科学学研究
Studies in Science of Science
ISSN 1003-2053,CN 11-1805/G3

《科学学研究》网络首发论文

题目：人工智能技术扩散中的知识产权治理——基于模型蒸馏的开放创新实践分析
作者：张惠彬，王怀宾
DOI：10.16192/j.cnki.1003-2053.20251020.001
收稿日期：2025-07-08
网络首发日期：2025-10-21
引用格式：张惠彬，王怀宾．人工智能技术扩散中的知识产权治理——基于模型蒸馏的开放创新实践分析[J/OL]．科学学研究.
https://doi.org/10.16192/j.cnki.1003-2053.20251020.001

DOI：10.16192/j.cnki.1003-2053.20251020.001 网络首发时间：2025-10-21 09:13:37 网络首发地址：https://link.cnki.net/urlid/11.1805.G3.20251020.1729.002

人工智能技术扩散中的知识产权治理

——基于模型蒸馏的开放创新实践分析

张惠彬，王怀宾

（西南政法大学民商法学院，重庆 401120）

摘　要：模型蒸馏作为人工智能技术扩散的核心技术，在实践中面临知识产权制度适配性不足的治理难题。开发者通过模型蒸馏获取教师模型输出可能构成著作权侵权，通过蒸馏手段提取相关数据可能侵犯商业秘密，而违反用户协议的蒸馏行为则存在违约或不正当竞争风险。针对上述问题，国际实践呈现两种典型模式：以 DeepSeek 为代表的开放创新范式，通过模型开源与宽松协议降低蒸馏侵权风险；以 OpenAI 为代表的防止竞争范式，则以技术闭源与严格协议禁止蒸馏行为。基于我国科技创新规律与产业升级需求，建议在尊重开源协议前提下，通过明确模型蒸馏的著作权法规则、限定反蒸馏协议的效力边界，系统化解技术扩散中的知识产权冲突，为人工智能技术创新营造友好制度环境。

关键词：模型蒸馏；知识产权治理；开放创新；制度适配

中图分类号：D923.4　　文献标识码：A

人工智能模型蒸馏系指在预先训练好的大型教师模型指导下，通过软化输出分布等策略将其知识迁移到轻量级学生模型中，从而实现模型压缩，以便手机、智能机器人等终端设备也可以低成本、高效率地部署模型[1]。模型蒸馏主要包括三种方法：软标签蒸馏、特征图蒸馏和关系蒸馏。软标签蒸馏是将教师模型的输出概率分布数据传递给学生模型，使学生模型在内容生成上与之对齐；特征图蒸馏是将反映教师模型训练数据集（输入）特征的中间层特征图传递给学生模型，使学生模型在模型中间层特征上与之对齐；关系蒸馏是将教师模型的样本间或特征间的关系数据传递给学生模型，使学生模型在推理过程上与之对齐。上述模型蒸馏过程可能涉及知识产权风险。学生模型开发者自动获取、复制教师模型的输出（技术上称之为“前向传播”步骤），有可能侵犯著作权。如果模型蒸馏对教

收稿日期：2025-07-08；**修回日期：**2025-09-22

基金项目：国家社会科学基金后期资助项目（24FFXB055）；2025 年重庆市研究生科研创新项目（CYB25157）；重庆市研究生教育教学改革研究一般项目（yjg250234）

作者简介：张惠彬（1984-），男，副教授，E-mail:179002478@qq.com。王怀宾（1996-），男，博士研究生。

师模型数据，包括输出分布概率、输入特征、推理过程数据进行了抄袭，则有可能侵犯著作权。上述数据如果构成商业秘密，则模型蒸馏有可能因属于不正当手段或违反保密义务获取商业秘密，还有可能因违反教师模型用户协议而构成违约。

当前产学界对模型蒸馏是否侵犯知识产权存在不同观点[2]。在实践中，DeepSeek 与 OpenAI 采取了不同的知识产权治理策略，其中 DeepSeek 尽可能对其模型进行开源，并在用户协议中允许用户自由蒸馏其模型；而 OpenAI 则对其关键模型进行闭源，并在用户协议中严格禁止模型蒸馏行为。综合现有观点争鸣和实践分歧，可以将模型蒸馏的知识产权保护模式分为开放创新型和防止竞争型，反映了不同的价值取向和产业背景。面对全球人工智能科技竞争格局愈演愈烈，我国应基于何种理念、选择何种知识产权治理模式化解人工智能模型蒸馏的知识产权侵权风险，以实现人工智能技术的创新与超越，成为亟须明确的问题。

1　人工智能模型蒸馏的知识产权风险识别

1.1　人工智能模型蒸馏的著作权侵权风险

模型蒸馏的著作权侵权风险是模型训练著作权侵权风险的延续。早期模型训练的著作权侵权风险系指开发者未经许可使用他人享有著作权的作品进行模型训练的著作权侵权风险。模型蒸馏的著作权侵权风险在此基础上演化出新特征，也即是学生模型开发者未经许可利用的是教师模型的输出（合成数据），而非他人享有著作权的作品。著作权侵权成立需要满足三个阶梯式要件。要件一，使用对象是受著作权法保护的作品。要件二，使用行为属于著作权法所规定的法定作品利用行为[3]。要件三，不存在合理使用情形。模型蒸馏的著作权侵权风险源自上述三个阶梯式要件在著作权法理论和实践层面的不确定性（见表 1）。

