知识蒸馏服务链下服务者责任归属的困境与治理

吴超超

安徽财经大学，中国·安徽 蚌埠 233000

摘要：随着知识蒸馏技术在模型轻量化领域的广泛应用，知识蒸馏技术作为深度学习模型轻量化的核心手段的同时，其服务链中多主体协作引发的责任归属问题日益突出。本文以知识蒸馏技术固有属性与法律框架的矛盾为切入点，系统探讨知识蒸馏服务链中服务提供者的责任认定困境及治理途径。通过技术可解释性增强、法律精细化适配与产业协同治理的三维路径，帮助克服知识蒸馏服务链中的“黑箱困境”，实现技术创新与公共利益的动态平衡，为医疗诊断、金融风控等敏感场景的合规落地提供理论支撑与实践参考。

关键词：知识蒸馏；服务链；责任归属；透明化治理

The Dilemma and Governance of Service Provider Liability Attribution in the Knowledge Distillation Service Chain

Wu Chaochao

Anhui University of Finance and Economics, China Anhui Bengbu 233000

Abstract: With the wide application of knowledge distillation technology in the field of model lightweighting, while it serves as a core approach for lightweighting deep learning models, the issue of responsibility attribution among multiple stakeholders in its service chain has become increasingly prominent. This paper takes the contradiction between the inherent attributes of knowledge distillation technology and the legal framework as the entry point, systematically exploring the responsibility determination predicament of service providers in the knowledge distillation service chain and the governance approaches. Through the three-dimensional path of enhancing technical explainability, fine-tuning legal adaptation, and industrial collaborative governance, the "black box predicament" of the knowledge distillation service chain can be overcome, achieving a dynamic balance between technological innovation and public interests, and providing theoretical support and practical references for the compliant implementation in sensitive scenarios such as medical diagnosis and financial risk control.

Keywords: Knowledge distillation; Service chain; Responsibility attribution; Transparent governance

0 引言

当前人工智能作为新一轮科技革命和产业变革的重要力量，无疑正在以前所未有的速度重塑着世界经济的版图与格局。近年来，生成式人工智能技术的突破正在深刻重塑着相关内容产业的整体格局，尤其是随着知识迁移机制对于模型优化能力的大幅度提升，通过模型优化与智能压缩的方法使模型算法应用的适应性效能逐渐增强。然而，当该项技术深度渗透至医疗诊断、金融分析以及新闻传播等涉及敏感信息的关键领域时，基于诸多原因所导致的决策偏差、信息泄露以及算法歧视等治理难题也在持续升级。本文则以此为基点从知识蒸馏服务链产业分工模式与知识蒸馏技术固有属性的双重维度切入，进而对知识迁移机制的服务供给者责任归属问题系统构建；基于相关司法判例与先进实践提出完备法律规范、完善技术监管、构建行业自律、优化企业守法维度下的多维治理路径。

1 知识蒸馏服务链中责任认定困境及相关理论基础

知识蒸馏服务链作为以知识蒸馏技术为核心来整合数据、算法、硬件及合规资源的系统性服务体系，其主要通过专业化分工与标准化流程的方式解决模型压缩、跨平台适配及规模化部署的难题，旨在实现深度学习模型从复杂到轻量的高效迁移。

1.1 知识蒸馏服务链的现状

知识蒸馏服务链一般包括数据供给方、教师模型开发方、蒸馏算法服务方、硬件适配方和终端部署平台等诸多角色，本文主要讨论的是知识蒸馏服务提供商的责任归属问题。知识蒸馏服务链的快速扩张使得模型轻量化技术从理论程度的模型研究转向具体实际应用的同时，随着服务提供者的责任认定问题的日益复杂程度，也暴露出了诸多矛盾。

1.2 知识蒸馏服务链中对服务提供商责任认定的必要性

知识蒸馏作为知识迁移的一种具体实现形式并非仅仅简单直接地复制模型数据或代码，而是通过“分析—模仿”机制模仿教师模型的输出来达到知识迁移的目的。知识蒸馏是一种教师——学生的训练结构，通常是已训练好的教师模型提供知识，学生模型通过蒸馏训练来获取教师的知识，它可以以轻微的性能损失为代价将复杂教师模型的知识迁移到简单的学生模型中。在数据合规重要性逐渐凸显的时代，无疑更应该考虑到当知识蒸馏服务商未经授权使用用户数据或未脱敏处理时所面临此类责任的情形，为此对于知识蒸馏服务链下 AIGC 服务提供商责任归属问题的认定也便显得越发重要了。

