

生成式人工智能训练数据著作权“合理使用”的边界认定与规则调适

陈星楠

宁波大学马克思主义学院, 浙江 宁波

收稿日期: 2026年9月11日; 录用日期: 2026年9月29日; 发布日期: 2026年10月23日

摘要

生成式人工智能的模型训练需要大规模利用已有作品, 此类利用能否构成合理使用, 直接决定训练数据的合规成本与权利人的受偿范围。既有研究与司法裁判分别以使用阶段、行为性质和主体目的划定边界, 三种划法结论互斥而标准无法通约, 根本原因在于划界依据取自产业政策权衡, 而非著作权的权利构造本身。著作权的排他效力仅及于作品的表达性利益, 边界的划定因而应以使用行为是否形成表达性替代为准。具体来说, 输入与训练阶段的非表达性利用不进入著作权权利范围, 输出端构成实质性相似的表达再现落入专有权范围, 模型记忆的评价位置在输出端而非训练端, 训练语料许可市场上的收益预期因循环论证与定价对象的缺失, 不属于三步检验法所称的合法权益。受制于复制权的过度扩张与《著作权法》第24条的封闭式列举, 该判准在现行法上尚无适用空间。我国宜将作品性使用规定为著作权侵权的成立要件, 并在《著作权法实施条例》中增设附机器可读权利保留机制的文本与数据挖掘例外。

关键词

生成式人工智能, 训练数据, 合理使用, 表达性替代, 作品性使用, 模型记忆

Delimiting the Boundary of Copyright “Fair Use” for Generative AI Training Data and Adapting the Rules

Xingnan Chen

School of Marxism Studies, Ningbo University, Ningbo Zhejiang

Received: September 11, 2026; accepted: September 29, 2026; published: October 23, 2026

Abstract

Training generative AI models requires the large-scale use of existing works. Whether such use can

文章引用: 陈星楠. 生成式人工智能训练数据著作权“合理使用”的边界认定与规则调适[J]. 法学, 2026, 14(10): 125-135. DOI: 10.12677/ojls.2026.1410286

constitute fair use directly determines the compliance costs of training data and the scope of compensation for right holders. In doctrine and judicial practice, boundaries have been drawn, respectively, by the stage of use, the nature of the act, and the purpose of the actor. The three approaches yield mutually exclusive conclusions, and their standards are incommensurable. The fundamental reason is that the basis for delimitation is drawn from industrial-policy balancing rather than from the rights structure of copyright itself. The exclusive effect of copyright extends only to the expressive interests of a work; the boundary should therefore be drawn according to whether the use effects an expressive substitution. Specifically, non-expressive uses at the input and training stages do not fall within the scope of copyright; at the output end, an expressive reproduction that is substantially similar falls within the scope of the exclusive rights; memorisation is to be assessed at the output end rather than at the training stage; and expectations of revenue in the market for licensing training corpora are not legitimate interests within the meaning of the three-step test, owing to the circularity of the argument and to the absence of any object that can be priced. Constrained by the over-expansion of the reproduction right and the closed enumeration in Article 24 of the Copyright Law, this criterion currently has no room for application under existing law. China should stipulate use as a work as a constitutive element of copyright infringement and add to the Implementing Regulations of the Copyright Law a text and data mining (TDM) exception accompanied by a machine-readable rights reservation mechanism.

Keywords

Generative Artificial Intelligence, Training Data, Fair Use, Expressive Substitution, Use as a Work, Model Memorisation

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

生成式人工智能的技术迭代建立在训练语料规模的持续扩张之上，但大量作品利用无法通过逐一授权完成，权利人与开发者之间的利益冲突随之进入司法程序。

2025 年是该类纠纷集中作出裁判的一年，而裁判所呈现的并非共识，恰恰是分歧。美国特拉华州联邦地区法院在 Thomson Reuters 诉 Ross Intelligence 案中认定，以他人法律要旨训练检索模型不构成合理使用；四个月后，加利福尼亚北区联邦地区法院在 Bartz 诉 Anthropic 案中把以合法购得图书训练大模型评价为高度转换性使用，却将盗版语料的下载与留存单独认定为侵权；随后，同一法院在 Kadrey 诉 Meta 案中同样支持了合理使用抗辩，审理法官却对前案的分析方法明确表示保留。¹德国方面，汉堡地区法院与汉堡高等法院先后驳回摄影师针对 LAION 提出的停止侵害请求，一审仅依科研例外作出判断，二审则认定一般例外与科研例外均可适用。²同一类技术事实，在不到两年的时间里被分别归入侵权、合理使用、科研例外与一般例外四种评价，分歧的根源显然不在事实认定。

我国的争议亦已显现。2024 年 6 月 20 日，北京互联网法院开庭审理全国首例涉 AI 绘画大模型训练著作权侵权案，被告以训练行为构成合理使用作为主要抗辩理由。³此前的裁判集中于输出端，北京互联

¹Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., No. 1:20-cv-613-SB (D. Del. Feb. 11, 2025), 该案已依 28 U.S.C. §1292(b)获准中间上诉，第三巡回上诉法院案号 No. 25-2153，一审程序中止；Bartz v. Anthropic PBC, No. 3:24-cv-05417 (N.D. Cal. Jun. 23, 2025)，该案于 2025 年 9 月以和解结案，其简易判决意见未经上诉审查；Kadrey v. Meta Platforms, Inc., No. 3:23-cv-03417-VC (N.D. Cal. Jun. 25, 2025).

