附录 A

指标、数据与研究方法

本附录说明如何阅读、核对和复用报告中的证据。第一章给出价值与指标框架，第二至第五章分别提供地图、榜单、技术全景与模型生态观察，第六章据此提出行动建议。核对一项结论时，应同时确认它观察谁、采用什么指标、覆盖什么时间与样本，以及哪些内容尚不能据此判断。

A.1 指标框架与三个观察视角

OSDGs 确立发展方向，OSDI 以四大支柱、十个维度、二十九项指标组织观察。框架的完整，不等于所有指标均已完成同等覆盖的测算。本报告不将不同单位、资产类型和证据状态合成一个总分，也不把指标角色解释为评分权重。

参与、贡献与影响力分别回答不同问题。它们可以共同解释生态，但不能互相替代；规模增长、位次变化和公共价值也需要分别取得证据。

观察视角	回答的问题	阅读时先核对
参与	谁进入了协作，参与基础怎样分布？	活跃判定、去重对象、平台与地域覆盖
贡献	谁在具体协作中投入了工作？	行为与关系范围、贡献算法、项目边界
影响力	谁在协作或复用网络中发挥连接作用？	网络范围、算法版本、主体映射与时间窗口

完整指标定义集中在指标字典。正文采用的实际统计对象、来源和比较范围，以相应图表及专题方法说明为准。指标字典中的关联文件是查证入口，不自动证明该指标已经完成计算。

A.2 统计周期、快照与平台范围

代码生态的主统计期为 2025 年 7 月 1 日至 2026 年 6 月 30 日。2022—2026 报告年度沿用上年 7 月至当年 6 月的约定；实现时保留起止边界及所用时区。资料的读取日期用于定位版本，不能延长观测期。

时间类型	本报告中的用法	比较边界
报告年度	代码生态的年度地图、榜单与历史比较	先核对平台、对象映射和统计规则是否一致
月度序列	技术类别变化、模型新增与月榜结构	使用图中实际月份；不补造缺失月份
时点快照与回溯	模型地域样本、开放材料及按创建时间回溯的存量	快照回溯不等于各年独立保存的历史全量

代码生态以 GitHub 为主要观测范围，Gitee、AtomGit、GitLab 在有明确来源和适用范围时作补充。GitCode 与 AtomGit 合并后的品牌统一使用 AtomGit；历史原表保留原始字段和来源日期，复用时记录对应关系，避免重复计入。

2.1 的 GitHub 地图、编制组五域统计和国内平台补充分别使用各自范围。国内平台补充量只用于平台覆盖比较，不再加到五域中国记录，也不按比例分配至省份。模型月度、年度回溯和十月补充快照，同样不能拼接为一个统一时间序列。

A.3 活跃对象、去重与归属

活跃开发者依据统计期内的有效协作识别，剔除已识别机器人与自动化服务账号。报告关注创建 Issue、Issue 评论、提交 PR、PR Review 评论、PR 合入等协作行为；Star、Fork 和点赞可作为各自用途的补充信号，不能直接当作上述协作工作量。

平台内以稳定账号标识去重，跨平台以核验后的身份映射连接。项目—仓库映射按报告年度保留有效边界，处理多仓库、镜像、转移与拆分。不同图表是否实际使用同一映射版本，需要随结果核对，不能仅因具备映射能力便假定所有表已经统一。

所在地、组织归属与项目归属是不同变量。开发者所在地来自公开资料的标准化结果，年度比较使用对应历史位置快照；它不等于国籍。多平台位置冲突先核验，无法消解的层级记为未知，仍保留其他可靠层级。

归属对象	本报告的解释	不直接解释为
开发者地域	公开所在地的观测属性	国籍、出生地或全部工作地点
项目与组织地域	相应榜单采用的公开组织关系	所有贡献者或使用者所在地
模型发布方地域	发布组织或账号的已关联归属	训练地点、部署地点或数据来源地

