此任務的模型：擴散模型。因為本身品質逐漸改善以及過程中一些爭議事件開始被大眾廣泛認知；在品質達到一定程度後，開始出現了掌控模型輸出結果的需求，加上開源模型的誕生，造就了各種新興技術的出現。現在，繪圖生成式 AI 進入了相較而言發展較平緩的時期，逐漸開始進入各式既有的應用場景中。⁴⁶同時也由於效率的突破，相關的應用場景如影片生成等，⁴⁷逐漸成為可能，不過大多都還待發展與完善。

透過了解繪圖生成式 AI 截至目前的發展歷史，我們可以更有依據的推測現今音樂生成式 AI 所處的時間段並更有根據的猜測其未來可能的發展方向。

⁴³ KohakuBlueleaf，〈LyCORIS - Lora beyond Conventional methods, Other Rank adaptation Implementations for Stable diffusion〉，參見 Github：<https://github.com/KohakuBlueleaf/LyCORIS/blob/main/README.md>，2023 年 12 月 14 日頁面資料。

⁴⁴ Barry Collins，〈Midjourney 5.1 Arrives - And It's Another Leap Forward For AI Art〉，參見 Forbes：<https://www.forbes.com/sites/barrycollins/2023/05/03/midjourney-51-arrivesand-its-another-leap-forward-for-ai-art/?sh=4bea0a5377cd>，2023 年 12 月 14 日頁面資料。

⁴⁵ Apple 研究員，〈三款我無法拒絕的 GPTs〉，參見 YouTube：https://www.youtube.com/watch?v=ZMuD_V3xKo0，2023 年 12 月 14 日頁面資料。

⁴⁶ 祝思齊，〈最近很火的 AI 作畫，遊戲開發者們已經在嘗試了〉，參見觸樂：<https://www.chuapp.com/article/288865.html>，2023 年 12 月 16 日頁面資料。

⁴⁷ Qbitai，〈Stable Video Diffusion 問世！全新 SVD 支援文字到影片、圖像到影片的繪製〉，參見電腦王：<https://www.techbang.com/posts/111279-stable-video-diffusion>，2023 年 12 月 16 日頁面資料。

四、為何 Suno AI Chirp 可能成為音樂生成式 AI 之 Midjourney ？

Midjourney 代表了繪圖生成式 AI 之品質與易用程度已經到達了大眾會去使用的程度，而由繪圖生成式 AI 之發展歷史來看，Midjourney 與其他繪圖生成式 AI 最大的差異為他改進了一些一般使用者在乎的點（如人聲之生成品質），並且因此獲得了更多人的關注。

而為何我們會將 Suno AI 定位為音樂生成式 AI 之 Midjourney 之可能人選，而不是我評價更高的 Stable Diffusion 呢？一個最重要的原因就是 Stable Diffusion 是開源的，並且作為繪圖生成式 AI 之基石模型，降低開發之門檻，使繪圖生成式 AI 進步速度能夠變得更快，而 Suno AI 並不是一個開源的音樂生成式 AI 模型，沒有辦法降低進入音樂生成式 AI 之開發之門檻。

Suno AI Chirp 通過人聲為主之生成模式與 V2 之品質改進，成功的引起不小的關注，短時間內搜尋熱度超越了以往任何一個文字轉音樂生成式 AI 且有維持的趨勢。縱使 Suno AI Chirp 不一定能夠成為像 Midjourney 一般的長青樹，其也肯定會在音樂生成式 AI 之發展歷史中留下重要的一頁。

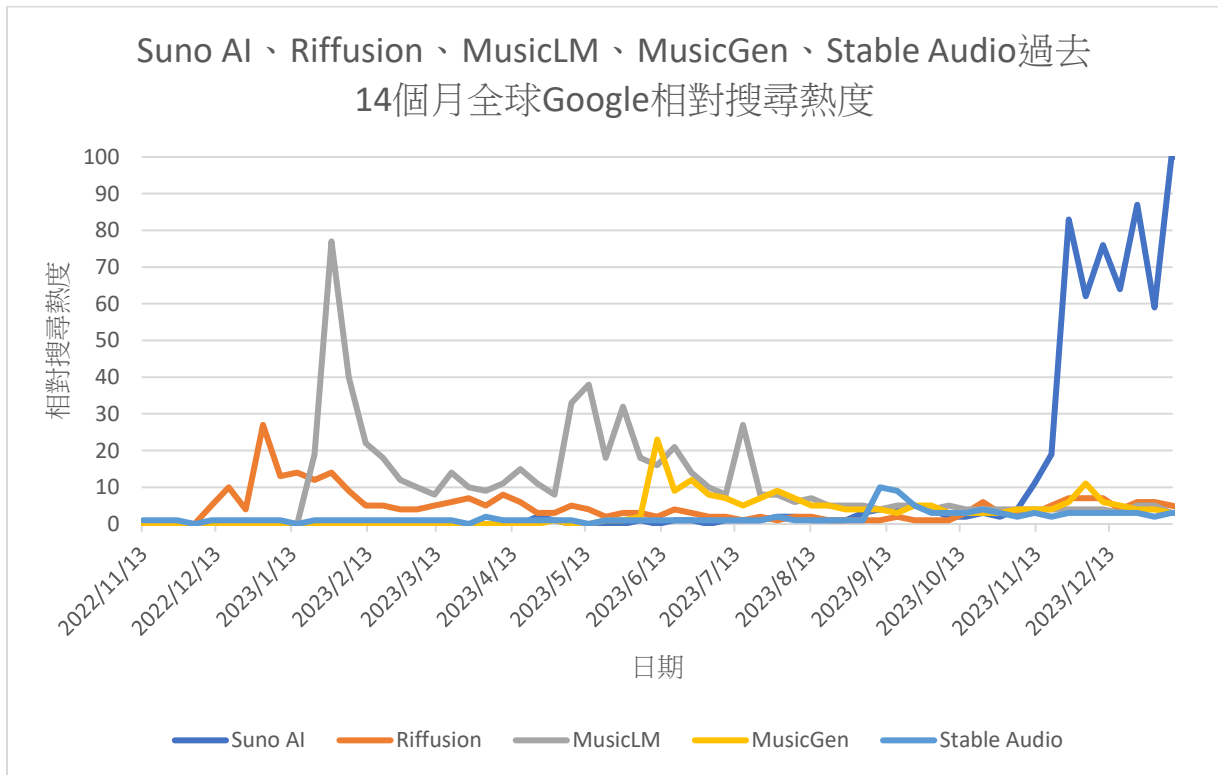


圖 14 文字轉音樂生成式 AI 相對搜尋熱度圖⁴⁸

⁴⁸ Google Trend，〈Suno AI、Riffusion、MusicLM、MusicGen、Stable Audio 過去 14 個月全球 Google 相對搜尋熱度〉，參見 Google Trend：<https://trends.google.com.tw/trends/explore?date=2022-11-18%202024-01-18&q=suno%20ai,Riffusion,MusicLM,MusicGen,Stable%20Audio>，2024 年 1 月 18 日頁面資料。

雖說改進後的 Riffusion 同樣採用以人聲為主生成模式且其生成品質個人認為比 Suno AI Chirp V2 還要好，但只能生成 12 秒且不能續寫、完成一首歌的限制多少影響到了人們投入與傳播的熱情，令其不可能以這個狀況成為文字轉音樂生成式 AI 之 Midjourney。

另一個決定 Suno AI Chirp 是不是文字轉音樂生成式 AI 之 Midjourney 之關鍵為是否能夠引起社會大眾更廣泛的注意。目前的一個爭議事件是海賊王作者尾田榮一郎分享其用 AI 做的歌曲，使部分網友不滿的事件，⁴⁹不過以 Google 搜尋熱度來看，在 Suno AI 日本引起的短暫旋風的主因並不是因為這件事，⁵⁰尾田榮一郎只是在這股旋風颳起之後跟風去使用了而已。

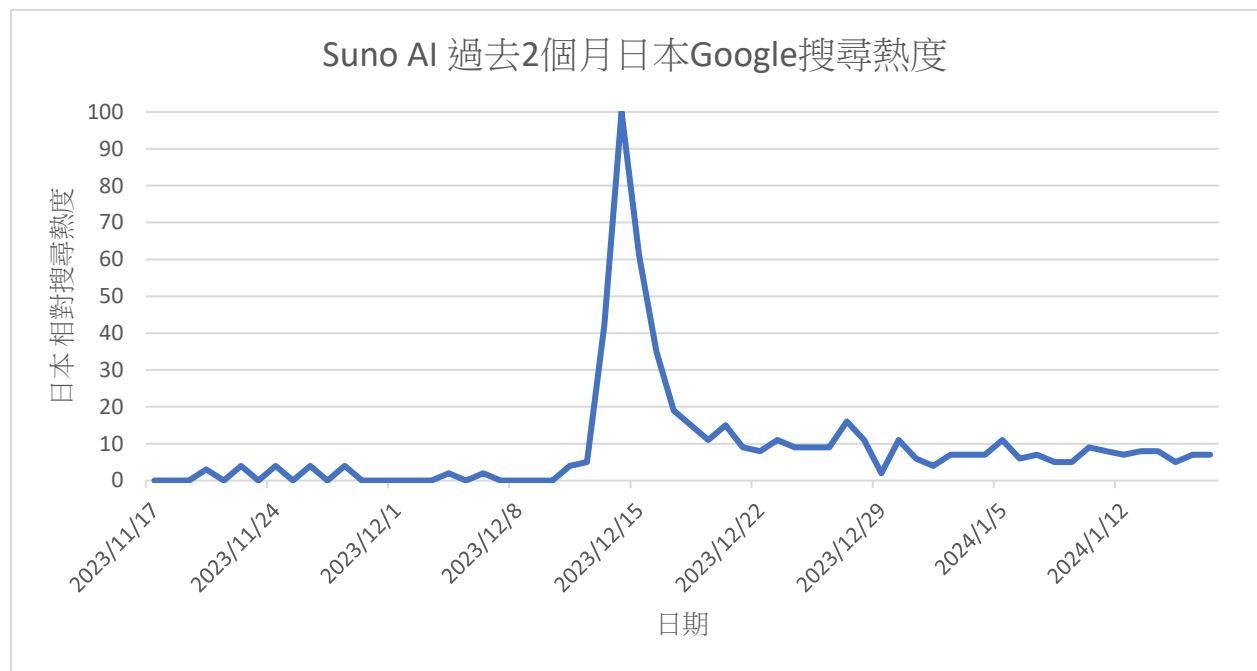


圖 15 Suno AI 日本 Google 相對搜尋熱度圖，⁵¹尾田榮一郎是在 2023 年 12 月 21 日分享他使用 Suno AI 做的歌曲

雖然有以上幾點替 Suno AI 佐證，不過畢竟 Suno AI 開始為人所知，截至目前 2024 年 1 月為止，也不過短短一兩個月，因此其究竟是否為音樂生成式 AI 之 Midjourney，還有待時間的證明。

