



Global Geochemical Atlas: 面向 AI Agent 的可复用科研 Skill

全球地球化学元素分布图谱 —— 把采集、标准化、质控、异常筛查与制图封装为可执行的 Science Skill 结果有来源，缺口有去向

AI4S Future ScienceSkills Hackathon · 气候与地球科学赛道 · 决赛08_词元代理人队



199,730

系测定正式入库
100% 记录级溯源
可回放原始件

45,753

个独立物理样品
四介质同一张表
直接联查

35 / 66

来源审计入门
31 个候选被门禁拦下

3,721

异常筛查候选
只筛选不处理
，判据随行

10 / 10

植入缺陷全部捕获
审计器先校准
再上岗 (RUN C)

393

系未过标准化门禁
知实保留
，不静默丢弃

REAL RUN · 08-19 23:45 → 08-20 02:38

全程自主 · 10 轮循环 · 总时长 2 小时 53 分

1 · Skill 的动机：图好画，可信难 MOTIVATION

一张漂亮的分布图，反而可能放大误判

分布图只能说明哪里高、哪里低，真正难的是解释高值——如西南高砷土壤：套通用阈值，自然高背景就被误判成「污染」。缺少方法、地质与时间的共同证据，图越漂亮，误判越有说服力。

分析化学之间
不同实验室、单位与测定方法，能直接比较吗？

地球化学之间
高值来自母岩与成土过程，还是异常富集？

环境科学之间
是长期自然高背景，还是工业排放的人为信号？

地球系统之间
何时出现、怎样演化，是否随治理而变化？

GGA 的回答：先建可比较、可追溯的统一底座，再用方法分层 × 地质背景 × 多期观测分开检验两类信号——证据不足就如实报缺口。

就算只求一张图，公开数据还有三道坎

- 数据世界极度破碎**
散落在 USGS、PANGAEA、FOREGS 等平台：有的给 API，有的只有 CSV，有的要翻论文附表。
- 凑齐了也不能直接合并**
mg/kg 与 wt% 并存；介质混杂；坐标基准、分析方法各不相同。
- 最磨人的是「未检出」**
仪器报告「低于检出限」不等于零：直接填 0 或减半代入，统计与异常判断就被悄悄带偏。

核心挑战：AI 的「幻觉」在科研里不可接受

单位算错、坐标颠倒，AI 依然能拼出一张**极具欺骗性的精美地图**。核心挑战不是「AI 不能画图」，而是：**全自动接管数据处理的同时，每一个数字都绝对可追溯、科学严谨。**

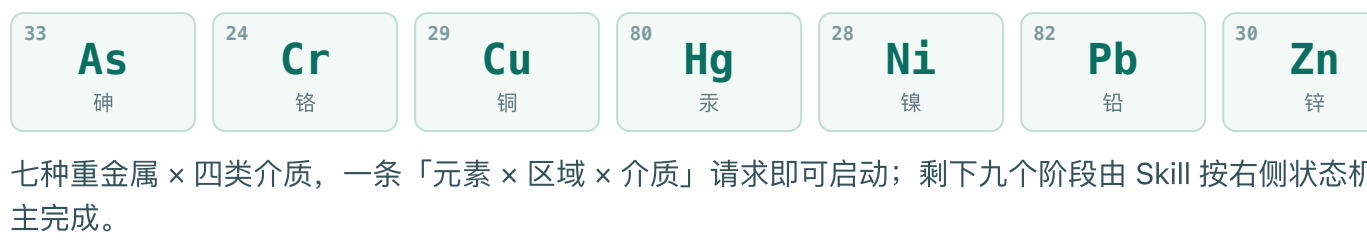
因此分工必须清晰：**人类给科学直觉、定边界 · Agent 编排与解释 · 确定性脚本**包揽全部数值计算。

