ICS 01.040.13
CCS Z 00

T

团　　体　　标　　准

T/CSES 179—2024

生态环境领域人工智能算法评估方法

Evaluation of artificial intelligence algorithms in the field of ecological environment

2024-12-25 发布　　2024-12-25 实施

中国环境科学学会　发布
中国标准出版社　出版

目　　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由生态环境部信息中心提出。

本文件由中国环境科学学会归口。

本文件起草单位：生态环境部信息中心、浙大启真未来城市科技（杭州）有限公司、天津市生态环境科学研究院、联通（浙江）产业互联网有限公司、广东省环境科学研究院、生态环境部环境规划院、浙江工业大学、浙江大学滨江研究院。

本文件主要起草人：陈晋音、张波、苏蒙蒙、范亚云、吴犇、毛永坚、李燕捷、王小涵、王立群、严金英、钟海林、蒋洪强、章俊岫、潘晓华、叶佩思、朱景熹。

引　言

中共中央、国务院印发的《数字中国建设整体布局规划》提出“建设绿色智慧的数字生态文明”。《中共中央 国务院关于全面推进美丽中国建设的意见》指出“深化人工智能等数字技术应用，构建美丽中国数字化治理体系，建设绿色智慧的数字生态文明”。为加快推进人工智能在生态环境领域的应用，保障算法的质量，促进公平竞争，加速创新和技术进步，亟需建立评估方法，为各部门、地区开展生态环境领域人工智能算法评估提供参考依据。

生态环境领域人工智能算法评估方法

1 范围

本文件规范了生态环境领域人工智能算法的评估体系、评估流程及方法。

本文件适用于指导生态环境领域人工智能算法开发方、用户方以及相关组织对生态环境领域人工智能算法开展的评估工作。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注明日期的引用文件，仅该日期对应的版本适用于本文件；不注明日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 20986 信息安全技术 网络安全事件分类分级指南

GB/T 41867 信息技术 人工智能 术语

GB/T 42888 信息安全技术 机器学习算法安全评估规范

T/CESA 1036 信息技术 人工智能 机器学习模型及系统的质量要素和测试方法

3 术语和定义

下列术语和定义适用于本文件。

3.1

人工智能 artificial intelligence

人工智能系统相关机制和应用的研究和开发。

[来源：GB/T 41867—2022，3.1.2]

3.2

对抗攻击 adversarial attack

通过在正常样本上添加难以察觉的微小扰动误导人工智能算法的攻击方法。

[来源：JR/T 0221—2021，3.8，有修改]

3.3

物理对抗攻击 physical adversarial attack

通过物理手段在真实世界构建对抗样本攻击人工智能算法的方法。

[来源：JR/T 0221，3.9]

3.4

模型后门攻击 model backdoor attack

在训练模型时植入后门，当触发后门条件时，模型会产生不正常的输出。

3.5

可解释性 explainability

系统以人能理解的方式，表达影响其（执行）结果的重要因素的能力。

注：可解释性理解为对“原因”的表述，而不是尝试以“实现必要的优势特性”做出争辩。

[来源:GB/T 41867—2022,3.4.3]

3.6

可靠性 reliability

实施一致的期望行为并获得结果的性质。

[来源:GB/T 41867—2022,3.4.4]

3.7

可控性 controllability

系统被人类或其他外部主体干预的性质。

[来源:GB/T 41867—2022,3.4.5]

3.8

鲁棒性 robustness

评估对象在任何情况下都保持其性能水平的特性。

[来源:GB/T 41867—2022,3.4.9,有修改]

3.9

测试元 test element

用于评估算法的具体指标,具有明确的测试目标和可量化的结果。

4 缩略语

下列缩略语适用于本文件。

API:应用程序接口(Application Programming Interface)

ASR:攻击成功率(Attack Success Ratio)

GPU:图形处理器(Graphics Processing Unit)

MAE:平均绝对误差(Mean Absolute Error)

MAPE:平均绝对百分比误差(Mean Absolute Percentage Error)

MSE:均方误差(Mean Square Error)

R^2:决定系数(Coefficient of Determination)

5 评估体系

5.1 概述

人工智能算法效果可能受到内外部因素影响。本文件按照 GB/T 42888 和 T/CESA 1036,结合生态环境领域应用场景,制定了一套生态环境领域人工智能算法评估体系,包含 5 个一级指标、11 个二级指标(见表 1)。

表 1 指标体系

一级指标	二级指标	重点要素
算法性能	精准性	回归任务精准性、分类任务精准性
	效能	执行效率、数据处理能力
可解释性	可解释性	模型复杂度、解释性能力、解释性质量
可控性	可控性	系统稳定性、参数可调性、运行状态的实时监控

表 1　指标体系（续）

一级指标	二级指标	重点要素
安全性	可靠性	中毒性攻击抵御可靠性、对抗性攻击抵御可靠性、物理对抗攻击抵御可靠性
	计算环境的鲁棒性	智能算法供应链的鲁棒性、分布式计算的鲁棒性、计算框架的鲁棒性
	保密性	数据敏感性与保密性、模型保密性、依赖信息保密性
维护性	兼容性	算法对数据格式的兼容性、算法对操作系统的兼容性、算法对其他软件的兼容性
	可维护性	算法迭代的更新频率、算法迭代的质量变化
	可移植性	算法对硬件设备的可移植性、算法对人工智能框架的可移植性
	可扩展性	算法水平扩展能力、算法垂直扩展能力