⁴⁹ 鯛魚，〈《尾田榮一郎分享 A I 作曲》反 A I 派網友氣炸了 喊出抵制航海王重製版動畫〉，參見宅宅新聞：<https://news.gamme.com.tw/1764631>，2024 年 1 月 17 日頁面資料

⁵⁰ Google Trend，〈Suno AI 過去 2 個月日本 Google 搜尋熱度〉，參見 Google Trend：[https://trends.google.com.tw/trends/explore?date=2023-11-17%2024-01-17&geo=JP&q=suno%20ai&hl=zh-TW](https://trends.google.com.tw/trends/explore?date=2023-11-17%202024-01-17&geo=JP&q=suno%20ai&hl=zh-TW)，2024 年 1 月 17 日頁面資料

⁵¹ Google Trend，〈Suno AI 過去 2 個月日本 Google 搜尋熱度〉，參見 Google Trend：<https://trends.google.com.tw/trends/explore?date=2023-11-17%2024-01-17&geo=JP&q=suno%20ai&hl=zh-TW>，2024 年 1 月 17 日頁面資料

五、文字轉音樂生成式 AI 之問題與未來展望

(一) 訓練資料問題與展望

目前常用來訓練音樂生成式 AI 之公開資料集有 **Free Music Archive** 等，這類公開資料及裡面收錄的歌曲許多已有一定年代，且製作品質不一定好。使用這樣的資料訓練出來的音樂生成式 AI，最多就是能夠生成與其品質接近的音樂，如果目標為想要生成與現今音樂品質相近之音樂，勢必得使用一些更近代之音樂做為訓練資料。然而，相比以個人創作者占多數之繪畫群體難以在法律層造成什麼威脅，音樂產業有著更多的大公司，更有可能興訟，如環球音樂就曾因為 **Anthropic AI** 未經授權使用其旗下歌詞來訓練 AI 模型提出告訴；⁵²嘗試從網路上蒐集高品質音樂來解決這個問題的公開音樂資料集 **DISCO-10M** 目前也遭遇到了一些法律問題，^{53 54}資料集被迫從網路上下架。

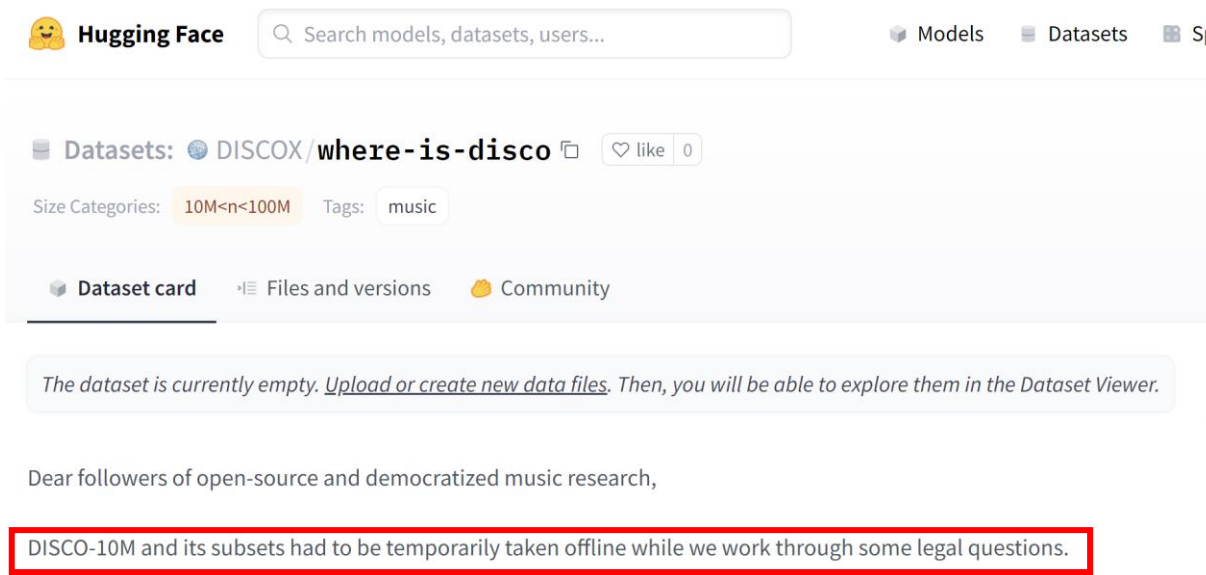


圖 16 DISCO-10M 公開音樂資料集面臨法律問題，被迫暫時下線⁵⁵

一個可行的解法是與音樂素材提供者合作，就像 **Stable Audio** 等所做的那樣。然而這些音樂素材提供者是否能夠提供時長足夠長、品質足夠好、風格足夠多元的音樂將會是個挑戰，如 **Stable Audio** 之合作音樂提供者 **AudioSparx**，旗下雖擁有超過 110 萬個音樂片段，

⁵² 林妍湊，〈Anthropic AI 助理擷取網路歌詞，遭環球音樂等唱片業者控告侵權〉，參見 iThome：
<https://www.ithome.com.tw/news/159388>，2023 年 12 月 15 日頁面資料。

⁵³ Lanzendörfer, Luca A et al. “DISCO-10M: A Large-Scale Music Dataset.” *arXiv.org* (2023): n. pag. Web.

⁵⁴ DISCO Research Group，〈where-is-disco〉，參見 Hugging Face：

<https://huggingface.co/datasets/DISCOX/where-is-disco>，2023 年 12 月 18 日頁面資料。

⁵⁵ 筆者於 2023 年 12 月 18 日於 <https://huggingface.co/datasets/DISCOX/where-is-disco> 所截

⁵⁶不過這次被 **Stable Audio** 使用的音樂素材已有超過 80 萬個，⁵⁷而其總時長也不過 2 萬小時左右。

繪圖生成式 AI 之模型之訓練資料中其實絕大多是一般的圖片，繪畫其實只是該模型之一小部分的樣貌。在模型規模擴大時，這些圖片同樣也能幫助到繪畫生成，或許音樂生成式 AI 也可以考慮通過引入更多的一般音訊檔案的方式，來擴充訓練資料。此外，有鑑於大模型所展現出的通用性，使用 **LoRA** 等技術進行精調以補足目前公開資料庫之曲風不夠廣泛之問題也是可行的方向。

（二）使用上可能存在之問題

一個可能會面臨到的問題是音樂生成式 AI 之使用者時間成本問題。相較繪畫可以一眼看到全貌，音樂總是需要聽一定時間才能了解到其整體架構與品質。在這個生出好的音樂仍需抽獎的現在，恐怕會降低人們使用音樂生成式 AI 的意願。此問題因為音樂先天的特性，恐怕沒有特別好的解法。我們能做的就是提升生成的效率以降低等待時間、提升生成品質以降低抽獎次數。

此外，由於音樂生成式 AI 品質的進步，如何能夠更好的控制其輸出結果之討論已開始被搬上檯面。⁵⁸相對繪畫修改沒有生成具有圖層之檔案還能直接塗改，以一般方法想要直接修改音樂可以說是相當困難的。因此，如何能夠提升音樂生成式 AI 之可控性將會是一個相當重要的議題。要想提升音樂生成式 AI 之可控性，一種的作法就是將繪圖生成式 AI 中正在使用的技術拿過來套，如 **Music Control-Net**。⁵⁹另外，由於在文字轉音樂生成式 AI 出現之前，音訊產業已有許多含有 AI 之產品存在，已積累了許多與 AI 之交互模式，值得生成式 AI 學習、借鑑。一個例子是拼接音樂生成 AI **SoundDraw**，其音樂生成的時候，支援用戶自由的選擇每段音樂之能量強度，以更貼近用戶之實際需求。

⁵⁶ AudioSparx 官網所寫，參見 <https://www.audiosparx.com/>，2023 年 12 月 16 日頁面資料。

⁵⁷ 陳曉莉，〈**Stability AI** 發表文字生成音樂 AI 工具 **Stable Audio**〉，<https://www.ithome.com.tw/news/158741>，2023 年 12 月 16 日頁面資料。

⁵⁸ Melechovsky, Jan et al. “Mustango: Toward Controllable Text-to-Music Generation.” *arXiv.org* (2023): n. pag. Web.

⁵⁹ Shih-Lun, Wu et al. “Music ControlNet: Multiple Time-Varying Controls for Music Generation.” *arXiv.org* (2023): n. pag. Web.



圖 17 SoundDraw 介面截圖⁶⁰

另一個例子是歌聲合成引擎 Synthesizer V，其 AI 歌唱功能之實現方式為在其合成算法之上，通過自動調整用戶也可修改之音高偏移參數，實現 AI 之便利與可控性共存。

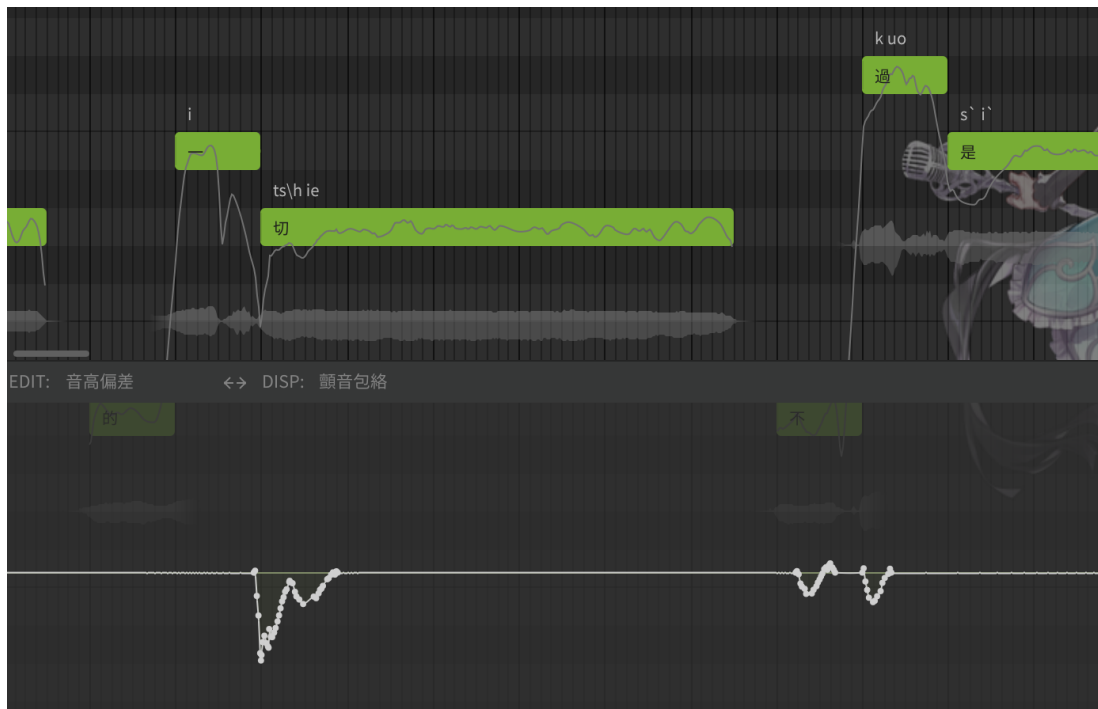


圖 18 Synthesizer V Studio Pro 1.9.0 介面截圖⁶¹

⁶⁰ 筆者於 2023 年 12 月 14 日於 https://sounddraw.io/create_music 所載

⁶¹ 筆者自行截圖

（三）繪畫生成式 AI 現在面臨之問題：音樂生成式 AI 會不會同樣陷入之？

1. 現存創作者抵制問題

2022 年 8 月，一款名為 **mimic**（模仿），主打學習特定創作者畫風之日本二次元人像繪圖生成式 AI 服務一經推出，即遭到以日本繪師為首之群眾的集體抗議，並被迫於推出隔日關停服務。⁶²



