Open Journal of Legal Science 法学, 2025, 13(7), 1432-1438
Published Online July 2025 in Hans. https://www.hanspub.org/journal/ojls
https://doi.org/10.12677/ojls.2025.137201

Hans汉斯

算法黑箱下污染预测责任主体与追溯困境研究

秦宝静

青岛科技大学法学院，山东 青岛

收稿日期：2025年5月23日；录用日期：2025年6月20日；发布日期：2025年7月21日

摘 要

本文聚焦算法黑箱导致的污染预测责任主体界定与损害追溯难题，从技术、法律与实践三维度展开分析。技术层面，深度学习模型的决策逻辑不可解释性与环境数据的动态复杂性，导致预测偏差的具体成因难以追溯；法律层面，现行环境法律对算法开发者、数据提供者、模型使用者等多元主体的责任划分模糊，传统因果关系证明规则难以适应多主体交织的责任链条；实践层面，污染损害的累积性、滞后性与跨区域特征，进一步加剧了责任追溯的现实困境。研究指出，破解上述难题需构建技术透明化、法律框架完善与协同治理相结合的多元规制路径，通过可解释AI技术创新、算法审计制度、分级责任制度及跨学科监管机制的协同作用。

关键词

污染预测，算法黑箱，责任主体，追溯困境，环境法

Research on the Subject of Pollution Prediction Liability and the Dilemma of Traceability under the Algorithm Black Box

Baojing Qin

College of Law, Qingdao University of Science and Technology, Qingdao Shandong

Received: May 23rd, 2025; accepted: Jun. 20th, 2025; published: Jul. 21st, 2025

Abstract

This paper delves into the challenges of identifying pollution prediction liability subjects and tracing damages in algorithmic black boxes and analyzes across technology, law, and practice. Technologically, deep learning models' non-interpretable decision-making and environmental data's

文章引用: 秦宝静. 算法黑箱下污染预测责任主体与追溯困境研究[J]. 法学, 2025, 13(7): 1432-1438.
DOI: 10.12677/ojls.2025.137201

dynamic complexity obscure prediction bias causes. Legally, current environmental laws lack clear liability demarcations for algorithm developers, data providers, and model users; traditional causality rules are ill-suited to multi-subject liability chains. Practically, pollution damages' accumulative, lagging, and cross-regional traits compound liability-tracing difficulties. The study proposes a multi-regulatory approach integrating technological transparency, legal enhancements, and collaborative governance, via explainable AI, algorithm audits, hierarchical liability, and interdisciplinary oversight.

Keywords

Pollution Prediction, Algorithm Black Box, Responsible Subject, Dilemma of Retrospect, The Environmental Law

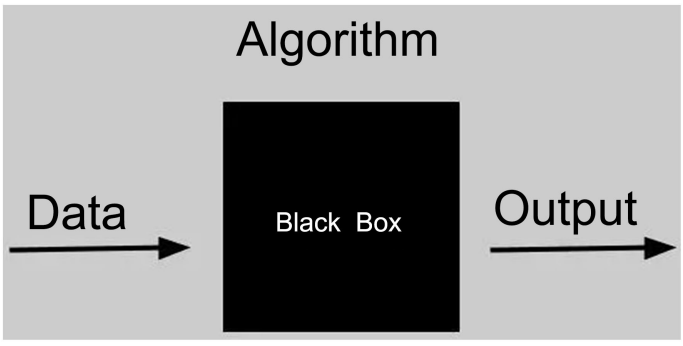
1. 引言

以深度学习为代表的算法技术因其内部运行逻辑的高度不透明性，导致当污染预测失误引发环境损害时，难以准确界定算法开发者、数据提供者、政府监管部门等主体的责任，传统环境法律框架在技术复杂性面前面临适配性危机。现行法律对算法决策的责任归属、因果关系证明等问题缺乏明确规制，使得污染损害追溯陷入技术不可解与法律无依据的双重困境[1]。

