

香港司法机构 法官及司法人员和支援人员 使用生成式人工智能的指引

目的

本指引旨在为法官及司法人员和司法机构支援人员提供有关在履行各项司法及行政职务时使用生成式人工智能的一般规则和指导原则。

在法院使用科技的一般原则

2. 司法机构的首要角色在于依法执行司法工作及进行审判、维护法治并捍卫个人权利。任何人工智能（包括生成式人工智能）的使用均应遵循在法院运作中使用科技的原则。具体而言，科技的作用在于支援司法机构更有成效和效率地履行其角色和职能，同时不损害司法独立、公正和问责的原则。科技的使用不得削弱司法职位的尊严和地位，或公众对司法工作的信任与信心。使用任何人工智能之前，法院必须了解和评估人工智能的能力和局限，并处理当中涉及的潜在风险。简而言之，法官及司法人员和支援人员使用任何人工智能都必须符合司法机构维护司法公正的首要职责。

生成式人工智能

3. 人工智能一般是指能够执行通常需要人类智慧的任务的电脑系统。生成式人工智能是人工智能中的一种，可以根据预先训练的数据集产生新的内容，例如文本、图像或其他媒体。生成式人工智能聊天机械人是一种利用生成式人工智能模拟线上人类对话的电脑程式。部分其他常见的人工智能相关用语载列于本指引的**附件**。

4. 虽然生成式人工智能有潜力成为具创意和创新的强大工具，但我们在采用该科技时应注意潜在的顾虑和挑战。特别是，不应以任何可能导致违反现行法律、规例或法院命令的方式使用生成式人工智能。

负责任使用生成式人工智能的指导原则

5. 在遵守本指引中所载的一般规则和指导原则的前提下，法官及司法人员和司法机构支援人员可在适当情况，审慎并负责任地于工作过程中使用生成式人工智能。

(A) 不得转授司法职能

6. 法官及司法人员应确保所有司法决定继续由他们独立并亲自作出，且在任何情况下均不得允许生成式人工智能代替其履行司法职能。换言之，法院必须确保任何生成式人工智能的使用是纯粹支援和利便履行——而非侵夺或干扰——其司法职能。

(B) 了解生成式人工智能的局限；检查以确保准确和问责

7. 生成式人工智能技术发展迅速，可供选择的产品日益增多。使用者必须了解所使用的特定模型的特点和局限。举例来说，目前许多生成式人工智能聊天机械人是以大型语言模型为基础，并根据其接收的指令和受训的数据，透过复杂的演算法生成新的文本（以及图像或其他媒体）。生成的输出结果是该模型基于资讯来源的文件和数据，而预测最有可能出现的字词和数据组合，尽管模型的回应方式可能存在随机因素。输出结果的质素取决于生成式人工智能聊天机械人的训练方式、训练数据的可靠度以及所输入的指令的质素。聊天机械人未必会提供来自权威数据库的答案。使用者应当注意，即使输入最佳的指令，其输出结果也可能不准确、不完整、误导或偏颇。例如，部分人工智能工具可能（以下并非详尽无遗）——

- (a) 编造虚构的案例、引称或引文，或提述不存在的法例、文章或法律文本——此风险是源于大型语言模型可以产生「幻觉」；
- (b) 就法律或其应用方式提供不正确或误导的资讯；

(c) 犯事实错误；及

(d) 即使资讯不正确，仍会在被问及时确认其为准确。

8. 法官及司法人员和支援人员必须注意所使用的生成式人工智能的能力和局限，并在工作中使用或依赖所获得的任何资讯之前进行检查和核实，以确保其准确和可靠。使用未经适当检查和核实的生成资讯，或会造成不公，并损害公众对司法机构的信心。

(C) 维护资讯安全；秉持保密及私隐原则

9. 为维护资讯安全，法官及司法人员和支援人员只应使用由司法机构提供的资讯科技装置（而非可能缺乏妥善保护资讯安全措施的个人装置）接达生成式人工智能工具。切勿连接司法机构提供的资讯科技装置至不可信的网络（包括 WiFi 网络），尤其是公共场所提供的网络。对于不明的 WiFi 网络，不应开启自动连接或登入。工作时使用司法机构电邮地址，以维护资讯安全。

10. 部分生成式人工智能聊天机械人会保留输入的资讯，并用以回应其他使用者的提问。除非所使用的是封闭型生成式人工智能，否则便应假定任何输入的资讯都可能被公开。法官及司法人员和支援人员不应在开放或公开的生成式人工智能聊天机械人中输入任何私人、机密或敏感的资讯；同时应确保输入的内容足够宽泛且已隐去姓名。如聊天机械人有聊天记录的选项，则应停用有关功能。请注意，某些人工智能平台可能会要求各种权限，使其可存取用以接达该等平台的资讯科技装置上的资讯；此等权限要求应当一律拒绝。

