

第八届中国研究生人工智能创新大赛

碳排放因子智能检索 MCP 服务器

(Carbon Factor Matcher)

项目文档

日期： 2026.08.31

团队名称： 零碳逻辑

参赛组别： 开放赛题——面向科学、工程和社会科学的“AI+X”

目 录

1	项目概况	2
1.1	背景和基础	2
1.2	场景和价值	3
1.3	所需支持	4
2	项目规划	5
2.1	整体目标	5
2.2	技术创新点	6
3	实施方案	7
3.1	技术可行性分析	7
3.2	技术细节	7
3.3	计划和分工	11
4	参考资料	12

1 项目概况

1.1 背景和基础

（一）时代背景

在“碳达峰、碳中和”目标与全球气候治理的推动下，碳排放核算的应用范围持续扩大：欧盟碳边境调节机制（CBAM）、新电池法规等已将产品级碳核算纳入国际贸易要求；国内碳市场持续扩容，ESG信息披露要求不断收紧。企业碳盘查、产品碳足迹核算与生命周期评价（LCA）的数据基础均为碳排放因子，即单位活动量的温室气体排放量（如 $\text{kgCO}_2\text{e/kWh}$ ）。

（二）行业现状

当前排放因子数据的获取和使用存在两方面突出问题。

其一，因子数据分散、检索成本高。权威排放因子分布于多个互不连通的数据库：欧盟 ELCD（493 条）、瑞士 ecoinvent（逾两万条）、美国 US LCI（532 条）、USDA LCA Commons（1,550 条）。各库格式不一（JSON-LD/ILCD、Excel 等），检索依赖专业客户端或网页表单，使用者需具备生命周期评价专业知识才能构造查询；完成一次因子核查常需在多个数据库之间反复比对，耗时从数分钟到数小时不等。

其二，大模型直接回答碳核算问题存在幻觉风险。用通用大模型直接询问“生产 1 kg PET 排放多少 CO_2 ”，模型给出的数字看似合理，但无出处、不可审计。碳核算结果属于合规材料，无法溯源的答案不能满足使用要求。如何使人工智能获得权威、可信、带质量评级的因子数据，是人工智能应用于双碳领域首先要解决的数据问题。

（三）项目起源与团队基础

本项目起源于一篇关于人工智能辅助碳排放因子检索的 SCI 论文。该研究提出，碳核算领域可以借助人工智能方法检索排放因子，从而提高因子匹配效率，但文中并未将这一思路实现为可用的工具。团队读后认同其观点，并注意到大模型工具链（MCP 协议）的兴起为该思路的落地提供了合适的技术条件，遂决定将其开发为实际可用的工具，即以 MCP 服务器的形式把权威排放因子数据库接入主流 AI 工具，使因子匹配能力从方法研究走向软件产品。

在团队基础方面，成员具备碳排放核算领域的业务经验与软件工程的开发能力；本系统已完成商业化发布并在线上运行，具备自动化测试与版本化发布体系，能够支撑项目的持续开发与维护。

1.2 场景和价值

（一）应用场景

本系统面向生命周期评价工程师、双碳与 ESG 咨询顾问、企业碳管理人员及碳核算软件开发人员。系统以 MCP（Model Context Protocol）服务器形式发布，用户在 Claude Desktop、Cursor 等支持 MCP 的工具中完成一次配置后，即可在对话中以自然语言（中英文均可）完成因子检索。系统提供四个工具，功能见表 1。

表 1: 系统提供的 MCP 工具

工具	功能
<code>factor_match</code>	混合检索与数据质量排序，返回带评级与出处的候选因子列表
<code>factor_search</code>	关键词检索，支持类别过滤
<code>factor_detail</code>	返回单个因子的完整元数据（数值、单位、地理范围、来源、年份、适用性）
<code>process_inventory</code>	清单级处理：物料清单去重、缺失因子相似替代、替代敏感性评级、透明度报告

以典型使用场景为例：用户在 AI 助手输入“帮我把这份物料清单配上碳排放因子，缺的给出替代方案”，系统即自动完成清单去重与逐条检索匹配，对无法精确匹配的物料按相似度给出替代因子并标注敏感性，最终输出列明全部假设与数据来源的透明度报告。

