香港司法機構
法官及司法人員和支援人員
使用生成式人工智能的指引

目的

本指引旨在為法官及司法人員和司法機構支援人員提供有關在履行各項司法及行政職務時使用生成式人工智能的一般規則和指導原則。

在法院使用科技的一般原則

2. 司法機構的首要角色在於依法執行司法工作及進行審判、維護法治並捍衞個人權利。任何人工智能（包括生成式人工智能）的使用均應遵循在法院運作中使用科技的原則。具體而言，科技的作用在於支援司法機構更有成效和效率地履行其角色和職能，同時不損害司法獨立、公正和問責的原則。科技的使用不得削弱司法職位的尊嚴和地位，或公眾對司法工作的信任與信心。使用任何人工智能之前，法院必須了解和評估人工智能的能力和局限，並處理當中涉及的潛在風險。簡而言之，法官及司法人員和支援人員使用任何人工智能都必須符合司法機構維護司法公正的首要職責。

生成式人工智能

3. 人工智能一般是指能夠執行通常需要人類智慧的任務的電腦系統。生成式人工智能是人工智能中的一種，可以根據預先訓練的數據集產生新的內容，例如文本、圖像或其他媒體。生成式人工智能聊天機械人是一種利用生成式人工智能模擬線上人類對話的電腦程式。部分其他常見的人工智能相關用語載列於本指引的*附件*。

4. 雖然生成式人工智能有潛力成為具創意和創新的強大工具，但我們在採用該科技時應注意潛在的顧慮和挑戰。特別是，不應以任何可能導致違反現行法律、規例或法院命令的方式使用生成式人工智能。

負責任使用生成式人工智能的指導原則

5. 在遵守本指引中所載的一般規則和指導原則的前提下，法官及司法人員和司法機構支援人員可在適當情況，審慎並負責任地於工作過程中使用生成式人工智能。

(A) 不得轉授司法職能

6. 法官及司法人員應確保所有司法決定繼續由他們獨立並親自作出，且在任何情況下均不得允許生成式人工智能代替其履行司法職能。換言之，法院必須確保任何生成式人工智能的使用是純粹支援和利便履行 —— 而非侵奪或干擾 —— 其司法職能。

(B) 了解生成式人工智能的局限；檢查以確保準確和問責

7. 生成式人工智能技術發展迅速，可供選擇的產品日益增多。使用者必須了解所使用的特定模型的特點和局限。舉例來說，目前許多生成式人工智能聊天機械人是以大型語言模型為基礎，並根據其接收的指令和受訓的數據，透過複雜的演算法生成新的文本（以及圖像或其他媒體）。生成的輸出結果是該模型基於資訊來源的文件和數據，而預測最有可能出現的字詞和數據組合，儘管模型的回應方式可能存在隨機因素。輸出結果的質素取決於生成式人工智能聊天機械人的訓練方式、訓練數據的可靠度以及所輸入的指令的質素。聊天機械人未必會提供來自權威數據庫的答案。使用者應當注意，即使輸入最佳的指令，其輸出結果也可能不準確、不完整、誤導或偏頗。例如，部分人工智能工具可能（以下並非詳盡無遺）——

(a) 編造虛構的案例、引稱或引文，或提述不存在的法例、文章或法律文本 —— 此風險是源於大型語言模型可以產生「幻覺」；

(b) 就法律或其應用方式提供不正確或誤導的資訊；

(c) 犯事實錯誤；及

(d) 即使資訊不正確，仍會在被問及時確認其為準確。

8. 法官及司法人員和支援人員必須注意所使用的生成式人工智能的能力和局限，並在工作中使用或依賴所獲得的任何資訊之前進行檢查和核實，以確保其準確和可靠。使用未經適當檢查和核實的生成資訊，或會造成不公，並損害公眾對司法機構的信心。