表 1　人工智能模型蒸馏的著作权侵权风险

Table 1　Copyright Infringement Risks in AI Model Distillation

涉及作品类型	1.教师模型输出	2.教师模型本身的诸多信息		
模型蒸馏方法	前向传播步骤	软标签蒸馏方法	特征图蒸馏方法	关系蒸馏方法
使用/获取的信息	模型输出	模型输出分布概率数据	模型输入（训练数据集）特征数据	（输入–输出）推理过程数据
著作权侵权要件一	模型输出是否构成作品	不属于作品	可能属于汇编作品	不属于作品

著作权侵权要件二	模型蒸馏使用模型输出是否法定作品利用行为？	属于法定作品利用行为
著作权侵权要件三	模型蒸馏使用模型输出是否属于合理使用？	是否属于合理使用？

就要件一而言，学生模型利用教师模式输出进行模型蒸馏，该教师模型输出的作品性质存在争议。“创作工具说”认为模型是用户创作的工具，模型输出是人类创作的结果，属于作品[4]。“自由意志决定表达说”认为用户无法准确预见模型输出的具体表达，模型输出不符合人类“智力成果”的要求，不属于作品[5]。就蒸馏所迁移的教师模型的三种数据而言，其中教师模型输出分布概率数据和推理过程数据通常不构成作品。这两类数据是人工智能对大量数据进行机器学习所生成的反映数据特征和模型推理规则的数据，难以反映开发者的智力劳动并为开发者所准确预见。模型输入的特征数据则有所不同，由于被训练数据集是由开发者汇编而成，这类数据集有可能构成汇编作品，其特征数据有可能与该汇编作品实质性相似。

就要件二而言，利用模型输出进行模型蒸馏是否属于“法定作品利用”行为也存在争议。“非表达性使用”观点认为模型训练并非理解作品的思想和表达，不属于传统著作权法意义上的“法定作品利用行为”[6]。而反对观点则认为，“非表达性使用”的解释路径会对著作权法体系造成冲击，破坏思想/表达二分法的内部逻辑[7]。并且学界对“非表达性使用”理论并未达成一致观点。

就要件三而言，利用教师模型输出进行模型蒸馏，以及基于模型蒸馏获取教师模型输入的特征数据是否属于合理使用也存在争议。反对者认为合理使用仅限公益性明显强于商业性的情形，当前主流商业性模型训练必须遵守法定许可[8]。赞同者主张大模型技术效果具有普惠性，模型训练对作品的使用既不会影响原作品的正常使用，也不会不合理地损害著作权人合法权益，应当属于合理使用[9]。

1.2 人工智能模型蒸馏的商业秘密侵权风险

侵犯商业秘密同样需要满足两个阶梯式条件。要件一，商业秘密需要满足秘密性、价值性和保密措施三个构成要素。未公开的作品如果具备商业价值，并且控制人采取了保密措施，则其信息内容同时构成商业秘密。要件二，行为人采取了不正当手段或违反了保密义务。不正当手段包括盗窃、贿赂、欺诈、胁迫、电子侵入或者其他不正当手段，违反保密义务则是违反保密义务或者违反权利人有关保守商业秘密的要求。模型蒸馏的侵害商业秘密风险同样源自上述两个阶梯式

条件在司法实践中的不确定性（见表 2）。

表 2　人工智能模型蒸馏的侵犯商业秘密风险

Table 2　Risks of Trade Secret Infringement in AI Model Distillation

涉及商业秘密类型	1.教师模型输出	2.教师模型本身的诸多信息				
模型蒸馏方法	前向传播步骤	软标签蒸馏方法	特征图蒸馏方法	关系蒸馏方法		
获取的信息	模型输出	模型输出分布概率	模型输入（训练数据集）特征	模型(输入-输出)推理过程	算法（如有）	源代码（如有）
侵犯商业秘密要件一	(公开访问)不构成商业秘密	可能构成商业秘密	可能构成商业秘密	可能构成商业秘密	构成商业秘密	构成商业秘密
侵犯商业秘密要件二	模型蒸馏是否属于不正当行为或违反保密义务？					

就要件一而言，教师模型的成果是否构成商业秘密存在争议。有观点认为模型生成的有价值的信息，也有可能是潜在的商业秘密[10]。但在公开访问背景下该信息是否具有秘密性和保密性存在争议。教师模型通常是公开访问的，任何用户都可以获得模型输出，因此难以满足秘密性要求。但如果教师模型是在控制者内部范围使用，则其输出内容有可能就具有了秘密性。在公开访问背景下，权利人主张模型本身而非模型输出构成商业秘密更具有可能性。（1）就模型输出的分布概率数据、输入到输出的推理过程数据、用于模型训练的数据集而言，开发者付出了高昂的数据获取成本，并且这些数据构成模型特征的系统架构，开发者通常以商业秘密进行保护。（2）就算法而言，我国商业秘密的司法解释以及美国司法实践明确算法可以构成商业秘密[11]。（3）就代码而言，尽管我国《计算机软件保护条例》将计算机程序的源代码与目标代码视为受著作权法保护的同一作品，但有观点认为源代码的公开程度低、隐秘性较强，同时其偏向于技术性特征而非表达功能，商业秘密保护更符合产业发展现实[12]。