5.2　算法性能

5.2.1　精准性

5.2.1.1　回归任务精准性

评估算法在回归问题上的预测精确程度，包括但不限于以下测试元：

a）　均方误差（MSE）；

b）　决定系数（R^2）；

c）　平均绝对误差（MAE）；

d）　平均绝对百分比误差（MAPE）。

5.2.1.2　分类任务精准性

评估算法在分类问题上的分类准确性，包括但不限于以下测试元：

a）　准确率；

b）　精确率；

c）　召回率；

d）　F1 分数。

5.2.2　效能

5.2.2.1　执行效率

评估算法的运行效率与资源使用情况，包括但不限于以下测试元：

a）　执行速度；

b）　资源利用率。

5.2.2.2　数据处理能力

评估算法应对大量数据或复杂任务的能力，包括但不限于以下测试元：

a）　吞吐量；

b）　并行处理能力；

c）　负载处理能力。

5.3 可解释性

5.3.1 模型复杂度

解释算法的复杂性和计算资源需求能力，包括但不限于以下测试元：

a） 模型参数数量；

b） 模型结构复杂度。

5.3.2 解释性能力

解释算法的输出预测结果，包括但不限于以下测试元：

a） 可视化效果评估；

b） 特征重要性评估。

5.3.3 解释性质量

评估算法对用户的交互性，包括但不限于以下测试元：

a） 解释的准确性；

b） 解释的完整性；

c） 解释的一致性。

5.4 可控性

5.4.1 系统稳定性

评估算法在面临操作环境变化时的稳定性表现，包括但不限于以下测试元：

a） 平均错误率；

b） 最大错误率；

c） 运行期间崩溃次数；

d） 崩溃后恢复时间。

5.4.2 参数可调性

评估算法应对不同参数设置变化的能力，包括但不限于以下测试元：

a） 参数调整范围；

b） 参数调整对性能的影响。

5.4.3 运行状态的实时监控

评估算法在运行过程中对信息资源的监控能力，包括但不限于以下测试元：

a） 状态信息更新频率；

b） 故障预警机制的有效性；

c） 资源利用率的优化程度；

d） 资源调度算法的合理性。

5.5 安全性

5.5.1 可靠性

5.5.1.1 中毒性攻击抵御可靠性

评估算法模型应对中毒攻击的可靠性，包括但不限于以下测试元：

a） 数据投毒攻击抵御能力；

b） 模型后门攻击抵御能力。

5.5.1.2 对抗性攻击抵御可靠性

评估算法模型应对对抗性攻击的可靠性，包括但不限于以下测试元：

a） 白盒对抗攻击抵御能力；

b） 灰盒对抗攻击抵御能力；

c） 黑盒对抗攻击抵御能力。

5.5.1.3 物理对抗攻击抵御可靠性

评估算法模型在面对物理对抗攻击时保持其性能和准确性的能力，包括但不限于以下测试元：

a） 有目标物理对抗攻击抵御能力；

b） 无目标物理对抗攻击抵御能力。

5.5.2 计算环境的鲁棒性

5.5.2.1 智能算法供应链的鲁棒性

评估算法模型应对各种不确定因素的能力，包括但不限于以下测试元：

a） 供应链完整性；

b） 组件来源可信性；

c） 供应链安全性。

5.5.2.2 分布式计算的鲁棒性

评估算法模型面对各种异常情况时，分布式算法保持稳定运行的能力，包括但不限于以下测试元：

a） 数据一致性能力；

b） 分布式结构安全性。

5.5.2.3 计算框架的鲁棒性

评估在各种异常情况下，计算框架保持稳定运行和正确执行任务的能力，包括但不限于以下测试元：

a） 算子安全性；

b） 框架库安全性；

c） API 安全性；

d） 编译器安全性。

5.5.3 保密性

5.5.3.1 数据敏感性与保密性

评估算法在处理敏感数据时的安全性与隐私保护能力，包括但不限于以下测试元：

a） 数据加密措施；

b） 数据访问控制；

c） 数据存储安全性。

5.5.3.2 模型保密性

评估算法在保护模型本身及其生成的数据时的安全性与隐私保护能力，包括但不限于以下测试元：

a） 模型参数保密性；

b） 模型文件加密；

c） 模型访问权限控制。

5.5.3.3 依赖信息保密性

评估算法在管理其依赖项安全性和隐私保护能力，包括但不限于以下测试元：

a） 依赖库和框架的保密性；

b） 依赖信息的访问控制；

c） 依赖信息的完整性保护。

5.6 维护性

5.6.1 兼容性

5.6.1.1 算法对数据格式的兼容性

评估算法在处理数据时的效率和准确度，包括但不限于以下测试元：

a） 数据格式兼容性；

b） 不同数据格式处理能力。

5.6.1.2 算法对操作系统的兼容性

评估算法在多种操作系统上运行的能力，包括但不限于以下测试元：

a） 操作系统兼容数量；

b） 跨平台操作便捷性。

5.6.1.3 算法对其他软件的兼容性

评估算法与第三方软件共同工作时的表现，包括但不限于以下测试元：

a） 第三方软件兼容性；

b） 软件更新适应性；

c） 国产化芯片的适配兼容性；

d） 人工智能框架的适配兼容性。

5.6.2 可维护性

5.6.2.1 算法迭代的更新频率