²LG Hamburg, Urteil vom 27.09.2024, Az. 310 O 227/23; OLG Hamburg, Urteil vom 10.12.2025, Az. 5 U 104/24.

³<https://mp.weixin.qq.com/s/cyskAz1cASBaNIYQpGpGsA>

网法院在(2023)京 0491 民初 11279 号案中就使用者对生成图片的独创性贡献作出认定,广州互联网法院在(2024)粤 0192 民初 113 号案中就服务提供者的合理注意义务作出认定,⁴输入端的判断标准至今未在裁判文书中获得正面表达。现行规范也没有提供实体答案。《生成式人工智能服务管理暂行办法》⁵第 7 条要求服务提供者使用具有合法来源的数据和基础模型,涉及知识产权的,不得侵害他人依法享有的知识产权,《生成式人工智能服务安全基本要求》⁶第 5.1 条与第 5.2 条分别就语料来源安全与语料内容安全提出要求,其中第 5.1 条明确使用开源语料应具有开源许可协议,使用自采语料不应采集他人已明确声明不可采集的语料。两项规范均以不得侵权为表述,而训练阶段的作品利用是否构成侵权本身即为待决问题,规范功能因此止于宣示。

应对方案已有多种。张平主张综合运用合同、著作权集体管理与法定许可等工具,给予“多元化、多层次的解决方案”^[1]。焦和平主张“在立法上增设‘人工智能创作’合理使用类型”^[2],张新宝、卞龙则主张构建“以合理使用为原则,以授权许可为补充”的保护模式^[3]。刘晓春将数据训练中的作品使用界定为“非作品使用行为”而排除在著作权权利范围之外^[4],李安更进一步,主张将“作品性使用作为版权侵权成立要件”之一^[5]。施小雪认为解决之道在于“重塑复制权,而不是重塑合理使用”^[6]。这些方案在结论的宽严上分歧明显,在方法上却取向相近,即先行判断应否为人工智能产业留出发展空间,再选取能够支撑该判断的教义学工具。划界标准因而依附于产业政策立场,立场发生变动,结论随之变动,规则的可预期性难以形成。训练数据合理使用的边界长期存在分歧,根本原因不在于宽严尺度尚未取得共识,而在于用以划定边界的标准本身缺乏独立的规范依据。确定一项不依赖产业政策立场的判准,并据此检讨现行规则的容纳能力与调适方向,是解决上述问题的前提。

2. 训练数据合理使用边界的三种划法及其局限

已有的划界努力大致沿三条路径展开,分别以作品利用发生的技术阶段、利用行为自身的性质以及利用人的身份与目的作为识别对象。三条路径各有其现实动因,也各自获得了司法上的部分印证,但 2025 年前后的几个判决恰好暴露出它们的共同弱点,即所选取的识别对象都来自著作权法之外,因而无法在不同技术形态与不同主体之间保持结论的一致。

2.1. 以使用阶段划界的输入输出二分

沿技术流程切分法律评价,是目前接受度最高的一种处理。数据收集与模型训练被归入输入端,用户指令触发的内容生成被归入输出端,前者因不利用作品的表达而获得宽容,后者因可能再现表达而受到严格审查。Bartz 诉 Anthropic 案的处理方式最能体现这一思路,审理法官在假定模型已经记忆了部分图书内容的前提下,仍然认定训练本身构成合理使用,并明确表示该判决不涉及模型在具体情形中是否输出侵权内容,训练与输出由此被分作两个独立的评价单元。⁷这样的切分与工程实践的时序高度吻合,取证与举证也相对容易,却把著作权法的判断建立在一个并不稳固的技术事实之上。Thomson Reuters 诉 Ross 案提供了反证。该案中被告使用 Westlaw 法律要旨训练法律检索模型,法院认定其中第 2243 条要旨构成直接侵权,四要素中分量最重的第一与第四要素均支持权利人;被告主张其对法律要旨的使用仅发生于训练查询的生成环节,属于不可避免的中间复制,审理法官驳回了这一抗辩,理由是支持中间复制免责的先例均涉及计算机代码,代码的功能性使得中间复制成为接触非受保护部分的唯一途径,而法律要旨是表达性内容,复制它并非接触思想的必要手段。⁸审理法官同时强调该案所涉为非生成式人工智

⁴北京互联网法院(2023)京 0491 民初 11279 号民事判决书;广州互联网法院(2024)粤 0192 民初 113 号民事判决书。

⁵https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

⁶<https://www.tc260.org.cn/tc260/jswj/202608/b40dff5e24ae41d1a3068a66422d3c43/files/89f0ff4da02249d3aae9320c71eaa27b.pdf>

⁷同注 1。

⁸同注 1。

能,这一限定恰恰暴露了阶段划法的困境,因为被告的复制行为在技术流程上与生成式模型的训练并无二致,结论的分歧只能来自阶段之外的因素。同样处于输入端的复制行为,在一案中因转换性而免责,在另一案中因表达性而担责,可见决定结论的并非阶段本身。技术演进还会使阶段的边界移动。检索增强生成在响应用户指令时重新调取原始语料,作品的表达并非在训练时被吸收,而是在输出时被直接读取,向量数据库与微调语料则把作品内容以可还原的形式留存于系统之内,输入与输出在时间上的先后不再对应于表达是否被再现。这一划法推至极致还会自我消解,倘若输入端因不构成侵权而无须例外规则,输出端又因侵权认定不妨碍技术发展而“无需为其设置合理使用规则”[7],合理使用制度在训练数据领域即告落空,而《中华人民共和国著作权法》⁹(后文简称《著作权法》)第24条第1款以在特定情形下可以不经著作权人许可、不向其支付报酬的表述,构成现行法上唯一的权利限制通道,绕开该条并未告诉法院在个案中应当依据什么作出裁判。

