



GLOBAL GEOCHEMICAL ATLAS

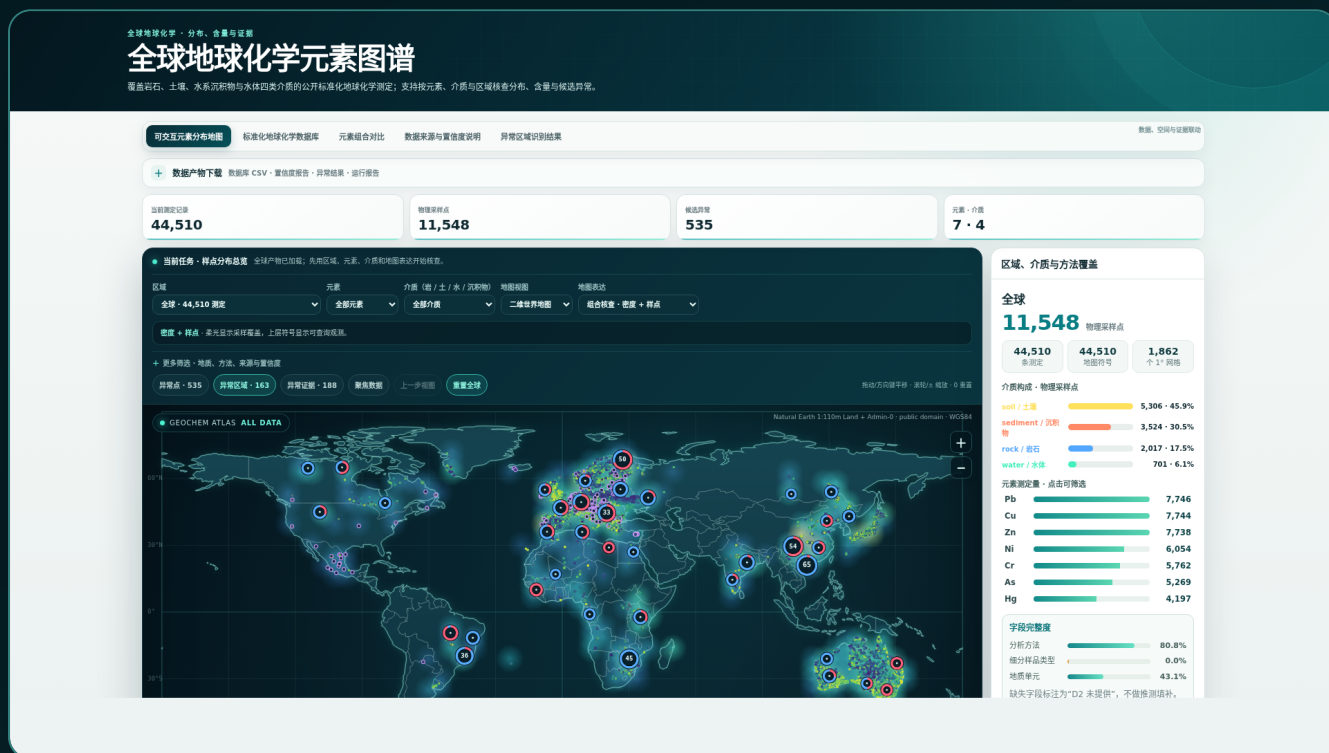
SKILL · global-geochemical-atlas

TEAM 0066 组

AI4S FUTURE SCIENCESKILLS HACKATHON · 气候与地球科学赛道

全球地球化学元素分布图谱 Skill 项目介绍

不是一张地图，而是一套可执行、可追溯、可复现的科研工作流



在线演示 <https://asimfish.github.io/global-geochemical-atlas-demo/>

Skill 仓库 <https://github.com/asimfish/global-geochemical-atlas-skill>

演示仓库 <https://github.com/asimfish/global-geochemical-atlas-demo>



一、赛题与核心方案

赛题要求：基于公开文献及开源数据平台，自动采集岩石、土壤、沉积物、水体等介质中的元素含量、采样坐标、地质背景与分析方法信息，完成单位统一、空间匹配、质量控制和来源追溯；支持按元素、区域、地质单元和样品类型生成全球或区域分布图、热力图及元素组合对比，识别异常富集与亏损区域。**最终交付物仅为一个完整、可复用的 Skill 文档。**