就要件二而言，模型蒸馏行为属于侵犯商业秘密行为还是正当的商业行为存疑。一方面，模型蒸馏作为一种新型信息获取方式，难以将其与“盗窃”等传统方式并列为不正当手段。相反，模型蒸馏使用的是可以公开访问的模型输出，模型蒸馏技术本身也是业界惯常技术，因此更符合反向工程的定义。另一方面，如果模型蒸馏不具有较高技术门槛和资金门槛，则可公开访问的教师模型本身的诸多数据信息会因其“容易获得”而不满足秘密性和保密性要求。当然，如果教师模型开发者通过设置技术措施或合同条款增加获取难度，则模型蒸馏行为有可能

失去正当性基础。

1.3 人工智能模型蒸馏的不正当竞争和违约风险

当教师模型开发者因法律规范的不确定而不诉诸著作权和商业秘密保护时，也还可以主张模型蒸馏构成不正当竞争或违约（见表 3）。

表 3 人工智能模型蒸馏的不正当竞争和违约风险

Table 3 Risks of Unfair Competition and Breach of Contract in AI Model Distillation

风险类型	1.不正当竞争风险		2.违约风险	
条件一	模型蒸馏导致教师模型竞争利益受损	“禁止自动抓取输出”用户协议	“禁止反向工程”用户协议	“禁止利用模型输出开发竞争品”用户协议
条件二	模型蒸馏违反商业道德	前向传播违反协议	模型蒸馏违反协议	模型蒸馏违反协议
保护利益	竞争利益扩张	著作权扩张	商业秘密扩张	竞争利益扩张

作为行为规制法，反不正当竞争之诉只需要原告证明被告的不正当竞争行为导致合法利益受损，而无需如著作权、商业秘密需要原告证明权利（益）的存在。反不正当竞争法极强的适用弹性及实用主义规制便利，促使其成为知识产权保护的兜底法，并在司法实践中不断扩张适用[13]。教师模型开发者如果声称其用户被学生模型掠夺，导致其垄断利益受损，这种利用教师模型输出开发竞争品的模型蒸馏行为有可能被认为违反商业道德的不正当竞争行为。

模型蒸馏的违约风险系指学生模型开发者违反教师模型用户协议进行模型蒸馏可能产生的法律风险。以闭源模式发布其教师模型的开发者，通常会在用户协议中对用户行为进行限制。例如 OpenAI 在用户协议中明确“禁止自动抓取输出”“禁止反向工程”“禁止利用模型输出开发竞争品”等行为。[①]尽管违约风险是合同风险而非知识产权风险，但违约行为所导致的损害通常与知识产权有关，这种用户协议是以契约的形式意定了面向公众的无形利益，因其要求所有用户必须同意用户协议才能使用服务。特别是在《反不正当竞争法》对知识产权补充保护背景下，用户协议通过限制用户对作品、商业秘密的正当获取和使用，无形之中加强了对知识产权的保护。例如用户协议禁止反向工程就实际上扩张了商业秘密保护的边界。

① OpenAI “Terms of Use”，2024 年 12 月 11 日，https://openai.com/policies/terms-of-use/.

2 人工智能模型蒸馏的知识产权风险治理模式

2.1 开放创新模式：DeepSeek 的模型开源和开放蒸馏协议

开放创新模式旨在通过开源知识产权运动实现对信息的有效利用和共享，降低后续创新的成本。开源知识产权运动最初集中在计算机软件的源代码开放。软件开源的产业背景在于，一方面企业可以从开源社区中获取技术支持和技术反馈，降低软件开发和维护成本；另一方面软件的快速迭代和竞争使得用户有定制化和改进软件的需求，“软件即服务”的理念促使软件版权开源以满足即时和多元需求[14]。产业实践要求源代码公开和宽松的版权环境。源代码可以作为作品进行保护，但著作权法并不要求著作权人公开软件源代码，我国《计算机软件著作权登记办法》也允许著作权人就源代码中的机密信息予以保密。②虽然公众可以从发布的软件中知悉目标代码，但从目标代码推理出源代码非常困难。即使是知悉源代码，但未经许可的使用、复制、修改和分发源代码可能侵犯著作权和商业秘密。据此，源代码开放+开源许可证构成开放创新理念的核心支柱。

基于开放共享理念，DeepSeek 试图通过源代码开放、开源许可证和开源用户协议构造模型蒸馏知识产权保护的开放创新模式。在著作权保护上，主张附许可内容和限制内容的自动化、宽松的版权开源许可协议；在商业秘密保护上，主张开源行动放弃教师模型诸多数据信息的商业秘密保护；在不正当竞争和违约上，主张通过相对宽松的用户协议彻底豁免模型蒸馏的违约风险和不正当竞争风险。