(C) 維護資訊安全；秉持保密及私隱原則

9. 為維護資訊安全，法官及司法人員和支援人員只應使用由司法機構提供的資訊科技裝置（而非可能缺乏妥善保護資訊安全措施的個人裝置）接達生成式人工智能工具。切勿連接司法機構提供的資訊科技裝置至不可信的網絡（包括 WiFi 網絡），尤其是公共場所提供的網絡。對於不明的 WiFi 網絡，不應開啟自動連接或登入。工作時使用司法機構電郵地址，以維護資訊安全。

10. 部分生成式人工智能聊天機械人會保留輸入的資訊，並用以回應其他使用者的提問。除非所使用的是封閉型生成式人工智能，否則便應假定任何輸入的資訊都可能被公開。法官及司法人員和支援人員不應在開放或公開的生成式人工智能聊天機械人中輸入任何私人、機密或敏感的資訊；同時應確保輸入的內容足夠寬泛且已隱去姓名。如聊天機械人有聊天記錄的選項，則應停用有關功能。請注意，某些人工智能平台可能會要求各種權限，使其可存取用以接達該等平台的資訊科技裝置上的資訊；此等權限要求應當一律拒絕。

11. 法官及司法人員和支援人員在使用生成式人工智能過程中處理個人資料時，應確保遵守《個人資料（私隱）條例》（第 486 章）的規定，包括該條例附表 1 所載的六項保障資料原則[1]。如使用

[1] 該六項保障資料原則，涵蓋個人資料由收集至銷毀的整個處理週期，包括：(1) 收集目的及方式；(2) 準確性及保留期間；(3) 資料的使用；(4) 資料的保安；(5) 透明度；及 (6) 查閱及改正。有關以上保障資料原則的詳情，請參閱個人資料私隱專員公署最近發表的《人工智能 (AI)：個人資料保障模範框架》附錄 A（網頁連結見下文註 3(a)）。

生成式人工智能處理司法或行政職務後，出現任何懷疑違反資訊安全或私隱的情況，相關的法官或司法人員應儘快向其法院領導報告有關事件，而相關的支援人員則應儘快向其上司報告有關事件，再由上司向有關部門主管報告。如懷疑有違反個人資料私隱的情況，應同時通知司法機構的保障資料主任（現為司法機構助理政務長（優質服務及資訊科技））。

(D) 防範侵犯版權和違反知識產權法例

12. 法官及司法人員和支援人員應避免以任何可能侵犯版權和違反知識產權法例的方式使用生成式人工智能。例如，將任何受知識產權保護的出版物上載至生成式人工智能聊天機械人以獲取摘要或分析，此舉可能會侵犯作者的版權。摘錄自原作品的輸出結果亦可能產生版權問題。使用者有責任在使用生成式人工智能時確保遵守版權及其他知識產權法例[2]。

(E) 覺察偏頗

13. 我們須知悉生成式人工智能聊天機械人是基於其受訓所用的數據集生成回應。訓練數據中的任何偏頗內容（包括文化或道德層面的偏頗內容）、地域焦點、錯誤資訊會無可避免地反映於生成的回應之中。法官及司法人員和支援人員應注意這點，並在使用或依賴生成的資訊前作出必要的更正。

(F) 承擔責任

14. 法官及司法人員和支援人員應緊記，對於以自身名義製作的任何材料，即使當中採用了從生成式人工智能取得的資訊，他們最終均須負上個人責任。

---

[2] 在香港，該等法例包括《版權條例》（第 528 章）、《防止盜用版權條例》（第 544 章）、《商標條例》（第 559 章）、《商品說明條例》（第 362 章）、《專利條例》（第 514 章）及《註冊外觀設計條例》（第 522 章）。

(G) 覺察法庭使用者使用生成式人工智能

15. 我們應意識到法庭使用者有可能在準備訴訟文件或材料時使用了生成式人工智能。雖然律師有專業責任確保向法庭呈交的任何材料（無論如何製成）必須準確恰當，但在適當情況下，法官及司法人員仍應提醒個別律師履行上述責任，以及確認他們已核實任何在生成式人工智能協助下蒐集的資料或引用的案例均為準確。