11. 法官及司法人员和支援人员在使用生成式人工智能过程中处理个人资料时，应确保遵守《个人资料（私隐）条例》（第 486 章）的规定，包括该条例附表 1 所载的六项保障资料原则¹。如使用

¹ 该六项保障资料原则，涵盖个人资料由收集至销毁的整个处理周期，包括：(1) 收集目的及方式；(2) 准确性及保留期间；(3) 资料的使用；(4) 资料的保安；(5) 透明度；及 (6) 查阅及改正。有关以上保障资料原则的详情，请参阅个人资料私隐专员公署最近发表的《人工智能 (AI)：个人资料保障模范框架》附录 A（网页连结见下文注 3(a)）。

生成式人工智能处理司法或行政职务后，出现任何怀疑违反资讯安全或私隐的情况，相关的法官或司法人员应尽快向其法院领导报告有关事件，而相关的支援人员则应尽快向其上司报告有关事件，再由上司向有关部门主管报告。如怀疑有违反个人资料私隐的情况，应同时通知司法机构的保障资料主任（现为司法机构助理政务长（优质服务及资讯科技））。

(D) 防范侵犯版权和违反知识产权法例

12. 法官及司法人员和支援人员应避免以任何可能侵犯版权和违反知识产权法例的方式使用生成式人工智能。例如，将任何受知识产权保护的出版物上载至生成式人工智能聊天机械人以获取摘要或分析，此举可能会侵犯作者的版权。摘录自原作品的输出结果亦可能产生版权问题。使用者有责任在使用生成式人工智能时确保遵守版权及其他知识产权法例²。

(E) 觉察偏颇

13. 我们须知悉生成式人工智能聊天机械人是基于其受训所用的数据集生成回应。训练数据中的任何偏颇内容（包括文化或道德层面的偏颇内容）、地域焦点、错误资讯会无可避免地反映于生成的回应之中。法官及司法人员和支援人员应注意这点，并在使用或依赖生成的资讯前作出必要的更正。

(F) 承担责任

14. 法官及司法人员和支援人员应紧记，对于以自身名义制作的任何材料，即使当中采用了从生成式人工智能取得的资讯，他们最终均须负上个人责任。

² 在香港，该等法例包括《版权条例》（第 528 章）、《防止盗用版权条例》（第 544 章）、《商标条例》（第 559 章）、《商品说明条例》（第 362 章）、《专利条例》（第 514 章）及《注册外观设计条例》（第 522 章）。

(G) 觉察法庭使用者使用生成式人工智能

15. 我们应意识到法庭使用者有可能在准备诉讼文件或材料时使用了生成式人工智能。虽然律师有专业责任确保向法庭呈交的任何材料（无论如何制成）必须准确恰当，但在适当情况下，法官及司法人员仍应提醒个别律师履行上述责任，以及确认他们已核实任何在生成式人工智能协助下搜集的资料或引用的案例均为准确。

16. 至于无律师代表诉讼人，他们大多可能没有能力核实生成式人工智能所提供的法律资讯，也可能不知道这些资讯容易出错。如诉讼人看来可能曾使用生成式人工智能拟备陈词或其他诉讼文件，法官及司法人员应向他查询，并了解他如何检查资讯的准确度。

两项指导规则

17. 上述各项指导原则可扼要归纳为以下两项指导规则——

规则 1 : 不得将司法职能转授予人工智能。在生成式人工智能聊天机械人输入资料时，
(关于输入) 注意资讯安全、保密及私隐的问题。应当意识到任何输入的内容皆有风险成为公共领域的资讯；以及

规则 2 : 对人工智能聊天机械人生成的输出结果
(关于输出) 保持警觉，尤其是事实的准确性、潜在的偏颇、侵犯知识产权等问题，并自行承担使用风险。使用者需对使用人工智能和最终成品负责。

潜在用途

18. 生成式人工智能在以下工作中可能有用——

- (a) **总结资讯：**虽然人工智能工具能够总结大量文字，但使用者仍需小心确保总结内容准确，并与原文内容意思贴合；
- (b) **撰写演辞 / 简报：**人工智能工具可用于筹划演辞、拟备发言要点大纲，以及就简报中可涵盖的各个主题提供建议；
- (c) **法律翻译：**以及
- (d) **行政工作：**人工智能工具在草拟电子邮件 / 备忘 / 书信方面虽然有用，但这些工具可以保留所输入的任何资料（包括姓名、电邮地址等），并有可能将这些资料透露给后来的使用者，因此需要小心留意。

19. 由于生成式人工智能聊天机械人受到日期范围、司法管辖区覆盖和可取阅的法律材料种类的局限，使用这些聊天机械人进行资料搜集需要格外谨慎。视乎所使用的人工智能模型和基础数据库的特点，如以这些聊天机械人完全取代其他方法来搜集法律资料，它们可能并不可靠。