评估算法的发展速度与适应变化的能力，包括但不限于以下测试元：

a） 迭代时间间隔能力；

b） 迭代代码变动量。

5.6.2.2 算法迭代的质量变化

评估算法在各次更新后的综合表现改善情况，包括但不限于以下测试元：

a） 迭代性能提升能力；

b） 迭代系统稳定能力。

5.6.3 可移植性

5.6.3.1 算法对硬件设备的可移植性

评估算法在不同硬件平台上运行时的适应性和效率，包括但不限于以下测试元：

a） 支持的硬件设备种类；

b） 跨硬件性能差异。

5.6.3.2 算法对人工智能框架的可移植性

评估算法在不同人工智能框架中运行的效率和适应性，包括但不限于以下测试元：

a） 支持的人工智能框架数；

b） 框架间的性能保持。

5.6.4 可扩展性

5.6.4.1 算法水平扩展能力

评估算法在通过增加更多的计算节点来提高系统的处理能力和容量，包括但不限于以下测试元：

a） 性能提升评估；

b） 无状态服务支持；

c） 自动识别与集成。

5.6.4.2 算法垂直扩展能力

评估算法在通过升级单个组件来提升性能的能力，包括但不限于以下测试元：

a） 硬件升级性能提升；

b） 最大算力限制。

6 评估流程

6.1 总则

生态环境领域人工智能算法评估流程见图 1。包括确定评估目标、制定评估方案、执行评估、汇总评估结论四个活动。算法评估实施案例见附录 A。

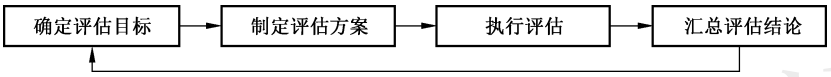

图 1　生态环境领域人工智能算法评估流程

6.2　确定评估目标

可运用以下步骤确定生态环境领域人工智能算法的评估目标。

a)　场景分析：分析生态环境领域人工智能算法的应用场景、运行环境与使用流程，既要考虑系统正常使用的情况，也要考虑可预见的异常情况。

b)　风险分析：通过算法场景分析、失效风险分析、历史运行数据分析、专家委员会评审等，调查评估算法失效可能产生的风险程度并确定风险等级。应按照 GB/T 20986 的要求，依据事件影响对象的重要程度与事件对业务损失、社会危害和潜在环境事件的严重程度来判定风险程度。风险程度对应的风险等级见表 2。

表 2　风险等级

危险严重性等级	描述
特别严重	风险事件发生在特别重要的事件影响对象上，并且： a)　导致特别严重的业务损失，或 b)　造成特别重大的社会危害，或 c)　导致特别重大突发环境事件
严重	风险事件发生在特别重要或重要的事件影响对象上，并且： a)　导致特别重要的事件影响对象遭受严重的业务损失或导致重要的事件影响对象遭受特别严重的业务损失，或 b)　造成重大的社会危害，或 c)　导致重大突发环境事件
较大	风险事件发生在特别重要或重要或一般的事件影响对象上，并且： a)　导致特别重要的事件影响对象遭受较大或较小的业务损失，或重要的事件影响对象遭受严重或较大的业务损失，或导致一般的事件影响对象遭受较大（含）以上级别的业务损失，或 b)　造成较大的社会危害，或 c)　导致较大突发环境事件
一般	风险事件发生在重要或一般的事件影响对象上，并且： a)　导致较小的业务损失，或 b)　造成一般的社会危害，或 c)　导致一般突发环境事件

c)　确定评估目标：根据算法失效的危险严重性等级，可建立人工智能算法的评估目标，见表 3。其中评估目标从高到低依次分为Ⅰ级、Ⅱ级、Ⅲ级、Ⅳ级四个级别。

表 3　生态环境领域算法评估目标等级划分

评估目标	目标等级描述
Ⅰ级	避免算法失效造成特别重大严重风险事件
Ⅱ级	避免算法失效造成严重风险事件
Ⅲ级	避免算法失效造成较大风险事件
Ⅳ级	避免算法失效造成一般风险事件

6.3　制定评估方案

根据项目实际情况选取测试元，确定测试元应达到的指标要求，并用专家打分法确定测试元权重，参考案例见表 A.2 和表 A.3。

6.4　执行评估

执行评估应按照评估方案对测试元逐一评估、形成分项评估结果、留存证明材料，包含以下内容：

a）　按照评估需求，开展评估；

b）　根据测试效果，对测试元进行分项评估，每项评估分值在 0～100 分，评估规则如表 4 所示；

c）　采用测试工具、经专家审核确定测试元得分，通过加权乘积之和逐级计算重点要素、二级指标、一级指标，确定评分结果；

d）　留存生态环境领域人工智能算法评估过程及结果的必要证明材料。

表 4　生态环境领域测试元评估等级划分

评估分值	描述
[0,45)	测试元明显未达到测试要求
[45,60)	测试元未达到测试要求
[60,80)	测试元符合测试基本要求
[80,90)	测试元优于测试要求
[90,100)	测试元明显优于测试要求

6.5　汇总评估结论

人工智能算法的所有测试元均通过评估，则算法通过评估并达到目标要求；否则未通过评估。

7　评估方法

7.1　算法性能

7.1.1　精准性

精准性涉及数据处理、分析及预测的准确性。在环境监测、污染防治等关键领域，通过细致的算法测试，评估算法在不同场景下的表现。生态环境领域算法的精准性评估方法见表 5。