1.3 知识蒸馏及相关理论基础——以国内 AIGC 服务者责任理论为鉴

大部分学者对于我国 AIGC 服务者责任的认定认为多数情况下（即非 AI 自身固有缺陷情况下）应当适用过错责任，即在服务者与被服务者都存在着一定过错的情况下依据各方的过错程度进行合理分责，从而有利于各方责任的精准认定。就生成式人工智能引发的侵权形态而言，除侵害个人信息权益之外该类侵权与一般侵权并没有本质差异，理应适用过错责任。该类学说对于过错认定标准也有着不同，一种是以现有的技术水平为标准，即从时间、行业、地域这三个不同的维度整体出发，在综合考量现有的技术水平后才对生成式人工智能服务者的过错进行判断。

相较于过错责任说中以 AIGC 服务者是否存在过错为标准不同，严格责任说则只要客观上造成损害结果，那么无论是否存在主观过错，都应该据此承担责任，即要求 AIGC 服务提供者无论其是否存在主观过错的产生，服务者都应当承担无过错责任。未来需进一步通过技术标准细化内容安全的分级以及相配套的保险机制，从而更好地缓解严格责任对企业的压力，实现技术创新与公共利益的动态平衡。

2 知识蒸馏技术视域下服务提供商责任归责机制的多维挑战

知识蒸馏作为深度学习模型压缩和迁移学习的重要方法早已在多个领域得到广泛应用，然而由于该技术本身所存在若干如信息传递的不完整性、模型自身安全性的缺乏以及模型架构的适配问题也无疑决定了知识蒸馏技术在场景应用上的局限性缺陷，这些缺陷不仅影响模型性能，还会引发安全、伦理以及工程应用等其他方面问题。

2.1 知识蒸馏服务链下产权归属的模糊性

知识蒸馏技术的产业化应用在催生多主体协作链条的同时，由于各环节主体的责任边界模糊，一旦在知识蒸馏过程中出现模型失效、数据泄露或伦理争议等问题时，其有关主体责任划分问题便会陷入困境。从现行著作权制度和原理视角考察，AI 模型的蒸馏不构成对源模型的著作权侵犯，即使蒸馏中复制或使用了在先的版权保护资料，亦属于合理使用的范围。目前，学界统一认为“数据抓取协议”难以从技术上达到实质性保护的目的。爬虫协议是国际互联网领域通行的商业惯例和行业规范，可作为商业道德判断的标准之一。中小企业若绕过或者违反被爬取方的爬虫协议，则可能被认定违反互联网领域商业道德，构成不正当竞争。AIGC 服务商在知识迁移过程中可以直接根据该协议来管控和处理数据，从而在不损害公共利益和他人合法利益的前提下有权自主决定数据内容、范围以及主体。

2.2 知识蒸馏服务链下责任追溯的复杂性

知识蒸馏是一种基于知识迁移的模型压缩技术，其核心目标在于将复杂教师模型中蕴含的隐式知识高效迁移至轻量学生模型，从而在保证模型性能的前提下降低计算复杂度与部署成本。同时由于现有的知识蒸馏服务提供商的监管体系尚不完善，监管部门在面对复杂技术和众多服务提供商时缺乏有效监管手段和专业技术能力，以至于没有足够能力实现全面、精准监管。即使在行业自律方面部分行业已经组织制定了相关规范和标准，但这些规范终究由于权威性和执行力不够导致对服务提供商约束有限，仍会出现部分服务提供商为了短期利益而忽视自身责任从而造成市场秩序混乱的情况。

2.3 知识蒸馏框架下模型内生性缺陷的负面传导效应

知识蒸馏技术通过将复杂教师模型的知识迁移到轻量级学生模型中的方法，在资源受到限制的环境中实现高效部署的同时，这一过程也会因 AI 系统固有的技术缺陷引发责任认定的困境。在知识蒸馏过程中，原教师模型中所存在的偏见内容可能会通过蒸馏过程进而被放大或者隐藏，而这种偏见在学生模型中会体现得更加具体，如在图像分类任务中学生模型的数据生成便会在达到表面高度准确率的同时也会继承并强化教师模型中的这种偏见。

3 知识蒸馏服务链中服务者归属问题的完善路径

在生成式人工智能时代发展迅猛的同时，知识迁移机制作为其关键底层技术极大地革新了内容生产模式，不仅使生成内容的产出效率得以大幅度提升，而且使生成形式

也越发多样。然而随着 AIGC 应用范围的不断扩大，对于知识迁移时 AIGC 服务提供商的责任归属问题仍处于复杂与模糊的现状便严重阻碍行业的稳定发展，所以当下从多个角度寻求知识迁移过程中 AIGC 服务提供商责任的归属问题无疑显得越发重要。