圖 19 *mimic* 服務之運作模式：使用者上傳圖片，生成類似畫風之人物圖像。⁶³

此次事件主要爭議點在於其以模仿為賣點，觸碰到了許多創作者的底線以及利益，許多畫風具辨識度的創作者擔憂利用該服務生成之圖像，若遭人惡意使用恐將危害他們的

⁶² 傅讖婷，〈AI 繪圖在全球大熱，*mimic* 卻在日本炎上！日本創作者為何對 AI 特別警戒？〉，參見數位時代：<https://www.bnext.com.tw/article/73439/after-the-release-of-the-ai-imitation-drawing-tool-mimic>，2023 年 12 月 15 日頁面資料。

⁶³ *mimic*（ミミック），〈無題〉，參見 X：<https://twitter.com/illustmimic/status/1564081818441109505>，2023 年 12 月 15 日頁面資料。

名譽與利益。此外，禁止 AI 學習的風潮也在某些繪畫群體之中蔚為風潮，⁶⁴然而，這類的宣言很可能無任何實際效力。事件發生地——日本的法律就表示，只要沒有不當損害著作權人利益，為訓練人工智慧而蒐集並使用之數據資料是屬合理使用之範疇，並不違反著作權法。⁶⁵雖然此處之不當損害著作權人利益之認定範圍沒有特別進行界定，但可以想見的是應該需要證據支持其利益遭侵犯，這在大多數的情形下都下不太可能。

而音樂生成式 AI 會不會遇到這樣的問題呢？我們可以由以下幾點來思考。一是音樂創作者之音樂風格，是否會像繪畫中的畫風一樣好辨認？這個問題的答案可能因觀察的對象而異，比如說有些人可能比較專注於作某一個曲風，或是旋律的寫作比較有自己的特色，如日本 VOCALOID 創作者傘村トータ，其作品絕大多數雖只有鋼琴伴奏搭配人聲，搭配富於同理的歌詞卻往往能打動人心，曲風辨識度很高。另一些創作者往往不限於單一個曲風，或是其曲風較大眾，因此很難以某一個特定聲響特色去辨識。綜合來說，音樂的曲風辨識度往往沒有繪畫明顯。

二是音樂創作者對於 AI 的接受程度，音樂創作者對於 AI 的接受程度普遍來說是比較高的。一個案例是孫燕姿等歌手採用 AI 技術模仿其聲線進行創作，體現了音樂創作者對於 AI 的接受；另一案例是許多不會作詞之創作者利用 ChatGPT 等文字生成式 AI 生成歌詞並用於作品或練習中，如在臺灣 VTuber 圈中小有名氣之音樂創作者 暗雨 Anyu；⁶⁶許多沒有經費請人製作封面圖的音樂創作者也會選擇使用 AI 生成之圖片或畫作當作自己樂曲的封面圖，如來自臺灣大學數位音樂創作社的 Caster 或是臺灣超強新銳音遊曲師暨音樂創作者暨職業圍棋選手 Salad Savior。^{67 68 69 70 71}

三是情感衝擊層面，此點勢必依舊存在，不過有繪圖生成式 AI 的案例在前，在心中有所預期的情形下，音樂生成式 AI 之突破所帶來的衝擊，應較繪圖生成式 AI 突破時低。

綜上所述，我們可以認為音樂生成式 AI 進展至可與現今創作者相提並論時，帶來的衝擊與創作者的抵制情緒比起繪圖來說將會更小。

⁶⁴ 觸圈，〈海外畫家集體抵制 AI 繪畫，Artstation 出現反 AI 繪畫浪潮〉，參見知乎：<https://zhuanlan.zhihu.com/p/592050105>，2023 年 12 月 15 日頁面資料。

⁶⁵ 日本著作權法第三十条の四，參見：<https://elaws.e-gov.go.jp/document?lawid=345AC0000000048>

⁶⁶ 暗雨 Anyu，〈星降る夜に / 暗雨 Anyu feat. 可不〉，參見 X：<https://twitter.com/Darkrain0320/status/1730301681664192561>，2023 年 12 月 15 日頁面資料。

⁶⁷ NTU DMCC，〈[社員作品] Caster - precession〉，參見 YouTube：<https://www.youtube.com/watch?v=PyIFwhlc2Rg>，2023 年 12 月 15 日頁面資料。

⁶⁸ Salad Savior，〈【2023 Dynamix Submission *Failed*】Pseudo-Church〉，參見 Youtube：<https://www.youtube.com/watch?v=G6mDoLMjN7g>，2023 年 12 月 15 日頁面資料。

⁶⁹ 蘿蔔@waifu_dog，〈Da Tenshi III〉，參見 pixiv：<https://www.pixiv.net/artworks/109234259>，2023 年 12 月 15 日頁面資料。

⁷⁰ Salad Savior，〈Profile〉，參見 X：<https://twitter.com/SaladSavior>，2023 年 12 月 15 日頁面資料。

⁷¹ Salad Savior 之介紹詞已獲本人之許可

2. 大眾抵制人工智慧生成之內容

雷亞遊戲是一間來自臺灣的遊戲公司，旗下知名作品有音樂節奏遊戲〈Deemo〉、〈Cytus〉等作。而雷亞旗下遊戲〈Cytus II〉於 2023 年 4 月推出之 5.0.3 更新中所釋出的圖片，⁷²由於畫風與市面上之繪圖生成式 AI 類似、細節有許多錯誤及不合理之處，加上雷亞先前曾在徵才平台上公開招聘 AI 美術設計師/AI 溝通師，⁷³被認為是由 AI 生成再加工修改而成，於網路上引起了軒然大波。以下為該引起爭議之圖 20：



圖 20 〈Cytus II〉中疑似使用 AI 輔助創作的畫作⁷⁴

Novel AI 之繪圖部分是由來自美國的 Anlatan 公司所開發並於 2022 年 10 月初正式公開的繪圖生成式 AI 服務，⁷⁵為首個獲得廣泛關注的日本二次元風格繪圖生成式 AI。而由於 Novel AI 的出現，以日本動漫風為主的知名繪畫交流網站 Pixiv 上開始出現大量的、由 Novel AI 產生的圖片；而又由於其圖片風格太過類、占據網站過多版面等因素，造成部分

⁷² Cytus，〈Cytus II 5.0.3 現已推出〉，參見 FaceBook：

<https://www.facebook.com/photo/?fbid=620118580148959&set=a.612600197567464>，2023 年 12 月 16 日頁面資料。

⁷³ 鄧舒澤，〈雷亞 AI 溝通師職缺引熱議，作為數位時代下的「魔法師」需要哪些關鍵技能？〉，參見未來商務：<https://fc.bnext.com.tw/articles/view/2612>，2023 年 12 月 15 日頁面資料。

⁷⁴ Cytus，〈Cytus II 5.0.3 現已推出〉，參見 FaceBook：

<https://www.facebook.com/photo/?fbid=620118580148959&set=a.612600197567464>，2023 年 12 月 16 日頁面資料。

⁷⁵ 吉川大貴，〈「とんでもなくハイクオリティー」話題の画像 AI「Novel AI」でひたすら二次元美少女と美少年を生成してみた〉，參見 ITmedia NEWS：

<https://www.itmedia.co.jp/news/articles/2210/03/news146.html>，2023 年 12 月 16 日頁面資料。

用戶的不滿。⁷⁶隨後，Pixiv 官方於同月月底發佈數個應對 AI 生成作品的相關功能，⁷⁷包括提供標註 AI 生成作品的選項、提供篩選與過濾 AI 生成作品之功能、專屬 AI 生成作品之排行榜，這才使這波抗議暫時告一段落。

這兩個事件，共通點在於原先都是以美術為主或美術重要程度很高的事物，在繪圖生成式 AI 出現以後出現了變化，而使部分用戶感到失望：真正令用戶感到失望的不是 AI 圖，而是不佳的品質。大量未經挑選、修正、缺乏設計的 AI 生成圖片湧現到網站上，淹沒了其他優秀的作品、影響了原先用戶的體驗；過去相當倚重美術之遊戲公司將沒有修正好的 AI 圖放進其遊戲中，等於是背叛了原先被其美術吸引的玩家。

此問題，在音樂生成式 AI 有所進展後，勢必也會發生在音樂上，惟程度可能有所不同。音樂分享平台勢必會開始出現音樂生成式 AI 生成之音樂作品，也一定將會有公司基於成本等考量，採用 AI 生成之音樂。

而解決方法、或降低影響的方法，大致有以下兩點。一是平台要針對使用 AI 生成之作品作出區隔，降低對於不喜歡 AI 生成作品用戶的影響；二是要認清自己的核心競爭力為何，不要在自己重要的地方減低成本。這不是代表不能採用 AI，是要確保最重要的地方品質不會下滑。而若音樂對於自己是不那麼核心，或甚至採用 AI 會使品質提升的話，那麼用戶也會樂見這樣的改進。

（四）音樂生成式 AI 未來可能造成之問題

1. 未來創作者的成長環境問題

音樂生成式 AI 的發展或許有能幫助未來創作者更快速的成長，但同時其也有可能成為未來創作者成長路上的絆腳石。

許多創作者創作的理由中，多少都含有想要被認同之想法，音樂生成式 AI 的進步，可能會降低創作者獲得之肯定；音樂創作的特殊性可能會消失，音樂創作者的自信心可能會受到影響；一些低成本的項目可能會開始採用 AI 生成之音樂，可能會使一些音樂製作人失去工作、練習之機會，更糟者可能被迫轉行；使用 AI 輔助創作的人可能會在其不擅長之地方產生對 AI 之依賴，並且由於並不理解，導致該處遲遲無法進步，難以更上一層樓；音樂創作者在創作之過程中之探索時常會成為其風格之基礎，音樂生成式 AI 的誕生可能會減少探索的需求，影響其風格之成形。

⁷⁶ fste，〈無題〉，參見 X：<https://twitter.com/ftrtn/status/1581107177066205186>，2023 年 12 月 18 日頁面資料。