2.2. 以行为性质划界的非作品性使用

把训练行为的性质而非位置作为识别对象,更接近问题的实质。这一思路在美国判例中已有较长的积累。Authors Guild 诉 Google 案中,被告扫描了两千余万册图书的全文并建立可检索的数据库,第二巡回上诉法院认定该行为具有高度转换性,理由是数字化的目的在于生成关于图书的信息而非提供图书的内容,片段展示的设计又使检索结果无法替代原书的阅读。¹⁰ A.V.诉 iParadigms 案中,论文查重系统完整存储了学生作品的全文用于比对,第四巡回上诉法院同样认定构成合理使用,因为系统从未向任何人展示这些作品的表达。¹¹两案的共同逻辑是,作品被完整复制这一事实本身不决定结论,决定结论的是复制之后表达有没有被再度呈现。以此观察数据训练,作品在其间被切分为词元、转化为向量并压缩为参数,模型所留存的主要是词语共现的概率分布而非文字的排列顺序,将其界定为非作品性使用因而具有相当的说服力。问题在于识别要素的选取。以利用的非特定性为标准,描述的是规模而非性质,海量且不指向单一作品的训练若在输出端复现了表达,规模并不能使之免责;针对某一部作品所作的词频统计或者风格量化分析虽然指向特定,却从未再现表达,指向的特定性同样不应使之落入侵权。以利用人的目的为标准,则会遭遇目的与效果的分离,参数过拟合与语料重复度过高都可能使一个以非表达为目的的训练在事实上产出替代性表达, Ross 案中被告的目的显然不是展示法律要旨,法院却仍然认定其使用与原作的用途相同。日本《著作权法》第30条之四在确立非享受目的利用规则的同时保留了但书,将特定情形排除在外,即依作品的种类与用途以及利用的样态,不当损害著作权人利益的情形。¹²其立法上的谨慎正是出于对这一落差的顾虑。就我国现行法而言,《著作权法》第52条第6项所称未经著作权人许可可以展览、摄制视听作品的方法使用作品或者以改编、翻译、注释等方式使用作品,第53条第1项所称未经著作权人许可复制、发行、表演、放映、广播、汇编、通过信息网络向公众传播其作品,均以作品被作为表达对象加以利用为前提,并未区分表达性利用与非表达性利用,非作品性使用要成为免责事由仍须立法配合,它相对于合理使用方案的优势主要停留在立法论层面。

2.3. 以主体目的划界的营利性限制

依利用人的身份分配宽严,是既有条文中可以直接推导出的做法。《著作权法》第24条第1款所列的12项具体情形中,第1项为个人学习、研究或者欣赏,第6项为学校课堂教学或者科学研究而少量复制,第八项的图书馆、档案馆等为陈列或者保存版本的需要复制本馆收藏的作品,都带有明显的非营利

⁹https://www.ncac.gov.cn/xxfb/flfg/flfg_532/202103/t20210309_50530.html

¹⁰Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).

¹¹A.V. ex rel. Vanderhye v. iParadigms, LLC, 562 F.3d 630 (4th Cir. 2009).

¹²<https://laws.e-gov.go.jp/law/345AC0000000048>

底色，欧盟《数字化单一市场版权指令》第3条与第4条建立的双轨结构又提供了现成样本，前者面向研究组织和文化遗产机构的科学研究目的并禁止以合同予以排除，后者面向一般主体却允许权利人声明保留。¹³德国 LAION 案两审的处理过程恰好显示出这条路径的不稳定。原告是一名摄影师，其上传至图库网站的照片被被告抓取并编入用于人工智能训练的开源数据集，图库网站的使用条款载有禁止以自动化程序抓取的表述。汉堡地区法院一审认定被告属于研究组织，其数据集免费向公众开放，商业公司事后使用该数据集不影响科研例外的适用，据此适用德国《著作权法》第60条d驳回诉请。¹⁴汉堡高等法院二审维持了结论，却把依据换成面向一般主体的第44条b，理由是权利人以自然语言写入使用条款的禁止抓取声明，在2021年的技术条件下不满足机器可读的要求，保留因而无效。¹⁵同一被告、同一行为，在两审中先后被归入非营利科研主体的例外与一般商业主体的例外，主体身份在结论中所起的作用因此难谓关键。移植到训练数据场景，这一做法的事实基础更为薄弱。学术机构与商业企业在语料获取上并无技术差别，当前具备规模的基础模型多由企业完成训练后再向研究者开放，主体身份还可以通过设立非营利实体或者产学合作的方式重新配置，Bartz 案中的被告是典型的商业公司，其训练行为仍被认定为合理使用，可见营利性在成熟的判断框架中也并非决定性因素。实定法上的依据同样不足，《著作权法》第24条第1款所列各项中仅有部分带有非营利限定，2020年修正后置于该款首部的概括性条件只要求不得影响作品的正常使用且不得不合理地损害著作权人的合法权益，并未把营利与否列为判断因素，将营利性提升为统一的划界标准超出了文义所能承载的范围。熊琦、张文窃主张“在机器学习的合理使用适用上破除‘非营利性’和‘适度性’局限”[8]，就破除本身而言方向可取，但破除之后仍须回答以什么替代营利性成为判断因素。