表 5　算法的精准性评估方法

重点评估要素	评估方法
回归任务精准性	a）均方误差（MSE）：计算预测值与实际值之差的平方的平均值。MSE 值越小，表示预测值与实际值之间的差异越小，回归任务精准性越高。 b）决定系数（R^2）：通过比较模型预测值与实际值的相关性来评估模型的拟合优度。R^2 值越接近 1，表示模型拟合效果越好，回归任务精准性越高。 c）平均绝对误差（MAE）：计算预测值与实际值之间绝对差值的平均值。MAE 越低，模型的准确性越高。 d）平均绝对百分比误差（MAPE）：计算预测值与实际值之间差值的绝对值除以实际值的平均值，以百分比表示误差
分类任务精准性	a）准确率：计算正确分类的样本数占总样本数的比例。准确率值越大，表示算法正确分类的能力越强，分类任务精准性越高。 b）精确率：计算真正为正样本占实际为正样本的比例。精确率值越大，表示算法预测为正样本的样本中真正为正样本的比例越高，分类任务精准性越高。 c）召回率：衡量模型正确预测正样本全度的能力，即正样本被预测为正样本占总的正样本的比例。召回率值越大，表示算法能找出正样本的能力越强，分类任务精准性越高。 d）F1 分数：综合考虑精确率和召回率，计算它们的调和平均值。F1 分数越大，表示算法在精确率和召回率之间取得的平衡更强，分类任务精准性越高

7.1.2　效能

效能关乎算法的运行效率，以及算法在解决实际问题的效果和能力。生态环境领域算法的效能评估方法见表 6。

表 6　算法的效能评估方法

重点评估要素	评估方法
执行效率	a）执行速度：使用计时器记录算法从开始执行到任务完成所需的总时间。总时间越小，表示算法的执行速度越快，性能越好。 b）资源利用率：监控算法运行时的内存占用、智能加速卡使用率，并计算其平均值。内存占用、智能加速卡使用率越低，表示算法的资源利用率越高，对系统资源的消耗越少
数据处理能力	a）吞吐量：记录算法在单位时间内成功处理的数据量或任务数。吞吐量越高，表示算法的数据处理能力越好。 b）并行处理能力：模型算法同时执行多个指令或任务的能力。并行处理能力的提升比例越高，说明算法在并行计算方面的优化越好。 c）负载处理能力：在不同负载条件下测试算法的性能表现，如高负载、低负载和中等负载情况。算法在不同负载条件下都能保持稳定的性能表现，说明其负载处理能力强

7.2　可解释性

可解释性是评估算法可靠性和适用性的重要维度。模型复杂度、解释性能力和解释性质量是构成生态环境领域算法可解释性的三个核心指标，涉及算法模型参数量和结构，可视化和特征重要性评估，以及解释质量的准确性、完整性和一致性。在关键领域如环境监测和污染防治，应进行严格的算法测试，以确保算法在各种场景下的表现能力。生态环境领域算法的可解释性评估方法见表 7。

表 7　算法的可解释性评估方法

重点评估要素	评估方法
模型复杂度	a）模型参数数量：记录模型中所有参数的数量。参数数量越多，表示算法的复杂度越高，可解释性越差。 b）模型结构复杂度：记录模型中的层级、深度和模型中神经元的连接方式。模型层级越多，深度越深，以及神经元的连接方式越多样，表明算法的结构复杂度越高，可解释性越差
解释性能力	a）可视化效果评估：记录算法可视化后信息密度和数据关系的表达。信息密度越大，表示算法的可视化效果越强，可解释性能力越好。 b）特征重要性评估：记录单个或多个特征输入算法后对预测结果的影响程度。影响程度越大，表示算法的特征重要性评估效果越强，可解释性能力越好
解释性质量	a）解释的准确性：在不同的输入参数或特定条件下，记录算法所生成的解释性结果。若解释结果清晰、准确，则解释的准确性满足。 b）解释的完整性：记录算法提供解释的涵盖范围。解释涵盖的范围越大，解释质量的完整性越高。 c）解释的一致性：在不同的输入或条件下，记录算法提供的解释结果。若解释结果保持相对稳定和一致，则解释的一致性满足

7.3　可控性

可控性是评估生态环境领域算法系统实际应用价值的关键指标，关乎算法系统的稳定性和可调性，以及涉及算法系统在运行时的情况。生态环境领域算法的可控性评估方法见表 8。

表 8　算法的可控性评估方法

重点评估要素	评估方法
系统稳定性	a）平均错误率：统计算法在连续运行期间出现的错误次数，并计算其平均值。平均错误率越小，说明算法在运行过程中出错的可能性越小，稳定性越高。 b）最大错误率：记录算法在特定条件下或特定时间段内的最大错误次数。最大错误率反映了算法在极端情况下的错误表现，其值越小，表示算法的稳定性和鲁棒性越好。 c）运行期间崩溃次数：监控算法在连续运行期间发生的崩溃次数。崩溃次数越少，说明算法的运行越稳定，用户体验越好。 d）崩溃后恢复时间：在算法发生崩溃后，记录从崩溃到恢复正常运行所需的时间。恢复时间越短，说明算法具备快速恢复的能力，对服务连续性的影响越低
参数可调性	a）参数调整范围：对算法参数进行调整，记录参数调整范围，范围越大，表明参数可调性越好。 b）参数调整对性能的影响：在调整算法参数过程中，记录算法性能变化情况。算法性能波动越小，说明算法参数调整越稳定
运行状态的实时监控	a）状态信息更新频率：记录算法在上一次状态更新到下一次状态更新的时间间隔。时间间隔越短，说明算法状态信息更新的越频繁，实时监控能力越强。 b）故障预警机制的有效性：在多次测试算法故障后，记录算法成功故障预警的次数。成功次数越多，表明算法故障预警机制越有效。 c）资源利用率的优化程度：记录资源优化前和优化后的利用比值。利用比值越低，表明算法资源利用率的优化程度越好。 d）资源调度算法的合理性：监控算法在调度资源后资源的分布情况。资源分布越平均，说明资源调度算法越合理