⁷⁷ pixiv 公告，〈應對 AI 生成作品的相關功能現已上線〉，參見 pixiv：<https://www.pixiv.net/info.php?id=8728>，2023 年 12 月 15 日頁面資料。

2. 生成式 AI 之大鍋飯與劣幣驅逐良幣問題

生成式 AI 大鍋飯問題是指若未來市場需求以修改 AI 所生成之作品為主流，且酬勞區別不大時，可能會導致的由於做得好或做得爛都差不多，因此喪失生產積極性的問題。以下為一假設，不代表未來的真實情形，也不代表繪圖生成式 AI 就一定無法克服細節錯誤這個問題，僅為了方便解說。

如果未來，依舊存在許多細節上的瑕疵，如下範例圖 21：



圖 21 銀髮教主 via AI 〈雨の中の銀髮少女〉⁷⁸



圖 22 為圖 21 之局部放大，可以看到其鈕扣及拉鍊都生成的有些不自然

要整修這些瑕疵往往費力耗時，最終結果卻往往差別甚微。包括委託者在內的絕大多數人，都難以分辨這些修正或未修正之作品之差異，因此用心與不用心勢必會獲得相近的酬勞，長久下來即可能降低生產積極性。

而劣幣驅逐良幣問題則是大鍋飯問題的加劇版本。當供給市場過剩（很有可能，因為入行門檻因為 AI 而降低），由於競爭造成的價格降低衝擊，最劇烈的即是這些更用心

⁷⁸ 銀髮教主，〈雨の中の銀髮少女〉，參見 pixiv：<https://www.pixiv.net/artworks/113640951>，2023 年 12 月 15 日頁面資料。

在每一個作品上面的人，因為其時間成本較高，在單價相近的情形下，最後肯定無法與其他沒有那麼用心的人比價格戰，因此最終只能選擇退出，造成整體繪畫品質下滑的問題。

生成式 AI 的誕生應該是要造福人類、降低進行創作之成本與阻礙並使創作品質進步，若造成整體品質下滑實在有違其初衷，因此我們須審慎的面對此類問題，確保生成式 AI 對於整體人類而言具有正面之效應。

（五）生成式 AI 發展時程預測、展望與解釋

1. 什麼時候會出現像 Stable Diffusion 一樣強大、開源的音樂生成模型？

強大開源模型的出現，能使生成式 AI 相關之開發成本大幅降低、促進創新，對於生成式 AI 知進步與生態建構而言有重大的意義。目前已經開源的 MusicGen，由於其生成效果不夠強大，沒有什麼知名的成果隨其產生；未來將開源的 Stable Audio，目前展現出來的效果也不構驚艷，加上未來開源時將換用其他訓練資料，品質很可能不生反降。目前現有及可預見的模型皆不可能，那麼，何時會出現像 Stable Diffusion 一樣強大、開源的文字轉音樂生成式 AI 模型呢？

同樣的簡單回顧一下 Stable Diffusion 的貢獻，Stable Diffusion 不僅效果足夠好，同時解決了算力與記憶體成本過高之問題，使其開源之意義更為重大。若假定 Suno AI Chirp 為音樂生成式 AI 之 Midjourney，那麼其背後之模型或許是以生成頻譜之 Stable Diffusion 為基礎，也就是說模型與成本上的問題或許已經有解，於是剩下的阻礙就剩下訓練資料了。目前公開的音樂資料集比較缺乏目前最頂尖的音樂，數量也不算龐大，加上音樂可能具有的版權，或許將會是未來開源音樂生成式 AI 遇到最大之阻礙。如果沒有現代的大型公開音樂資料集，那麼音樂生成式 AI 的市場將被擁有大量音樂版權或是不在乎音樂版權的公司壟斷，對於社會與音樂生成式 AI 的發展來說都不是好事，因此建立無或較無版權疑慮的當代大型公開音樂資料庫有其必要。

2. 為什麼音樂生成式 AI 發展的比繪圖生成式 AI 緩慢？未來速度會不會加快？

音樂生成式 AI 發展比繪圖生成式 AI 慢有許多可能的理由，其中目前音樂領域中比較沒有像 LAION 那樣巨大的資料集，⁷⁹而且音樂的版權問題相較繪畫、圖像之版權問題須更謹慎的對待，或許會使訓練資料之質與量降低；先前音樂生成式 AI 沒有哪個模型具有明顯的優勢，屬於探索的階段，針對音樂生成式 AI 之研究方向不集中，也使音樂生成式 AI 的發展比起先前找到擴散模型的繪圖生成式 AI 看起來慢上許多；音樂生成式 AI 中類似 Stable Diffusion 那樣的基石模型尚未出現，開發應用、技術開發之成本、門檻仍居高不

⁷⁹ Romain Beaumont，〈LAION-5B: A NEW ERA OF OPEN LARGE-SCALE MULTI-MODAL DATASETS〉，參見 laion.ai：<https://laion.ai/blog/laion-5b/>，2023 年 12 月 14 日頁面資料。

下。這些原因，都可能是音樂生成式 AI 發展的比繪圖生成式 AI 緩慢的原因。現在 Suno AI Chirp 的出現，或許將會終結音樂生成式 AI 對於模型的探索，加速音樂生成式 AI 的發展。

依照目前之情勢來看，音樂領域缺乏公開資料庫仍舊是目前最大的問題，一個試圖解決此問題的嘗試是先前有提及之 DISCO-10M，⁸⁰其以 YouTube 為音樂資料來源，⁸¹並且通過多層的過濾以保證音樂的品質，若能成功或將能很大程度的解決音樂生成式 AI 目前面臨之訓練資料問題。然而，DISCO-10M 正面臨一些法律問題，資料集也被迫暫時下線。缺乏大量優質公開訓練資料的音樂生成式 AI 將很難發展出強大的開源 AI，而缺乏強大的開源會使投入開發、研究技術的門檻無法降低，而閉源的音樂生成式 AI 如 Suno AI Chirp 也會因為缺乏競爭與沒辦法吸收開源社區的研究成果而發展的不像繪圖生成式 AI 那麼快。

（六）法律層面之展望

1. 現行生成式 AI 之訓練過程是否違反著作權法、是否為合理使用？

在生成式 AI 蓬勃發展的今日，經常會有創作者主張生成式 AI 未經授權使用其作品進行訓練，侵犯了其之著作權或表明禁止 AI 學習其作品，如作家崔布雷（Paul Tremblay）和阿瓦德（Mona Awad）對 OpenAI 提告或是在數位藝術網站 ArtStation 中爆發的抗議行為。^{82 83}

所以，究竟使用從網路上蒐集的資料訓練生成式 AI 模型有沒有構成侵權呢？要探討這個問題，我們得先從最初制定著作權法之理由與生成式 AI 原理說起。

根據我國著作權法第 1 條「為保障著作人著作權益，調和社會公共利益，促進國家文化發展，特制定本法。本法未規定者，適用其他法律之規定。」可以很清楚地認識到，著作權法的設立之宗旨在於平衡創作者權益以及公眾之利益，⁸⁴在保證創作積極性的前提下也能使大眾、社會受益。

生成式 AI 之原理為機器學習。機器學習的宗旨簡單來說就是透過大量的輸入輸出尋找一個適當的函數，以實現對新輸入的有效預測。其過程中包含了選擇合適的模型結構、

⁸⁰ Lanzendörfer, Luca A et al. “DISCO-10M: A Large-Scale Music Dataset.” *arXiv.org* (2023): n. pag. Web.

⁸¹ DISCO-10M 資料集不含音樂檔案本身，需從 YouTube 下載

⁸² Ashley Capoot, 〈Authors sue OpenAI, allege their books were used to train ChatGPT without their consent〉, 參見 CNBC: <https://www.cnbc.com/2023/07/05/authors-sue-openai-allege-chatgpt-was-trained-on-their-books.html>, 2023 年 12 月 16 日頁面資料。

⁸³ 遊戲葡萄, 〈繪師怒氣爆發, 用「抵制 AI 繪圖」洗版抗議〉, 參見科技新報: https://technews.tw/2022/12/18/artstation_no-to-ai-generated-images/, 2023 年 12 月 16 日頁面資料。

⁸⁴ 著作權法（民國 111 年 06 月 15 日）。參見：<https://law.moj.gov.tw/LawClass/LawAll.aspx?PCode=J0070017>

調整參數，最終目標是使模型具有良好的泛化能力，而不僅僅是對訓練數據的擬和，因此可以認為其沒有重製等一般意義上的侵權行為。

綜上所述，我們可以知道若訓練之最終目的為供社會大眾使用，應在不損害創作積極性的前提下，在必要的範圍內將用於 AI 訓練之資料蒐集與使用劃入合理使用之範疇。

2. 若台灣認定生成式 AI 之訓練非合理使用會有什麼問題

生成式 AI 通過大量數據，習得了音樂或繪畫中的特質並能運用在生成新的作品上，其實本質上與人類藝術家學習過程相當接近，若要認定現行生成式 AI 之訓練為非合理使用，可能會使法律自相矛盾，造成爭議與對法律的信任危機。

另外，目前已經有許多開放原始碼及參數權重之模型在網路上，已經被很多人下載了很多次，在世界各地有無數份的原始檔。縱使全世界都認定生成式 AI 之訓練非合理使用，也無法阻止這些模型具續被使用、繼續「侵害」創作者的權利。更何況，有些國家已經在法律層面上認可使用數據進行訓練為合理使用之範疇，如日本於 2018 年修訂之著作權法三十條之四之二，就明文表示了從大量作品提取等大量資訊中提取構成訊息的語言、聲音、圖像等要素相關的信息，並進行比較、分類等分析的行為為合理使用之範疇，在必要的範圍內可以自由使用。⁸⁵遵守臺灣法律之企業或模型將因授權素材有限、成本更高，在與來自這些國家的生成式 AI 模型之競爭中將難以取得優勢，錯失建立下一個重大的產業根基的機會。

還有一個問題是在生成式 AI 誕生初期，許多公司或模型都會公佈其使用的訓練資料來源。或許是因為需要的訓練的資料即越來越多，難以全數取得授權又怕被批評未經授權，最後越來越多公司或模型選擇不公開自己訓練資料的來源。然而，這樣的行為很有可能違反了著作權中的著作人格權。根據我國著作權法第 66 條，縱使使用他人作品可能屬合理使用之範疇，也不能侵害他人之著作人格權。

然而，只要我們的法律認可將資料用於訓練生成式 AI 屬合理使用之範疇，就能令人放心的公開訓練資料集，很大程度的解決這個問題。同時，公開資料集也能促進 AI 相關研究的交流與進步，可謂一舉多得。

⁸⁵ MONOLITH LAW OFFICE，〈ネット上の画像のクローリングは著作権法違反？機械学習の法的問題を解説〉，參見 monolith.law：<https://monolith.law/corporate/copyright-machine-learning/#%E5%B9%B3%E6%88%9030%E5%B9%B4%E8%91%97%E4%BD%9C%E6%A8%A9%E6%B3%95%E6%94%B9%E6%AD%A3%E3%81%AE%E8%83%8C%E6%99%AF>，2023 年 12 月 16 日頁面資料。