3. 表达性替代作为统一的划界依据

三种划法所依据的技术阶段、利用目的与主体身份，都不是著作权法自身的概念，判准若要摆脱产业政策的牵引，只能回到权利构造中寻找。著作权的排他效力究竟及于何处，决定了作品利用行为的评价起点，而三步检验法第三步在训练语料场景中所遭遇的循环，则说明这一起点若不予明确，例外规则本身也无法运作；模型记忆与累积性损害则构成对该判准的直接检验。

3.1. 著作权的排他效力仅及于表达性利益

《著作权法》第3条将作品界定为文学、艺术和科学领域内具有独创性并能以一定形式表现的智力成果，《中华人民共和国著作权法实施条例》¹⁶(后文简称《著作权法实施条例》)第2条重申了同一界定，即具有独创性并能以某种有形形式复制的智力成果。受保护的是表达而非思想，这一点在思想表达二分法的长期实践中已经确立。排他效力的边界因此不由使用行为的规模、频次或者商业价值划定，而由该行为是否触及作品的独创性表达划定。临时复制的处理最能说明这一点，数字传输过程中作品必然被完整地固定于内存之中，从物理形态上看复制无可否认，多数法域仍将其排除于复制权之外，理由并非复制的时间短暂，而是这类固定不面向传播，不使作品在受众处获得独立的呈现，因而不触及著作权所保护的那部分利益。同样的道理适用于阅读、浏览与检索，一个人反复研读某部小说并归纳其叙事技法，无论投入多少时间，都不构成对该小说著作权的利用。Google 图书案与 iParadigms 案之所以在完整复制的事实下仍认定合理使用，Bartz 案之所以在假定模型已经记忆部分内容的前提下仍认定训练具有转换性，依据的都是这一层判断，即著作权人真正有权控制的是表达在他处的再度呈现，而非表达曾经进入过何

¹³<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019L0790>

¹⁴https://www.gesetze-im-internet.de/urhg/_60d.html

¹⁵https://www.gesetze-im-internet.de/urhg/_44b.html

¹⁶<http://xzfg.moj.gov.cn/front/law/detail?LawID=598>

种系统。据此可以提出一项统一的划界依据，即表达性替代，指使用行为是否使作品的独创性表达在他处重新出现并被受众感知，从而在著作权法所保护的表达性利益上对原作品形成替代。该判断不依赖技术形态，不依赖利用人的身份，也不依赖对产业前景的预测，它所追问的只是权利所保护的对象有没有受到侵害。

3.2. 三步检验法第三步的循环论证与许可市场的经济分析

以表达性替代作为划界依据，最直接的障碍来自《著作权法》第24条第1款首部的概括性条件，其中“不得合理地损害著作权人的合法权益”一句，在训练语料场景中会陷入循环。权利人主张损害的通常不是作品的销售或者传播市场，而是本可以向模型开发者收取的语料许可费，这项收益预期的形成却完全依赖于结论本身，只要认定训练需要获得许可，许可费市场随即产生，市场的存在又反过来证明未经许可的训练造成了损害。Ross案第四要素的说理提供了这一循环的司法样本，法院认定被告不仅在法律检索产品市场上与原告直接竞争，还损害了原告将其法律要旨用于人工智能训练数据的潜在许可市场，而该潜在市场在诉讼之前并不存在，它恰恰是本案侵权认定所创设的。这一循环在美国判例法上早有警觉，美国地质学会诉德士古公司案的异议意见即指出，以许可市场的存在论证市场影响要素，等于让结论充当自己的前提。¹⁷ Kadrey案则从另一侧显示了同一问题，审理法官提出模型大规模产出同类内容会稀释原作所在市场的理论，认为这才是权利人最有力的主张，只因原告未能就此举证而判其败诉。市场稀释若被普遍接受，任何一种新的内容生产方式都可以被认定为损害了在先作品的市场，合理使用制度将无从成立。问题的关键在于合法权益的范围如何确定。三步检验法源自《伯尔尼保护文学和艺术作品公约》¹⁸第9条第2款，该款准许成员国在某些特殊情况下允许复制作品，但以不损害作品的正常使用且无不故侵害作者的合法利益为条件，其第三步所要保护的是著作权人基于法定专有权可以获得的正常收益，而不是任何在事实上可能实现的收入机会。著作权人有权禁止他人复制、发行、改编、传播其作品，并由此在这些利用方式上取得对价，这是法律已经承认的收益来源；至于作品被用于统计分析、语言建模或者风格量化所可能产生的付费需求，属于法律尚未赋权的领域，把这部分预期计入合法权益，等于用第三步反向创设一项专有权，而该检验法的功能本是限制专有权而非扩张专有权。

这一结论在经济分析上亦可获得支持。合理使用所处理的，正是交易成本高于交易收益而使许可无从达成的市场失灵情形。¹⁹训练语料恰在此列。权利人数以亿计且大量作品权属不明，逐一磋商的成本远高于单一许可可能带来的收益；更为根本的是定价对象的缺失，模型性能取决于语料整体的规模与多样性，单一作品的边际贡献趋近于零且事后无从计量，按作品数量或者字符数计费只是对不可计量之贡献所作的拟制。其分配后果亦与初衷相悖，有能力实际收取许可费的是持有大规模作品库的出版集团与平台，分散的创作者因谈判成本无从进入交易，许可费反而构成新进入者的固定成本；加之作品多完成于模型训练成为现实之前，事后授予的收益不产生任何事前激励。据此，该项收益预期既非法定专有权项下的正常收益，也不具备经济上的可实现性。

