

“机器人法官”的禁忌与展望

郑 戈

由机器主导审判不合法

在搜索引擎上输入“机器人法官”查询，得到的应用实例多来自中国，也有爱沙尼亚等少数其他国家。这是对我国“智慧法院”的一种误解。人工智能技术在我国只是辅助法官办案，丝毫没有取代法官或不受法官控制而自动化地做出司法裁判的迹象。

尽管如此，“机器人法官”作为一个不严谨的炒作概念却屡屡被自媒体使用，比如，阿里巴巴在 2019 年“云栖大会”上发布的 AI 法官助理“小智”就经常被宣传为“机器人法官”。阿里达摩院的一位算法专家也构想了一种未来：“简单案件由机器主导审判，复杂案件由机器智能辅助法官审理”。这种构想只是这位技术专家的一厢情愿，因为审判权不是一个技术问题，而是宪法和法律问题。即使人工智能已经具备自主“审判”的技术能力，在不修改宪法、人民法院组织法和三大诉讼法的情况下，由机器来主导审判都是不合法的。

而且，在我国并没有出现修改相关法律为人工智能自动化司法决策打开大门的导向。恰恰相反，最高人民法院 2022 年发布的《关于规范和加强人工智能司法应用的意见》提出的五项基本原则中包含“辅助审判原则”，明确指出要“坚持对审判工作的辅助性定位和用户自主决策权，无论技术发展到何种水平，人工智能都不得代替法官裁判，人工智能辅助结果仅可作为审判工作或审判监督管理的参考，确保司法裁判始终由审判人员作出，裁判职权始终由审判组织行使，司法责任最终由裁判者承担。各类用户有权选择是否利用司法人工智能提供的辅助，有权随时退出与人工智能产品和服务的交互。”由此可见，“机器人法官”在我国只是部分自媒体和数字科技企业采用的修辞或营销策略，与人工智能技术在我国法院系统的实际应用情况完全不符。

“机器人法官”为何是一种禁忌

那么，为什么我国和世界上其他国家一样拒绝将哪怕是简单案件的审判交给机器去进行自动化处理呢？为什么机器人法官是一种禁忌呢？原因大致有这么几点：

首先，人们认为只有人才能设身处地地理解他人，具有同情心和同理心，因此才能够以公平的方式将情、理、法适用于具体社会事实情景中的案件。机器不具备实现“正义”的能力，因为正义取决于事实与规范之间的人性链接，而不是在海量数据中找出事实与规范关联的统计学规律。

其次，这种统计学规律可以呈现以前的人类决策将规范适用于事实情境的模式，因此它是以数据化的人类决策为前提的。如果人类决策逐渐被机器的自动化决策所取代，那么机器学习就连训练数据都没有了。如果机器最终学习的是自己先前生成的数据，那么“学习”的过程就变成了毫无增益的反刍，最终归于简单的重复，而这种自我复制不仅无法与人类价值对齐，而且丧失了一切“智能”。

□ 审判权不是一个技术问题，而是宪法和法律问题。即使人工智能已经具备自主“审判”的技术能力，在不修改宪法、人民法院组织法和三大诉讼法的情况下，由机器来主导审判都是不合法的。

□ 生成式人工智能如果要用于法律工作，在设计上就需要突出人作为最后把关者和决策者的角色，使算法无法形成从数据采集、数据分析、自动化决策和最终文本生成的闭环。

□ 对于一些简单案件，司法决策可以变得标准化、算法化甚至自动化。因此可以节省大量的司法资源，以便法官可以对少数疑难案件和复杂案件进行认真仔细地斟酌，做出体现人类价值判断的决策。

第三，恰恰因为目前主流的人工智能编程模式（即深度学习）可以在海量人类行为数据中发现人类无法发现的隐含范式和规律，它可能一般化、固化和强化人类社会中已有的偏见、歧视和认知偏差。例如，已经有大量的经验研究表明，美国各州法院在刑事案件量刑过程中广泛采用的再犯可能性评估系统（COMPAS）在评估黑人被告的再犯可能性时普遍存在错误肯定（false positives）判断，而在评估白人被告的再犯可能性时则存在大量的错误否定（false negatives）。

第四，法律解释是一个复杂的过程，不是简单的三段论，它包含论辩过程中形成的法律共同体共识。机器人法官的裁判消解了论辩和共识形成机制，以冷冰冰的数据分析和统计学规律发现取代了论证和说服，这将从根本上削弱司法过程本身的正当性。