3. 生成式 AI 是否會影響創作積極性

能夠促進創作積極的社會因素有許多，其中名與利之保障佔了相當重要的成分，也就是說，生成式 AI 是否可能扼殺創作積極性可由其對創作者之名與利之侵害程度來窺視。

若以目前，生成式 AI 多未表揚（Credit）訓練資料之原作者們的情形來說，生成式 AI 有可能會影響創作者獲得名聲。具體來說，過去的美術創作者若發明了某種畫法、改動了某種畫風，影響後續創作者甚多，縱使沒有因此獲利也能夠藉此收穫歷史地位。⁸⁶而以現在生成式 AI 的情形來看，過去的藝術家除特定名氣高者如畢卡索、梵谷等會出現在提示詞中，基本上沒有被提及。未來開創新畫法者，最有可能獲得知名度的方式是透過 LoRA 等 fine-tune 模型的傳播，與過去傳統意義上的名聲有所不同，仍需待時間檢驗其對於創作者的吸引力如何。

生成式 AI 更直觀的是會影響到創作者的金錢利益，假設市場對於創作的的需求沒有大幅度的提升，生成式 AI 低成本的特性，必然使其在市場中具有一定地位。此外，部分或全部工作也會將其 workflow 開始融入生成式 AI，這帶來的更低之入行門檻將造成供給競爭更加激烈、可取代性增加，很有可能伴隨著相對低的酬勞。

綜上所述，雖說生成式 AI 的誕生沒有必然伴隨著創作積極性的下降，但我們同樣的不能確保這件事不會發生。因此，建立、修改制度以預防其之發生仍然有其必要性。

創作是創新的前提，而創新是推動領域進展的原動力，由此可見創作積極性的重要性。我們不妨做個思想實驗，來進一步體會缺乏創新的嚴重性。假設在那個貝多芬仍能夠聽到聲音的時候，音樂生成式 AI 就被訓練出來並且被廣泛使用，而人們就此失去了創作積極性，未來數百年都了無創新，那麼現在的音樂會長什麼樣子呢？當然，巴哈、莫札特、貝多芬的音樂有其美妙之處，但現在的音樂就會少了許多可能性、少了許多色彩，這將會是多麼的令人惋惜啊！

4. 建立表揚（Credit）機制與分潤機制

續方才所提到的，表揚創作者於生成式 AI 的產物中的貢獻、給予應得名譽者應得之名譽，有其重要性與必要性存在。除了確立合理使用促使公開訓練資料來源以外這種一視同仁之作法外，也須根據其實際貢獻分配名譽，才能夠體現促進創新積極性的意義。一個例子為目前正在開發的 HiiMeta，其展現出了類似的思考，可供參考借鑒。

⁸⁶ 抖抖村，〈你的畫風來自哪個流派?|動畫繪畫風格演進系列〉，參見 YouTube：<https://www.youtube.com/watch?v=SGWUvd5ncqA>，2023 年 12 月 16 日頁面資料。

HiiiMeta 為北京光合界科技有限公司正開發中的 AIGC 共創社區，⁸⁷由於筆者未能取得內部測試資格，以下內容僅從包括其董事長所撰寫之文章等其它資訊來源推敲而成，不代表其成品效果。然因此處僅欲借用其思考，故應無大影響。

HiiiMeta 自己定位自己為一個「集藝術風格確權、授權和使用為一體的 AI 藝術生態」⁸⁸。確權部分，其將風格提供者稱為開源藝術家，開源藝術家將風格授權給其他使用者，而其他使用者使用平台生成之圖像在平台上即會留下所謂「基因脈絡」。其於宣傳片中展示了其概念，如下圖 20：

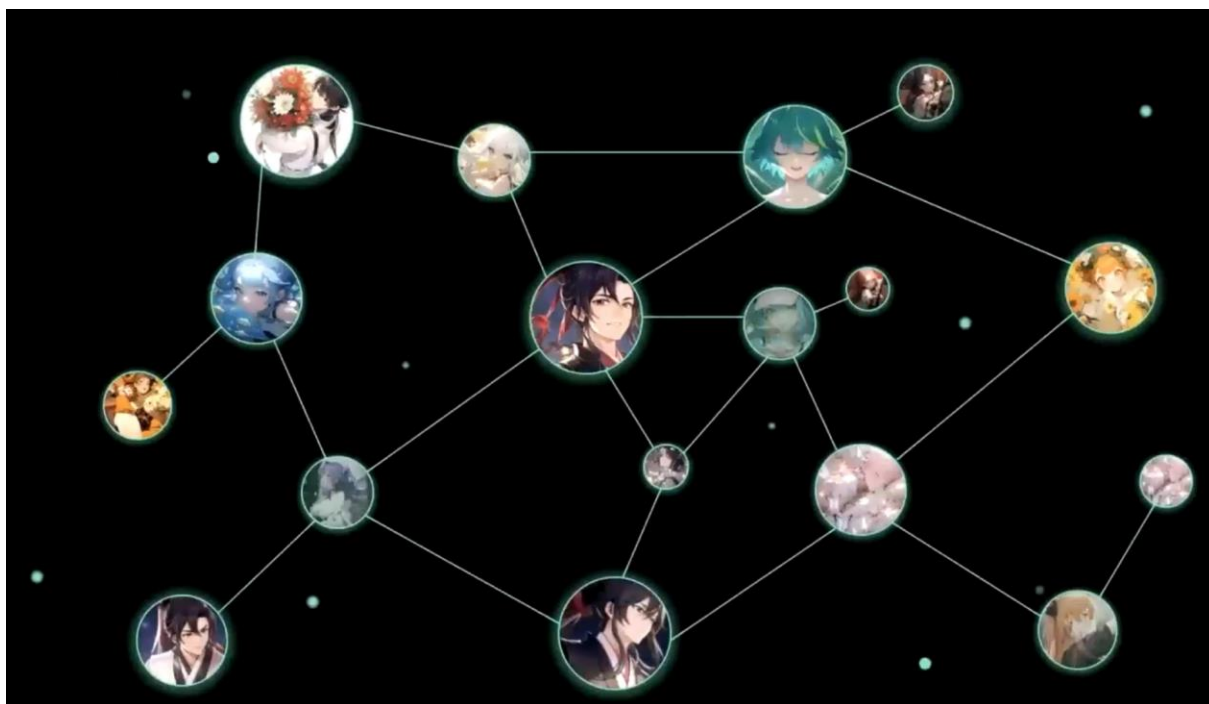


圖 23 HiiiMeta 畫風「基因脈絡」示意圖⁸⁹

續名利之給予部分，給予創作者合理的利益分配同樣是至關重要的。細節部分，要根據其作品實際對於生成式 AI 生成作品之貢獻來進行分配，才能夠體現制度建立之目的：促進創新的意義。

實際案例方面，剛剛所提及之 HiiiMeta 同樣有關於這方面的規劃可供參考借鑒。在 HiiiMeta 中，生成圖像之費用部分會根據風格來源進行分潤，只要一直被後續使用者使用

⁸⁷ 企查查，〈北京光合界科技有限公司〉，參見 qcc.com：

<https://www.qcc.com/firm/85d9963fa1358b1221fc02fa78cfd9e.html>，2023 年 12 月 16 日頁面資料。

⁸⁸ 魔王醬，〈無題之回答〉，參見知乎：<https://www.zhihu.com/question/568804205/answer/2777680843>，2023 年 12 月 16 日頁面資料。

⁸⁹ AICan 的 B 站號，〈AICanTalk#13 | HiiiMETA – AIGC 協助 IP 共創社區〉，參見 bilibili（0 時 9 分 56 秒）：www.bilibili.com/video/BV1Pz4y1N7Xu，2023 年 12 月 16 日頁面資料。

就能一直產生收入，而交了授權費用的人們就能在規定的範圍內使用所生成之圖片，如下圖 21：

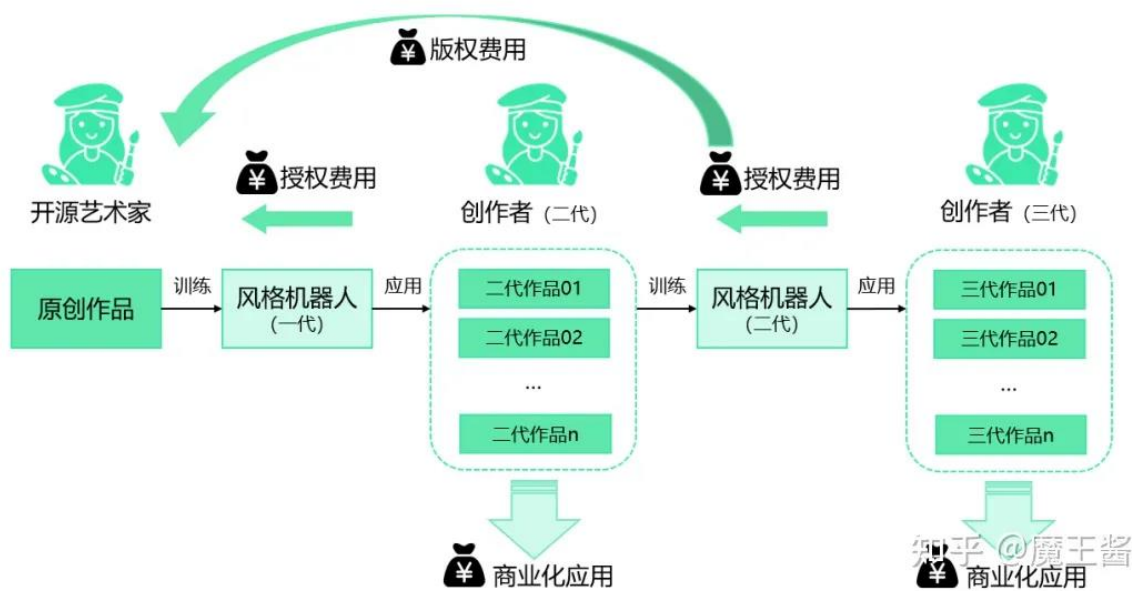


圖 24 HiiiMeta 分潤示意圖^{90 91}

而因為建立分潤制度之初衷為保障創作積極性以促使創新，因此無論最終具體分潤細節怎麼設計，需將其創新程度納入考量，且應依作品誕生時間有所減少。

5. 建立 AI 生成內容之正盜版制度與意識

在前一章節，我們提到了要建立分潤機制，這是否與更早敘述之「模型不會消滅」、
「遵守法律取得授權之台灣企業、模型將缺乏競爭力」產生矛盾呢？這裡的回答是既是又不是。

不是的部分在於我們主張為訓練使用資料為合理使用，及代表至少在同樣的情形下，模型品質沒有落後的理由，落後即為本身模型架構等其它地方具有問題；生成式 AI 的前沿進展同樣不會被影響，只是在應用落地時會有些變化。是的部分在於確實在同樣的條件下，會造成更高的成本，但是的地方又不是是因為此主張之前提為我們必須建立 AI 生成內容之正盜版制度與意識。