法国竞争管理局2024年3月对谷歌所作的二亿五千万欧元罚款常被引为权利人享有数据市场利益的例证，但该案的规范基础是新闻出版者邻接权项下的谈判义务与其此前所作承诺的违反，处罚理由之一在于谷歌以出版者内容训练生成式服务时未予告知，属于竞争法上对谈判行为的规制，与著作权法上合理使用的判断并不同层。²⁰

¹⁷American Geophysical Union v. Texaco Inc., 60 F.3d 913 (2d Cir. 1995) (Jacobs, J., dissenting); 另参见 Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 592 (1994)，该案认定著作权人对批评性使用不享有受保护的演绎市场。

¹⁸<https://www.wipo.int/wipolex/zh/treaties/textdetails/12214>

¹⁹Wendy J. Gordon, Fair Use as Market Failure: A Structural and Economic Analysis of the Betamax Case and Its Predecessors, 82 Colum. L. Rev. 1600 (1982).

²⁰Autorité de la concurrence, Décision n° 24-D-03 du 15 mars 2024 relative au respect des engagements figurant dans la décision n° 22-D-13 du 21 juin 2022, <https://www.autoritedelaconcurrence.fr/fr/decision/relative-au-respect-des-engagements-figurant-dans-la-decision-de-lautorite-de-la-0>.

3.3. 模型记忆与累积性损害的处理

把训练定性为非表达性利用，不应以将该过程理想化为纯粹的统计分析为代价。抽取实验已经表明，以特定前缀提示可使模型逐字输出训练语料中的文本，记忆量随模型参数规模与样本在语料中的重复次数而增长，扩散模型亦可在生成结果中复现训练集里的具体图像。²¹惟记忆在著作权法上的意义须分三层。参数中的统计编码不构成可独立感知的复制件，亦无从为受众直接读取；可提取性以提示者已经持有原文为前提，模型并未因此成为受众接触原作的替代渠道；惟有输出与在先作品构成实质性相似时，表达才在他处重新出现并被受众感知。Google 图书案与 iParadigms 案的处理正是在这一层展开的，决定结论的并非系统内部存储了什么，而是有没有向公众呈现。记忆问题的安放位置因而在输出端而非训练端，它所要求加强的是输出环节的控制与责任，而非把训练行为整体重新定性。

但一概以训练端无涉表达处理，会忽略记忆程度在个案之间的差异，判准因而需作两处细化。其一是可归责性，语料去重、样本加权与训练轮次控制均可显著降低复现概率，开发者对此具有支配力，未采取合理技术措施而致复现的应认定有过错；已履行去重与输出过滤义务的，就偶发的相似输出不承担赔偿责任。其二是临界情形，单一作品在语料中的重复次数显著偏高，或者模型在通常提示条件下即可稳定复现足以构成独创性表达的连续片段的，应当推定训练环节本身构成表达性利用，由开发者就语料构成与去重措施承担反证责任。Kadrey 案所称的累积性损害则应按可归因性分流。泛化的市场稀释源于内容生产成本的整体下降，与任何一部在先作品的表达之间并无可追溯的因果关系，若据以认定损害，思想表达二分将在市场影响的名义下被取消；可追溯至特定作品集合的定向训练替代，则仍依本判准处理。

3.4. 依表达性替代重新归置三类行为

判准一旦确定，训练流程中的作品利用行为可以据此重新归置。数据收集、语料清洗、词元切分与参数训练构成一个连续的技术过程，其间作品的表达被拆解为统计特征，模型对外呈现的是概率分布而非文字序列，这类利用不使表达在他处重现，属于著作权利范围之外的非表达性利用，无须借助合理使用抗辩即可获得正当性，将其安置于合理使用之内反而要以承认其原本落入专有权为代价。输出端生成内容自身能否获得著作权保护，尚存争议。王迁以创作是“直接决定表达性要素的行为”为由否定其可版权性^[9]，徐小奔则主张对生成内容“与人类作品给予相同的制度评价”^[10]。该争议关乎生成物的权属，与训练阶段的利用行为是否落入专有权范围分属不同层面，不影响此处的归置。输出端的判断则回到既有规则，模型根据用户指令生成的内容若与在先作品构成实质性相似，表达即在他处重新出现并被受众感知，替代由此成立，无论该内容出自记忆化复现还是出自参数偶合，都应按侵权处理，赔偿责任的有无则依前节所述可归责性的区分确定。介于两端之间的是风格与题材层面的相似，模仿某位画家的笔触或者某位作家的句式而不触及具体表达，属于思想层面的借鉴，不构成替代，这一结论与人类创作者研习前人作品后形成个人风格的处理并无二致。真正需要单独安排的是针对个别作者的定向训练，以某一位在世作者的全部作品为语料训练专用模型并对外提供服务，即便单次输出未必构成实质性相似，模型整体已经在该作者的表达市场上形成持续性替代，Ross 案的结论正可以在这一层获得解释，被告以竞争对手的编辑成果为语料训练同类检索工具，其产品在功能上直接替代原产品，替代的成立与该案是否属于生成式人工智能并无关系。此种情形应当允许权利人事先声明排除，其正当性不在于训练行为本身的性质，而在于替代效果的可预见性。

²¹Nicholas Carlini *et al.*, Quantifying Memorization Across Neural Language Models, in Proceedings of the 11th International Conference on Learning Representations (2023); Nicholas Carlini *et al.*, Extracting Training Data from Diffusion Models, in Proceedings of the 32nd USENIX Security Symposium (2023).