7.4 安全性

7.4.1 可靠性

可靠性涉及算法模型的训练、测试等过程。在环境监测、污染防治等关键领域，通过算法模型的能力，应对各种数据、模型攻击方法，评估算法模型在不同攻击方法下的可靠性。生态环境领域算法的可靠性评估方法见表9。

表9 算法的可靠性评估方法

重点评估要素	评估方法
中毒性攻击抵御可靠性	a) 数据投毒攻击抵御能力：计算成功攻击使算法模型对输入数据进行错误分类的比例。比例越小，表示被成功攻击的算法模型越少，数据投毒攻击抵御能力越强 b) 模型后门攻击抵御能力：计算成功攻击的后门样本与全部后门样本的比值。比值越小，表示后门攻击成功率(ASR)越低，模型后门攻击抵御能力越强
对抗性攻击抵御可靠性	a) 白盒对抗攻击抵御能力：计算成功攻击的白盒对抗样本与全部白盒对抗样本的比值。比值越小，表示白盒对抗攻击成功率(ASR)越低，模型白盒对抗攻击抵御能力越强。 b) 灰盒对抗攻击抵御能力：计算成功攻击的灰盒对抗样本与全部灰盒对抗样本的比值。比值越小，表示灰盒对抗攻击成功率(ASR)越低，模型灰盒对抗攻击抵御能力越强。 c) 黑盒对抗攻击抵御能力：计算成功攻击的黑盒对抗样本与全部黑盒对抗样本的比值。比值越小，表示黑盒对抗攻击成功率(ASR)越低，模型黑盒对抗攻击抵御能力越强
物理对抗攻击抵御可靠性	a) 有目标物理对抗攻击抵御能力：计算成功攻击的有目标物理对抗样本与全部有目标对抗样本的比值。比值越小，表示有目标物理对抗攻击成功率(ASR)越低，模型有目标物理对抗攻击抵御能力越强。 b) 无目标物理对抗攻击抵御能力：计算成功攻击的无目标物理对抗样本与全部无目标对抗样本的比值。比值越小，表示无目标物理对抗攻击成功率(ASR)越低，模型无目标物理对抗攻击抵御能力越强

7.4.2 计算环境的鲁棒性

算法计算环境的鲁棒性涉及算法模型的训练、测试等过程。在环境监测、污染控制等关键领域，通过全面考察计算，评估算法计算环境在不同场景下的鲁棒性。生态环境领域算法计算环境的鲁棒性评估方法见表10。

表10 算法计算环境的鲁棒性评估方法

重点评估要素	评估方法
智能算法供应链的鲁棒性	a) 供应链完整性：统计计算环境中供应链的实际环节数量与理论环节数量的比值越大，表示供应链的环节数量越多，供应链完整性越高。 b) 组件来源可信性：通过信誉度、稳定性、质量等评估组件或组件的供应商。组件稳定性、质量等越高，商家信誉度、合规性越好，组件来源可信度越高。 c) 供应链安全性：通过供应链透明度、供应商信誉度、安全合规性、灾难恢复与业务连续性等评估供应链的安全性。供应链透明度越高、供应商信誉度越高、安全合规性越好、灾难恢复与业务连续性越有效等，则供应链的安全性越高

表 10　算法计算环境的鲁棒性评估方法（续）

重点评估要素	评估方法
分布式计算的鲁棒性	a）数据一致性能力：通过数据一致性级别、读写数据的时延、系统吞吐量等评估数据一致性能力。数据一致性级别越高、系统读写数据时延越低、系统吞吐量越大等，则数据一致性能力越强。 b）分布式结构安全性：通过数据安全性、数据保密性、访问控制策略严密性、数据备份与恢复能力等评估分布式结构安全性。数据安全保密性越好、访问控制策略严密性越高、数据备份与恢复能力越强等，则分布式结构安全性越好
计算框架的鲁棒性	a）算子安全性：通过访问控制和权限管理严密性、安全验证规范性、日志和监控完整性等评估算子安全性。访问控制和权限管理严密性越好、安全验证规范性越强、日志和监控完整性越高等，则算子安全性越强。 b）框架库安全性：通过漏洞管理和修复能力、身份验证和访问控制严密性、日志和监控完整性等评估框架库安全性。漏洞管理和修复能力越强、身份验证和访问控制严密性越好、日志和监控完整性越高等，则框架库安全性越好。 c）API 安全性：通过访问控制和权限管理严密性、输入验证和过滤机制、持续安全监控设置等手段评估 API 安全性。访问控制和权限管理严密性越好、输入验证和过滤机制越完整、设置了持续安全监控等，则 API 安全性越好。 d）编译器安全性：通过编译器的代码审查功能、漏洞披露与修复技术、运行保护能力等方面评估编译器安全性。编译器的代码审查功能越好、漏洞披露与修复技术越成熟、运行保护能力越强等，则编译器安全性更好

7.4.3　保密性

保密性是评估生态环境领域算法在处理敏感数据时安全性的关键指标，涉及算法在数据存储、传输和处理过程中的加密和访问控制能力。在需要严格数据保护的生态监测、物种多样性分析等场景中，通过评估算法的数据加密强度、认证机制以及权限设置来确保其保密性能符合要求。生态环境领域算法的保密性评估方法见表 11。

