

# 生成式人工智能模型训练在欧美版权法上的定性

□ 中国应用法学研究所互联网司法研究中心主任 宋建宝

生成式人工智能模型在开发之初，就像一个婴儿，不具备任何关于世界的先验知识。训练是生成式人工智能模型获得智能和创造力的唯一途径。训练实际上就是生成式人工智能模型的学习过程，从海量的数据中学习语言规律、视觉模式、概念联系等，并通过分析数据、调整内部参数，最终构建对世界的认知和理解。训练完成后，生成式人工智能模型能够根据用户的提示词，利用学到的模式和知识，创造出新颖、合理、多样化的内容。由此可见，生成式人工智能模型在训练过程中必然使用大量受版权保护的作品，而且这些作品往往是高质量、不可或缺的训练数据。

按照版权法的基本原理，未经著作权人许可，使用他人受版权保护作品原则上都构成侵权。同时，各国版权法对权利都会作出一些限制，规定某些未经许可使用他人作品的行为可以获得侵权的责任豁免。

未经权利人许可，将他人受版权保护作品用于生成式人工智能模型训练，这种行为是否构成版权侵权，是否能够获得版权法上的侵权豁免，是当前开发生成式人工智能模型过程中无法回避的一个现实问题。本文通过考察欧盟和美国的版权法及相应制度，分析生成式人工智能模型训练在欧美版权法上的法律定性以及是否可以获得版权侵权豁免，以期借鉴。

## 生成式人工智能模型训练构成初步版权侵权

### 模型训练准备工作至少侵害权利人的复制权

为了训练人工智能模型，人工智能模型开发者首先需要做好训练前的一些准备工作，其中重要的任务就是收集数据和整理数据。开发者在收集数据过程中，在大多数情况下，就是从公开可用的数据源（网站、数据库等）下载数据，但是无论来源如何，都必然复制数据，形成数据备份。随后，开发者将根据实际需求对数据进行整理。就受版权保护的作品来说，开发者通过以下方式将对作品反复进行复制：（1）下载作品；（2）在存储介质间传输作品；（3）将作品转换为特定格式；（4）创建或修改数据集；（5）经筛选后将作品纳入数据集。由此可见，开发者在训练人工智能模型之前的数据收集和整理都需要大量复制作品。因此，人工智能模型训练准备工作至少直接侵害权利人的复制权。

### 模型训练过程至少侵害权利人的复制权

模型训练过程也需要反复复制之前整理好的各种数据备份。首先，模型的训练速度和训练规模要求开发者在训练前下载数据集并复制到高性能存储设备中，以便提高

训练效率。其次，在训练期间，数据集将被分批“投喂”给人工智能模型，这就需要临时复制数据集。第三，训练过程中向模型提供示例所用的数据、衡量模型相对于预期输出的性能所用的数据以及迭代更新权重所用的数据都包含此前复制的数据。就受版权保护作品来说，这些过程都将复制作品或作品的实质部分。因此，人工智能模型训练过程也至少直接侵害权利人的复制权。

综上所述，生成式人工智能模型训练，无论是训练前的准备工作，还是模型训练过程本身，都必然大量复制受版权保护作品，都至少直接侵害权利人的复制权。但是，利用受版权保护作品训练人工智能模型，其使用作品的范围和规模是前所未有的，远远超出传统版权法的制度涵盖范围和立法者的预见能力范围。因此，问题的关键就在于，构成版权侵权的生成式人工智能模型训练能否满足版权法宽宥而获得侵权豁免。

## 欧盟版权法下模型训练版权侵权的责任豁免

生成式人工智能模型训练过程中使用受版权保护的作品，构成欧盟版权法上的复制行为并无争议，问题在于是否有任何特定的欧盟版权限制可以适用于生成式人工智能模型训练。目前在欧盟有关诉讼中，作为被告的模型开发者提出了可能适用于人工智能模型训练版权侵权豁免的三种版权限制，具体包括：（1）临时复制；（2）科研性文本与数据挖掘；（3）商业性文本与数据挖掘。