4. 现行法容纳该判准的规范障碍

判准的两端对现行法提出了不同的要求，一端要求非表达性利用停留在权利范围之外，另一端要求具备替代可能的利用获得可操作的例外空间。我国现行著作权法在这两处都缺少接口，复制权的构成要件过于宽泛使前者无从落地，第 24 条的封闭构造使后者无从展开，实践中用以填补空白的司法政策则在效力与融贯性上都存在疑问。

4.1. 复制权的过度扩张吞并了非表达性利用

《著作权法》第 10 条第 1 款第 5 项将复制权界定为以印刷、复印、拓印、录音、录像、翻录、翻拍、数字化等方式将作品制作一份或者多份的权利，数字化一词的加入使任何形式的固定都可能落入其中。这一扩张有其历史来由，复制权诞生于印刷术时代，当时的复制件本身即是流通标的，控制复制等于控制传播，二者在事实上重合，无须在构成要件上加以区分。进入数字环境后，复制不再必然指向传播，浏览网页产生的缓存、传输过程中的内存暂存、备份系统的镜像存储都是纯粹的技术环节，若一并纳入专有权，复制权将异化为专有阅读权。多数法域已对此作出回应，欧盟《信息社会版权指令》第 5 条第 1 款豁免了属于技术过程必要组成部分且不具有独立经济价值的临时复制，²²法国《知识产权法典》第 L.122-5 条第 6 项对过渡性与附属性临时复制作了同样处理。²³我国立法未设对应条款，《著作权法实施条例》第 5 条仅界定了时事新闻、录音制品、录像制品、录音制作者、录像制作者与表演者六项用语，未就复制的构成要件作出限定，爬取、缓存、清洗与向量化这一系列中间行为在文义上均可被认定为复制。后果在举证结构上尤为明显，即便法院认同这些行为不触及表达，也只能通过合理使用抗辩加以处理，而抗辩的成立以侵权构成作为前提，被上诉人须先承认行为落入专有权，再逐项证明符合例外条件，非表达性利用因此被迫先进入权利范围再谋求退出。Ross 案中被告的中间复制抗辩之所以被轻易驳回，一个技术性的原因正是该抗辩在体系上属于合理使用第一要素项下的论点，法院只需认定其用途与原作相同即可终止讨论，无须回答复制行为本身是否触及表达。

4.2. 《著作权法》第 24 条的封闭式列举与兜底条款的双重锁闭

《著作权法》第 24 条第 1 款以 12 项具体情形加一项兜底的方式列举合理使用，第 13 项为法律、行政法规规定的其他情形，创设新类型的权限被明确保留给法律和行政法规，法院无从在个案中自行增设。训练数据的利用也无法通过解释纳入既有各项，第 1 项限于个人学习、研究或者欣赏，训练行为的主体与规模均不符合；第 6 项限于学校课堂教学或者科学研究中的少量复制，“少量”的限定与语料规模直接冲突；第 12 项针对以阅读障碍者能够感知的无障碍方式提供，与训练场景无涉。修法路径同样不畅，2020 年第三次修正刚刚完成，我国著作权法的修订周期历来以十年计，短期内再度启动的可能性有限，而模型训练的合规需求已经现实存在，等待修法意味着在相当长的时期内将问题留给个案裁量。2020 年修正后置于该款首部的概括性条件虽然为解释留下了余地，但它的功能是对已列明的情形施加限制，而非独立的授权依据，把它当作创设新例外的通道，与该款的结构安排不合。相形之下，德国法在转化欧盟指令时于《著作权法》中分设第 44 条 b 与第 60 条 d 两项文本与数据挖掘例外，为 LAION 案提供了可直接援引的规范依据，两审法院的分歧仅在于适用哪一条，而非有无可适用之条，这一对照说明规范供给的有无比规范尺度的宽严更为基础。

4.3. 司法政策中混合标准的效力局限

规范供给不足的空白，实践中主要由司法政策填补，其效力与体系融贯性都存在疑问。最高人民法院

²²<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32001L0029>

²³https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000037388886

院 2011 年印发《关于充分发挥知识产权审判职能作用推动社会主义文化大发展大繁荣和促进经济自主协调发展若干问题的意见》²⁴提出,对于著作权法未明确规定的利用作品方式,在考虑使用作品的目的与性质、被使用作品的性质、被使用部分的数量与质量、对作品潜在市场或价值的影响等因素后,可以认定为合理使用。该意见实际上将美国《版权法》²⁵第 107 条的四要素与三步检验法后两步组合在一起,形成一套混合标准,为个案说理提供了工具。问题在于,四要素判断法生长于判例法体系,其运作依赖法官在长期个案积累中形成的类型化经验,Ross 案与 Bartz 案在同一年内就相同性质的行为得出相反结论,正说明四要素本身并不自动产生确定性;而三步检验法是对成文法上权利限制的外部约束,其功能在于防止例外被过度扩张。两者的规范功能并不相同,简单叠加会使判断因素之间缺乏位阶,法官在个案中得以通过选择侧重何种要素而决定结论。效力层级的问题更为直接,该意见属于司法政策文件,其效力不足以突破《著作权法》第 24 条第 13 项对创设权限的保留,法院援引它认定新的合理使用类型,在方法上仍是以政策解释填补立法留白。