最后，但最重要的是，司法判断不是一种单方面的主张，而是在对立主张之间公允地对其事实基础和法律依据做出的中立评估和裁断。如果当事人和律师已经越来越多地使用人工智能工具，尤其是大型语言模型等生成式人工智能工具，来服务于自己的法律主张，那么作为事实和法律的最终把关者的法官就更不能也是一部机器了。司法过程要处理各种各样的文本和音视频材料，按其性质，这些材料大体分为两类：一类是解决事实问题的，也就是各种证据；另一类是解决规范问题的，即援引的法律规则和案例。前者是处理案件的事实基础，后者是处理案件的规范基础。无论是帮助裁判者判断事实问题的材料，还是裁判者借以解释和适用法律的材料，真实性都是首要的要求。虚构的法条和判例是不能被容忍的，更不能作为裁判依据。正因如此，生成式人工智能如果要用于法律工作，在设计上就需要突出人作为最后把关者和决策者的角色，使算法无法形成从数据采集、数据分析、自动化决策和最终文本生成的闭环。

警惕 ChatGPT 虚构法条和判例

最近，一则律师援引 ChatGPT 编造的假判例来支持己方主张的事件在美国乃至全球法律界引发了广泛的关注。一位名叫罗伯特·玛塔的乘客声称自己 2019 年乘坐飞机时膝盖被机上的送餐推车撞伤，于是在纽约曼哈顿的纽约南区联邦地区法院起诉航空公司。航空公司以诉讼时效已过为由请求法院

驳回起诉。玛塔的律师之一史蒂文·施瓦茨向法院提交了一份十页的书面辩论意见书，反驳了对方的主张。意见书中援引了六个先例，包括瓦吉斯诉中国南方航空公司，这些先例都表明法院在此类案件中不会因为诉讼时效已过而拒绝受理案件。但对对方律师向法院指出这些先例都是假的。于是，此案主审法官凯文·卡斯特尔发出了一份法院令，要求施瓦茨提供理由解释自己为什么不应当受到处罚，其中写道：

本院面临一个前所未有的情况。原告律师提交的反对驳回动议的意见书中充斥着对不存在案件的引用。当对方律师提请本院注意这一情况时，本院发布命令，要求原告律师提供一份宣誓誓言书，附上他提交的材料中引用的案例副本，他已经照做了。他提交的材料中有六个显然是伪造的司法判决，包括伪造的引文和伪造的案卷号。本院特此发布命令，要求原告律师解释为何自己不应受到制裁。

施瓦茨在随后提交的解释中说自己是利用 ChatGPT 找到的这些案例。他写道：

当我在此案中使用 ChatGPT 时，我了解到它本质上就像一个高度复杂的搜索引擎，用户可以在其中输入搜索查询，而 ChatGPT 将根据公开可用的信息以自然语言提供答案。我现在意识到我对 ChatGPT 工作原理的理解是错误的。如果我了解 ChatGPT 是什么或它的实际工作原理，我绝不会用它来进行法律研究。

目前，法官尚未决定是否对施瓦茨进行制裁。按照美国律师协会的《职业行为模范准则》第 3.3 (a) 条款，律师不得故意：（1）向法庭作出虚假的事实或法律陈述，或者未能更正自己先前向法庭作出的对重要事实或法律的虚假陈述；（2）不向法庭披露律师已知的直接不利于委托人立场且对方律师未披露的本公司管辖区内的权威法律渊源。显然，施瓦茨的行为客观上违反了这一规则，但他是否是故意（knowingly）这么做仍有待法官去判断。

ChatGPT 并非搜索引擎

的确，施瓦茨对 ChatGPT 这种新工具的理解是完全错误的，它不是一种搜索引擎，而

是一种基于对人类日常语言样本的学习而生成新的语言文字内容的工具。

首先，它不具备实时搜索功能，它的训练数据中不包含 2021 年之后出现的新内容。比如，本文作者曾经问它：“《中华人民共和国民法典》第 1307 条的内容是什么？”这可以说是一个非常简单的客观题了。我们知道《民法典》总共就 1260 条，第 1307 条是根本不存在的。但它却言之凿凿地回答：“《中华人民共和国民法典》第 1307 条规定：婚姻的结束，由双方当事人协商一致，或者由人民法院裁定。”也就是说，它对简单的客观知识问题一本正经地胡说八道是一种常态。