为守住严谨，我们定了四条底线

- 数字不过模型的手**
全部数值出自确定性脚本，模型只做编排与解读
- 门禁不过就停机**
条件不满足如实报缺口，不用近似值硬算
- 验收员先被考核**
审计器先抓齐 10 类植入缺陷，才有资格验收产物
- 每个缺口都记账**
未达标项逐条登记去向，不静默丢弃

研究者只需要提出一个任务

「使用公开数据生成**西欧 · 土壤 · 砷**分布图，标出候选富集区域，给出数据来源与工作流置信度。」



七种重金属 × 四类介质，一条「元素 × 区域 × 介质」请求即可启动；剩下九个阶段由 Skill 按右侧状态机自主完成。

与「常规 AI 出图」的五个不同

常规 AI 出图	本 Skill
数据获取 ✗抓到什么用什么，来源不留痕	✓准入审计：许可 / 版本 / SHA-256 全存档
数值计算 ✗由语言模型直接「给数」	✓全部出自确定性脚本，模型只做编排与解读
失败处理 ✗静默跳过，或用近似值填补	✓失败即关闭：517 项任务分 202 组定向入队
结果验证 ✗验证脚本跑通即宣布通过	✓审计器先以植入缺陷校准召回率，再采信
结论边界 ✗直接下「污染 / 超标」结论	✓只报告筛查候选，成因结论留给研究者