5. 规则调适的三重安放

判准所要求的三件事,分别落在不同的规范层次上。非表达性利用的排除属于侵权构成要件问题,具备替代可能的利用需要实定法上的例外通道,而两端之间的灰色地带无法由实体规则穷尽,只能借助程序性义务加以承接。三者依次对应前述三种划法各自留下的空缺,合并起来构成一套可以实际运作的规则安排。

5.1. 在侵权成立要件中确立作品性使用

非表达性利用要在权利范围之外获得安置,其路径在于在侵权认定的入口而非出口加以处理。可行的做法是把作品性使用确立为著作权侵权的成立要件,即被诉行为须以再现作品的独创性表达为内容,方有进入侵权判断的资格,爬取、清洗、词元切分与参数训练因不具备这一属性而在构成要件层面即被排除,无须权利人另行举证,也无须被诉人援引合理使用抗辩。这一要件在解释论上已有生长空间,《著作权法》第 52 条与第 53 条所列各项侵权行为,无论复制、发行、出租、展览还是信息网络传播,均以作品被作为表达对象加以利用为前提,把使用作品理解为表达性利用,属于对既有文义的限缩解释而非扩张,与思想表达二分法的既有实践也相互支撑。与重塑复制权构成要件的方案相比,作品性使用要件的适用面更宽,复制权的重塑只能解决固定环节的定性,而训练流程还涉及改编、汇编与信息网络传播,逐项改造各权利的构成要件成本过高,在侵权认定的共同入口设置一道过滤,可以一次性覆盖全部权利。该要件同时能够纠正现行举证结构的偏差,Bartz 案的走向已经显示,一旦训练行为整体被推入合理使用的评价框架,讨论就会集中于转换性与市场影响这类高度弹性的因素,而语料来源合法与否这一更为清晰的事实反而退居其次,最终以七百余万册盗版图书的留存行为单独承担责任。把作品性使用前置,可以使表达是否被再现这一实质问题优先于政策权衡获得回答。可能的反对意见是,在构成要件层面直接排除非表达性利用,会使著作权人彻底丧失与模型开发者谈判的筹码。这一顾虑值得重视,但它所设想的筹码来自一项著作权法从未赋予的专有权,即控制他人以非表达方式分析自己作品的权利,以谈判地位为由承认该项权利,与前述对三步检验法第三步循环论证的批评落入同一错误。权利人的议价基础应当另有来源,权利保留机制使其能够事先排除针对自身作品的定向训练,来源合法性门槛使盗版语料无

²⁴<http://gongbao.court.gov.cn/Details/90857967e518c766c368851b1b705a.html?sw=%e5%85%b3%e4%ba%8e%e5%85%85%e5%88%86%e5%8f%91%e6%8c%a5%e7%9f%a5%e8%af%86%e4%ba%a7%e6%9d%83%e5%ae%a1%e5%88%a4%e8%81%8c%e8%83%bd%e4%bd%9c%e7%94%a8%e6%8e%a8%e5%8a%a8%e7%a4%be%e4%bc%9a%e4%b8%bb%e4%b9%89%e6%96%87%e5%8c%96%e5%a4%a7%e5%8f%91%e5%b1%95%e5%a4%a7%e7%b9%81%e8%8d%a3%e5%92%8c%e4%bf%83%e8%bf%9b%e7%bb%8f%e6%b5%8e%e8%87%aa%e4%b8%bb%e5%8d%8f%e8%b0%83%e5%8f%91%e5%b1%95%e8%8b%a5%e5%b9%b2%e9%97%ae%e9%a2%98%e7%9a%84%e6%84%8f%e8%a7%81>

²⁵<https://www.wipo.int/wipolex/zh/legislation/details/3923>

法进入例外，两项安排都指向具体可主张的事实，比一项笼统的许可请求权更具可执行性。真正的难点在于该要件的证明分配，被诉行为是否再现表达，权利人在模型内部无从查证，若一概由权利人承担举证责任，要件将沦为免责通道，故需在程序上要求服务提供者就其训练流程与语料处理方式作出说明，并在前述临界情形下就去重与过滤措施承担反证责任。

5.2. 在《著作权法实施条例》中增设文本与数据挖掘例外

具备替代可能而仍值得宽容的利用，需要一条实定法上的例外通道，《著作权法》第24条第1款第13项已经预留了这一接口，行政法规是现阶段成本最低的载体。修改《著作权法实施条例》增设文本与数据挖掘例外，既不触动法律的修订周期，又能满足《著作权法》第24条第1款第13项对规范位阶的要求。立法技术上宜采结果导向而非条件限制的模式，欧盟《数字化单一市场版权指令》第3条与第4条以主体资格、访问合法性与保留声明等条件层层设限，条件之间的组合关系复杂，LAION案两审法院在适用哪一条例外上的反复即由此而来；日本《著作权法》第30条之四则以利用是否以享受著作物所表现的思想或情感为目的作为唯一标准，并以但书排除不当损害权利人利益的情形，条文简洁而适应性强。以此为参照，增设的条款可以规定为：为分析、统计或者训练计算机模型的目的使用作品，未向公众再现该作品独创性表达的，可以不经著作权人许可，不向其支付报酬，但不得影响作品的正常使用，也不得不合理地损害著作权人的合法权益。鉴于记忆的客观存在，其中“未向公众再现”应同时包含采取语料去重与输出控制等合理技术措施的义务。条款还应附设权利保留机制，允许著作权人以显著方式声明其作品不得用于模型训练。声明的形式要件必须同步明确，LAION案二审提供了直接的教训，权利人以自然语言写入网站使用条款的禁止抓取声明被认定不满足机器可读的要求，保留因而归于无效，我国若不同步建立机器可读的声明标准，保留机制在诉讼中将同样落空。声明的效力范围亦应加以限定，宜限于以该权利人作品为主要语料的定向训练，避免保留声明被普遍作出而使例外条款整体虚置。