就侵犯著作权风险而言，DeepSeek 采取授权更多、限制更少的开源许可证，开发者既可以直接挪用已开源的教师模型的代码、权重、数据集等作品，也可以通过模型蒸馏的方式间接挪用，而不必担心著作权侵权风险。DeepSeek 在 Hugging Face 上一共开源了 68 个模型和一个数据集，③其中 R1 和 R1-Zero 模型代码和权重都采用 MIT 许可证，其余模型权重采用 DeepSeek 许可证，但代码采用的是 MIT 许可证，这两种许可证都属于 Permissive 许可证类型。在同样遵循原始版权声明保留的基础上，上述许可证较 Copyleft 许可证类型赋予了其他开发者更多的自由，开发者无需将其衍生品同样开源并继承开源许可证，这也是其被视为最宽松开源许可证的重要特征。MIT 许可证并非没有任何限制，其免除了版权人对其他开发者因使用其开源内容造成损害的法律责任，同时也没有如

② 参见《计算机软件著作权登记办法》第十二条。

③ 参见 Hugging Face 开源网站：https://huggingface.co/deepseek-ai.

Apache 2.0 许可证一样明确包含专利授权和商标授权条款。而 DeepSeek 许可证则更贴近主流的 Apache 2.0 许可证，额外包含专利授权条款和专利报复条款。

DeepSeek 的开源许可证仅解决了其他开发者直接挪用，以及通过模型蒸馏方式间接挪用其开源模型成果的问题。而对于模型蒸馏行为需要使用到的模型输出的著作权侵权风险，则由用户协议加以豁免。《DeepSeek 用户协议》第 4.2 条规定，“我们将本服务输出的内容的任何权利、所有权和利益（如有）归属于您。您可将本服务的输入与输出应用于广泛的使用场景中，包括个人使用、学术研究、衍生产品开发、训练其他模型（如模型蒸馏）等。”④该用户协议不仅将模型输出的所有权利归属于用户，而且还特别说明可以将模型输出用于训练其他模型。

就侵犯商业秘密风险而言，DeepSeek 一方面通过模型开源放弃模型本身诸多数据的商业秘密保护；另一方面通过用户协议放弃模型输出的商业秘密保护。上述用户协议将模型输出的所有权利（益）转让给用户，是考虑到即使模型输出构成商业秘密，用户也可以基于该协议使用模型输出进行模型蒸馏，而不担心侵犯开发者的商业秘密。就违约和不正当竞争风险而言，上述用户协议的兜底条款重申了其他开发者模型蒸馏的自由，之所以是重申而非补充，是因为模型开源和权益转让已经足以豁免模型蒸馏的知识产权侵权风险。但兜底条款也有助于在条款术语解释存在歧义时有利于模型蒸馏方。

2.2 防止竞争模式：OpenAI 的模型闭源和禁止蒸馏协议

防止竞争模式旨在通过强知识产权保护巩固企业核心竞争力，确保收回前期投资收益的同时，防止其他主体进入市场进行竞争。与传统有体财产不同，智力成果的非竞争性和非排他性特征使其具有溢出效应，个人的创造性活动能使他人或社会受益，而受益者无需付出成本，产权化则可以将无形智力成果的外部收益内部化[15]。外部收益内部化是通过限制竞争的方式实现的，也即是通过制止他人的“搭便车”行为限制竞争，为自己赢得竞争优势。这一法律经济学分析构成了知识产权激励学说的理论基石。知识产权强保护带来的反竞争模式，通过使创造者独占其智力成果的全部社会价值，激励其进行持续创新，从而实现社会整体福利增长[16]。这是那些拥有人工智能先发优势的技术公司采取模型闭源和严格用户协议措施的重要理念支撑。

④ 参见《DeepSeek 用户协议》：https://cdn.deepseek.com/policies/zh-CN/deepseek-terms-of-use.html.

基于防止竞争理念，OpenAI 一方面将其主要模型闭源，另一方面通过严格的用户协议限制其他开发者的模型蒸馏行为。在著作权保护上，通过模型闭源限制其他开发者直接挪用教师模型本身的诸多数据，以及在著作权法规则不明确的情况下通过用户协议限制其他开发者利用模型输出进行模型蒸馏，以补充保护著作权；在商业秘密保护上，通过模型闭源将模型本身的诸多数据以商业秘密加以保护，并通过用户协议限制其他开发者的反向工程行为，以扩张商业秘密权益的边界；在不正当竞争与违约上，以诸多用户协议限制其他开发者的模型蒸馏行为，强化对其先发优势这一无形竞争利益的保护。

就著作权保护而言，模型闭源意味着其他开发者不得未经许可直接挪用其模型（构成作品的）诸多数据，也不得通过模型蒸馏方式间接获取并使用这些数据。著作权法在利用模型输出进行模型蒸馏是否侵犯著作权这个问题上，仍然不够明确。就商业秘密保护而言，模型闭源意味着 OpenAI 并不想将构成模型核心竞争力的模型本身的诸多数据公之于众，至于仅仅采取模型闭源是否足以确保这些数据构成商业秘密，仍然存在争议。这些争议导致 OpenAI 对模型蒸馏行为的风险规制不足，由此诉诸用户协议的合同法规制对其无形竞争利益进行补充保护。