2. 算法黑箱与污染预测理论

2.1. 算法黑箱

算法黑箱(见图 1)是指人工智能系统在数据输入与决策输出之间的处理过程呈现高度不透明性，其内部运行逻辑难以被人类直接理解或解释的技术现象[2]。这类系统通常依赖复杂的数学模型与计算架构，尤其是深度学习、神经网络等机器学习算法，其决策过程涉及海量数据的非线性变换与多层级参数调优，导致即使是设计者也难以完全追溯具体决策的形成路径。算法黑箱的核心特征在于输入输出的可观测性与中间过程的不可解释性并存，这种技术特性在提升系统预测精度的同时，也引发了决策逻辑不透明带来的责任界定难题。



图片来源：作者自绘。

Figure 1. Algorithm Black Box
图 1. 算法黑箱

在污染预测领域，算法黑箱表现为模型对污染物浓度、扩散路径等环境要素的预测结果可被获取，但模型如何通过历史数据训练形成特定预测结论的具体机制却难以被清晰解构，进而导致环境损害发生时责任主体的追溯障碍。

2.2. 污染预测算法

污染预测算法是基于环境科学与数据科学的交叉融合，通过对大气、水、土壤等污染物相关数据的分析处理，实现对污染状态或趋势预测的数学模型集合。其核心在于构建污染物浓度与影响因子之间的映射关系，常见的模型包括多元线性回归、时间序列分析、机器学习算法等。以多元线性回归模型为例，其数学表达式为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon$$

其中 y 代表污染物浓度，x_1，x_2，…，x_n 为影响污染物的自变量如风速、湿度、工业排放量等，β_0，β_1，…，β_n 为模型参数，ε 为误差项。时间序列模型中的 ARIMA 自回归积分滑动平均模型则通过处理数据的时间依赖性进行预测，其公式可表示为：

$$\left(1-\varphi_1 B-\cdots-\varphi_p B^p\right)(1-B)^d y_t = \left(1-\theta_1 B+\cdots+\varphi_q B^q\right)\varepsilon_t$$

其中 B 为滞后算子，p 为自回归阶数，d 为差分阶数，q 为滑动平均阶数，ε_t 为白噪声序列。当采用深度学习算法如循环神经网络 RNN 或卷积神经网络 CNN 时，污染预测模型通过多层神经元的加权连接与激活函数变换捕捉数据中的复杂模式，其输出层的计算形式为

$$y = \sigma\left(W_h h_t + b\right)$$

其中 W_h 为权重矩阵，h_t 为隐藏层状态，b 为偏置项，σ 为激活函数。

3. 环境法与数字技术融合

3.1. 大数据环境风险预警

大数据技术整合大气监测、水质传感、卫星遥感等多源异构数据实现对污染事件的早期识别与趋势预测。

物联网设备与传感器网络持续采集污染物浓度、气象参数、工业排放等海量数据，经清洗、标准化处理后输入预警系统；在分析阶段，机器学习算法通过挖掘数据间的关联关系，识别污染形成的关键驱动因素，将风速、降水量等自然变量与工业产值、能源消耗等社会经济指标结合，构建污染扩散预测模型[3] [4]。

大数据环境风险预警存在显著的应用边界：环境数据的时空异质性与动态变化特征对数据采集的完整性和准确性提出极高要求，部分偏远地区或复杂地形区域的监测数据缺失导致模型预测偏差；数据共享机制的不完善使得政府、企业、科研机构之间存在信息孤岛。当算法依赖的训练数据存在偏差或噪声时，导致风险预警结果失真，而算法黑箱特性使得这种失真的具体成因难以被有效追溯。

3.2. 现行法律框架算法应用规制

我国现行环境法律体系对数字技术在污染预测中的应用尚未形成系统性规制框架，相关规定散见于《中华人民共和国环境保护法》《中华人民共和国大气污染防治法》等法律法规中。

《中华人民共和国环境保护法》要求重点排污单位如实公开环境信息，为数据采集与模型构建提供了基础数据源，但未明确数据应用于算法训练时的权利义务边界；《中华人民共和国大气污染防治法》