（二）与现有方案的对比

本系统与现有方案的对比如表 2 所示。

表 2: 本系统与现有方案的对比

方案	检索效率	结果可溯源	质量评估	可集成性
专业客户端手工检索 (openLCA 等)	低, 需专业知识构造查询	高	需人工判断	差
通用搜索引擎	低, 结果非结构化	无保证	无	差
通用大模型直接问答	高	不可溯源 (存在幻觉风险)	无	高
本系统	高 (自然语言直达)	每条结果带来源、年份、地理范围	五维质量评级	基于 MCP 标准协议, 主流 AI 工具即插即用

(三) 发布与商业化情况

系统已完成商业化发布。官方网站 (<https://zerocarbonlogic.com>) 提供产品介绍与购买入口; 分发渠道包括 PyPI、npm 与 Smithery 平台 (<https://smithery.ai/servers/nikeandocean/carbon-factor-matcher>)。收费模式为免费与买断相结合: 免费档每日提供 300 次查询; Pro 档一次性买断, 解锁混合检索、质量评级与 ecoinvent 数据集; 跨境支付与许可密钥自动发放流程已上线运行。系统自 2026 年 6 月中旬起在上述平台发布, 截至 2026 年 8 月底, Smithery 平台累计使用超过 3,600 次, npm 包累计下载约 1,800 次, 近一个月 npm 与 PyPI 合计下载约 650 次。

在社会价值方面, 权威因子库的使用门槛不仅在于专业软件与专业训练, 还在于数据授权费用: ecoinvent 的因子数据需按年付费授权, 商业单用户年费在 1,500 欧元以上, 中小企业难以支撑。本系统使用户以一句自然语言即可完成因子检索, 中小咨询机构与中小企业无需承担高昂的年度授权费用, 即可获得可信的碳核算数据。

1.3 所需支持

在算力方面, 现阶段无需 GPU 训练资源, 语义编码模型 (text2vec-base-chinese, 约 400 MB) 在 CPU 上推理即可支撑单机与云端部署; 后续若开展中文碳领域嵌入模型微调, 预计需要单张高端 GPU 与数天的训练时间。

在数据方面，希望获得更多官方因子库的对接授权，特别是中国本土生命周期评价数据库（如 CLCD）与官方发布的区域电网因子，以提升对国内核算场景的覆盖。

在应用与合作方面，希望与碳排放咨询机构、企业碳管理团队开展试点应用，在实际核算场景中检验并改进系统；同时希望获得云资源支持，保障线上服务的稳定运行与后续扩容。

2 项目规划

2.1 整体目标

本项目的总体目标是建设面向 AI 智能体的碳核算数据基础设施，使各类 AI 智能体能够通过标准协议获取可信、可溯源、带质量评级的排放因子数据。阶段目标与当前进展见表 3。

表 3: 阶段目标与当前状态

阶段	目标	状态
初赛阶段	四库 8,404 条因子整合；混合检索与质量排序引擎；四个 MCP 工具；多渠道发布与商业化流程；消融实验验证架构设计	已完成
决赛及后续	中文本土因子库接入；检索模型中文领域优化；多因子不确定性表达；面向智能体的计算规范文档（集成 ILCD 手册等官方文档核心）；覆盖更多 AI 工具生态	规划中

在后续工作中，项目计划提供一份服务于本 MCP 服务器的规范文档（spec）。该文档将集成 ILCD 手册等官方文档的核心内容，包括因子适用条件、单位约定与核算方法要求，供调用方智能体在因子检索之外获取领域知识，指导其在具体核算场景中合理选择因子并完成排放计算。

2.2 技术创新点

(一) 现有技术对比调研

现有技术路线及其不足见表 4。

表 4: 现有技术路线对比

技术路线	代表	核心不足
专业数据库客户端检索	openLCA、ecoinvent 官方客户端	关键词精确匹配为主；跨库不互通；需专业训练
通用搜索引擎	Google / Bing	结果非结构化，无质量与适用范围元数据
通用大模型直接问答	ChatGPT 等	结果不可溯源，不可用于合规核算
纯向量 RAG 检索	常规 RAG 方案	依据语义相似度召回，缺少领域质量约束；在间接描述（词汇失配）场景下准确率低，本实验实测最优模型 Precision@1 仅 38%（见第 3 节）

(二) 本项目技术创新点

在协议方面，系统基于 MCP 标准协议连接大模型与排放因子数据库，使支持该协议的 AI 工具（Claude、Cursor 等）无需改造即可获得因子检索能力，融入现有 AI 智能体生态。

在检索方法上，系统采用两阶段混合检索：对查询进行关键词打分（权重 0.3）与语义向量相似度计算（权重 0.7）并线性融合；第一阶段以关键词粗筛缩小候选集，第二阶段计算向量相似度，在保持召回水平的同时降低嵌入计算量。