AI 生成內容之正盜版制度與意識雖說聽起來冠冕堂皇，其實不過也就是有表揚（Credit）、有分潤的綜合體而已。不過就是這個語意的細微轉變，將造成我們對此事完

⁹⁰ 魔王醬，〈AI 煉丹黑魔法：當開源藝術家遇見 HiiiMeta！〉，參見知乎：
<https://zhuanlan.zhihu.com/p/584958306>，2023 年 12 月 16 日頁面資料。

⁹¹ 魔王醬為北京光合昇科技有限公司 CEO 陳定媛在知乎上的暱稱

全不同的態度。以正盜版為名，代表我們將一切未表揚、未分潤的生成式 AI 認定為盜版、為非法的行為，若不如此，那麼我們辛苦建立起的生態將因無實際誘因而面臨失敗的必然結局；任何行為，特別是商業行為在使用生成式 AI 時，必須掂量掂量其使用盜版生成式 AI 被揭穿之法律後果。因此，我們須建立起合適的辨別方法以區分正盜版之生成式 AI，如透過中央資料庫儲存所有經正版途徑產生之 AI 生成物。

為使得大眾使用正版之生成式 AI，正版之生成式 AI 必須足夠便利且價格不能夠造成太大的負擔。同時也須盡力打擊盜版、使獲取盜版的成本增加。我們同樣也需在大眾中推廣生成式 AI 之正盜版意識，使其認知到此制度帶來之益處將遠大於所需額外付出之成本。在有眾多軟體前例以及正版意識抬頭的現在，這應該不會是一件太困難的事。當然，若事態要上升到此層級，那肯定不是只有臺灣的事了，而臺灣也能趁機於此領域取得話語權，為將來產業的發展打下良好的基礎。

（七）小結

本章節討論了許多音樂生成式 AI 現在及未來可能會碰到、造成的問題，其中包括有可能會影響到強大開源模型誕生的訓練資料與版權問題、可能會影響到未來使用者使用意願之問題、影響到能否與原先音樂創作社群和平共處的創作者抵制問題、影響到大眾接受程度的大眾抵制問題、影響到未來音樂創作者成長速度與心態的創作者成長環境問題、影響到生產與創作積極性的生成式 AI 大鍋飯與劣幣驅逐良幣問題、影響到未來音樂生成式 AI、開源模型發展及企業競爭力的著作權法合理使用問題和版權要求嚴格導致著作人格權受損等等的問題。

本章節同樣藉由這些問題探討了未來音樂生成式 AI 的一些展望，包括音樂生成式 AI 與強大開源模型的發展速度可能受版權問題所限、著作權法的未來、生成式 AI 之分潤模式的建立、生成式 AI 正盜版意識建立等等。

本文希望透過這些廣泛的探討，激起社會對生成式 AI 可能造成的問題的重視，並且能採取必要的手段以保證生成式 AI 造成之損害不會大過其所帶來之貢獻。

六、結論

本文通過介紹幾個文字轉音樂生成式 AI 的方式，簡單的回顧了現今文字轉音樂生成式 AI 的進展，同時透過參考繪圖生成式 AI 之歷史之方式，提出了 Suno AI Chirp 可能是音樂生成式 AI 之 Midjourney 之論述。此外，我們也通過參考音樂生成式 AI 本身的特性與繪圖生成式 AI 之歷史，提出了數個音樂生成式 AI 正在面臨、可能會面臨或可能造成的問題，供社會大眾及相關人士參考借鑒並引起對生成式 AI 可能造成的危害之重視。一個例子是音樂生成式 AI 正面臨缺乏大型現代公開音樂資料集之問題，可能會造成強大且開源之音樂生成式 AI 遲遲無法誕生，或將使音樂生成式 AI 被壟斷且進展緩慢。

文字轉音樂生成式 AI 正在成為逐漸成為現實，然而，在使音樂生成式 AI 造福人類的路上仍舊有許多的阻礙，等待著我們去拔除；同時，我們也必須審慎面對那些音樂生成式 AI，或甚至是任何生成式 AI 可能造成之危害，確保這股強大的力量能引導人類邁向更好的未來。

附件

一、FAD ∞ 實驗

此表目的為實驗看看將一般的音樂拿去給 Frechet Audio Distance Toolkit 算 FAD ∞ 分數大概會得到多少，參考資料集為 FMA Pop，embedding 使用 CLAP-2023。使用 FMA Pop 最為參考資料集是因為 Azalea Gui 等之論文 Adapting Frechet Audio Distance for Generative Music Evaluation 中有提到 FMA Pop 是相對合適的一個參考資料集；使用 CLAP-2023 的主因是我們認為他的敘述更貼切我們的需求，而且算力與記憶體需求較低，可以更方便的進行實驗。選擇歌曲的標準主要以筆者個人熟悉度為主。

Song ID	Song Title	Artist	FAD ∞ Score(CLAP-2023)
1	KING	Kanaria	728.0633041
2	テオ	Omoi	704.2573982
3	リアライズ	柊マグネタイト	767.6725828
4	crystallized	Camellia	594.6822076
5	Synthesis	tn-shi	666.7573990
6	Cyberfantasia	tn-shi	677.0225954
7	PARTY in PSPACE	tn-shi	750.8964258
8	ONESELF	Twinfield	712.1205848
9	METEOR	DIVELA	657.3310782
10	Everything Goes On	Porter Robinson	584.0240500
11	Look at the Sky	Porter Robinson	498.3902578
12	I REALLY WANT TO STAY AT YOUR HOUSE	ROSA WALTON HALLIE COGGINS	548.3821032
13	Whats up Pop	Capchii	742.6438598
14	Beyond Sunshine	ABlackCat	719.8819497
15	Beyond Apocalypse	ABlackCat	693.3618967
16	Re: Lonely Walker	ABlackCat	696.1439689
17	108 課綱	ABlackCat	713.6801300
18	奴役	ABlackCat	679.1677210
19	weird2	ABlackCat	713.6473818
20	weird1	ABlackCat	868.0619992
21	sf	ABlackCat	1064.2952914
22	記憶	ABlackCat	781.3558224
23	little star cover	ABlackCat	887.0728447

24	Train cover	ABlackCat	974.5865690
25	Cat war cover	ABlackCat	1028.1665367
27	一千光年	いよわ	487.7119416
28	きゅうくらりん	いよわ	510.9566992
29	ジェリーフィッシュ	Yunomi	586.1005566
30	まにまに	r-906	561.3273668
31	リメンシア	せきこみごはん	588.0430106
32	くうになる	MIMI	651.8785725
33	Arcahv	Feryquitous	545.1971093
34	Altale	Sakuzyo	513.8717177
35	Destr0yer	Sakuzyo	535.071078
36	Lapis	SHIKI	686.271701
37	Hifumi Daisuki (Remix)	Quilt heron	629.8852127
38	Aoharu Band Arrange	Blue Archive	603.5618063
39	Constant Moderato	Blue Archive	686.4337044
40	RE Aoharu	Blue Archive	606.5639948

二、各音樂生成式 AI FAD ∞ 分數表

此表為將 AI 生成的音樂拿去給 Frechet Audio Distance Toolkit 算 FAD ∞ 分數大概會得到多少，參考資料集為 FMA Pop，embedding 使用 CLAP-2023。音樂來源為筆者使用官方或第三方服務生成，或在官方網站下載他人生成之音樂，沒有取得內測資格的 MusicLM 音樂來源為網路上的試用影片。⁹²

生成式 AI 名稱	FAD ∞ Score (CLAP-2023)
Suno AI	693.7563793
	654.8034944
	687.2502145
	532.5300641
	776.9902604
Riffusion	740.8942284
	697.9160189
	720.0964292
	644.4673583
	700.1148202
Stable Audio	720.4391815
	718.1362554
	614.7838708
MusicGen	692.1298676
	761.2384022
	759.6227184
MusicLM	842.0924543
	764.4076926
	778.733547

⁹² Mike Russell，〈Google MusicLM: Turn Text into Mind-Blowing Music! (AI GENERATED MUSIC IS REAL)〉，參見 Youtube：<https://www.youtube.com/watch?v=IwtXUcbs7aI&t=1369s>，2024 年 1 月 17 日夜面資料。

參考資料

書籍與期刊著作

- Agostinelli, Andrea et al. “MusicLM: Generating Music From Text.” *arXiv.org* (2023): n. pag. Web.
- Borsos, Zalan et al. “AudioLM: A Language Modeling Approach to Audio Generation.” *IEEE/ACM transactions on audio, speech, and language processing* 31 (2023): 2523–2533. Web.
- Copet, Jade et al. “Simple and Controllable Music Generation.” *arXiv.org* (2023): n. pag. Web.
- Defferrard, Michaël et al. “FMA: A Dataset For Music Analysis.” *arXiv.org* (2017): n. pag. Web.
- Gui, Azalea et al. “Adapting Frechet Audio Distance for Generative Music Evaluation.” *arXiv.org* (2023): n. pag. Web.
- James Seale Smith et al. “Continual Diffusion: Continual Customization of Text-to-Image Diffusion with C-LoRA.” *arXiv.org* (2023): n. pag. Web.
- Ji, Shulei, Jing Luo, and Xinyu Yang. “A Comprehensive Survey on Deep Music Generation: Multi-Level Representations, Algorithms, Evaluations, and Future Directions.” *arXiv.org* (2020): n. pag. Web.
- Lanzendörfer, Luca A et al. “DISCO-10M: A Large-Scale Music Dataset.” *arXiv.org* (2023): n. pag. Web.
- Le, Quoc V. “Building High-Level Features Using Large Scale Unsupervised Learning.” *2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013.* 8595–8598. Web.
- Melechovsky, Jan et al. “Mustango: Toward Controllable Text-to-Music Generation.” *arXiv.org* (2023): n. pag. Web.
- Rombach, Robin et al. “High-Resolution Image Synthesis with Latent Diffusion Models.” *arXiv.org* (2022): n. pag. Web.
- Schneider, Flavio et al. “Moûsai: Text-to-Music Generation with Long-Context Latent Diffusion.” *arXiv.org* (2023): n. pag. Web.

- Zhang, Lvmin, Anyi Rao, and Maneesh Agrawala. “Adding Conditional Control to Text-to-Image Diffusion Models.” *arXiv.org* (2023): n. pag. Web.