未知地域保留在适用的全球总量中，并在地域分析中单列；图表仅采用可识别样本时，需写明分母。中国省级展示保留台湾省、香港和澳门的相应记录，沿用已核对的地域底图与映射。未覆盖地区不解释为没有活动。

A.4 活跃度、贡献度与网络影响力

OpenDigger 的 Activity 在开发者—仓库—月份层面，将各类协作次数按权重相加后开方；仓库活跃度再汇总各开发者的结果。它描述行为强度，不是开发者人数，也不是本报告的综合发展分数。

协作行为	官方方法列示权重
创建 Issue	2.2235
Issue 评论	0.5252
提交 PR	4.0679
PR Review 评论	0.7427
PR 被合入	2.0339

上述权重依据 OpenDigger 官方方法说明，核对日期为 2026 年 10 月 22 日。它们不代替各数据文件的实际生产记录，也不是二十九项 OSDI 指标的权重。月度人数不能直接相加为年度去重人数；活跃度、计数与网络得分应分别保留单位和汇总方式。

社区 OpenRank 使用项目内 Issue、PR 等协作关系观察贡献；全域 OpenRank 通过开发者与仓库的协作网络观察影响，并保留历史继承。只有记录了算法类型、版本、输入范围、时间窗口和主体聚合方式，才能判断两组得分是否可比。历史继承也意味着，不能仅凭 OpenRank 大于零便认定对象在当期活跃。

项目、企业与基金会的汇总需要核对有效映射、交叉归属和聚合规则，不默认将多个仓库得分相加为一个项目分数。人才观察采用的年度全域 OpenRank 前 5% 规则，应固定比较集合和并列处理；不将这个阈值自动用于其他主体。

方法来源：Activity、社区 OpenRank、全域 OpenRank。

A.5 地图、榜单与技术全景的比较边界

第二至第四章观察同一生态的不同侧面。地图强调地域分布，榜单强调已界定主体的相对位置，全景强调功能与类别。它们共享来源记录和对象映射的要求，但不因此具有相同的样本与分母。

研究板块	已采用的观察方式	必须保留的边界
GOSO · 第二章	参与、贡献、影响、人才、资产、多样性与治理的主题地图	区分全域、已识别地域、五域统计和专题样本；地域集合不随意相加
GOSR · 第三章	行政区划、企业、基金会、项目、新势力与模型榜单	说明主体、榜单深度、筛选和并列规则；样本第一不自动等于所有对象中的第一
GOSL · 第四章	4.1 的分类、月度比较与贡献观察；4.2—4.6 的全景与研究框架	分类说明功能，不代表质量评分；后五节不据全景图补造排名或演变结论

2.2 的贡献来源图使用开发者所在地到项目归属地的有向记录，未列来源份额不分摊给国家。网络中的底图纹理、方向光点和统一线宽增强仅用于阅读，不增加贡献者、事件或贡献数值。

月度位次、得分与份额是三种不同读数。位次上升并不必然代表绝对得分增长；Top20、Top100、Top1000 相互包含，不能相加。列表外的对象不自动记为零分或最后一名。

技术分类保留来源、版本和归类规则。一个项目可能承担多个功能，跨类别统计只有在说明是否互斥及如何去重后才能汇总。Agent Infra 与 Model Infra 的对比，应沿用各图声明的对象及分母。