规定了政府部门在大气环境质量预测预警中的职责，却未对预测模型的技术标准、透明度要求及责任归属作出具体规定[5] [6]。

在责任认定层面，传统环境法律以谁污染谁治理为核心原则，强调行为与损害结果之间的直接因果关系，而算法决策的介入使得污染预测与损害发生之间的因果链条呈现多主体参与、多因素交织的复杂形态。现行法律对算法开发者、使用者、数据提供者等多元主体的责任划分缺乏明确界定，导致当污染预测失误或环境损害发生时，难以依据现有法律条款准确追溯责任主体[7]。

4. 算法黑箱下责任主体界定困难

4.1. 决策不透明

在污染预测场景中，深度学习模型通过多层神经网络自动提取数据特征并构建预测规则，其决策过程涉及海量参数的复杂运算，甚至伴随模型在训练过程中的自我优化与调整，形成人类认知难以穿透的决策黑箱。

当预测结果与实际环境损害存在偏差时，由于缺乏可解释性，无法明确是算法设计初始的逻辑缺陷、训练数据的偏差，还是模型运行过程中的参数异常导致的错误预测。在工业废气排放预测模型中，若因训练数据未能全面覆盖特殊气象条件下的污染物扩散规律，导致模型在极端天气时作出错误预警，此时难以判断责任应归咎于算法开发者对数据局限性的预见不足，还是数据提供者未能完整提供历史监测数据[8]。

4.2. 主体多元性

污染预测算法的运行涉及多个参与主体，包括算法开发者、数据提供者、模型使用者以及平台运营者等，各主体在不同环节发挥作用，形成复杂的责任链条。算法开发者负责模型架构设计与参数调优，数据提供者为模型训练提供基础环境数据，政府或企业等使用者基于预测结果制定环境管理决策，而平台运营者则提供算法运行的技术支撑环境。当污染预测出现偏差并导致环境损害时，各主体往往以理由主张免除自身责任，导致事实上呈现无真正责任承担者的现象，这不仅是对社会公共利益的损害和漠视，社会公众因环境污染事件造成严重物质损失和人身健康受到侵害而得不到相关责任者的现象也时有发生。

5. 污染损害追溯的现实困境与成因

5.1. 技术维度

污染损害追溯在技术层面面临算法决策逻辑不可解的核心障碍。人工智能污染预测模型，尤其是深度学习算法，依赖多层神经元网络的非线性变换处理环境数据，其内部参数调优与特征提取过程呈现高度自动化与自适应性，形成人类难以解读的复杂决策路径。

环境数据的多源性与动态性加剧了追溯难度，大气污染物扩散模型需整合气象、工业排放、交通流量等多维度数据，任一环节的数据质量缺陷或传输延迟均引发预测偏差。

5.2. 法律维度

现行法律体系在污染损害追溯中面临归责原则与因果关系证明的双重困境。传统环境法律以直接因果关系为归责基础，要求损害结果与污染行为之间存在明确可证的线性联系，而算法决策的介入使因果链条演变为多主体、多因素交织的网状结构。某化工企业依据污染预测算法调整减排策略后仍发生超标排放，法律上难以判断是算法开发者的模型误差、数据提供者的监测数据缺失，还是企业对预测结果的不合理应用导致损害发生。

5.3. 实践维度

污染损害的累积性与长期滞后性加剧了实践中的追溯困境。环境污染物如持久性有机污染物、重金属离子等，其危害效应通常需经过数年甚至数十年的积累才会显现，而在此期间污染预测算法经历多次迭代升级，数据采集设备也更新换代，导致损害发生时难以还原事发时的算法运行状态与数据输入环境[9]。跨区域、跨介质污染事件中，责任链条涉及多个行政区域的监管部门、不同行业的企业主体以及多元技术服务提供商，各主体在数据共享、责任认定中存在协作壁垒。实践中常出现地方政府以依据算法预测合规决策为由规避监管责任，企业以遵循技术建议为由推卸减排义务，损害追溯陷入多头管理、无人担责的治理困境[10]。