### 临时复制例外的适用探讨

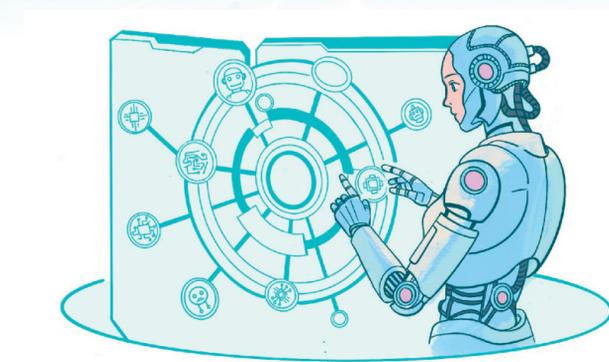
欧盟《信息社会指令》第5条是整个指令中最重要且最复杂的条款，也是欧盟版权法的核心条款，该条款定义了数字环境下版权保护的边界，通过规定一系列强制性和选择性的例外与限制，在保护创作者权利与保障信息自由流通之间寻求平衡。

第5条第1款规定了临时复制例外，这是《信息社会指令》中唯一的强制性例外，欧盟所有成员国都必须实施。临时复制例外允许短暂或附带的复制行为，此等复制行为是技术实施过程中不可或缺的，其唯一目的是使作品在网络中通过中间媒介在第三方之间进行传输，但是该复制行为须本身没有独立的经济意义。

生成式人工智能模型训练和训练过程中复制受版权保护作品，这些复制行为显然不是附带的或伴随发生的，先于训练中的复制行为是独立的，训练过程中的各种复制行为是在从事人工智能训练的个人或组织自主控制下反复复制作品的一个过程。因此，生成式人工智能模型训练使用受版权保护作品，不能满足临时复制例外的法律要求。

### 文本与数据挖掘例外的适用探讨

随着数字技术的发展应用，需要对各种数据进行计算或分析，这就是数据挖掘。数据挖掘是一种利用自动



武凡熙 作

化计算技术，从大规模、结构化或非结构化的数据集中发现新模式、新趋势、新关联、新知识的计算过程。数据挖掘的关键流程通常包括数据采集、数据清洗与预处理、数据转换与可视化。由此可见，数据挖掘多个步骤都涉及对原有数据的复制。具体到受版权保护作品来说，也同样需要进行复制，例如数据采集、数据清洗、数据预处理、数据转换等过程都需要复制作品。为解决数据挖掘技术与版权保护的冲突，欧盟《数字单一市场版权指令》将文本与数据挖掘作为一项版权保护例外，并针对科研性文本与数据挖掘和一般性文本与数据挖掘分别确立了不同的规则。

科研性文本与数据挖掘例外。《数字单一市场版权指令》第3条规定了科研性文本与数据挖掘，这是欧盟成员国必须通过国内法实施的一项强制性例外。科研性文本与数据挖掘例外适用主体包括大学、公立研究实验室等科研机构以及为科研机构提供协作的图书馆、博物馆等文化遗产管理机构。

科研性文本与数据挖掘例外可以适用的行为，顾名思义就是以科学研究为目的，对合法获取的作品或数据进行复制和提取的行为。同时，科研性文本与数据挖掘例外还需要满足一些法定条件，例如目的仅限“科学研究”、副本须保存在安全环境中、科研机构需要确保其仅用于科研验证、须通过合法途径获得、禁止规避技术措施等。因此，在欧盟境内，从事生成式人工智能训练的非商业性科研机构，如果因使用受版权保护作品而被指控侵犯版权的，可以援引科研性文本与数据挖掘例外作为侵权责任的豁免依据。当然，这些科研机构需要严格遵守科研性文本与数据挖掘例外的各项法定条件。

一般性文本与数据挖掘例外。《数字单一市场版权指令》第4条规定了一般性文本与数据挖掘例外，这也是欧盟成员国必须通过国内法实施的一项强制性例外。一般性文本与数据挖掘例外适用于任何主体，包括任何个人或组织。一般性文本与数据挖掘例外对行为目的不作限制，允许对合法获取的作品或数据的任何复制和提取。同样，一般性文本与数据挖掘例外也需要满足一些法定条件，具体

有：（1）必须通过合法途径获得数据来源；（2）权利人的权利保留条款优先，如果权利人通过机器可读方式明确保留文本与数据挖掘权利，则一般性文本与数据挖掘例外自动失效，需要另行获得额外授权；（3）副本仅能用于文本与数据挖掘分析，不可用于其他目的；（4）可以保存文本与数据挖掘副本直至分析所需期限届满，但需要证明其必要性。