5.3. 以来源披露与输出过滤义务承接灰色地带

判准两端之间的灰色地带，靠实体规则无法穷尽，需要程序性义务加以承接。蔡琳所构造的“语料来源的可追溯义务”是其中最为基础的一项^[11]。《生成式人工智能服务管理暂行办法》第7条已经要求使用具有合法来源的数据，第19条进一步规定主管部门可以要求提供者就训练数据来源、规模、类型、标注规则等作出说明，但这些要求指向行政监管，尚未转化为民事诉讼中的举证规则。可行的调适是将其接入侵权程序，权利人初步举证其作品可能被用于训练后，由服务提供者就语料构成与处理方式承担说明责任，说明不能的，推定该作品被用于训练。来源合法性还应当成为前述例外条款的适用前提，Bartz案中同一被告的两类行为得到相反评价，区别仅在于图书是购买而来还是自盗版站点下载而来，这一区分与训练行为的转换性无关，却直接决定了责任的有无，将其确立为例外适用的门槛，可以在不改变判准的前提下阻断以盗版语料训练的行为。输出端的过滤义务承担另一半功能，服务提供者应当对语料中的高重复样本作降重处理并保存记录，在指令端拦截明显指向复现特定作品的请求，在输出端对与在先作品构成实质性相似的内容加以识别与阻断，履行上述义务的，就偶发的相似输出不承担赔偿责任。两项义务共同作用，使定向训练与记忆化复现这两类最易产生替代的情形在事前获得防控，在事后获得举证支点，例外条款中的权利保留声明也因此获得可执行的技术依托。

6. 结语

生成式人工智能训练数据的合理使用之争，表面上争的是尺度宽严，实际上争的是依据何种标准划线。2025年前后美国、德国的几个判决在半年之内给出四种不同的评价路径，恰恰说明以技术阶段、利用目的或者主体身份划线，都是把著作权法之外的因素引入权利边界的判断，标准随产业格局与个案事

实浮动,规则的可预期性无从形成。回到著作权的权利构造,排他效力仅及于作品的表达性利益,划线的依据只能是使用行为是否在这一利益上对原作品形成替代。依此判断,训练阶段的非表达性利用本不在权利范围之内,输出端的实质性相似本不受例外庇护,真正需要制度安排的只是二者之间的狭窄地带。模型记忆与累积性稀释并不否定该判准,前者的评价位置在输出端,后者以能否归因于特定作品的表达为限。作品性使用要件解决非表达性利用的排除路径,实施条例中的例外条款解决灰色地带的正当性来源,来源披露与输出过滤义务解决替代风险的事前防控与事后证明。

需要说明的是,表达性替代判准所回答的只是著作权法内部的边界问题。创作者在数字经济中的价值分配、训练语料所凝结的智力劳动应否获得回馈,是产业政策与分配制度需要处理的议题,著作权法既缺少相应的工具,也不宜为承担该功能而扩张自身的权利范围。合理使用制度在训练数据领域的功能因此被限缩到一个更小的范围,但也正因为范围明确,它才可能从政策权衡的对象转变为可以反复适用的裁判规则。

参考文献

- [1] 张平. 人工智能生成内容著作权合法性的制度难题及其解决路径[J]. 法律科学(西北政法大学学报), 2024, 42(3): 18-31.
- [2] 焦和平. 人工智能创作中数据获取与利用的著作权风险及化解路径[J]. 当代法学, 2022, 36(4): 128-140.
- [3] 张新宝, 卞龙. 生成式人工智能训练语料的著作权保护[J]. 荆楚法学, 2024(5): 77-90.
- [4] 刘晓春. 生成式人工智能数据训练中的“非作品性使用”及其合法性证成[J]. 法学论坛, 2024, 39(3): 67-78.
- [5] 李安. 机器学习的版权规则: 历史启示与当代方案[J]. 环球法律评论, 2023, 45(6): 97-113.
- [6] 施小雪. 重塑复制权: 生成式人工智能数据训练的合法化路径[J]. 东方法学, 2024(6): 70-83.
- [7] 魏远山. 生成式人工智能训练数据的著作权法因应: 确需设置合理使用规则吗? [J]. 图书情报知识, 2025, 42(1): 78-88.
- [8] 熊琦, 张文竊. 人工智能生成内容著作权规制的全球趋向与本土路径[J]. 知识产权, 2024(11): 59-76.
- [9] 王迁. 再论人工智能生成的内容在著作权法中的定性[J]. 政法论坛, 2023, 41(4): 16-33.
- [10] 徐小奔. 论人工智能生成内容的著作权法平等保护[J]. 中国法学, 2024(1): 166-185.
- [11] 蔡琳. 生成式人工智能语料来源可追溯义务的法律构造[J]. 江苏社会科学, 2025(2): 161-169, 243-244.