《英国司法人员使用人工智能指南》发布及摘要

作者：袁杜娟

2025 年 4 月 15 日，英国首席大法官卡尔男爵夫人等司法界政要联合发布《英国司法人员使用人工智能指南》修订版。作为 2023 年 12 月首版指南的首次修订，此次更新主要源于近一年多来人工智能技术与应用的迅猛发展，以及英国本国法院和欧洲、中国等地法院在人工智能应用领域积累的新实践经验。修订版在保留原版框架的基础上新增了部分内容，现将全文编译呈现，供各位研究参考。

1.简介

本次修订版指南旨在为司法人员使用人工智能（AI）提供协助，它修改并取代了 2023 年 12 月发布的指南文件。

该指南说明了与使用人工智能相关的主要风险和问题，以及一些降低这些风险的建议，该指南还包含了人工智能潜在应用场景的示例。

司法人员或代表司法机构的组织使用人工智能时，必须符合司法机关维护司法行政完整性的首要义务。

本指南还介绍了一款个人人工智能工具 —— 微软的 “Copilot 聊天”，目前司法人员可通过电子司法系统在其设备上使用该工具。

本指南适用于由首席大法官和法庭高级庭长负责管辖的所有司法人员，以及他们的书记员、司法助理、法律顾问、法律官员和其他辅助人员。本指南在线发布，旨在提升透明度、公开性和公信力。

2.常用术语

Algorithm（算法）：一组用于执行任务（如计算和数据分析）的指令，通常借助计算机或其他智能设备实现。

Alignment（对齐）：人工智能系统与组织目标或道德伦理保持一致的任何过程，例

如与 “去偏见” 相关的原则。

Artificial Intelligence (AI)（人工智能）：能够执行通常需要人类智能才能完成的任务的计算机系统。

AI Agent（人工智能代理）：一种使用人工智能的软件程序，它能感知周围环境、处理信息，并根据输入的信息采取行动从而实现目标。

AI Prompt（人工智能提示词）：向人工智能系统发出的输入或指令，旨在生成特定的回应或结果。提示词通常以文本形式呈现，如今许多聊天机器人也接受语音提示。

Generative AI（生成式人工智能）：一种能生成文本、图像、声音和计算机代码等新内容的人工智能形式。有些生成式人工智能工具还可用于执行操作。

Generative AI Chatbot（生成式人工智能聊天机器人）：一种利用生成式人工智能模拟在线人类对话的计算机程序。目前公开可用的例子包括 ChatGPT、谷歌 Bard 和 Meta AI等。

Hallucination（幻觉）：人工智能模型生成的错误或误导性结果。这些错误可能由多种因素导致，包括训练数据不足、模型的统计特性、模型做出的错误假设，或用于训练模型的数据中存在的偏见等。

Large Language Model (LLM)（大型语言模型）：大型语言模型是一种人工智能模型，通过海量文本上接受训练，该模型学会预测句子中下一个最佳词语或词语片段。ChatGPT 和Bing Chat使用的是 OpenAI 的大型语言模型。

Machine Learning (ML)（机器学习）：人工智能的一个分支，它利用数据和算法模仿人类的学习方式，逐步提高准确性。通过统计方法，算法被训练进行分类或预测，并在数据挖掘项目中发现关键见解。

Natural Language Processing（自然语言处理）：对计算机系统进行编程，使其能够理解和生成人类的语音与文本。使用算法探寻句子和段落构建中的语言模式，以及词语、语境和结构如何共同作用产生意义。其应用包括语音转文本转换器、文本摘要在线工具、聊天机器人、语音识别和翻译等。

Responsible AI（负责任的人工智能）：该人工智能在设计、开发、部署时遵循特定

的价值观，例如可信、符合伦理、透明、可解释、公平、稳健以及维护隐私权。

Technology Assisted Review (TAR)（技术辅助审查）：在披露过程中用于识别潜在相关文件的人工智能工具。在技术辅助审查中，机器学习系统先通过律师手动识别相关文件所创建的数据进行训练，然后该工具利用习得的标准从海量的披露数据集中识别其他类似文件。