16. 至於無律師代表訴訟人，他們大多可能沒有能力核實生成式人工智能所提供的法律資訊，也可能不知道這些資訊容易出錯。如訴訟人看來可能曾使用生成式人工智能擬備陳詞或其他訴訟文件，法官及司法人員應向他查詢，並了解他如何檢查資訊的準確度。

兩項指導規則

17. 上述各項指導原則可扼要歸納為以下兩項指導規則 ——

規則 1 （關於**輸入**）	：	不得將司法職能轉授予人工智能。在生成式人工智能聊天機械人輸入資料時，注意資訊安全、保密及私隱的問題。應當意識到任何輸入的內容皆有風險成為公共領域的資訊；以及
規則 2 （關於**輸出**）	：	對人工智能聊天機械人生成的輸出結果保持警覺，尤其是事實的準確性、潛在的偏頗、侵犯知識產權等問題，並自行承擔使用風險。使用者需對使用人工智能和最終成品負責。

潛在用途

18. 生成式人工智能在以下工作中可能有用 ——

(a) **總結資訊**：雖然人工智能工具能夠總結大量文字，但使用者仍需小心確保總結內容準確，並與原文內容意思脗合；

(b) **撰寫演辭／簡報**：人工智能工具可用於籌劃演辭、擬備發言要點大綱，以及就簡報中可涵蓋的各個主題提供建議；

(c) **法律翻譯**；以及

(d) **行政工作**：人工智能工具在草擬電子郵件／備忘／書信方面雖然有用，但這些工具可以保留所輸入的任何資料（包括姓名、電郵地址等），並有可能將這些資料透露給後來的使用者，因此需要小心留意。

19. 由於生成式人工智能聊天機械人受到日期範圍、司法管轄區覆蓋和可取閱的法律材料種類的局限，使用這些聊天機械人進行資料蒐集需要格外謹慎。視乎所使用的人工智能模型和基礎數據庫的特點，如以這些聊天機械人完全取代其他方法來蒐集法律資料，它們可能並不可靠。

20. 如所使用的生成式人工智能只是基於概率，而非基於對文字間任何細微差異和語境的理解來生成文本，且無法仔細審視從數據中識別的規律，這樣就有可能得出不準確或偏頗的結論，因此不適合用作法律分析。除非有生成式人工智能模型經證實能夠保護機密、限閱及私隱的資訊，並有足夠的內置檢查和核實機制確保資料準確可靠，**否則不建議使用生成式人工智能進行法律分析**。

指引的進一步更新

21. 本文為司法機構首套發出關於利用生成式人工智能協助執行司法與行政職務的指引，此前已參考其他司法管轄區法院、政府資訊科技總監辦公室，以及個人資料私隱專員公署近期發出的指

引。[3] 我們將緊貼世界各地生成式人工智能的最新發展，以及其他司法管轄區法院公布的任何新指引，以期在有需要時更新本指引。

查詢

22. 有關本指引的查詢，請聯絡總司法行政主任（資訊科技事務處）趙金泉先生（電話：2867 2669）或高級系統經理（資訊科技事務處（技術支援））李偉文先生（電話：2886 6895）。

司法機構政務處
2024 年 7 月

---

[3] 這些指引包括（按發出日期載列，由最近期開始）：

(a) 個人資料私隱專員公署於 2024 年 6 月 11 日發布的《人工智能 (AI)：個人資料保障模範框架》(https://www.pcpd.org.hk/chinese/resources_centre/publications/files/ai_protection_framework.pdf)；

(b) 加拿大聯邦法院於 2023 年 12 月 20 日發布的《關於法院使用人工智能的過渡性原則與指引》(https://www.fct-cf.gc.ca/en/pages/law-and-practice/artificial-intelligence)；