核心方案：一个 Skill，一次运行，交付全部五项输出

自动采集 ACQUIRE

15 个公开来源经准入审计接入；岩石 / 土壤 / 沉积物 / 水体四类介质；每条测定携带逐记录证据。

科学处理 STANDARDIZE

单位统一 mg/kg、坐标统一 WGS84、删失值保留检出限 qualifier、GLiM 地质单元匹配、五维置信度评分。

研究输出 VISUALIZE

全球 / 区域分布图与热力图、元素组合对比、robust-z 异常富集 / 亏损候选，全部自包含可交互。

真实运行指标（expanded run，全程可复现）

44,536

采集原始测定（逐记录证据 100%）

44,510

可上图测定 · 11,548 个物理采样点

15 / 15

来源通过准入审计 · 六大洲覆盖

0.874

全库综合置信度（五维加权）

535

robust-z 记录级候选异常

15 / 15

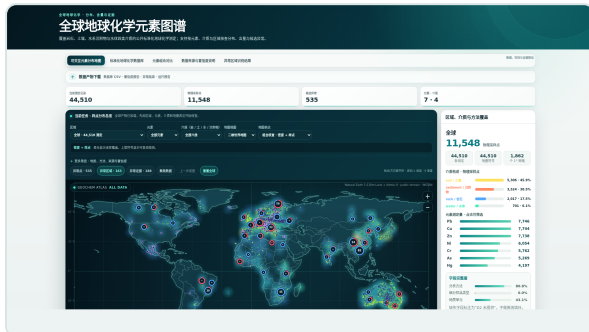
产物契约文件逐项验证通过

数据来源（15 个公开来源 · 覆盖六大洲与全球淡水）

区域	来源	介质
欧洲（7）	FOREGS 地球化学基线 ×6 · GEMAS	土壤 / 沉积物 / 水体 / 腐殖质等六介质
非洲（2）	AfSIS Phase I · PANGAEA 北非	土壤
亚洲（2）	TPDC 中国山地土壤 · GSJ 日本	土壤 / 河流沉积物
大洋洲 · 南极（2）	NGSA 澳大利亚 · GEOROC 南极板内	地表样品 / 火山岩
全球 · 岩石圈（2）	GEMStat (UNEP) · GEOROC 太古宙克拉通	全球淡水 / 岩石



二、赛题五项交付（真实运行产物）



交付 1

可交互元素分布地图

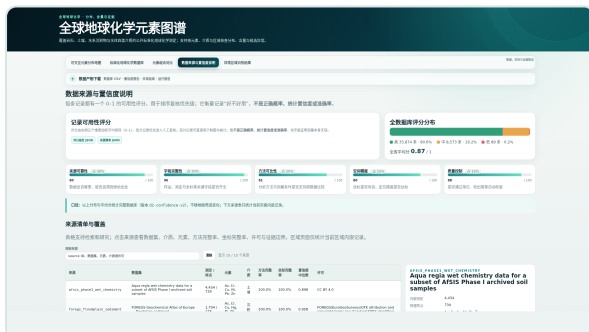
按元素、区域、地质单元、样品类型筛选；热力图表达采样密度，空白区域如实显示数据缺口；任意点位直达原始证据。同一页面内置分布地图、数据库、组合对比、来源置信、异常识别五个研究视图，另附三维点阵地球仪；产物为自包含单文件，断网即点即用。