3.法院和法庭中负责任地使用人工智能的指南

3.1了解人工智能及其应用

在使用任何人工智能工具前，需确保对其功能和潜在局限性有基本了解。一些关键局限性包括：

1)公共人工智能聊天机器人的答案并非来自权威数据库。它们基于接收的提示词和训练数据，通过算法生成新文本。这意味着人工智能聊天机器人的输出是模型预测的最可能的词语组合（基于其作为源信息的文档和数据），但未必是最准确的答案。

2)与互联网上的其他信息类似，人工智能工具或许能帮助找到你已知正确但手边暂无的资料，但不适用于研究无法验证的新信息。它们更适合用于获取确认，而非直接提供准确事实。

3)输出内容的质量取决于与人工智能工具的交互方式（包括输入的提示词性质）及底层数据集质量，可能包含错误信息（无论是否故意）、选择性数据或过时数据。即便提示词质量极高，输出信息仍可能不准确、不完整、具有误导性或存在偏见。

4)目前可用的大型语言模型似乎是基于互联网公开内容训练的。它们对 “法律” 的理解往往严重倾向于美国法律，尽管部分模型声称能区分英美法律差异。

3.2维护私密性和隐私

不得向公共人工智能聊天机器人输入任何非公开信息，也不得输入私人或机密信息。输入公共人工智能聊天机器人的任何内容，都应被视为已向全球公开。

当前公开可用的人工智能聊天机器人会记录所有提问及输入信息，这些信息可能被

用于回应其他用户的查询。因此，你输入的任何内容都可能被公开。

若公共人工智能聊天机器人提供关闭聊天历史的选项（目前 ChatGPT 和谷歌 Bard 支持此功能，其他部分聊天机器人不支持），应予以关闭。这能防止你的数据被用于训练模型，且 30 天后对话将被永久删除。但即便关闭历史记录，仍需假定输入的数据已被披露。

注意，部分人工智能平台（尤其是手机应用）可能请求各种权限以访问设备上的信息，此时应拒绝所有此类权限。

若不慎泄露机密或私人信息，应立即联系你的主管法官和司法办公室。若泄露信息包含个人数据，需将其作为数据事件上报。向司法办公室报告数据事件的详细方式可参考以下链接：司法机构数据泄露通知表

你应当确信，所有人工智能都可能公开你输入其中的任何内容。

3.3确保可靠性和准确性

在使用或依赖人工智能工具提供的信息前，必须先核实其准确性。

人工智能工具提供的信息可能不准确、不完整、具有误导性或过时，即便声称反映英国法律，实际也可能并非如此。

人工智能工具可能：编造虚构的案例、引文或引语，或提及不存在的法规、文章或法律文本；提供关于法律内容或适用方式的错误或误导性信息；存在事实错误

3.4警惕偏见

基于大型语言模型的人工智能工具根据训练数据集生成回应。其输出信息不可避免地反映训练数据中的错误和偏见，这些cup五和偏见可能通过 “对齐策略” 有所缓解。

需始终意识到这种可能性，并主动纠正偏见。英国司法学院编写的《平等待遇法官手册》（Equal Treatment Bench Book）可能对此提供帮助。

3.5保障安全

使用人工智能时应遵循维护个人及法院 安全的实践做法，具体如下：

1)使用工作设备（而非个人设备）访问人工智能工具；

2)使用工作邮箱地址；

3)若订阅了付费人工智能平台，优先使用付费版本（付费订阅通常比免费版本更安全）。但需注意，部分第三方公司从其他平台授权使用人工智能技术，其信息使用的可靠性较低，应避免使用。