在排序方法上，系统参照 ILCD 数据质量体系建立五维数据质量评级（技术代表性 0.3、来源可靠性 0.4、地理代表性 0.1、时间代表性 0.1、因子类型 0.1），并将质量得分与混合相关性得分按 5:5 融合，使领域数据质量参与最终排序。

在系统架构上，消融实验（4 种嵌入模型 × 2 种配置，50 个词汇失配用例）表明，由

大模型对候选因子进行语义复核可将 Precision@1 一致提升 6~16 个百分点。据此，系统将最终因子选择交由调用方 AI 智能体完成：服务端返回带质量评级与出处的排序候选列表，而非单一结果。该设计使服务端无需调用任何大模型接口，降低了推理成本，且选择过程对用户可见、可干预，便于合规审计。

在工程实现上，系统支持中英文单位别名归一（如 kWh ↔ 千瓦时 ↔ 度），适配中文用户查询英文因子库的使用场景。

3 实施方案

3.1 技术可行性分析

在数据获取方面，ELCD 由欧盟委员会联合研究中心公开提供（JSON-LD/ILCD 格式），US LCI 与 USDA LCA Commons 均可公开获取，ecoinvent 数据已完成授权导入与统一化预处理。四库共 8,404 条因子经字段级清洗（名称、类别、数值、单位、地理范围、来源、年份、适用性、五维质量评级）后统一托管于云端对象存储，经 CDN 与 API 分发，终端用户无需下载数据库。数据层采用适配器模式，可接入新的数据源。

在行业知识方面，五维质量评级框架源自 ILCD 手册的数据质量体系（1= 优、2= 中、3= 差），权重按领域惯例设定并支持配置调整；单位制与地理编码（ISO 3166）等按国际标准处理，保证与主流生命周期评价软件的互操作性。

在算力方面，全流程可在 CPU 上运行：嵌入模型本地缓存，向量按因子索引懒加载；云端服务无 GPU 依赖、无状态，具备低成本规模化部署的条件。

在工程方面，系统已完成发布并在线上运行，具备自动化测试体系、版本化发布流程（PyPI、npm、Smithery 多渠道）与用量统计能力。

3.2 技术细节

（一）系统架构

系统分为三层。AI 工具层为 Claude Desktop、Cursor 等支持 MCP 协议的工具，作为用户入口并完成最终因子选择；服务层为 MCP 服务器，提供四个工具，内含混合搜索引擎、质量评级模块、用量统计与授权管理；数据层包括云端统一因子库（8,404 条，经 CDN 与 API 分发）与本地 JSON-LD 适配器（支持离线部署、私有化部署与接入新数据源）。

(二) 检索管线

检索管线分为两个阶段。第一阶段为混合召回：对查询进行关键词打分（对名称、类别、适用范围字段加权）与语义向量相似度计算，按式 (1) 线性融合；先以关键词粗筛候选集、再计算嵌入相似度，以控制计算量。嵌入模型为 text2vec-base-chinese，在 CPU 上推理，向量按因子索引缓存。

$$\text{hybrid} = 0.3 \times \text{keyword} + 0.7 \times \text{embedding} \quad (1)$$

第二阶段为质量重排：将混合相关性得分与数据质量得分按 5:5 融合，见式 (2)。其中质量得分由五维评级加权得到（1= 优、2= 中、3= 差，数值越小质量越高）。系统最终返回排序后的候选列表，由调用方智能体结合用户语境完成最终选择。

$$\text{final} = 0.5 \times \text{hybrid} + 0.5 \times \text{quality} \quad (2)$$

(三) 实验与效果指标

为评估系统效果，项目构建了三份标注基准，共 83 个用例：标准集 17 例（直接名称查询）、复杂集 16 例（复合条件查询）、词汇失配集 50 例（查询采用功能或用途描述而非材料名，如以 “producing the most abundant metal on Earth” 指代铝，用于考察语义理解能力；含中英文对照查询）。

实验一考察大模型参与候选选择的作用，采用消融实验设计：4 种嵌入模型（BERT 768 维、Sentence-BERT-MiniLM 384 维、gte-large 1024 维、text2vec-base-chinese 768 维） $\times$ 2 种配置（纯向量检索；向量召回前 10 候选后由大模型语义重排，重排模型为 deepseek-chat, temperature 取 0.1），在词汇失配集 50 个用例上比较 Precision@1、MRR、NDCG@5 与 Recall@5，并按 2,000 次重采样计算 95% bootstrap 置信区间。结果见表 5 与图 1。