交付 2

标准化地球化学数据库

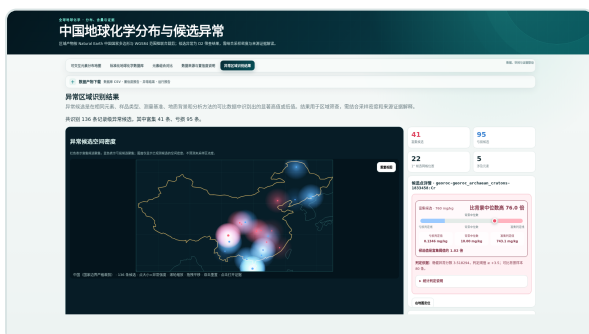
40+ 字段专业长表，一行 = 样品 × 分析物 × 测定；原始值与标准化值双列并存、换算因子独立成列；删失值保留 qualifier 不填补为 0；语义对齐 ODM2，字段映射对照 EarthChem / USGS / IGSN。下载的 geochemistry.csv 即交付物本体。



交付 3

数据来源与置信度说明

每条记录按来源可靠性 30% / 字段完整性 20% / 方法可比性 20% / 空间精度 15% / QC 质量 15% 加权评分，全库综合 0.874，高置信 80.6%。来源清单附许可与覆盖说明；证据链到底：原始值 → 来源 DOI / 表 / 行 → SHA-256 manifest。



交付 4 / 交付 5

异常区域识别结果 + 可复用 Skill 文档

异常三层呈现：地图红环（富集）/ 蓝环（亏损）标记、候选列表一键定位、判据卡片解释“为什么被标记”；判定口径 robust z ≥ 3.5，背景组限定同元素 + 同介质 + 同测量基准。**交付 5 即比赛最终提交物**：完整、可复用的 Skill 文档（SKILL.md + references + scripts + fixtures/tests）。

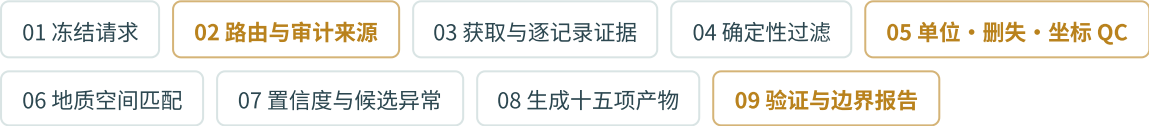


三、Skill 专业设计

四层可执行结构（区别于“长提示词”）

SKILL.md 状态机与决策门 九阶段按序执行，每步有明确输入、输出与失败状态，不允许跳步。	references/ 科学知识库 数据格式、科学规则与来源准入独立成文，按需加载。	scripts/ 确定性执行器 单位、坐标、QC、异常、地图全由可测试脚本执行，核心零第三方依赖。	fixtures + tests 可复现可评测 真实数据切片+哈希重放；有 / 无 Skill 对照评测证明真实提升。
---	--	---	---

九阶段执行状态机（前向确定执行）



金色为决策门：来源准入不过整源拒绝；背景组 $n < 20$ 不做异常判定；出图验证不过拒绝交付。全部失败收敛到 8 个稳定状态码（invalid_input、insufficient_background、needs_human_review 等），Agent 与 CI 均可编程处理。

核心机制：可审计 iteration loop（反向证据修复）

VALIDATE（十五项产物逐文件验证）→ BACKLOG（问题显式登记 iteration_backlog.csv）→ ROUTE（按问题码定向修复：补来源证据或补确定性规则）→ RERUN + DIFF（重跑并按稳定 item_id 比较），最多 2 轮，问题不降即停并转人工复核。本次真实运行显式登记 **114,807** 项：自动处置 89,207 · 人工复核 25,352 · 科学限制 248，**零静默删除**；修复生成新运行，原始证据与标准数据历史不可变。

跨平台复用

采用 Agent Skills 公共文件子集：OpenCode 隔离评测与 Skill 挂载已实测；Codex、命令行与 CI 通过普通文件+命令接口直接调用，零专有 API。模型只做编排与解读，科学计算全部由确定性脚本完成——换任何兼容 Agent，结果一致。