在欧盟，对于那些不能适用科研性文本与数据挖掘例外的生成式人工智能模型训练，只能援引一般性文本与数据挖掘例外作为侵权责任的豁免依据。由于一般性文本与数据挖掘例外规定了比较严苛的种种限制，在实践中被告很难获得成功的抗辩。

为应对数字技术应用、保障数字经济健康发展，欧盟充分平衡技术创新与著作保护，专门创制出版权保护的文本与数据挖掘例外，确实能够为相关研究和技术创新提供一定的法律保障。但是，文本与数据挖掘例外中权利保留声明与合法获取这两大核心限制将显著制约文本与数据挖掘例外的实际效果。在权利保留声明制下，权利人通过在网站或数据中嵌入文本与数据挖掘保留声明，可以比较容易地排除文本与数据挖掘例外的适用。这势必严重削弱文本与数据挖掘例外的价值，尤其影响商业和创新型中小企业。合法获取作为适用条件之一，禁止规避技术保护措施，势必大大压缩文本与数据挖掘例外可以适用的空间。这些争议与挑战对于生成式人工智能模型训练也同样如此。未来的法律演进可能围绕文本与数据挖掘例外的权利保留声明的合理性、合法获取的必要性开展讨论。

## 美国版权法下模型训练版权侵权的责任豁免

美国版权法在其演进中，不断适应新技术发展应用，既推动技术进步，又保持对创作的激励，使得美国的创意产业一直处于全球领先地位。目前在有关诉讼中，权利人起诉被告的人工智能训练行为侵犯其版权时，被告往往以合理使用为由提出抗辩，请求获得版权侵权豁免。

合理使用制度是美国司法实践中

为了平衡权利人利益与社会公众利益而发展出来的一套裁判规则体系，后被美国1976年《著作权法》吸收而规定于第107条。合理使用制度采取了要素叠加方式，具体包括：（1）使用行为的目的和特点；（2）作品的性质；（3）作品被使用的数量和实质性；（4）对作品的潜在市场或价值的影响。这就是通常所说的合理使用制度“四要素”。

### 使用行为的目的和特点

使用作品的目的和特点，是判断使用行为是否构成合理使用的一个重要因素。在司法实践中，法院通常强调转换性使用和商业性这两个核心要件。

在判断转换性使用时，核心问题在于被告使用作品是单纯取代原告作品，还是通过赋予原告作品新的目的或不同特性，并以全新表达、意义或信息改变原告作品，增添新价值。因此，转换性使用通常在市场上不会替代原告作品，反而更可能实现版权法的立法宗旨。同时，转换性程度越高，商业性等不利因素在合理使用认定中的权重就越低。

商业性使用审查的核心在于，如果被告使用原告作品获取经济利益却规避付费，则可能构成不公。被告使用原告作品进行人工智能模型训练是商业性的还是非商业性的，关键是要审查在没有支付通常应当支付许可费的情况下被告是否从使用作品行为中有所获益。有的判例在考虑被告使用行为的目的和特点时，还会考虑被告是否合法获取原告作品。利用盗版或非法获取材料进行模型训练的行为性质比较严重，即使被告出于合理使用目的也不例外，因此非法获取行为本身即影响使用性质认定。

### 受版权保护作品的性质

这里所谓作品的性质，主要是指作品的原创性。作品的原创性高则保护范围大，作品的原创性低则保护范围小。作品保护范围的大小与认定合理使用尺度的宽严应当保持对应关系。使用保护范围越大的作品构成著作侵权的可能性越大，如果从严适用合理使用制度，那么绝大多数情形都将不构成合理使用。这将造成保护范围大的作品几乎不适用合理使用制度的事实。使用保护范围越小的作品构成著作侵权的可能性越小，如果从宽适用合理使用制度，那么绝大多数情形都将构成合理使用。这将造成保护范围小的作品几乎不受著作保护的事实。人工智能模型训练通常混合使用表达性作品与功能性作品、已公开与未公开作品，当所涉作品表达性越强或属于未公开作品时，受版权保护作品的性质这个要素将显著不利于合理使用认定。