表 11　算法的保密性评估方法

重点评估要素	评估方法
数据敏感性与保密性	a）数据加密措施：算法在存储和传输数据时采用的加密技术，包括加密协议的类型、密钥管理机制以及加密强度级别。采用的加密技术越先进、密钥管理越严格、加密级别越高，则数据的安全性和保密性越好。 b）数据访问控制：算法实施的数据访问权限设置，包括用户身份验证、角色基础的访问控制、访问控制列表等机制的部署和细粒度。访问控制机制越全面，权限划分越细致，表明算法在保护数据不被未授权访问方面的效能越强。 c）数据存储安全性：检验算法在数据存储阶段采取的安全措施，包括数据备份机制、存储加密技术以及物理和逻辑安全策略的实施情况。若算法提供的数据备份频率高，存储加密技术先进，且有严格的物理和逻辑安全策略，则其数据存储安全性较强

表 11　算法的保密性评估方法（续）

重点评估要素	评估方法
模型保密性	a）模型参数保密性：评估模型参数的保护措施，包括参数加密、访问控制以及在模型分享或发布时的保密策略。模型参数的加密程度越高，访问权限划分越精确，则模型参数保密性越强。 b）模型文件加密：检查算法在保存和分发模型文件时所采用的加密措施，包括文件加密标准、传输过程的加密协议以及密钥管理策略。若模型文件使用强加密标准，传输过程中有安全保障，且密钥管理严格，则认为模型文件的加密保护较为充分。 c）模型访问权限控制：考察算法实施的模型访问权限管理机制，包括用户身份验证、权限分配、权限审核流程和操作日志监控等。如果算法支持细粒度的权限配置，能够根据用户角色限制对模型的访问和操作，且有完备的审计跟踪记录，则认为其模型访问权限控制较为严格
依赖信息保密性	a）依赖库和框架的保密性：分析算法所依赖的库和框架的保密措施，包括它们的访问权限设置、加密特性以及是否对外部暴露敏感信息。如果依赖的库和框架提供了良好的安全性能，定期更新修复漏洞，并且对敏感数据处理有明确的安全策略，则认为其保密性较强。 b）依赖信息的访问控制：检查算法及其依赖库和框架的访问权限设置，包括版本管理系统的权限配置、依赖信息文件的读取权限以及对这些信息的修改和更新机制。若算法实现了对依赖信息的严格访问控制，确保只有授权用户能够获取、修改或更新这些信息，则认为其访问控制措施得当。 c）依赖信息的完整性保护：审查和验证算法所依赖的库和框架的发布来源，检查依赖信息文件以确认其未经未授权篡改。若依赖信息能够与官方或可信源发布的信息匹配，且无迹象表明文件被篡改，则认为完整性保护措施有效

7.5　维护性

7.5.1　兼容性

兼容性是评估生态环境领域算法适应性的关键指标，涉及算法在各种硬件、软件及操作系统上的运行能力。在多设备、跨平台的环境管理任务中，通过全面的兼容性测试，评估算法在不同系统和设备上的稳定性和运行效率。生态环境领域算法的兼容性评估方法见表 12。

表 12　算法的兼容性评估方法

重点评估要素	评估方法
算法对数据格式的兼容性	a）数据格式兼容性：统计算法能够兼容和处理的数据格式种类数量，以及支持的编码标准数量。两者数量越大，表示算法对不同数据格式和编码标准的适应性越强，灵活性越高。 b）不同数据格式处理能力：通过测试算法处理不同数据格式的数据集，记录完成相同任务所需的时间及结果的正确率。时间越短，表示算法处理该数据格式的效率越高；正确率越高，表示算法处理该数据格式的准确性越好
算法对操作系统的兼容性	a）操作系统兼容数量：统计算法能够在其上运行无误的操作系统类型及各类型的不同版本数量。兼容的操作系统类型及版本数量越多，表示算法的适应性和泛用性越强。 b）跨平台操作便捷性：评估算法在不同操作系统平台上进行安装、配置和使用的难易程度，以及算法在各平台上的表现是否一致。安装和配置过程越简单，用户在不同平台间切换使用时的学习成本越低，表示算法的跨平台操作便捷性越高；各平台上算法表现的差异越小，表示其跨平台一致性越好

表 12　算法的兼容性评估方法（续）

重点评估要素	评估方法
算法对其他软件的兼容性	a)　第三方软件兼容性：统计算法能够与之无缝集成和协作的第三方软件的种类及其支持的版本数量。若算法能与更多类型的第三方软件兼容，并且支持它们的多个版本，则说明其适应性和集成能力更强。 b)　软件更新适应性：监测算法在依赖的第三方软件发生更新或变更时，其运行稳定性和功能完整性的表现，记录出现问题的频率及解决问题所需的时间。若算法在面对第三方软件更新或变更时，仍能保持稳定运行且功能不受影响，或在问题出现后能迅速恢复，则说明其适应性更强。 c)　国产化芯片的适配兼容性：评估算法在国产化芯片上的性能和稳定性，确保算法能够有效地利用国产化硬件资源。各国产化芯片上算法表现的差异越小，表示其兼容国产化芯片的能力越好。 d)　人工智能框架的适配兼容性：评估算法在国产、人工智能框架中的表现，确保算法能够在这些框架上高效运行。各人工智能框架上算法表现的差异越小，表示其兼人工智能框架的能力越好