5.4. 责任主体模糊与追溯困境的互动机理

5.4.1. 技术黑箱下的双向制约

算法黑箱的决策不透明性既导致责任主体难以界定，又直接加剧追溯困境。一方面，深度学习模型的参数调优与特征提取过程缺乏可解释性，使得污染预测失误时，无法厘清算法开发者、数据提供者、使用者等主体的具体过错模型设计缺陷、数据偏差或决策滥用，形成主体责任真空；另一方面，责任主体的模糊状态进一步阻碍追溯路径，当损害发生时，各主体常以技术不可解释为由推诿责任，导致追溯证据链因主体间的责任切割而断裂。某化工园区废气预测模型因训练数据未覆盖特殊气象条件，导致极端天气下污染扩散预测偏差，此时开发者与数据提供者均可能以技术局限性逃避责任，而追溯者难以证明具体环节的过错归属。

5.4.2. 法律关系的交叉影响

现行法律对多元主体的责任划分模糊，与追溯困境形成恶性循环。传统环境法以直接因果关系为归责基础，但算法决策介入后，污染损害的因果链条演变为数据采集，模型训练，决策执行的多阶段复合结构。若法律未明确各阶段主体的义务边界数据提供者的完整性义务、开发者的风险告知义务，则追溯时易出现责任链条脱节。企业依据污染预测算法调整减排策略后仍发生超标排放，法律难以判断是算法误差、数据缺失还是企业执行不当所致，导致追溯因因果关系证明不足而失败。反之，追溯困境又会强化主体责任的模糊性，由于损害难以追溯至具体主体，各主体更倾向于规避风险，形成无主体担责的治理僵局。

5.4.3. 实践场景中的联动效应

污染损害的累积性、跨区域性与主体责任的复杂性相互叠加。在跨区域污染事件中，多个行政区域的监管部门、企业及技术服务主体均可能参与污染预测流程，但若某省环保部门使用的预测模型因邻省数据提供者未共享关键污染源信息而失效，导致污染扩散至下游区域，此时责任主体涉及跨省数据共享方、模型开发者与本地监管者，追溯需协调多方法律关系，而各地法律适用标准的差异进一步加剧追溯难度。实践中这种主体多元性，追溯复杂性的联动效应，常导致污染损害长期无法得到救济。

6. 算法黑箱下的责任规制路径

6.1. 技术透明化

构建可解释人工智能与全流程算法审计制度，推动可解释 AI 技术创新，要求污染预测算法在输出预测结果的同时，提供决策依据的可视化分析，如关键数据特征的权重排序、异常预测结果的归因说明，确保技术专家与法律主体能够理解算法逻辑。建立算法审计制度，由第三方机构对污染预测模型的训练数据质量、架构设计合理性、参数调整过程进行定期审查，形成可追溯的技术文档记录。针对大气污染

预测模型，审计机构需验证训练数据是否涵盖当地主要污染源特征、极端气象条件下的参数适配机制是否完备，并将审计结果作为责任追溯的技术依据。

6.2. 法律框架完善

算法开发者需对模型设计缺陷与技术验证不足承担过错责任，数据提供者需对因数据偏差或缺失导致的预测失误负责，模型使用者需对不合理依赖算法或决策执行不当承担后果责任。

引入《民法典》过错推定的归责原则，在污染损害发生时，推定技术优势方存在过错，除非其能证明已履行充分的技术审查与风险告知义务。在责任承担框架之内融入过错推定原则，不仅能够使得相关主体于环境污染损害结果发生之后承担环境损害侵权责任，填平和赔偿对环境公共利益和私人利益造成的物质损害和精神损害，更重要的是在算法黑箱开发者和使用者地位、信息不对等的情况下，合理加重算法黑箱开发者的污染预测责任，有助于从源头减轻环境污染损害的发生率，并明确具体的为司法机关调查取证环境污染事件提供涉案数据。