### 作品被使用的数量和实质性

从数量角度来说，被使用部分在受保护整体作品中占比越大，不构成合理使用的可能性越大。从实

质性角度来说，被使用部分对于整体作品的实质性越重要，不构成合理使用的可能性越大。对于人工智能模型训练整个过程来说，有的步骤需要复制作品的全部内容，有的步骤可能仅需要复制作品的部分内容。既然存在复制作品全部内容的潜在市场或价值的影响，这就无须再具体细分哪些步骤复制了作品的全部内容，哪些步骤复制了作品的部分内容。

### 使用行为对受保护作品的潜在市场或价值的影响

美国联邦最高法院认为，第四个因素无可争议地构成合理使用制度中的首要因素，是判断使用行为是否构成合理使用的最重要的因素。认定使用行为是否构成合理使用，需要考虑使用行为对受保护作品的潜在市场或价值的影响。在任何情况下，最重要的就是判断使用行为与权利人就其作品所拥有的专有市场是否构成竞争关系，重点评估使用受版权保护作品训练人工智能模型对原告作品的市场影响，包括直接销售损失、市场稀释效应及许可机会流失等。虽然版权人需要承担初步举证责任以指明相关市场范围，但在应对合理使用抗辩时无须自行提供经验数据。

判定合理使用需要综合考量相关情况，平衡多项法定因素。人工智能模型训练过程中对受版权保护作品的各种使用方式可能是转换性的，这主要取决于使用了哪些作品、来源如何、目的为何以及对输出施加了何种控制，所有这些都会影响版权人的市场。当人工智能模型被部署用于分析或研究等目的时，其输出内容一般不会替代训练中所使用的作品。

2025年6月，美国加州北区联邦地区法院就Anthropic案作出一审判决，这是美国法院首次使用受版权保护作品训练人工智能模型是否构成合理使用作出判决。该判决将Anthropic的人工智能模型训练划分为数据获取、数据存储、数据训练、逐一分析、各自定性，最终得出Anthropic利用合法取得的受版权保护作品训练其人工智能模型构成合理使用，主要原因就在于使用受版权保护作品训练人工智能模型具备高度转换性，并且没有直接替代原作品的市场。当然，这只是美国一家联邦初级法院在个案案件中的裁判观点，后续有待更高级别的法院就这个问题形成具有法律约束力的判例。



# 澳门家事调解制度的新发展

□ 杨银 杨刚

澳门特别行政区（以下简称澳门）为构建非对抗性多元化的家事争议解决机制，于2025年6月11日审议通过的《家事案件调解制度》，将于2026年1月1日起正式实施，这标志着澳门家事调解步入常态化、规范化和制度化的新阶段。

## 调解制度的规范化演进

澳门初级法院设有专门的家事法庭集中处理家事案件，法院查明案件事实后依法作出裁判，该裁判结果具有权威性、确定性和可执行性。但是，家事案件采取对抗性的诉讼程序不利于缓和家庭矛盾，澳门倡导法官在家事案件诉讼程序中召集当事人举行会议试行调解，形成“先调后审”的惯例。

随着家事案件的多样化和复杂化，澳门特别行政区立法会以《民事诉讼法典》的调解程序为基础，在《家事案件调解制度》中引入调解制度，意在诉讼前化解家事纠纷，实现案件分流，以缓解讼累。《家事案件调解制度》不仅借鉴大陆法系的家事调解经验，还结合澳门本土的传统家事文化及司法实

践，确立社会工作局作为家事调解的主管机构，明确调解范围、调解程序、调解员选任与回避等内容，相较于《民事诉讼法典》的调解条款，新的调解制度更具有强制性、专业性和灵活性。

## 调解制度的范围限定与强制适用

家事案件兼具人身属性和财产属性，涉及婚姻、亲子、抚养、赡养和财产分割等法律关系。澳门法院每年审理的千余宗家事案件中大部分属于诉讼离婚，以及离婚中需要协议的亲权行使、抚养给付和分配家庭居所事宜。鉴于此，《家事案件调解制度》采取类型化列举的立法模式，将调解范围限定为四类：1.诉讼离婚；2.亲权的行使；3.抚养的给付；4.分配家庭居所。该家事调解制度通过类型化列举的立法模式划定调解范围，避免因概括性条款引发适用范围争议，体现立法者审慎把控家事调解制度的适用边界。