OpenAI 于 2024 年 12 月 11 日发布其最新一版用户协议。[⑤]该协议禁止以下事项。第（1）项旨在重申其闭源模式，也即是禁止未经许可复制、修改和分发服务，否则构成著作权侵权；第（2）项是以合同形式禁止反向工程，如果未经许可或不符合法定条件进行逆向工程，则构成违约或侵犯商业秘密。第（3）项则是反爬虫约定，也即是禁止用户自动抓取数据。在中外司法实践中，违反反爬虫约定通常会被认为是侵权或违约[17]。第（4）项则声明，绕过技术保护措施获取作品有可能违反著作权法；第（5）项则与 DeepSeek 协议完全相反，完全禁止了其他开发者利用模型输出进行模型蒸馏的可能。

OpenAI 用户协议的核心是反爬虫、反“反向工程”、反蒸馏三类限制，目的是以合同形式保护其知识产权法未明文规定的新竞争利益。其中反爬虫协议是开发者对大量数据集合的控制利益；反“反向工程”协议是开发者对模型诸多数据信息的商业秘密利益的扩张；反蒸馏协议则是开发者对模型输出的著作权利益的扩张，也即是即使利用模型输出进行模型蒸馏属于合理使用，也不得随意使用。

⑤ See OpenAI “Terms of Use”，December 11, 2024，https://openai.com/policies/terms-of-use/

可以看出，基于防止竞争理念的用户协议设置，是通过限制用户合理、正当地接触数据、获取商业秘密、使用模型输出（作品）的自由，来扩张其无形竞争利益。在美国司法实践中，法院常常会基于合同自由认定该限制性约定有效，由此构成对教师模型竞争利益的保护[18]。

表 4　开放创新与防止竞争模式比较

Table 4　Comparison of Open Innovation and Anti-Competition Models

知识产权保护模式	开放创新模式（DeepSeek）		防止竞争模式（OpenAI）	
模型开源与否	（部分）模型代码、权重开源	减轻侵犯著作权风险	模型闭源	增加著作权侵权风险
开源内容	模型代码、权重和数据集			
开源协议	MIT 许可证/deepseek 许可证			
模型是否属于商业秘密	开源模型放弃商业秘密	减轻侵犯商业秘密风险	闭源模型保护商业秘密	增加侵犯商业秘密风险
用户协议限制蒸馏与否	允许模型蒸馏	减轻不正当竞争和违约风险	禁止爬取模型输入和输出 禁止反向工程 禁止利用模型输出进行蒸馏	增加不正当竞争和违约风险
输入/输出的知识产权归属	归属用户	减轻著作权侵权风险	归属用户	减轻著作权侵权风险
蒸馏其模型的知识产权风险	较小		较大	

3　技术理想与商业现实的调和：人工智能模型蒸馏的知识产权风险治理

3.1　开放创新模式更符合我国产业需求和发展规律

（1）加速技术普惠与创新生态构建

模型蒸馏的开放创新知识产权保护模式有助于加速技术普惠。《政府工作报告（2025）》指出，“持续推进‘人工智能+’行动，将数字技术与制造优势、市场优势更好结合起来，支持大模型广泛应用，大力发展智能网联新能源汽车、人工智能手机和电脑、智能机器人等新一代智能终端以及智能制造装备。”⑥为解决复杂任务，大模型通常需要大量计算资源和存储空间，难以适应手机等移动或嵌入式设备的消耗、计算和存储需求。而模型蒸馏能够显著减少模型功耗，并提高运行效率，方便将模型嵌入到各种终端产品中[19]。大模型开发门槛高，通常是科技巨头的游戏。通过“教师”模型向“学生”模型传递知识，模型蒸馏可以减少

⑥ 新华社.政府工作报告——2025 年 3 月 5 日在第十四届全国人民代表大会第三次会议上[N], 2025 年 3 月 12 日, https://www.news.cn/politics/20250312/a71e63d66967404e8e644f9753c65fc9/c.html.

对原始数据的依赖，降低开发成本并保障隐私和数据安全[20]。中小企业参与竞争的门槛降低，进一步加速大模型技术普惠。

（2）塑造行业标准与规则话语权

开放创新模式通过开源社区和技术共享，能更快形成事实性行业标准和规则话语权。例如 DeepSeek 的开源策略不仅为其赢得广泛的用户基础，而且还吸引众多初创公司通过 API 调用模型进行技术开发。这一开源策略使得 DeepSeek 的“思维链模板”有取代 Prompt 工程成为新的行业事实标准的趋势，促使其他竞争者持续跟进并受到这一软性行业标准的约束。安卓系统的成功经验表明，基于技术开源构建的庞大生态系统较之封闭的强知识产权保护更加有效。这一开源生态系统不仅可以降低开发成本、促进创新，而且能够依托庞大的开发者社区自发维持生态系统稳定。而开源者则通过广告、应用销售和其他服务获取收入。开源生态系统建立事实性行业标准时，可以通过标准、服务垄断收回利润，最终实现技术理想和商业现实的调和。