7.5.2　可维护性

可维护性是评估生态环境领域算法长期适用性和可更新性的关键指标，涉及算法的代码结构、文档完整性及模块化设计。在面对环境变化和新技术挑战的背景下，通过定期的维护和升级，评估算法的修复效率。生态环境领域算法的可维护性评估方法见表 13。

表 13　算法的可维护性评估方法

重点评估要素	评估方法
算法迭代的更新频率	a)　迭代时间间隔能力：测量算法从一次迭代开始到下一次迭代开始之间的时间间隔。时间间隔越短，表示算法的迭代速度越快，能够更快地响应环境变化和数据更新。 b)　迭代代码变动量：记录算法在每次迭代中代码的改变程度，包括新增、修改或删除的代码行数。变动量越少，表示算法迭代时对原代码结构的影响越小，维护成本和迭代风险越低
算法迭代的质量变化	a)　迭代性能提升能力：通过基准测试或性能评估，比较算法迭代前后在执行效率、准确率或其他关键性能上的提升百分比。提升百分比越大，表示算法每次迭代带来的性能改善越显著。 b)　迭代系统稳定能力：通过运行系统稳定性测试，记录算法迭代前后系统的故障频率和恢复时间。若迭代后故障频率降低且恢复时间缩短，表明算法的改进增强了系统的稳定性

7.5.3　可移植性

可移植性是评估生态环境领域算法适应性和广泛适用性的关键指标，涉及算法在不同计算环境、操作系统及硬件配置上的运行能力。在多平台部署和跨设备应用的情境中，通过考察算法依赖管理的简便性以及对不同系统架构的支持程度来评估其可移植性。生态环境领域算法的可移植性评估方法见表 14。

表 14　算法的可移植性评估方法

重点评估要素	评估方法
算法对硬件设备的可移植性	a）支持的硬件设备种类：列出算法能够兼容并优化运行的硬件设备类型，如不同制造商的处理器、特定型号的图形处理单元(GPU)、各式传感器等，并统计数量。支持的硬件设备种类数量越多，表明算法的适应性和通用性越强，能够部署在更多样化的硬件环境中。 b）跨硬件性能差异：在不同类型的硬件设备上运行相同的算法，记录并比较其性能，如处理速度、吞吐量和资源占用等。若算法在不同硬件上的性能差异较小，表示其具有良好的跨平台性能一致性；若差异较大，则说明算法对特定硬件的优化程度不一
算法对人工智能框架的可移植性	a）支持的人工智能框架数：统计算法能够与之兼容并优化运行的人工智能框架数量。支持的人工智能框架数越多，表明算法的适应性和兼容性越强，对开发者的选择提供更多灵活性。 b）框架间的性能保持：在多个不同的人工智能框架上实施同一算法，对比其在各框架上的执行效率、准确性和稳定性等关键性能。若算法在不同框架间表现出相似的性能水平，说明其具有良好的跨框架性能一致性；若性能差异显著，则需针对特定框架进行优化

7.5.4　可扩展性

可扩展性是评估生态环境领域算法在面对增长的数据量和计算需求时的适应性和灵活性的关键指标。在数据处理量不断增加和计算资源需求变化的情境中，通过考察算法在增加计算资源时的性能提升能力，以及在升级单个组件时的性能改善情况来评估其可扩展性。生态环境领域算法的可扩展性评估方法见表 15。

表 15　算法的可扩展性评估方法

重点评估要素	评估方法
算法水平扩展能力	a）性能提升评估：评估算法在增加计算资源时，性能是否能够相应地线性或接近线性地提升。如果性能提升显著且与资源增加成正比，则表明算法具有良好的水平扩展能力。 b）无状态服务支持：检查算法是否支持无状态服务以便于横向扩展。如果新增节点对现有操作的影响很小或没有影响，并且系统能够平滑地处理负载均衡，则表明算法具有良好的无状态服务支持能力。 c）自动识别与集成：验证新增加的资源能否被系统自动识别并有效利用，确保新资源能够无缝集成到现有系统中。如果系统能够自动识别和配置新增资源，并且这些资源能够立即投入使用而无需人工干预，则表明算法具有良好的自动识别与集成能力
算法垂直扩展能力	a）硬件升级性能提升：算法在通过升级单个组件来提升性能的能力。如果性能提升明显且与硬件增强成正比，则表明算法具有良好的垂直扩展能力。 b）最大算力限制：确定单个实例支持的最大算力限制，并分析随着硬件增强带来的性能改善比例。单个实例支持的最大算力越高，且性能改善比例越大，算法的垂直扩展能力越好