A.6 模型供给、派生网络与开放材料

第五章连接供给、复用与开放条件，采用多套独立证据。模型仓库数不等于独立模型家族数；累计下载不等于用户数、部署量或性能；公开可见也不等于已满足特定开放定义。

证据组	本版范围与时间	正确的比较方式
5.1 供给	2026-09-13 可见模型的年度回溯；2026-10-16/17 的平台与组织补充快照；独立月度新增序列	组织地域样本与全平台分母分别标注，不由快照差异直接解释增长
5.2 派生网络	2026-10-16 关系快照，5,777 个节点、5,111 条声明关系	基座指向派生；非随机样本，不代表全平台；年份选择不生成虚构网络历史
5.2 跨国流向	原报告中双端国家可识别的跨国子集，共 10,943 对关系	方向为基座来源国到派生发布国；不能解释为人员、资金或部署流动
ModelRank 榜单	第三章模型月榜及 5.2 的 2024-07—2026-08 国家席位序列	按各榜单的完整模型 ID、月份和榜单深度比较；席位份额不是得分份额
5.3 开放观察	2026-09-14 提取的 870 个报告样本；2026-10-17 的 884 条补充记录及独立固定版本案例	两套样本未完成身份对齐，差值不代表净增长；案例不推及总体

ModelRank 已用于本报告的模型榜单和相关观察。报告保留所用月榜结果及来源，不以下载量或自行假定的权重替代其得分，也不声称已在本报告中复算全部上游算法。跨月归一化得分不直接计算增长率；月榜空白表示榜单未覆盖，不能据此判断首次发布或退出。

派生关系来自发布者明确声明，不等于训练过程已独立核验；缺少声明也不证明模型原创。模型关系数、派生模型数和去重模型数分别计数。跨国图中“双端国家可识别比例”与“跨国子集占比”的分母不同，不能互换。

开放观察以具体发布版本为单位，分别核对材料完整性与许可条件。L 与 O 的释义、源平台判断和独立案例分别保留；待核实不等于不存在，材料可见不等于运行验证完成。报告分类、OSI 开源 AI 定义和 DPGA 正式认可的区别，集中见附录 B。

A.7 缺失、示意与可比性

保留真实的证据状态，是跨地区、跨年度比较的前提。缺失、未覆盖和真实为零分别展示；图件为便于理解而使用的示意，也应与统计结果分开。

状态	含义	呈现原则
观测为零	在所述范围与规则下得到零值	保留零值与适用分母
未知或待核实	归属、材料或判断证据尚不足	单列，不自动归零或判为不存在
未覆盖	平台、样本或榜单未包含该对象	不推断其活动、等级或退出
示意或模拟	用于解释关系、交互或研究设想	就近明确标注，不纳入实测合计或研究发现

分母为零时不计算比例或增长率。比例同时说明分子、分母、对象和时间；按类别汇总前核查重叠，跨期比较前核查平台、映射、地域快照、算法与样本变化。必要时固定可比样本，或将不同口径并列展示。

图中的面积、颜色、线宽、节点大小和空间距离，只按图例赋予含义。关系示意和流动动画不能证明因果；缺少历史记录时，不用动画补造演变。第六章的连线表达建议路径，不表示数量、权重或已经验证的政策效果。

行为、网络和材料数据提供观察线索，不能单独证明经济收益、治理质量或公共成效。相关判断还需要具体使用、维护、受益和评价记录。

A.8 来源记录、查证与复用

从图表回查时，依次确认图号与结论、对应底表或冻结快照、指标及计算口径、来源与时间范围，再核对算法、对象映射和展示转换。读取日期、原始统计期和报告版本应分别记录。

每份可复用材料至少说明：来源与获取日期，统计起止或快照日期，主体与样本，指标及单位，缺失处理，转换与聚合规则，以及对应的文件或版本标识。数值表保留原始精度，图面可采用一致的显示精度；显示取整不反向改写源数据。

数据文件清单与当前仓库同步维护，不在正文写死文件数量。原报告数据、补充平台快照、官方案例与示意材料分别登记；文件入库、口径对齐、计算可复现和结论复核是不同状态。方法说明的修订不代表历史数据已全部重算。

Web 与 PDF 使用同源正文和图件数据。交互筛选用于探索，PDF 则明确所选范围并保留必要比较；两版的方向、单位、分母、样本状态与结论一致。复用时注明报告版本、章节或图号、统计范围及原始来源。

集中查证：指标字典、数据文件清单、来源登记。模型开放条件参照见附录 B。