（3）应对外部技术封锁与竞争压力

在中美科技竞争的背景下，开放创新模式不仅能加速技术追赶、构建自主生态，还能通过开源协作打破美国技术垄断。人工智能技术领域的竞争主要表现为中美的竞争。其中美国拥有技术资源和先发优势，并通过技术出口限制竞争对手发展[21]。在知识产权领域，美国时常滥用“301 调查”与商业秘密法的域外适用效力，对中国科技和制造业展开歧视性打击[22]。美国《芯片和科学法案》等法案将中国列为威胁国家安全或知识产权的“外国对手”，并设置了各种“护栏”条款防止中国从美国技术发展中受益。而美国系列芯片封锁措施对中国人工智能技术发展造成了不利影响。美国技术闭源的产业背景是雄厚的技术资源和先发优势，基于强知识产权保护的防止竞争模式有助于维持其在人工智能技术领域的垄断地位。中国在芯片、算力、数据等人工智能发展核心资源受制于美国单边措施，而且人工智能技术较美国起步较晚，如果继续坚持内部技术闭源策略很可能扼杀众多初创企业创新活力。并且模型蒸馏降低了对外部芯片、算力、数据的依赖，显著提高我国人工智能产业应对外部技术封锁的议价能力。

3.2 我国应构建知识产权友好型的产业环境促进开放创新模式发展

（1）尊重并明晰人工智能模型本身的开源协议

无论采取何种知识产权保护模式，都需要首先尊重权利人关于模型本身的开源协议。首先，开源的内容必须是权利人有权开源的。通常，开源内容包括作为计算机软件作品的模型代码和可能构成作品的数据集。然而，模型权重是模型自发从数据中学习到的参数集合，不应被视为人类创作的作品。因此，开源模型权重通常被看作是放弃其商业秘密保护，而不受开源协议的限制。

其次，开源协议通常明确了许可内容和限制条件，对于未明确的内容应当做有利于开放创新的解释。大多数开源协议都没有明确许可解除条件和商业使用限制条件，但是中途撤回开源许可、要求商业收费，以及新增使用限制条款是显著的商业不道德和违反诚信原则的行为，不符合开源知识产权运动的基本精神。在罗盒开源软件著作权侵权案中，法院即否定了原告撤回开源许可并转向闭源收费的行为，体现了司法裁判中对开放创新精神的支持。⑦此外，作为特殊著作权许可合同，被许可人违反开源协议构成违约的同时，其合法使用作品的法律基础丧失，后续继续使用则构成了著作权侵权。

（2）明确人工智能模型蒸馏的著作权法规则

就模型蒸馏中对模型输出的利用，以及通过模型蒸馏迁移教师模型输入的特征数据的著作权风险而言，应基于开放创新理念，尽可能豁免其著作权侵权风险。

首先，司法实践可以将利用教师模型输出进行模型蒸馏视为我国司法实践中不侵犯著作权的“临时复制”行为。所谓“临时复制”系指在网络传输或计算机运行过程中演示、运行、传送或储存计算机程序和数据时随之发生的暂时性复制行为[23]。认定“临时复制”需要满足临时性、技术过程所必要且没有独立经济价值的要件。⑧首先，模型蒸馏过程中对教师模型输出的复制是模型蒸馏技术所不可或缺的，并且这些复制的数据在蒸馏完成后会被立即删除，满足临时性、技术过程所必要的要件。其次，这种复制是为模型蒸馏这一合法使用目的服务的。对模型输出的临时复制本身不具备独立的经济意义，也不会对著作权人的合法利益造成损害。就像网页缓存和转码是互联网发展的必要技术一样，模型蒸馏也是人工智能技术扩散的必要基础。

其次，将通过模型蒸馏迁移教师模型输入特征数据的行为解释为合理使用。2011 年最高人民法院的指导意见提出，在促进技术创新和商业发展确有必要的

⑦ 参见广州知识产权法院（2019）粤 73 知民初 207 号民事判决书。
⑧ 参见欧盟《信息社会版权指令》第 5 条第 1 款。

特殊情形下，可以参考美国“合理使用四要素”来认定合理使用行为。（1）就作品使用的性质和目的而言。模型蒸馏对教师模型输入数据特征的迁移通常不是原样挪用具体表达，而是根据蒸馏程度的不同迁移特征知识，由此实现模型知识的迁移和模型规模的压缩等技术目的；（2）就被使用作品的性质而言。教师模型所抓取的网络数据非常之多，尽管不能判断独创性高低和公共领域信息的详细比例，但可以想象类似经典文学艺术作品等高独创性作品终究是少数，更何况经过海量数据的稀释。（3）被使用部分的数量和质量。学生模型虽然对教师模型进行了压缩，但是模型蒸馏的一个典型特征是规模压缩但不显著降低模型质量。因此学生模型迁移教师模型知识时通常保证了迁移质量。（4）使用对作品潜在市场或价值的影响。对于教师模型输入特征数据因模型蒸馏而导致的价值降低或潜在市场减损而言，短期内很难评估。上述四要素分析中第一、二项有利于学生模型开发者，第三项有利于教师模型开发者，第四项则难以评估，需要在实践中综合考虑彼此的竞争关系和具体使用情况。基于开放创新理念，模型蒸馏对教师模型输入特征数据的迁移更应当认定为合理使用。至于教师模型开发者的竞争利益可以通过反不正当竞争法进行个案保护。