若存在潜在安全漏洞，处理方式参见上文第 2 点。

3.6承担责任

司法人员对以其名义生成的材料负有个人责任。

法官通常无需说明作出判决前的研究或准备工作，只要遵循本指南，生成式人工智能可以作为潜在的辅助工具。

若书记员、司法助理、法律官员、法律顾问或其他工作人员在为你工作时使用人工智能工具，你应与他们沟通，确保其规范使用并采取风险缓解措施。若使用 Dom 1 笔记本电脑，还需获得 HMCTS 服务经理的批准。

3.7注意诉讼参与人可能使用人工智能工具

法律专业人士使用某些人工智能工具已有相当长的时间且未出现问题，例如，技术辅助审查已成为电子披露流程的常规手段。除法律领域外，人工智能的许多应用已普及，如搜索引擎的自动补全、社交媒体的内容筛选、图像识别和预测文本等。

所有代理律师应对提交给法院的材料负责，并负有确保其准确性和适当性的职业义务。只要负责任地使用人工智能，法律代表无需特别说明其使用情况，但这取决于具体情况，

不过，在法律行业尚未完全熟悉这些新技术的阶段，有时需要提醒律师履行相关义务，并确认他们是否已独立核实过借助人工智能聊天机器人生成的研究内容或案例引述的准确性。

无律师代理的诉讼当事人现在开始使用人工智能聊天机器人。这可能是部分当事人获取建议或帮助的唯一途径。这些当事人往往缺乏独立验证人工智能聊天机器人提供的法律信息的能力，且可能未意识到其存在错误风险。若发现诉讼文书可能由人

工智能聊天机器人生成，应询问相关情况、核实是否已进行准确性检查（若有），并告知当事人需对提交给法院 的内容负责。下文将举例说明可能由人工智能生成法律文本。

人工智能工具现已被用于生成虚假材料，包括文本、图像和视频。法院向来需要处理不同复杂程度的伪造及伪造指控，法官应意识到这一新可能性及深度伪造技术带来的潜在挑战。

4.Copilot Chat——司法人员设备上的安全人工智能工具

一款私人人工智能工具 —— 微软的 “Copilot Chat”，目前已通过电子司法系统在司法人员的设备上可用。本指南中的所有内容均适用于 Copilot Chat的使用。

该工具可通过 Edge 浏览器或微软 365 Copilot 应用程序访问。该工具提供企业级数据保护，并在微软 365 的隐私与安全框架内运行。当登录电子司法账户后，你向 “Copilot Chat” 提交的数据是安全的，不会被公开。

关于 Copilot Chat的 IT 支持指南，可在司法学院学习网站上查阅。

5.示例：生成式人工智能在法院和法庭中的潜在用途与风险

5.1具有潜在用途的任务

1)人工智能工具能够总结大量文本。但与任何总结内容一样，需注意确保其准确性。

2)人工智能工具可用于撰写演示文稿，例如提供值得涵盖的主题建议。

3)人工智能可执行行政任务，包括撰写、总结和优先处理电子邮件，转录和总结会议内容，以及起草备忘录。

5.2不建议的任务

1)法律研究：人工智能工具不适用于研究无法独立验证的新信息。它们仅在帮助回忆已知正确的资料方面可能有一定用处。

2)法律分析：当前的公共人工智能聊天机器人无法生成有说服力的分析或推理。

5.3可能表明内容由人工智能生成的迹象

1)提及不熟悉的案例，或引用不常见的案例（有时是美国案例）；

2)就同一法律问题引用不同体系的判例法；

3)提交的文书与你对该领域法律的一般理解不一致；

4)提交的文书使用美式拼写或引用海外案例；

5)提交的内容（至少表面上）看似极具说服力、文笔流畅，但仔细检查后会发现明显的实质性错误。

上海办公室
+86 021-64179501
linkking@linkking.cn
上海徐汇区宜山路425号光启城大厦1609室

洛杉矶办事处
+86 021-64179501
la@linkking.cn
美国加利福尼亚州丹维尔市安瑟里姆路4号

东京办事处
+86 021-64179501
tyo@linkking.cn
日本〒102-0093东京都千代田区平河町一丁目6番11号エクシール平河町302号