附 录 A
（资料性）
算法评估实施案例

评估准备见表 A.1。人工智能评估指标见表 A.2。算法性能评估结果见表 A.3。算法评估结论见表 A.4。

表 A.1 评估准备

算法名称	湿地地物分类算法		
算法说明	湿地地物分类算法是基于正射遥感影像数据信息进行地物分类的技术。其通过无人机采集湿地的高精度正射遥感影像，使用基于大模型的算法进行地物分类，分为森林、灌丛、草地、水体、建筑等类型。 无人机将定期巡航获取目标区域的正射遥感影像数据，用户提供需要分类的类型并进行识别，算法将产出语义识别结果并生成分类影像结果。该系统将应用于西溪湿地的影像识别中用于产出生态侧应用的基础数据。		
场景分析	算法运行条件	a） 本地： • 硬件设备：智能摄像头； • 操作系统：Linux； • 人工智能框架：PaddlePaddle Mobile； • 本地设备通过网络接入云端。 b） 云端： • 硬件设备：GPU Nvidia P4； • 操作系统：Ubuntu 20.04； • 人工智能框架：PaddlePaddle Serving	
	算法运行模式	无人机搭载的正射传感器对湿地地区进行影像采集。无人机可以灵活飞行于不同高度和角度，获取高分辨率、高精度的正射遥感影像数据。采集到的正射遥感影像数据通过无线传输方式发送至云端服务器。在云端，对原始影像进行预处理以消除图像中的各种误差和干扰，提高图像质量。无人机摄像头捕捉影像，将正射遥感影像进行预处理，之后发送云端服务器中部署的地物分类算法，该算法已经基于大模型进行校正和预训练。最后，经过算法识别输出湿地地物类型的语义识别结果	
	正常运行场景	a） 算法周期性接收无人机传回的影像； b） 应用分类算法，返回识别结果	
	可预见的异常场景	a） 植被茂密遮挡导致分类难以识别遮盖的物体； b） 影像质量不佳导致阴影之类的影响识别结果； c） 一些分类类型外的地物如堆料等问题难以识别	
风险分析	算法失效序号	算法失效说明	识别方法
	1	植被遮挡导致的遮盖地物未识别	增加倾斜摄影支持
	2	影像阴影导致切割误差	增加巡航次数合成高质量影像
	3	分类外的堆料问题	头脑风暴

表 A.1　评估准备（续）

算法名称	湿地地物分类算法		
危险严重性等级评估	算法失效序号	后果	危险严重性等级
	1	导致影像边缘切割模糊，漏分部分地类边缘影像	一般
	2	导致违规建设等问题未被识别，可能造成一般后果	一般
	3	导致湿地旱化问题未被识别，可能造成一般后果	一般
确定评估目标	危险严重性等级说明		评估目标
	基于地物分类的识别算法将被应用到西溪湿地正射遥感解译工作中去。算法失效可能导致：①影像边缘切割模糊，导致影响生态价值核算结果。②湿地内的违规建设或者旱化问题未被及时发现，所以归为一般级		Ⅳ

表 A.2　人工智能评估指标

一级指标	选择的测试元
算法性能	a）　精准性：准确率、精确率、召回率、F1 分数； b）　效能：执行速度、资源利用率、吞吐量、并行处理能力、负载处理能力
可解释性	可解释性：模型参数数量、模型结构复杂度、可视化效果评估、特征重要性评估、解释的准确性、解释的完整性、解释的一致性
可控性	可控性：最大错误率、运行期间崩溃次数、崩溃后恢复时间、参数调整范围、参数调整对性能对影响、状态信息更新频率、故障预警机制的有效性、资源利用率的优化程度、资源调度算法的合理性
安全性	a）　可靠性：数据投毒攻击抵御能力、模型后门攻击抵御能力、白盒对抗攻击抵御能力、灰盒对抗攻击抵御能力、黑盒对抗攻击抵御能力、有目标物理对抗攻击抵御能力、无目标物理对抗攻击抵御能力； b）　计算环境的鲁棒性：供应链完整性、组件来源可信性、供应链安全性、数据一致性能力、分布式结构安全性、算子安全性、框架库安全性、API 安全性、编译器安全性； c）　保密性：数据加密措施、数据访问控制、数据存储安全性、模型参数保密性、模型文件加密、模型访问权限控制、模型参数保密性、模型文件加密、模型访问权限控制、依赖库和框架保密性、依赖信息访问控制、依赖信息完整性保护
维护性	a）　兼容性：数据格式兼容性、不同数据格式处理能力、操作系统兼容数量、跨平台操作便捷性、第三方软件兼容性、软件更新适应性； b）　可维护性：迭代时间间隔能力、迭代代码变动量、迭代性能提升能力、迭代系统稳定能力； c）　可移植性：支持的硬件设备种类、跨硬件性能差异、支持的人工智能框架数、框架间的性能保持； d）　可扩展性：性能提升评估、无状态服务支持、自动识别与集成、硬件升级性能提升、最大算力限制

表 A.3　算法性能评估结果

一级指标	算法性能				
二级指标	重点要素	测试元	评估工作	测试元分值	测试元权重
精准性	分类任务精准性	准确率	准确率应达到 90%以上	90	0.15
		精确率	精确率应达到 85%以上	85	0.15
		召回率	召回率应达到 80%以上	85	0.15
		F1 分数	F1 分数应达到 80%以上	85	0.05
效能	执行效率	执行速度	非实时应用场景下，单张图像(1 000×1 000 像素)的处理时间应小于 100 ms(硬件满足算力需求)	80	0.1
		资源利用率	CPU、GPU、内存空余量应大于 20%	85	0.1
	数据处理能力	吞吐量	IPS(每秒处理的图像数)≥10	90	0.1
		并行处理能力	从单线程、单 GPU 到多线程、多 GPU 环境下，扩展效率应达到 80%	90	0.1
		负载处理能力	高负载下，性能下降不超过 10%	90	0.1
一级评估结果	86.75				

表 A.4　算法评估结论

一级指标名称	指标评分	权重	算法评估结果
算法性能	86.75	0.2	84.15
可解释性	85	0.2	
可控性	85	0.2	
安全性	81	0.2	
维护性	83	0.2	
评估结论	生态环境领域算法通过评估，并达到Ⅳ级目标要求		