（3）限缩反蒸馏协议的效力边界

对于闭源环境下模型蒸馏的侵犯商业秘密、不正当竞争、违约风险，需要通过限制反蒸馏协议的效力进行化解。就侵犯商业秘密而言，如果教师模型是在内部受控环境下被使用，则模型输出和模型本身的诸多数据信息都构成商业秘密。此时模型蒸馏对教师模型控制者商业秘密的窃取是毫无疑问的。如果教师模型是以公众访问的形式公开发布，则教师模型输出不构成商业秘密。即使如此，教师模型本身的诸多数据信息仍然属于商业秘密，此时通过模型蒸馏获取这些商业秘密可以归为反向工程，不侵犯公开发布的（闭源）教师模型开发者的商业秘密。而基于开放创新理念，模型蒸馏也不应视为违反商业道德的不正当竞争行为，除非其影响原模型的使用时才需加以规制。

就违约风险而言，无论是反爬虫、反“反向工程”还是反蒸馏兜底协议，当承认模型蒸馏属于促进产业开放创新的对信息的合理接触、反向工程和合理使用时，对这类行为的合同限制相当于权利人通过私人约定重新定义了知识产权保护的法定内容和边界，从而打破了立法者确立的利益平衡关系[24]。据此，上述协议

都可以视为对法定的公众行动自由的限制和对法定的权利（益）边界的扩张。特别是在开放创新理念下，这一协议的效力应当得到限缩。当然，由于该协议本身属于合同自由的范畴，在否定该协议的效力时也需要适当考虑权利人的其他利益。基于开放创新理念，首先排除反蒸馏兜底协议的法律效力；其次，反爬取协议和反“反向工程”协议不得限制以模型蒸馏为目的数据爬取和反向工程，因其属于实现模型蒸馏的必要且合理的步骤；最后，模型蒸馏过程不得突破开发者为实现反蒸馏目的而设置的技术措施，也不得高频访问、过度挤占平台服务器资源从而影响模型的正常服务。

参考文献:

[1] Gou J, Yu B, Maybank S J, et al. Knowledge distillation: A survey[J]. International Journal of Computer Vision, 2021, 129(6): 1789-1819.

[2] Hrdy C A. Trade secrecy meets generative AI[C]//Rutgers Law School Research Paper Forthcoming," Disrupting AI" Symposium Issue of the Chicago Kent Law Review, Forthcoming. 2025.

[3] 涂藤.机器学习的著作权侵权判定：超越非表达性使用理论[J].政治与法律,2024,(10):162-176. Tu T. Copyright infringement determination in machine learning: Beyond the non-expressive use theory [J]. Politics and Law, 2024, (10): 162-176.

[4] 崔国斌.人工智能生成物中用户的独创性贡献[J].中国版权,2023,(06):15-23. Cui G B. Users' original contributions in ai-generated works [J]. China Copyright, 2023, (06): 15-23.

[5] 王迁.三论人工智能生成的内容在著作权法中的定位[J].法商研究,2024,41(03):182-200. Wang Q. Three discussions on the legal status of ai-generated content under copyright law [J]. Law and Business Research, 2024, 41(03): 182-200.

[6] 陶乾.基础模型训练的著作权问题：理论澄清与规则适用[J].政法论坛,2024,42(05):152-164. Tao Q. Copyright issues in foundation model training: Theoretical clarification and rule application [J]. Political and Legal Forum, 2024, 42(05): 152-164.

[7] 易继明.大模型语料训练合理使用问题研究[J].中国版权,2024,(06):5-26. Yi J M. A study on the rational use of large-scale model corpus training [J]. China Copyright, 2024, (06): 5-26.

[8] 蔡元臻.机器学习著作权法定许可的适用基础与规则构建[J].知识产权,2024,(11):77-93. Cai Y Z. The application basis and rule construction of statutory licensing for machine learning copyright [J]. Intellectual Property Rights, 2024, (11): 77-93.

[9] Lemley, M. A., Casey, B. Fair learning[J]. Tex. L. Rev., 2020, 99: 102-181.

[10] Levine D S. Generative artificial intelligence and trade secrecy[J]. J. Free Speech L., 2023, 3: 559.

[11] 许娟.生成式人工智能的“三经九纬”法治新模式[J].西南政法大学学报,2024,26(03):140-158. Xu J. A new model of “Three classics and nine weaves” Rule of law in generative artificial intelligence [J]. Journal of Southwest University of Political Science and Law, 2024, 26(03): 140-158.

[12] 刘汉霞.对计算机程序保护中“同一作品”原则的质疑——兼评《著作权法(修订草案送审稿)》第 5 条第 15 项[J].知识产权,2016,(06):54-61. Liu H X. Questioning the principle of “identical works” In the protection of computer programs: A commentary on article 5, item 15 of the draft revision of the copyright law [J]. Intellectual Property Rights, 2016, (06): 54-61.

[13] 陈耿华.反不正当竞争法规制界限之反思[J].法学,2025,(03):158-177. Chen G H. Reflections on the regulatory boundaries of anti-unfair competition laws [J]. Law Science, 2025, (03): 158-177.

[14] 张平.开放创新的知识产权应用机制[J].知识产权,2024,(06):3-17.

[15] Frischmann B M. Evaluating the demsetzian trend in copyright law[J]. Review of Law & Economics, 2007, 3(3): 649-677.

[16] Demsetz H. Towards a theory of property rights[J]. The American Economic Review, 1967, 57(2): 347-359.