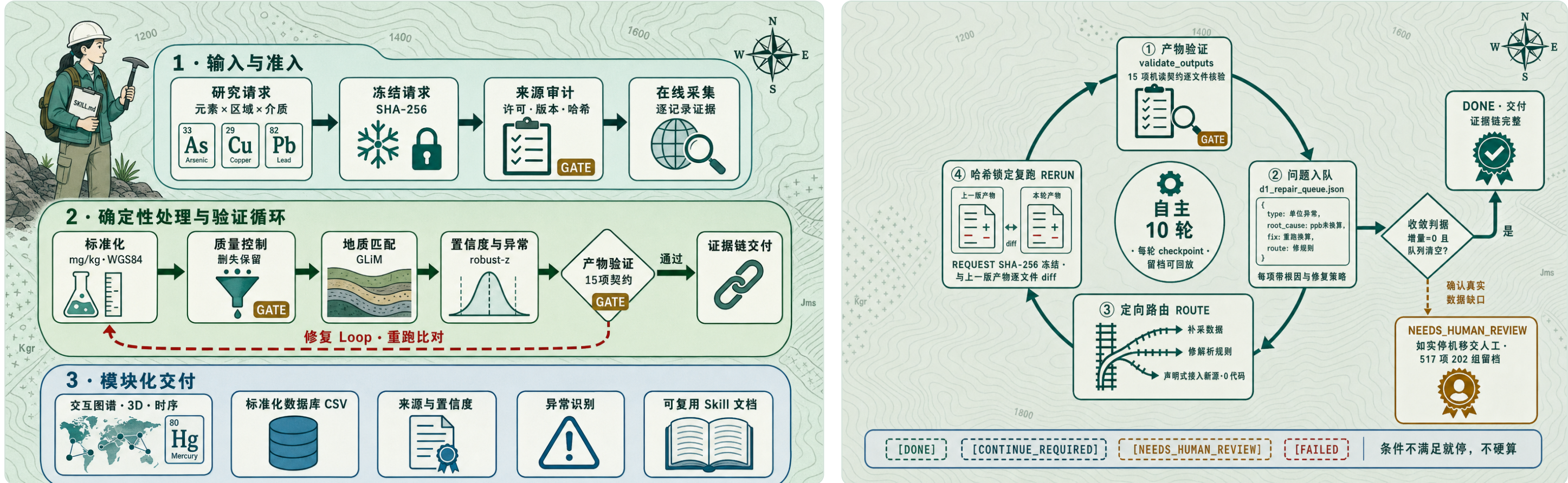
应用场景

近看 矿产勘查靶区初筛 · 环境地球化学评估 · 公开数据标准化 · 可视化教学

远看 证明一种范式：Agent 统筹全局、状态机锁定流程、确定性脚本做固定操作，可复制到水质、土壤碳等场景。

2 · Skill 设计：模型负责编排与解读，确定性负责计算与验证 SKILL DESIGN

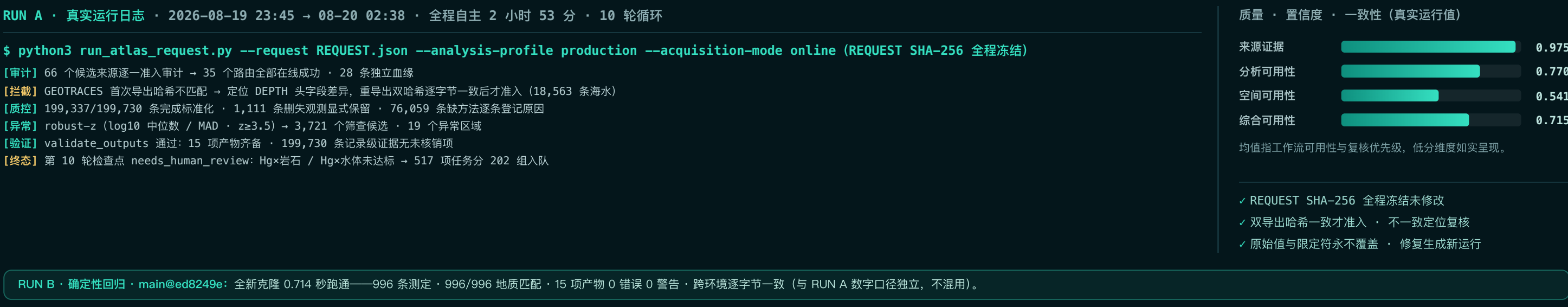
一图看懂 Skill 内核：SKILL.md 九阶段状态机压成三段流水线（左）× 修复 Loop：根因入队 · 三路走向派发 · 哈希锁定复跑与收敛判据（右）



每个数字如何变得可信：Skeptic × Referee 对抗式审计 · 可反查的「记录护照」



3 · 真实运行：上述设计的一次全球尺度执行 REAL RUN



4 · 交付与发现 OUTPUT

赛题要求，逐条落地

自动采集 ACQUIRE	岩石	土壤	沉积物	水体	元素含量	采样坐标	地质背景	分析方法
科学处理 STANDARDIZE	单位统一	空间匹配	质量控制	来源追溯	原始值永不覆盖			
研究输出 VISUALIZE	全球与区域分布图	热力图	元素组合对比	异常富集 / 亏损识别	按元素 / 区域 / 地质单元 / 样品类型			

同一次 Skill 运行 · 同时产出五项交付 ↓

- 可交互元素分布地图
- 标准化地球化学数据库
- 数据来源与置信度说明
- 异常区域识别结果

⑤ 可复用 Skill 文档 + 3D 地球 · 时间演变图 (附加)

Auto-Research 赛题之外再进一步：在图谱之上自动提出选题 → 统计试点 → 成稿论文的全自动科研流程 (成果见下方 DATA → 管线)