3.1 构建知识蒸馏服务链下服务提供商的评估体系

知识蒸馏作为一种模型压缩技术，其方法的核心在于“知识”的设计、提取和迁移方式的选择，通常不同类型的知识来源于网络模型不同组件或位置的输出。通过让较小的学生模型模仿较大的教师模型的输出来提升性能，教师模型提供者开发原始的大模型，学生模型使用者负责具体有关应用，中间的服务提供商所需要扮演的角色则是负责提供蒸馏的技术或平台。与教师模型提供者只负责开发原始模型不同，知识蒸馏服务提供商更专注于如何将这个大模型压缩成小模型，同时学生模型使用者可能不具备蒸馏的技术能力，所以需要依赖服务提供商来完成这一步为此需要建立更全面的知识蒸馏服务提供商评估体系，将数据安全防护采用动态加密技术而非简单脱敏的方法贯穿始终，从而在传输过程中即使被截获也无法还原，给数据安全穿上真正的“隐身衣”。

3.2 加强知识蒸馏服务链中责任认定的透明化治理

在法律层面上，应当细化《生成式人工智能服务管理暂行办法》中有关“责任溯源标识”的义务条款，对于服务商呈现给用户的输出结果应附有“责任溯源标识”并标明其教师模型来源、训练数据授权情况和蒸馏环节各参与主体等信息；对于涉及医疗诊断等场景的应用来说，蒸馏模型输出的诊断建议要同步具备相关联的原始教师模型的资质认证编号和数据合规审查报告，便于监管和用户追溯到相关责任主体。

3.3 对知识蒸馏服务链下服务者一般审查义务的排除

考虑到 AIGC 技术的不可预测性和广泛传播性，由于 AIGC 服务者对算法训练、数据输入和内容生成机制具有主导权，因此也应当有能力通过、风险标注等技术手段来充分预防侵权风险。为促进 AIGC 产业形成合理规范并且不断创新发展的良好环境，整个行业需要通力合作，从而做到规范有序、管理得当以及监管得严。对于企业的技术和管理人员，需要定时地进行法律合规教育培训和伦理教育，以提升服务商充分了解知识迁移过程中产生的潜在责任风险；此外，还需尽快建立高效的事故责任紧急应对响应机制以便在发生侵权、数据泄露等责任事件发生时，可以迅速启动应急预案并配合有关部门调查、处理，从而有效通过及时采取下架内容、封存数据、告知用户等方式妥善应对责任危机。

3.4 建立行业自律规制和协同合作环境

要充分发挥行业协会的示范协调能力，制定具有权威性、可操作性的自律公约并要求服务提供商在知识迁移的使用过程中严守法制和道德底线，切实对数据的标注过程、使用过程以及模型优化过程中的相关行为做好规范。因此，要利用先进技术对知识迁移的整个过程进行实时、动态监控，对整个知识迁移过程实现全面监控，即使出现问题也能很快精准地找到责任环节。同时建立常态化的技术审查机制，监管部门要定期全面审查评估服务提供商的知识迁移算法、模型参数等核心技术，重点关注其安全性、合规性和伦理合理性，对于监管过程中所发现的技术安全隐患、合规问题或伦理风险等问题应当立即责令限期整改，实现从源头预防责任风险。

4 结语

生成式人工智能技术应用的深入发展推动着模型压缩和能力迁移等新方法的出现，有效促进基于知识蒸馏对大模型进行瘦身并训练使用的效率，同时使得知识蒸馏相继进入医疗诊断、金融风控等领域，实现相应的商业落地并取得巨大的经济效益，在一定程度无疑上促进了产业发展。本文围绕知识蒸馏的技术特点及实际产业应用情况，重点研究知识蒸馏服务提供方在模型训练过程中、训练数据调用时、输出结果的各个环节中应该承担怎样的责任问题，并就我国现有的制度规则不完善的方面提出相应改进方向，提出以提高技术的可解释性、明晰规则的法律逻辑适用性以及健全多方协同共治体系为基础的方法和解决手段。

参考文献：

[1] 黄震华，杨顺志，林威等. 知识蒸馏研究综述[J]. 计算机学报，2022,45(3):625-627.

[2] 王利明. 生成式人工智能侵权的法律应对[J]. 中国应用法学，2023(5):32-38.

[3] 林秀芹. DeepSeek 模型蒸馏的著作权法正当性重勘. 知识产权，2025(4):108-110.

[4] 陈淑婷. 中小企业爬取数据的正当性及规制路径. 华南理工大学学报，2024,26(4):103-104.

[5] 周险峰，尹文沛. 基于知识蒸馏技术的教学优化：DeepSeek 的教学应用与反思 . 湖南科技大学学报，2025,28(2):1-5.

[6] 邵仁荣，刘宇昂，张伟等 . 深度学习中知识蒸馏研究综述，2022,45(8):1642-1647.