表 5: 消融实验结果（词汇失配集，50 用例）

模型	配置	Precision@1	MRR	NDCG@5	Recall@5
BERT (768d)	Embedding Only	24%	0.320	0.346	46%
BERT (768d)	LLM-Enhanced	40%	0.452	0.460	50%
Sentence-BERT (384d)	Embedding Only	32%	0.378	0.396	48%
Sentence-BERT (384d)	LLM-Enhanced	46%	0.476	0.484	52%
gte-large (1024d)	Embedding Only	38%	0.431	0.438	50%
gte-large (1024d)	LLM-Enhanced	50%	0.524	0.529	56%
text2vec-base (768d)	Embedding Only	10%	0.127	0.139	18%
text2vec-base (768d)	LLM-Enhanced	16%	0.167	0.169	18%

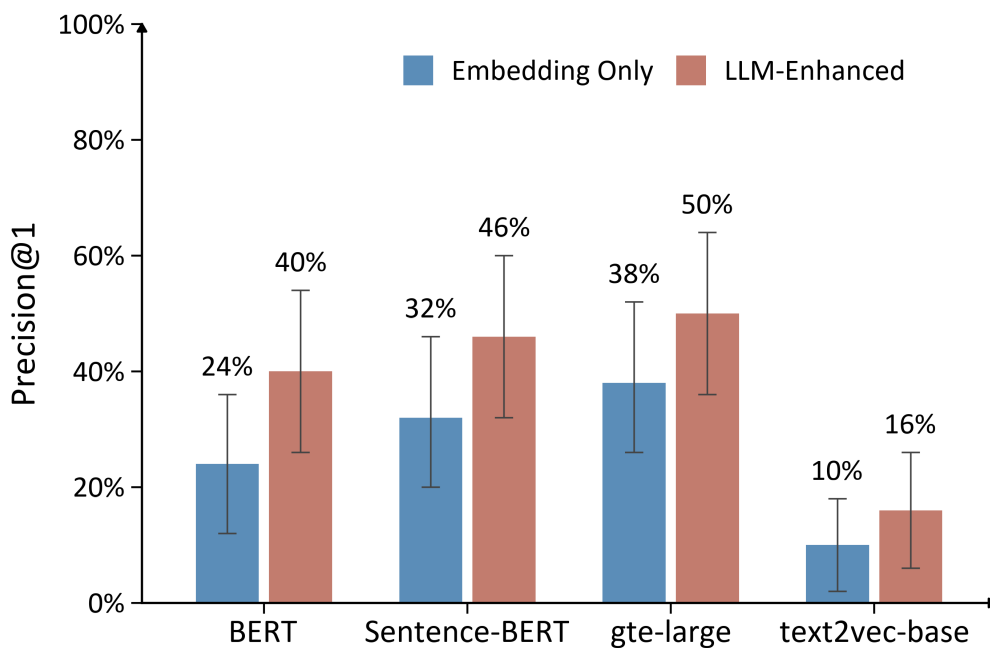


图 1: 大模型参与候选选择对 Precision@1 的提升（误差棒为 95% bootstrap 置信区间）

实验结果表明，词汇失配是纯向量检索的普遍不足，效果最好的 gte-large 其 Precision@1 也仅为 38%；引入大模型语义复核后，四种模型的 Precision@1 一致提升 6~16 个百分点，且召回率不低于纯向量检索。大模型重排的单次延迟为 1.4~2.2 秒，嵌入检索为毫秒级，主要开销来自大模型调用。基于上述结果，系统采用服务端排序候选、调用方智能体最终选择的架构，在获得复核精度收益的同时，服务端不依赖大模型接口。

需要说明的是，text2vec-base-chinese 为中文模型，在英文查询与英文因子名构成的跨语言测试集上绝对指标偏低，反映出跨语言检索能力是后续的优化方向。生产配置面向中文用户选用该模型，消融实验则覆盖多种模型，以考察复核机制的通用性。