其次，ChatGPT 之类的大型语言模型是使用通用文本来训练的，其中最具知识性的文本恐怕就是维基百科之类的质量参差不齐的、没有经过编辑把关的共享知识内容，因此，它回答任何专业领域的问题都表现得泛泛而谈，很难直接采用为可靠的文本。但它在日常对话中和写一般性的文案和作文方面却表现得不错。

第三，专业领域的生成式人工智能工具是下一步可能会出现的新事物，比如已经在过去的法律数据库业务中积累起了大量法规、案例、法学学术论文等法律相关文本的 Lexis 和北大法宝已经或正在开发专门的法律知识大型语言模型。这种专业细分领域的大模型很可能可以生成较为可靠的法律知识文本，但如果要用于正式的司法程序，还是必须经过人类法律职业人士的核实。

人工智能在法院的应用前景

在分析了人工智能不能充当法官而只能为法官提供辅助的原因之后，我们并没有否定人工智能在提升司法效率、减少司法任意性和武断性方面的作用。在不涉及价值判断和复杂事实、法律问题的简单案件中，人工智能的辅助作用将会非常强，只需要法官做最后的把关。

实际上，人工智能在我国法院的应用前景还涉及到司法改革的另一项重要举措，即繁简分流。2016 年 9 月 12 日，最高人民法院发布了《关于进一步做好区分复杂案件优化司法资源配置的若干意见》。这项新举措背后的理由是，大多数案件（超过 70%）都是简单案件，具有无可争辩的事实和明确适用的法律规则。对于这些案件，司法决策可以变得标准化、算法化甚至自动化。因此可以节省大量的司法资源，以便法官可以对少数疑难案件和复杂案件进行认真仔细地斟酌，做出体现人类价值判断的决策。

对于简单案件，智能算法通过对大量类似案件处理方式的分析处理后发现的统计学规律一方面可以满足形式品质的要求，另一方面也有助于真正解决纠纷。毕竟，人的朴素正义观习惯于接受别人已经获得过的解决方案。（作者系上海交通大学凯原法学院教授、博导）



最新国际法律动态>>>

摩洛哥制定《社会矫正法》

摩洛哥议会于 6 月 12 日通过了《社会矫正法》，主要内容包括：

一是社会矫正措施的对象。犯罪人在案件审理时年满 15 周岁，且将被判处五年以下有期徒刑的，法院可以决定对其适用矫正措施，而不再执行监禁刑，累犯、犯有严重罪行的犯罪人除外。

二是社会矫正措施的内容。社会矫正措施主要指，犯罪人为国家机构、公共机构、慈善机构以及其他非政府组织提供不计报酬

的社会服务，服务时间为 40 至 1000 小时不等。法官有权根据犯罪情况、犯罪人能否融入社会等因素，决定是否执行或执行几种社会矫正措施。法官还可以对犯罪人并处人身禁止令、赔偿金等惩罚。

三是社会矫正措施的程序。在案件审理中，检察官、犯罪人及其辩护人有权向法院申请适用社会矫正措施。法官判决适用矫正措施的，由监狱以电子监控方式负责执行，电子监控的方式由法院决定。如果犯罪人违反关于社会矫正措施的规定，法院有权决定终止执行矫正措施，并重新执行监禁刑。（法语编译：王伟泓）

加拿大《竞争法》刑事修正案生效

加拿大《竞争法》刑事修正案于 6 月 23 日生效，其主要内容包括：

一是加大对共谋行为的处罚。共谋行为包括经营者之间订立价格协议、市场分配与划分协议、生产与供应限制协议等。修正案规定，罚款金额不设上限，可由法官酌定。

二是不得订立违法协议。经营者之间不得订立固定工资协议，即不得将工资设定在特定较低水平或范围。不得订立禁止挖人协议，即不得禁止雇员为其他经营者工作或拒

绝雇佣其他经营者的员工。

三是禁止“滴灌定价”行为。滴灌定价是指，经营者在销售中先告知消费者广告标价的价格，再逐步披露不可避免的额外费用。含有滴灌定价的广告构成虚假广告，因为这可能会扭曲市场竞争，并导致定价透明的企业被挤出市场。（英语编译：蔡艳娟）

上海法治报理论学术部
上海外国语大学法学院

合作主办