20. 如所使用的生成式人工智能只是基于概率，而非基于对文字间任何细微差异和语境的理解来生成文本，且无法仔细审视从数据中识别的规律，这样就有可能得出不准确或偏颇的结论，因此不适合用作法律分析。除非有生成式人工智能模型经证实能够保护机密、限阅及私隐的资讯，并有足够的内置检查和核实机制确保资料准确可靠，否则不建议使用生成式人工智能进行法律分析。

指引的进一步更新

21. 本文为司法机构首套发出关于利用生成式人工智能协助执行司法与行政职务的指引，此前已参考其他司法管辖区法院、政府资讯科技总监办公室，以及个人资料私隐专员公署近期发出的指

引。³ 我们将紧贴世界各地生成式人工智能的最新发展，以及其他司法管辖区法院公布的任何新指引，以期在有需要时更新本指引。

查询

22. 有关本指引的查询，请联络总司法行政主任（资讯科技事务处）赵金泉先生（电话：2867 2669）或高级系统经理（资讯科技事务处（技术支援））李伟文先生（电话：2886 6895）。

司法机构政务处
2024 年 7 月

³ 这些指引包括（按发出日期载列，由最近期开始）：

- (a) 个人资料私隐专员公署于 2024 年 6 月 11 日发布的《人工智能 (AI)：个人资料保障模范框架》(https://www.pcpd.org.hk/chinese/resources_centre/publications/files/ai_protection_framework.pdf)；
- (b) 加拿大联邦法院于 2023 年 12 月 20 日发布的《关于法院使用人工智能的过渡性原则与指引》(<https://www.fct-cf.gc.ca/en/pages/law-and-practice/artificial-intelligence>)；
- (c) 英国司法机关法院与审裁处于 2023 年 12 月 12 日发出的《人工智能 – 给司法人员的指引》(<https://www.judiciary.uk/wp-content/uploads/2023/12/AI-Judicial-Guidance.pdf>)；
- (d) 新西兰法院于 2023 年 12 月 7 日发出的《关于法院与审裁处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) 政府资讯科技总监办公室于 2023 年 8 月发出的《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) [注：由于政府资讯科技总监办公室已于 2024 年 7 月改组为数字政策办公室，有关指引最新版本可于上述连结查阅]；以及
- (f) 最高人民法院于 2022 年 12 月 8 日发布的《关于规范和加强人工智能司法应用的意见》(<https://www.court.gov.cn/fabu/xiangqing/382461.html>)。

部分常见的人工智能相关用语

大型语言模型：

大型语言模型是一种人工智能模型，透过利用海量文本进行训练，学习预测句子中下一个最佳单字或单字的部分。生成式人工智能聊天机械人通常使用大型语言模型来生成对「指令」的回应，例子有 ChatGPT 和 Bing Chat。

基于转换器的生成式预训练模型 (“GPT”)：

基于转换器架构的大型语言模型，可生成文本。它会预先经过训练，以预测文字中的下一个词元来学习语言模式。在预先训练之后，GPT 模型可以透过重复预测其预计随后会出现的词元来产生类似人类的文本。GPT 模型通常可经过微调，以减少幻觉或有害行为，或以会话格式编排输出结果。

词元：

在自然语言处理中，词元是人工智能处理的文字单位，通常代表一个单字或单字的部分。然而，词元并没有固定的字元或单字长度。相反，词元可以根据语言和内容的复杂性而有所不同。

机器学习：

人工智能的一个分支，利用数据和演算法模仿人类学习的方式，并会逐渐提高准确度。演算法透过统计学方法训练，以进行分类或预测，以及在数据开采项目中发掘关键的见解。

深度学习：

人工智能中一种模仿人脑的功能，学习人脑如何组织和处理资讯作出决策。这一种机器学习可以在监督下从零散的数据中学习，而不需依赖只能执行单一特定任务的演算法。

数据开采：

整理大型数据集，以识别可改善模型或解决问题的规律的规律的过程。

自然语言处理：

使电脑理解人类的口头和书面语言的一种人工智能。自然语言处理使装置拥有辨认文字及说话等功能。

技术辅助审阅：

作为披露文件过程的一部分，用以识别潜在相关文件的人工智能工具。在技术辅助审阅中，机器学习系统根据由律师人工识别相关文件所建立的数据进行训练；然后该系统利用学习到的准则，从庞大的披露数据集中识别其他类似的文件。

指令：

输入生成式人工智能聊天机械人的简短指示，以取得所需的答案 / 输出结果。

检索增强生成：

检索增强生成是用于人工智能及自然语言处理的技术，旨在透过整合来自不同来源的外部资讯，提升生成文本的质素。

幻觉：

人工智能系统将虚假资讯如同事实般呈现，即谓人工智能幻觉。

OpenAI：

OpenAI 是一家美国人工智能公司。该公司从事人工智能的研究，并在过去十年开发了数个人工智能模型及服务，包括 GPT-3、ChatGPT 及 Dall-E。