实验二评估 process_inventory 工具的清单处理效果，在本地库（ELCD 与 ecoinvent，共 7,742 条）上开展三组对照实验，以不做清单处理为基线，以去重加缺失替代为处理组。去重实验使用 20 条物料（含复数、大小写与同义表述，5 组真重复），处理组 Precision 与 Recall 均为 1.000，重复记录全部检出且无误删（基线 Recall 为 0）。缺失替代实验在无嵌入的保守模式（按类别回退）下进行，替代使平均绝对误差改善 8.4%，千单位排放偏差改善 77.1%（由 19,143.5 kgCO₂e 降至 4,381.7 kgCO₂e），所有替代因子均自动标注敏感性等级（本组为高），保证结果可审计。端到端实验使用 15 条真实清单（含 2 组重复与 1 个缺失因子），以 ELCD 真实因子值为对照，清单总排放计算偏差由基线的 11.7% 降至 0.0%，共去重 2 条、替代 3 条。这里的 0.0% 表示处理后结果与按真实因子直接计算完全一致，表明清单处理环节未引入额外误差；该实验不涉及排放因子数据本身的适用性评价。

实验三评估完整管线（混合检索与质量排序）的检索质量。在标准基准 17 例（含对抗性用例）上实测，Recall@1 为 52.9%，Recall@5 为 64.7%，MRR 为 0.665。同时对 6 个含 CO₂ 数值的算例量化了因子错配的排放影响：错配情形的平均排放影响为 595.6 kgCO₂e/千单位（正确匹配为 0），最严重个案（国别电网因子错配）为 1,521 kgCO₂e/千单位。该结果从应用层面说明，排放因子检索不宜由服务端直接返回单一结果，而应返回带质量评级与出处的候选列表，由掌握用户语境的调用方智能体复核选择。完整评估日志已归档备查。

（四）预期技术指标与实测对照

预期技术指标与实测结果的对比如表 6 所示。

表 6: 预期技术指标与实测结果对照

指标	目标	实测
词汇失配场景 Precision@1 (纯向量)	≥ 35%	38% (gte-large)
词汇失配场景 Precision@1 (加大模型复核)	≥ 45%	50% (gte-large)
大模型复核带来的 Precision@1 提升 (4 模型)	一致为正	+6 ~ +16 pp
数据覆盖	≥ 5,000 条、多库	8,404 条 / 4 库
清单去重 Precision / Recall	≥ 0.95	1.000 / 1.000
清单端到端排放偏差	≤ 5%	11.7% → 0.0% (处理后)
标准集检索质量 (Recall@5 / MRR)	——	64.7% / 0.665
单次检索响应	秒级	检索毫秒级；含大模型复 核 1.4-2.2 s

3.3 计划和分工

开发计划见表 7。

表 7: 开发计划

阶段	时间	里程碑
需求调研与首版发布	2026.06.14–06.30	痛点验证、数据源调研、在 Smithery、npm 与 PyPI 发布首版
核心研发	2026.07.01–07.31	混合检索与质量排序引擎、MCP 工具、基准构建与消融实验
产品化	2026.08.01–08.15	云端数据分发、商业化流程（支付与许可密钥自动发放）
大赛冲刺	2026.08.16–09.01	材料整理、效果复测、初赛提交
后续规划	2026.09–	中文库接入、检索模型领域优化、计算规范文档编制

在团队分工方面，本项目由 Nike 独立开发完成，其担任项目负责人，负责项目规划、检索算法设计、系统开发、实验与评估以及产品化与文档的全部工作。

4 参考资料

[1] European Commission, Joint Research Centre. European Reference Life Cycle Database (ELCD). <https://eplca.jrc.ec.europa.eu/ELCD9/>

[2] Wernet G, Bauer C, Steubing B, et al. The ecoinvent database version 3 (part I): overview and methodology. The International Journal of Life Cycle Assessment, 2016, 21(9): 1218–1230.

[3] National Renewable Energy Laboratory. U.S. Life Cycle Inventory Database. <https://www.nrel.gov/lci/>

[4] USDA National Agricultural Library. LCA Commons. <https://www.lcacommons.gov/>

[5] Anthropic. Model Context Protocol Specification. 2024. <https://modelcontextprotocol.io/>

[6] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL-HLT, 2019: 4171–4186.

- [7] Reimers N, Gurevych I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. EMNLP-IJCNLP, 2019: 3982–3992.
- [8] Wang W, Wei F, Dong L, et al. MiniLM: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. NeurIPS, 2020.
- [9] Li Z, Zhang X, Zhang Y, et al. Towards general text embeddings with multi-stage contrastive learning. arXiv:2308.03281, 2023.
- [10] shibing624. text2vec: Chinese text vectorization (text2vec-base-chinese). <https://github.com/shibing624/text2vec>
- [11] DeepSeek-AI. DeepSeek-V3 Technical Report. arXiv:2412.19437, 2024.
- [12] European Commission. ILCD Handbook: General guide for life cycle assessment — Detailed guidance. 2010.