(c) 英國司法機關法院與審裁處於 2023 年 12 月 12 日發出的《人工智能－給司法人員的指引》(https://www.judiciary.uk/wp-content/uploads/2023/12/AI-Judicial-Guidance.pdf)；

(d) 新西蘭法院於 2023 年 12 月 7 日發出的《關於法院與審裁處使用生成式人工智能的指引》(https://www.courtsofnz.govt.nz/going-to-court/practice-directions/practice-guidelines/all-benches/guidelines-for-use-of-generative-artificial-intelligence-in-courts-and-tribunals/)；

(e) 政府資訊科技總監辦公室於 2023 年 8 月發出的《人工智能道德框架指引》（第 1.3 版）(https://www.digitalpolicy.gov.hk/en/our_work/data_governance/policies_standards/ethical_ai_framework/doc/Ethical_AI_Framework.pdf) [註：由於政府資訊科技總監辦公室已於 2024 年 7 月改組為數字政策辦公室，有關指引最新版本可於上述連結查閱]；以及

(f) 最高人民法院於 2022 年 12 月 8 日發布的《關於規範和加強人工智能司法應用的意見》(https://www.court.gov.cn/fabu/xiangqing/382461.html)。

附件

部分常見的人工智能相關用語

大型語言模型：
大型語言模型是一種人工智能模型，透過利用海量文本進行訓練，學習預測句子中下一個最佳單字或單字的部分。生成式人工智能聊天機械人通常使用大型語言模型來生成對「指令」的回應，例子有 ChatGPT 和 Bing Chat。

基於轉換器的生成式預訓練模型 ("GPT")：
基於轉換器架構的大型語言模型，可生成文本。它會預先經過訓練，以預測文字中的下一個詞元來學習語言模式。在預先訓練之後，GPT 模型可以透過重複預測其預計隨後會出現的詞元來產生類似人類的文本。GPT 模型通常可經過微調，以減少幻覺或有害行為，或以會話格式編排輸出結果。

詞元：
在自然語言處理中，詞元是人工智能處理的文字單位，通常代表一個單字或單字的部分。然而，詞元並沒有固定的字元或單字長度。相反，詞元可以根據語言和內容的複雜性而有所不同。

機器學習：
人工智能的一個分支，利用數據和演算法模仿人類學習的方式，並會逐漸提高準確度。演算法透過統計學方法訓練，以進行分類或預測，以及在數據開採項目中發掘關鍵的見解。

深度學習：
人工智能中一種模仿人腦的功能，學習人腦如何組織和處理資訊作出決策。這一種機器學習可以在監督下從零散的數據中學習，而不需依賴只能執行單一特定任務的演算法。

數據開採：
整理大型數據集，以識別可改善模型或解決問題的規律的過程。

自然語言處理：
使電腦理解人類的口頭和書面語言的一種人工智能。自然語言處理使裝置擁有辨認文字及說話等功能。

技術輔助審閱：
作為披露文件過程的一部分，用以識別潛在相關文件的人工智能工具。在技術輔助審閱中，機器學習系統根據由律師人工識別相關文件所建立的數據進行訓練；然後該系統利用學習到的準則，從龐大的披露數據集中識別其他類似的文件。

指令：
輸入生成式人工智能聊天機械人的簡短指示，以取得所需的答案／輸出結果。

檢索增強生成：
檢索增強生成是用於人工智能及自然語言處理的技術，旨在透過整合來自不同來源的外部資訊，提升生成文本的質素。

幻覺：
人工智能系統將虛假資訊如同事實般呈現，即謂人工智能幻覺。

OpenAI：
OpenAI 是一家美國人工智能公司。該公司從事人工智能的研究，並在過去十年開發了數個人工智能模型及服務，包括 GPT-3、ChatGPT 及 Dall-E。