《家事案件调解制度》规定，当事人就上述四类家事案件向法院提起诉讼或非诉讼程序之前，必须先向社会工作局申请调解，否则法院将驳回其起诉或申请。此外，为缓解强制调

解程序与诉讼时效的冲突，《家事案件调解制度》规定，自当事人向社会工作局申请调解之日起至该局发出调解证明书之日止，上述四类家事案件的除斥期间和时效期间中止计算，法院采取保全措施后要求提起诉讼或非诉讼程序的期间也中止计算，为家事案件实现调解优先提供制度保障。

## 调解程序的分层设计

为保障调解工作的有序开展，《家事案件调解制度》从案件流程的管理角度，形成四阶段分层调解体制。

当事人提出申请。当事人向社会工作局提出家事调解申请时，须载明争议类型和争议说明等重要内容。其中，申请人身份证明、婚姻登记证明、未成年人出生证明等文件，由社会工作局依托《个人资料保护法》构建的政务资料共享机制直接调取，减轻了当事人提交资料的程序负担。同时，社会工作局收到家事调解申请后，应于6个工作日内指派一名家事调解员，由调解员采取合理方式通知双方当事人调解会议的时间和地点。若无法通过指定的方式通知申请人，视为放弃申请，但不影响重新提交调

解申请；若被申请人身份不明、下落不明、拒绝在收信回执上签名或拒绝接收信件，则调解程序结束并由社会工作局向申请人出具调解证明。

召开调解会议。该阶段的程序设计具有三大亮点：一是调解会议的灵活性。《家事案件调解制度》规定，首次调解会议应于指定调解员后20日内举行，整个调解程序须在60日内完成，若调解员有正当理由且经双方当事人同意，可以最多延长首次调解会议20日和整个调解程序30日。此外，调解会议的灵活性还体现在调解员有权自主决定调解程序的方法与步骤，但是不得向当事人强加任何协议或对调解结果作出任何承诺。二是参与主体的扩展性。调解员可以邀请双方当事人的血亲、姻亲或其他适宜人员参加调解会议，契合传统的家族协商文化。三是缺席处理的差异化。如果当事人有正当理由由缺席调解会议或者有理由相信其可能达成和解协议，该调解制度允许调解员变更调解会议的时间或地点；否则，调解程序即行结束，且该方将在后续诉讼程序中承担更高比例的诉讼费用。

调解结束。当出现以下情形时，

调解员应结束调解程序：当事人达成和解协议；当事人无正当理由缺席调解会议；被申请人存在《家事案件调解制度》第六条第四款规定的情形；其他不宜继续调解的情形。家事调解程序结束后，若双方当事人就部分或全部争议事项达成和解，调解员须拟定和解协议，经当事人共同签署后提交社会工作局；若未达成和解协议，调解员须制作调解报告并提交给社会工作局。该调解报告不能在后续诉讼、仲裁或行政程序中作为证据使用，亦不能作为司法上诉或行政申诉之标的。

司法衔接。社会工作局收到调解报告或和解协议后，应向申请人出具调解证明书。若当事人对《家事案件调解制度》划定的四类家事纠纷提起诉讼或非诉讼程序无法提供调解证明书时，法院将驳回其起诉或申请。同时，当事人在法官召集的首次会议结束前未对和解协议提出异议，法官经审查和解协议标的、行为资格以及协议内容是否违反强制性规定和公序良俗，审查后将和解协议载入会议记录。

## 调解员的选任与规制

《家事案件调解制度》仅规定两类主体具有家事调解员资格：一是社会工作局从事社会工作的公共行政人员，二是经《社会工作专业资格制度》专业资格认可的社会工作者。

同时，《家事案件调解制度》对调解员的行为进行了规制，以充分保障调解程序及调解结果的公正性。一是回避制度，若调解员与案件当事人之间存在亲属关系、债权债务关系、雇佣关系以及其他影响调解公正性的情形时，调解员应主动回避或者由当事人声明拒绝参与有关家事调解程序。二是执业限制，禁止调解员在《家事案件调解制度》所规定的四类家事案件中担任代理人、鉴定人、社会报告制作者或其他任何身份协助任一当事人。三是保密义务，除非存在法定特殊情形外，家事调解程序的参与者应对调解过程中获得的信息承担保密义务。

（作者单位：澳门科技大学法学院；重庆市酉阳土家族苗族自治县人民法院）