網路新聞與其他資料

- L. A. Hiller, Jr., and L. M. Isaacson, *Illiac Suite for String Quartet, New Music Edition, Vol. 30, No. 3*, Bryn Mawr, Pa., Theodore Presser Co., 1957.
- Ed Lauder, 〈Aiva is the first AI to Officially be Recognised as a Composer〉, 參見 AI BUSINESS : <https://aibusiness.com/verticals/aiva-is-the-first-ai-to-officially-be-recognised-as-a-composer>, 2023 年 12 月 15 日頁面資料。
- AIVA, *Op.24 “I AM AI”*, 2017
- 中央通訊社, 〈AI 聊天機器人 ChatGPT 用戶數成長速度創紀錄〉, 參見中央通訊社 : <https://www.cna.com.tw/news/ait/202302020181.aspx>, 2023 年 12 月 15 日頁面資料。
- 36 氪, 〈Midjourney : AIGC 現象級應用, 一年實現 1000 萬使用者和 1 億美元營收〉, 參見 yahoo!新聞 : <https://reurl.cc/GKzvED>, 2023 年 12 月 15 日頁面資料。
- Stability AI, 〈Stable Audio: Fast Timing-Conditioned Latent Audio Diffusion〉, 參見 stability.ai : <https://stability.ai/research/stable-audio-efficient-timing-latent-diffusion>, 2023 年 12 月 15 日頁面資料。
- 林亦 Lyi, 《聊天就能程式! 我用 GPTs 做了個自己的數字化身》, 參見 YouTube : <https://www.youtube.com/watch?v=pjKD3coLju0>, 2023 年 12 月 15 日頁面資料。
- DigiLog 作者群, 〈將聲音視覺化後重新作曲? 全新的 AI 音樂生成項目 Riffusion〉, 參見 DigiLog : <https://digilog.tw/posts/1633>, 2023 年 12 月 15 日頁面資料。
- Kyle Wiggers, 〈AI-generating music app Riffusion turns viral success into \$4M in funding〉, 參見 TechCrunch : <https://techcrunch.com/2023/10/17/ai-generating-music-app-riffusion-turns-viral-success-into-4m-infunding/>, 2023 年 12 月 15 日頁面資料。
- Web3 天空之城, 〈AI 繪畫何以突飛猛進? 從歷史到技術突破, 一文讀懂火爆的 AI 繪畫發展史〉, 參見微信 QQ : <https://mp.weixin.qq.com/s/LsJwRMSqPXSjSyhoNgzi6w>, 2023 年 12 月 15 日頁面資料。

- Hank Campbell , 〈 Singularity Update: 16,000 Processors To Identify A Cat 〉 , 參見 Science 2.0 :
https://www.science20.com/science_20/singularity_update_16000_processors_identify_cat-91926 , 2023 年 12 月 18 日 頁面資料 。
- OpenAI , 〈 DALL-E: Creating images from text 〉 參見 openai.com :
<https://openai.com/research/dall-e> , 2023 年 12 月 15 日 頁面資料 。
- OpenAI , 〈 CLIP: Connecting text and images 〉 參見 openai.com :
<https://openai.com/research/clip> , 2023 年 12 月 15 日 頁面資料 。
- Koalapon , 〈 What's happening here? (Disco Diffusion) 〉 , 參見 Reddit :
https://www.reddit.com/r/MediaSynthesis/comments/shmggh/whats_happening_here_disco_diffusion/ , 2023 年 12 月 15 日 頁面資料 。
- Yubin , 〈 All Midjourney Versions [V1-V5.2] Compared: The Evolution of Midjourney 〉 , 參見 Aituts.com : <https://aituts.com/midjourney-versions/> , 2023 年 12 月 15 日 頁面資料 。
- Rachel Metz , 〈 AI won an art contest, and artists are furious 〉 , 參見 CNN Business :
<https://edition.cnn.com/2022/09/03/tech/ai-art-fair-winner-controversy/index.html> , 2023 年 12 月 16 日 頁面資料 。
- Papageno , 〈 A zsűrit is megvezette a mesterséges intelligenciával készített műalkotás 〉 , 參見 PAPAGENO : <https://papageno.hu/intermezzo/2022/09/mi-colorado-kepzomuveszet/> , 2023 年 12 月 16 日 頁面資料 。
- 李沁儒 , 〈 吳淡如 AI 炎上事件懶人包！AI 畫作＝電腦繪圖嗎？她處理態度挨轟「理直氣壯糾正專業」 〉 , 參見風傳媒 : <https://www.storm.mg/lifestyle/4731458> , 2023 年 12 月 14 日 頁面資料 。
- 林彥銘 , 〈 吳淡如為「AI 繪圖」風波道歉！稱無意浪費社會資源 加碼自首「這件事」 〉 , 參見 newstalk : <https://newstalk.tw/news/view/2023-02-16/857845> , 2023 年 12 月 18 日 頁面資料 。
- Henley , 〈 吳淡如『AI 繪圖』事件惹議！一分鐘告訴你 AI 繪圖和電繪到底差在哪 〉 , 參見電癩少女 : <https://agirls.aotter.net/post/61828> , 2023 年 12 月 16 日 頁面資料 。

- Qualcomm ，〈Snapdragon 8 Gen 3 Mobile Platform〉，參見 qualcomm.com：
<https://www.qualcomm.com/products/mobile/snapdragon/smartphones/snapdragon-8-series-mobileplatforms/snapdragon-8-gen-3-mobile-platform>，2023 年 12 月 14 日頁面資料。
- Priyanshu Anand ，〈AI by the people, for the people - How Stability AI is democratizing AI〉，參見 Thredo AI：
<https://www.thredo.com/resources/ai-by-the-people-for-the-people-how-stability-ai-is-democratizing-ai>，2023 年 12 月 14 日頁面資料。
- KohakuBlueleaf ，〈LyCORIS - Lora beYond Conventional methods, Other Rank adaptation Implementations for Stable diffusion〉，參見 Github：
<https://github.com/KohakuBlueleaf/LyCORIS/blob/main/README.md>，2023 年 12 月 14 日頁面資料。
- Barry Collins ，〈Midjourney 5.1 Arrives - And It's Another Leap Forward For AI Art〉，參見 Forbes：
<https://www.forbes.com/sites/barrycollins/2023/05/03/midjourney-51-arrivesand-its-another-leap-forward-for-aiart/?sh=4bea0a5377cd>，2023 年 12 月 14 日頁面資料。
- Apple 研究員，〈三款我無法拒絕的 GPTs〉，參見 YouTube：
https://www.youtube.com/watch?v=ZMuD_V3xKo0，2023 年 12 月 14 日頁面資料。
- 祝思齊，〈最近很火的 AI 作畫，遊戲開發者們已經在嘗試了〉，參見觸樂：
<https://www.chuapp.com/article/288865.html>，2023 年 12 月 16 日頁面資料。
- Qbitai，〈Stable Video Diffusion 問世！全新 SVD 支援文字到影片、圖像到影片的繪製〉，參見電腦王：
<https://www.techbang.com/posts/111279-stable-video-diffusion>，2023 年 12 月 16 日頁面資料。
- 林妍濤，〈Anthropic AI 助理擷取網路歌詞，遭環球音樂等唱片業者控告侵權〉，參見 iThome：
<https://www.ithome.com.tw/news/159388>，2023 年 12 月 15 日頁面資料。
- DISCO Research Group，〈where-is-disco〉，參見 Hugging Face：
<https://huggingface.co/datasets/DISCOX/where-is-disco>，2023 年 12 月 18 日頁面資料。
- <https://www.audiosparx.com/>，2023 年 12 月 16 日頁面資料。
- 陳曉莉，〈Stability AI 發表文字生成音樂 AI 工具 Stable Audio〉，
<https://www.ithome.com.tw/news/158741>，2023 年 12 月 16 日頁面資料。

- 傅謙婷，〈AI 繪圖在全球大熱，mimic 卻在日本炎上！日本創作者為何對 AI 特別警戒？〉，參見數位時代：<https://www.bnext.com.tw/article/73439/after-the-release-of-the-ai-imitation-drawing-tool-mimic>，2023 年 12 月 15 日頁面資料。
- mimic (ミミック)，〈無題〉，參見 X：<https://twitter.com/illustmimic/status/1564081818441109505>，2023 年 12 月 15 日頁面資料。
- 觸圈，〈海外畫家集體抵制 AI 繪畫，Artstation 出現反 AI 繪畫浪潮〉，參見知乎：<https://zhuanlan.zhihu.com/p/592050105>，2023 年 12 月 15 日頁面資料。
- 暗雨 Anyu，〈星降る夜に / 暗雨 Anyu feat. 可不〉，參見 X：<https://twitter.com/Darkrain0320/status/1730301681664192561>，2023 年 12 月 15 日頁面資料。
- NTU DMCC，〈[社員作品] Caster - precession〉，參見 YouTube：<https://www.youtube.com/watch?v=PyIFwhlc2Rg>，2023 年 12 月 15 日頁面資料。
- Salad Savior，〈【2023 Dynamix Submission *Failed*】Pseudo-Church〉，參見 Youtube：<https://www.youtube.com/watch?v=G6mDoLMjN7g>，2023 年 12 月 15 日頁面資料。
- 蘿蔔@waifu_dog，〈Da TenshiIII〉，參見 pixiv：<https://www.pixiv.net/artworks/109234259>，2023 年 12 月 15 日頁面資料。
- Salad Savior，〈Profile〉，參見 X：<https://twitter.com/SaladSavior>，2023 年 12 月 15 日頁面資料。
- Cytus，〈Cytus II 5.0.3 現已推出〉，參見 FaceBook：<https://www.facebook.com/photo/?fbid=620118580148959&set=a.612600197567464>，2023 年 12 月 16 日頁面資料。
- 鄧舒澤，〈雷亞 AI 溝通師職缺引熱議，作為數位時代下的「魔法師」需要哪些關鍵技能？〉，參見未來商務：<https://fc.bnext.com.tw/articles/view/2612>，2023 年 12 月 15 日頁面資料。
- 吉川大貴，〈「とんでもなくハイクオリティー」話題の画像 AI「Novel AI」でひたすら二次元美少女と美少年を生成してみた〉，參見 ITmedia NEWS：

<https://www.itmedia.co.jp/news/articles/2210/03/news146.html>，2023 年 12 月 16 日頁面資料。