当污染预测模型连续出现重大偏差时，司法机关和社会公众可要求开发者提供模型训练日志、数据清洗记录等技术文件，保证社会公众对环境污染事件的知情权，提高司法实践效率并节省司法机关的经济成本和人力成本，若无法提供则推定其存在过失并承担相应责任。

在《环境保护法》《大气污染防治法》等法律法规的基础之上细化相关法规规章，就数字技术应用于环境法出台司法解释或者地方性法规，明确人工智能污染预测系统的法律定位、风险评估程序与责任分担机制，合理利用《民法典》规定的责任认定原则，填补现有法律对算法决策责任的空白。

6.3. 协同治理机制

应对责任追溯困境需构建跨学科专家参与的动态监管体系。建立由环境科学家、计算机工程师、法律专家组成的多元共治平台，在污染预测算法的立项、应用、评估各环节提供跨领域专业支持。

技术层面利用区块链技术实现数据采集、算法运行、决策执行的全流程证据保全，确保损害发生时责任链条可追溯；法律层面建立环境公益诉讼与生态损害赔偿的快速响应通道，允许环保组织、检察机关基于技术存证数据提起责任认定诉讼。此外，强化企业与公共部门的协同义务，要求算法使用者定期公开预测模型的关键技术参数与决策效果评估报告，接受社会监督；通过真实案件检验预测模型预测结果的真实性和有用性，有序完善更新预测模型系统使之与现有的环境污染水平动态匹配。

7. 结论

算法黑箱下的污染预测责任主体界定与追溯困境，本质上是数字技术的复杂性与法律确定性需求之间的深层对话体现。本文通过对技术特性、法律框架与实践场景的交叉分析，揭示了算法决策不透明、主体多元性、损害追溯技术壁垒等核心问题，并从技术透明化、法律责任重构与协同治理机制三方面提出规制路径。核心目的是在保障算法创新活力与维护环境法秩序之间寻求平衡，为生成式人工智能时代的生态环境治理提供同时包含技术可行性与法律正当性的解决方案。这一法律规制框架不仅有助于破解算法黑箱引发的责任困境，更旨在构建适应数字经济特征的新型环境法秩序，推动环境污染治理从事后追责向风险预防的法律范式转型。

参考文献

[1] 孙海波, 谢辉, 陈嘉, 等. 污染土地再开发中的环境风险与责任[J]. 上海国土资源, 2017, 38(1): 79-82, 86.

[2] 张继平, 潘易晨, 孔凡宏, 等. 政治晋升激励视角下我国海洋陆源污染治理的研究[J]. 中国海洋大学学报: 社会科学版, 2017(4): 20-26.

[3] 杨枫, 许伟宁, 陈灿林, 陈海燕, 刘东萍, 于会彬, 宋永会. 珠江口鸡啼门水道汛期污染溯源: 基于水质变化及DOM 特征分析[J]. 环境科学研究, 2024, 37(12): 2687-2697.

[4] 许雄飞, 李瑶, 刘桢, 谭菊. 浅析监测数据在环境应急事件污染源追溯中的应用[J]. 环境与可持续发展, 2018, 43(1): 86-88.

[5] 何圣华. 基于 Google 地图 API 的水质模拟与污染源追溯系统的设计与开发[J]. 电脑与信息技术, 2017, 25(1): 11-13.

[6] 孟祥琪. 滇池水域重金属污染的历史追溯及生态风险评价[D]: [硕士学位论文]. 昆明: 昆明理工大学, 2016.

[7] 孙倩, 李妍, 李明月, 等. 以典型新兴污染物追溯自然水体的污水来源[C]//中国化学会. 第 30 届学术年会摘要集, 第二十六分会: 环境化学. 大连: 中国化学会, 2016.

[8] 王淼. 从鱼生态追溯水源污染[J]. 防灾博览, 2024(5): 58-61.

[9] 陈振飞. 水体污染源追溯与水环境治理研究[J]. 清洗世界, 2024, 40(3): 142-144.

[10] 黄尚辉. 监测数据在环境应急事件污染源追溯中的应用[J]. 皮革制作与环保科技, 2022, 3(3): 67-69.