[17] 崔国斌.网络反爬虫措施的法律定性[J].中国法律评论,2023,(06):157-174. Cui G B. Legal characterization of anti-crawling measures on the internet [J]. China Legal Review, 2023, (06): 157-174.

[18] Socal Diesel, Inc. v. Extrasensory Software, Inc., No. B290062, 2022 WL 702427, (Cal. Ct. App. Mar. 9, 2022).

[19] Chen Y, Zheng B, Zhang Z, et al. Deep learning on mobile and embedded devices: State-of-the-art, challenges, and future directions[J]. ACM Computing Surveys (CSUR), 2020, 53(4): 1-37.

[20] 邵仁荣,刘宇昂,张伟,等.深度学习中知识蒸馏研究综述[J].计算机学报,2022,45(08):1638-1673. Shao R R, Liu Y , Zhang W , et al. A review of knowledge distillation in deep learning [J]. Journal of Computers, 2022, 45(08): 1638-1673.

[21] 宫云牧.技术权力视角下的中美欧人工智能技术竞争[J].欧洲研究,2025,43(01):24-52+165+6. Gong Y M. The competition in artificial intelligence technology between china, the united states, and europe from the perspective of technological power [J]. European Studies, 2025, 43(01): 24-52+165+6.
[22] 李雨峰,刘明月.美国商业秘密法域外适用的扩张与中国应对——以《2022 年保护美国知识产权法》为中心[J].知识产权,2023,(08):106-126. Li Y F, Liu M Y. The expansion of the extraterritorial application of u.s. Trade secret law and china's response: Focusing on the protecting american intellectual property act of 2022 [J]. Intellectual Property Rights, 2023, (08): 106-126.
[23] 商建刚.数据训练的著作权法分析[J].法学论坛,2025,40(02):68-78. Shang J G. Copyright law analysis of data training [J]. Legal Forum, 2025, 40(02): 68-78.
[24] Lemley M A. Intellectual property and shrinkwrap licenses[J]. S. Cal. L. Rev., 1994, 68: 1239-1294.

Intellectual Property Governance in the Diffusion of Artificial Intelligence Technology

— An Analysis of Open Innovation Practices Based on Model Distillation

Zhang Huibin, Wang Huaibin

(Southwest University of Political Science and Law, School of Civil and Commercial Law, Chongqing 401120)

Abstract: Artificial intelligence model distillation, as a key technology driving lightweight deployment of large models, serves as the core pathway for achieving “AI+” industrial integration. However, during its technological diffusion, it faces governance challenges stemming from inadequate intellectual property system adaptability. AI model distillation refers to the process of transferring knowledge from a pre-trained large teacher model to a lightweight student model through strategies such as softening the output distribution. This achieves model compression, enabling cost-effective and efficient deployment on terminal devices like smartphones and intelligent robots. Model distillation can be categorized into three approaches: soft label distillation, feature map distillation, and relational distillation. These correspond respectively to the distribution probability data output by the teacher model, the feature data input to the teacher model, and the inference process data from input to output within the teacher model. These data may potentially hold intellectual property rights. The distribution probability data output by the teacher model, the feature data input to it, and the inference process data from input to output may constitute works. If student model developers improperly replicate these data through model distillation, they may infringe copyright. If the aforementioned data constitute trade secrets, model distillation may infringe trade secrets by constituting improper means or violating confidentiality obligations. Distillation practices violating user agreements carry risks of breach of contract or unfair competition. Regarding these issues, international practice presents two typical models: the open innovation model, exemplified by DeepSeek, aims to achieve effective information utilization and sharing through open-source intellectual property initiatives, thereby reducing subsequent innovation costs. For copyright protection, it advocates automated, permissive open-source copyright licenses with licensed and restricted content; for trade secret protection, it proposes that open-source initiatives waive trade secret protection for much of the teacher model's data. Regarding unfair competition and breach of contract, it advocates for relatively lenient user agreements to completely exempt model distillation from breach of contract and unfair competition risks. The anti-competition model, exemplified by OpenAI, prohibits distillation through closed-source technology and strict agreements. Regarding copyright protection, it restricts other developers from directly appropriating the trainer model's data through closed-source models. It also supplements copyright

protection by limiting other developers' use of model outputs for distillation via user agreements when copyright law rules are ambiguous. For trade secret protection, it safeguards the model's data as trade secrets through closed-source models and expands the scope of trade secret rights by restricting reverse engineering by other developers via user agreements. Regarding unfair competition and breach of contract, user agreements restrict model distillation by other developers, strengthening protection of their intangible competitive advantage—first-mover advantage. Open innovation pathways accelerate technological accessibility, foster innovation ecosystems, control industry standards and regulatory discourse, and counter external technological blockades and competitive pressures. Thus, they better align with China's industrial needs and scientific development patterns. China should cultivate an intellectual property-friendly industrial environment to promote open innovation models. It is recommended that, while respecting open-source agreements, China systematically resolve intellectual property conflicts in technology diffusion by clarifying copyright law rules for model distillation and defining the scope of anti-distillation agreements. This will foster a supportive institutional environment for AI technological innovation.

Keywords: model distillation; intellectual property governance; open innovation; institutional adaptation