- fste，〈無題〉，參見 X：<https://twitter.com/ftrtn/status/1581107177066205186>，2023 年 12 月 18 日頁面資料。
- pixiv 公告，〈應對 AI 生成作品的相關功能現已上線〉，參見 pixiv：<https://www.pixiv.net/info.php?id=8728>，2023 年 12 月 15 日頁面資料。
- 銀髮教主，〈雨の中の銀髮少女〉，參見 pixiv：<https://www.pixiv.net/artworks/113640951>，2023 年 12 月 15 日頁面資料。
- Romain Beaumont，〈LAION-5B: A NEW ERA OF OPEN LARGE-SCALE MULTI-MODAL DATASETS〉，參見 laion.ai：<https://laion.ai/blog/laion-5b/>，2023 年 12 月 14 日頁面資料。
- Ashley Capoot，〈Authors sue OpenAI, allege their books were used to train ChatGPT without their consent〉，參見 CNBC：<https://www.cnbc.com/2023/07/05/authors-sue-openai-allege-chatgpt-was-trained-on-their-books.html>，2023 年 12 月 16 日頁面資料。
- 遊戲葡萄，〈繪師怒氣爆發，用「抵制 AI 繪圖」洗版抗議〉，參見科技新報：https://technews.tw/2022/12/18/artstation_no-to-ai-generated-images/，2023 年 12 月 16 日頁面資料。
- MONOLITH LAW OFFICE，〈ネット上の画像のクローリングは著作権法違反？機械学習の法的問題を解説〉，參見 monolith.law：<https://monolith.law/corporate/copyright-machine-learning/#%E5%B9%B3%E6%88%9030%E5%B9%B4%E8%91%97%E4%BD%9C%E6%A8%A9%E6%B3%95%E6%94%B9%E6%AD%A3%E3%81%AE%E8%83%8C%E6%99%AF>，2023 年 12 月 16 日頁面資料。
- 抖抖村，〈你的畫風來自哪個流派?動畫繪畫風格演進系列〉，參見 YouTube：<https://www.youtube.com/watch?v=SGWUvd5ncqA>，2023 年 12 月 16 日頁面資料。
- 企查查，〈北京光和界科技有限公司〉，參見 qcc.com：<https://www.qcc.com/firm/85d9963fa1358b1221fc02fa78cfdd9e.html>，2023 年 12 月 16 日頁面資料。

- 魔王醬，〈無題之回答〉，參見知乎：
<https://www.zhihu.com/question/568804205/answer/2777680843>，2023 年 12 月 16 日頁面資料。
- AICan 的 B 站號，〈AICanTalk#13 | HiiiMETA—AIGC 協助 IP 共創社區〉，參見 bilibili（0 時 9 分 56 秒）：www.bilibili.com/video/BV1Pz4y1N7Xu，2023 年 12 月 16 日頁面資料。
- 魔王醬，〈AI 煉丹黑魔法：當開源藝術家遇見 HiiiMeta！〉，參見知乎：
<https://zhuanlan.zhihu.com/p/584958306>，2023 年 12 月 16 日頁面資料。
- 日本著作權法第三十條の四，參見：<https://elaws.e-gov.go.jp/document?lawid=345AC0000000048>
- 著作權法（民國 111 年 06 月 15 日）。參見：
<https://law.moj.gov.tw/LawClass/LawAll.aspx?PCode=J0070017>
- 鯛魚，〈《尾田榮一分享 A I 作曲》反 A I 派網友氣炸了 喊出抵制航海王重製版動畫〉，參見宅宅新聞：<https://news.gamme.com.tw/1764631>
- Google Trend，〈Midjourney 2022 年 6 月 16 日至 2023 年 6 月 16 日台灣搜尋熱度〉，參見 Google Trend：<https://trends.google.com.tw/trends/explore?date=2022-06-16%202023-06-16&geo=TW&q=%2Ffg%2F1t4b6xmjk&hl=zh-TW>，2023 年 12 月 19 日頁面資料。
- Google Trend，〈Suno AI、Riffusion、MusicLM、MusicGen、Stable Audio 過去 14 個月全球 Google 相對搜尋熱度〉，參見 Google Trend：
<https://trends.google.com.tw/trends/explore?date=2022-11-18%202024-01-18&q=suno%20ai,Riffusion,MusicLM,MusicGen,Stable%20Audio>，2024 年 1 月 18 日頁面資料。
- Google Trend，〈Suno AI 過去 2 個月日本 Google 搜尋熱度〉，參見 Google Trend：
<https://trends.google.com.tw/trends/explore?date=2023-11-17%202024-01-17&geo=JP&q=suno%20ai&hl=zh-TW>，2024 年 1 月 17 日頁面資料
- Mike Russell，〈Google MusicLM: Turn Text into Mind-Blowing Music! (AI GENERATED MUSIC IS REAL)〉，參見 Youtube：
<https://www.youtube.com/watch?v=IwtXUcbs7aI&t=1369s>，2024 年 1 月 17 日夜面資料。

引用圖片來源

- 圖 1 Riffusion 舊官網截圖..... 5
DigiLog 作者群，〈將聲音視覺化後重新作曲？全新的 AI 音樂生成項目 Riffusion〉，參見 DigiLog：<https://digilog.tw/posts/1633>，2023 年 12 月 15 日頁面資料。
- 圖 2 Riffusion 新官網截圖 5
筆者於 2023 年 12 月 15 日於 <https://riffusion.com> 所截
- 圖 3 MusicLM 模型架構圖 6
Agostinelli, Andrea et al. “MusicLM: Generating Music From Text.” arXiv.org (2023): n. pag. Web.
- 圖 4 MusicGen 詞符交錯模式示意圖 7
Copet, Jade et al. “Simple and Controllable Music Generation.” arXiv.org (2023): n. pag. Web.
- 圖 5 Stable Audio 模型架構圖 7
Stability AI，〈Stable Audio: Fast Timing-Conditioned Latent Audio Diffusion〉，參見 [stability.ai](https://stability.ai/research/stable-audio-efficient-timing-latent-diffusion)：<https://stability.ai/research/stable-audio-efficient-timing-latent-diffusion>，2023 年 12 月 16 日頁面資料。
- 圖 6 Suno AI Chirp 網頁版截圖 8
筆者於 2023 年 12 月 17 日於 <https://app.suno.ai/create/> 所截
- 圖 7 能分類貓臉的神經網路之視覺化 12
Le, Quoc V. “Building High-Level Features Using Large Scale Unsupervised Learning.” 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013. 8595–8598. Web.

圖 8 經預訓練後之 StyleGAN2 生成的日本動漫風格人物頭像 12
Fathy Rashad, 〈Generating Anime Characters with StyleGAN2〉, 參見 Medium :
<https://towardsdatascience.com/generating-anime-characters-with-stylegan2-6f8ae59e237b>, 2023 年 12 月 15 日頁面資料。

圖 9 DALL-E 所生成的圖像, 文字提示詞為 an illustration of a baby daikon radish in a tutu walking a dog 13
OpenAI, 〈DALL-E: Creating images from text〉參見 openai.com :
<https://openai.com/research/dall-e>, 2023 年 12 月 15 日頁面資料。

圖 10 koalapon 於網路論壇 Reddit 分享其使用 Disco Diffusion 生成之圖像 13
Koalapon, 〈What's happening here? (Disco Diffusion)〉, 參見 Reddit :
https://www.reddit.com/r/MediaSynthesis/comments/shmggh/whats_happening_here_disco_diffusion/, 2023 年 12 月 15 日頁面資料。

圖 11 傑森·艾倫 via Midjourney 〈Théâtre d'Opéra Spatial〉 14
Papageno, 〈A zsűrit is megvezette a mesterséges intelligenciával készített műalkotás〉, 參見 PAPANENO : <https://papageno.hu/intermezzo/2022/09/mi-colorado-kepzmuveszet/>, 2023 年 12 月 16 日頁面資料。

圖 12 吳淡如 via Midjourney 〈櫻花貓少女〉 15
Henley, 〈吳淡如『AI 繪圖』事件惹議! 一分鐘告訴你 AI 繪圖和電繪到底差在哪〉, 參見電癩少女 : <https://agirls.aotter.net/post/61828>, 2023 年 12 月 16 日頁面資料。

圖 13 臺灣 Midjourney 相對搜尋熱度, 可以看到吳淡如 AI 炎上事件發生之 2023 年 2 月搜尋熱度突然竄升 15
Google Trend, 〈Midjourney 2022 年 6 月 16 日至 2023 年 6 月 16 日台灣搜尋熱度〉, 參見 Google Trend : [https://trends.google.com.tw/trends/explore?date=2022-06-16%2023-06-16&geo=TW&q=%2Ffg%2F11t4b6xmqq&hl=zh-TW](https://trends.google.com.tw/trends/explore?date=2022-06-16%202023-06-16&geo=TW&q=%2Ffg%2F11t4b6xmqq&hl=zh-TW), 2023 年 12 月 19 日頁面資料。

圖 14 文字轉音樂生成式 AI 相對搜尋熱度圖	18
Google Trend，〈Suno AI、MusicLM、Stable Audio、Riffusion、MusicGen 過去 12 個月全球 Google 搜尋熱度〉，參見 Google Trend： https://trends.google.com.tw/trends/explore?q=suno%20ai,MusicLM,Stable%20Audio,Riffusion,MusicGen ，2023 年 12 月 19 日頁面資料。	
圖 15 Suno AI 日本 Google 相對搜尋熱度圖	19
Google Trend，〈Suno AI 過去 2 個月日本 Google 搜尋熱度〉，參見 Google Trend： https://trends.google.com.tw/trends/explore?date=2023-11-17%2024-01-17&geo=JP&q=suno%20ai&hl=zh-TW ，2024 年 1 月 17 日頁面資料	
圖 16 DISCO-10M 公開音樂資料集面臨法律問題，被迫暫時下線	20
筆者於 2023 年 12 月 18 日於 https://huggingface.co/datasets/DISCOX/where-is-disco 所截。	
圖 17 SoundDraw 介面截圖	22
筆者於 2023 年 12 月 14 日於 https://sounddraw.io/create_music 所截	
圖 18 Synthesizer V Studio Pro 1.9.0 介面截圖	22
筆者自行截圖	
圖 19 mimic 服務之運作模式：使用者上傳圖片，生成類似畫風之人物圖像。	23
mimic（ミミック），〈無題〉，參見 X： https://twitter.com/illustmimic/status/1564081818441109505 ，2023 年 12 月 15 日頁面資料。	
圖 20 〈Cytus II〉中疑似使用 AI 輔助創作的畫作	25
Cytus，〈Cytus II 5.0.3 現已推出〉，參見 FaceBook： https://www.facebook.com/photo/?fbid=620118580148959&set=a.612600197567464 ， 2023 年 12 月 16 日頁面資料。	
圖 21 銀髮教主 via AI 〈雨の中の銀髮少女〉	27
銀髮教主，〈雨の中の銀髮少女〉，參見 pixiv： https://www.pixiv.net/artworks/113640951 ，2023 年 12 月 15 日頁面資料。	

圖 22 為圖 21 之局部放大，可以看到其鈕扣及拉鍊都生成的有些不自然..... 27
銀髮教主，〈雨の中の銀髮少女〉，參見 pixiv：
<https://www.pixiv.net/artworks/113640951>，2023 年 12 月 15 日頁面資料。

圖 23 HiiiMeta 畫風「基因脈絡」示意圖 32
AICan 的 B 站號，〈AICanTalk#13 | HiiiMETA – AIGC 協助 IP 共創社區〉，參見
bilibili (0 時 9 分 56 秒)：www.bilibili.com/video/BV1Pz4y1N7Xu，2023 年 12 月 16
日頁面資料。

圖 24 HiiiMeta 分潤示意圖 33
魔王醬，〈AI 煉丹黑魔法：當開源藝術家遇見 HiiiMeta！〉，參見知乎：
<https://zhuanlan.zhihu.com/p/584958306>，2023 年 12 月 16 日頁面資料。