时间演变分析 为异常识别加上时序维度：同一格网早 / 晚期含量对比、采样史回顾——区分富集来自母质背景，还是工业排放等新近输入 (图见左侧)。

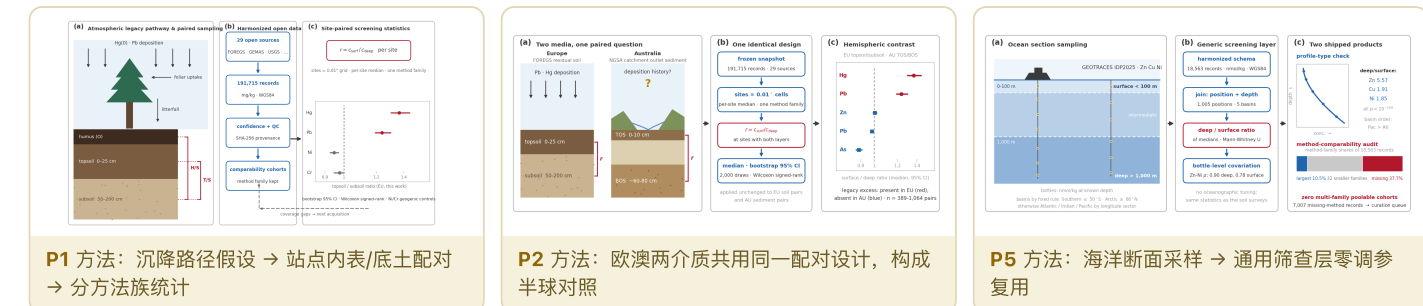
覆盖矩阵：四介质全线打通

	As	Cr	Cu	Hg	Ni	Pb	Zn
土壤 75,537 条 · 15,293 样品							
沉积物 85,699 条 · 12,959 样品							
水体 29,203 条 · 12,318 样品							
岩石 12,291 条 · 4,763 样品							

DATA → PAPERS: 图谱之上的自动科研

拿到全球图谱只是起点：Auto-Research 管线从同一冻结快照自动提出选题、跑统计试点、写成论文——**5 篇**全部成稿 · 合计 **46 页 · 24 图 9 表** · 每篇回答一个领域内明文悬置的问题 · 全过 a-g 七项质量审计——证明这份数据据真的能支撑科学发现。

- P1 欧洲土壤 Hg/Pb 遗留富集：按方法分层检验**
Hg 表 / 底土 1.36 × Pb 1.24 · 地质对照 Ni/Cr ≈1 10 页 · 4 图
- P2 表层富集并非全球常态：澳大利亚检验**
澳 Pb 0.982 (1,064 站) vs 欧 1.24 9 页 · 5 图
- P3 同一份土样，两种方法可差三倍**
Cr 2.94x · Cu 反向 0.86 · 误差 18.2-11.0% 9 页 · 5 图
- P4 全球碎图的地面验证：共址一致性**
26 共址单元同址一致到 2% · 73% 落在 2 倍带 9 页 · 5 图
- P5 零调参恢复已知海洋学：GEOTRACES 审计**
Zn 深 / 表 5.57x · 太平洋 24.5x > 大西洋 13x 9 页 · 5 图



同类数据库对比：此前不存在的一层

大型汇编库在覆盖广但保留发表原格式，专题库标准化好但限于单介质单大洲；GGA 不比规模，建的是它们之间缺少的一层。研究者的六个真实问题，逐一对照：

Q1 拿到能直接分析的？ 全部统一 - mg/kg + WGS84 + 129 字段，四介质直接联查，免溯源清洗	Q2 不同方法能合并吗？ 逐条标注方法族 (缺失 36.3% 留痕)，P3 给出留痕输出的换算模型
Q3 数字可信可追溯吗？ 100% 级 SHA-256，任何数字可回放至原始下载件	Q4 能合规复用再发布吗？ 记录级许可证 100% 覆盖，可按行筛出可再分发子集
Q5 缺口在哪、下步测什么？ 缺口输出为机器可读采集优先级队列，直接支持采样规划	Q6 单条记录可置信？ 五维置信度评分 100% 覆盖，随记录进入筛选与敏感性分析

对比对象 (官方页面核验, 2026-08-20): GEOROC 4,270 万值 / 70.7 万样品 (按发表原样汇编); EarthChem 联邦 ≈5,000 万值 (各库保持原口径); USGS NGDB 152.4 万样品 / 5,900 万测定 (108 种方法并存, 无统一层); GEOTRACES 123 航次 / 220 万值 (仅海水); GEMAS / FOREGS / NGS / LUCAS / AFIS / CGB (各大洲单库); GGA 补的正是 OneGeochemistry 国际倡议指出的缺口：数据碎片化在数千个库中、几乎无法复用——跨源、跨介质、记录级可审计的统一层。