四、创新点

01 证据优先 (Evidence-first) 的记录级溯源

地图上任何一个点，都能追到全库唯一 record_id → 原始值 / 单位 / 限定符 → 来源原生记录 → DOI / 表 / 行定位 → 文件与运行 SHA-256。原始值永不覆盖，标准化值与换算因子另列保存；每个关键结论绑定来源定位。

02 LLM 编排 × 确定性科学计算的严格分工

语言模型只负责解析意图、路由来源与组织交付；单位换算、坐标、QC、异常阈值、哈希等低自由度任务全部由可测试脚本执行。跨模型复用结果一致，从根上抑制“模型幻觉进入科学结论”。

03 Profile 驱动的可复现研究视图

用户问题冻结为 visualization profile，确定性渲染并独立验证；输入、profile 与输出三重哈希锁定——同输入 + 同 profile = 同哈希，任何人可复现任何一次运行。

04 可审计 iteration loop：修复证据，不修改历史

问题显式登记、按问题码定向路由、重跑对比、最多两轮自动收敛；需要专家判断的进入人工门禁，删失等真实科学限制不许猜值“修复”。质量控制从口号变成可检查的产物。

05 失败可编程：8 个稳定失败状态

“未检索到 ≠ 不存在”；失败停在当前阶段并返回状态码、事实与下一步，绝不用下游可视化补造上游证据。Agent 与 CI 都能围绕失败状态编排重试与人工介入。

06 自包含交付与全谱可视化

交互图谱（五视图）与三维点阵地球仪均为单文件产物，无需服务器与外网，评审现场断网照样全功能交互；异常候选带统计判据卡片，可直读“比背景中位数高多少倍”。



五、应用场景与使用方式

地质调查与矿产勘查初筛

按元素组合与区域快速生成研究视图，robust-z 富集候选连同判据与复核链路输出，作为野外核查与靶区论证的前置筛选（候选 ≠ 矿化定论，边界明确）。

环境地球化学基线与风险筛查

多介质基线整编（土壤 / 沉积物 / 水体），亏损与富集双向候选支持重金属污染、生态风险的初步圈定与持续监测比对。

多源数据整编与 FAIR 数据服务

把分散异构的公开数据整编为带证据链与置信度的标准长表，直接服务于科研数据管理、机构数据平台与论文可复现性要求。

地学教学与科普展示

自包含交互图谱与三维地球仪零部署即用，适合课堂演示全球元素分布、采样覆盖与异常识别方法论。

Agent 科研工作流范式复用

四层结构 + 状态机 + iteration loop 是通用骨架：更换 references 与 scripts 即可迁移到水质、土壤碳、生物地球化学等领域，沉淀为机构级 Skill 资产。

跨 Agent 平台协作

同一 Skill 可被 OpenCode、Codex、CLI、CI 等直接调用；团队内不同模型、不同工具链共享同一科学契约与产物标准。

使用方式

① Agent 挂载（OpenCode / Codex 等，自动读取 SKILL.md 进入状态机）：

```
opencode run -m qwen3.8-max "全球 Cr·Cu·Ni·Pb·Zn 元素图谱与异常筛查"
```

② 直接命令行（无需 Agent，确定性复现）：

```
python scripts/run_atlas_request.py --request request.json --analysis-profile production
python scripts/validate_outputs.py --output-dir OUTPUT_DIR
```

③ 在线演示：<https://asimfish.github.io/global-geochemical-atlas-demo/>（12 页演示 + 可交互产物 + 3D 地球仪 + PPTX 下载）

口径声明：置信度衡量工作流可用性，不等于测量正确概率；哈希证明链路一致，不自动证明测量真实；high / low 仅为可比背景组内候选方向，不表述为污染、矿化或成因结论；热力图仅表达采样密度，不插值元素浓度。