

CodeGraphRAG: 基于LLM的大规模代码分析 及在 Eagle DevAgent 中的应用

李明宇 | 中科院计算所-京兆人公司 联合研发团队

科技生态圈峰会 + 深度研习

—1000+ 技术团队的选择



敦煌站

K+思考周®研习社

时间: 2025.08.29-30



上海站

K+金融专场

时间: 2025.10.17-18



香港站

K+思考周®研习社

时间: 2025.11.25-26



K+峰会详情



上海站

AI+研发数字峰会

时间: 2025.05.17-18



北京站

AI+研发数字峰会

时间: 2025.08.08-09



深圳站

AI+研发数字峰会

时间: 2025.11.28-29



AIDD峰会详情



李明宇

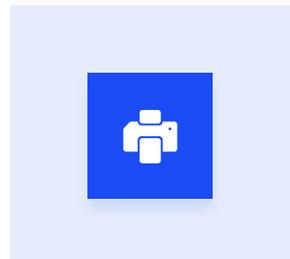
中科院计算所高级工程师 / Eagle DevAgent 作者

中科院计算所高级工程师，近20年软件研发与产业化工作经历。历任国家重点实验室课题组负责人、创业公司CTO、上市公司事业群技术总监和首席技术专家。荣获中国软件协会“优秀CTO”等多项荣誉，任中国新一代IT产业联盟分委会秘书长、全国高校人工智能与大数据创新联盟专家委员、北京开源创新委员会委员、中国开源软件创新大赛总决赛评委。

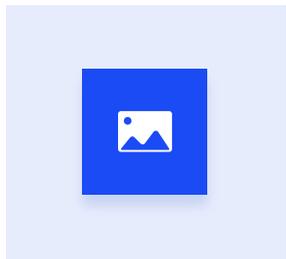
01

你在面向ChatGPT编程吗？

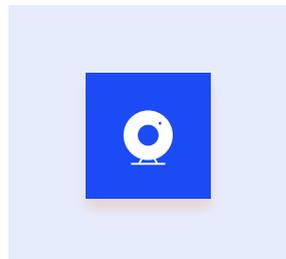
▶ 你在面向 编程吗?



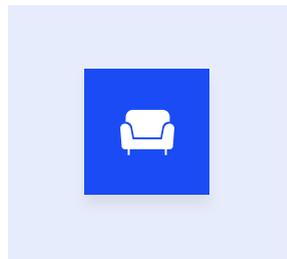
→ 面向
Google编程



→ 面向
Stack Overflow编程



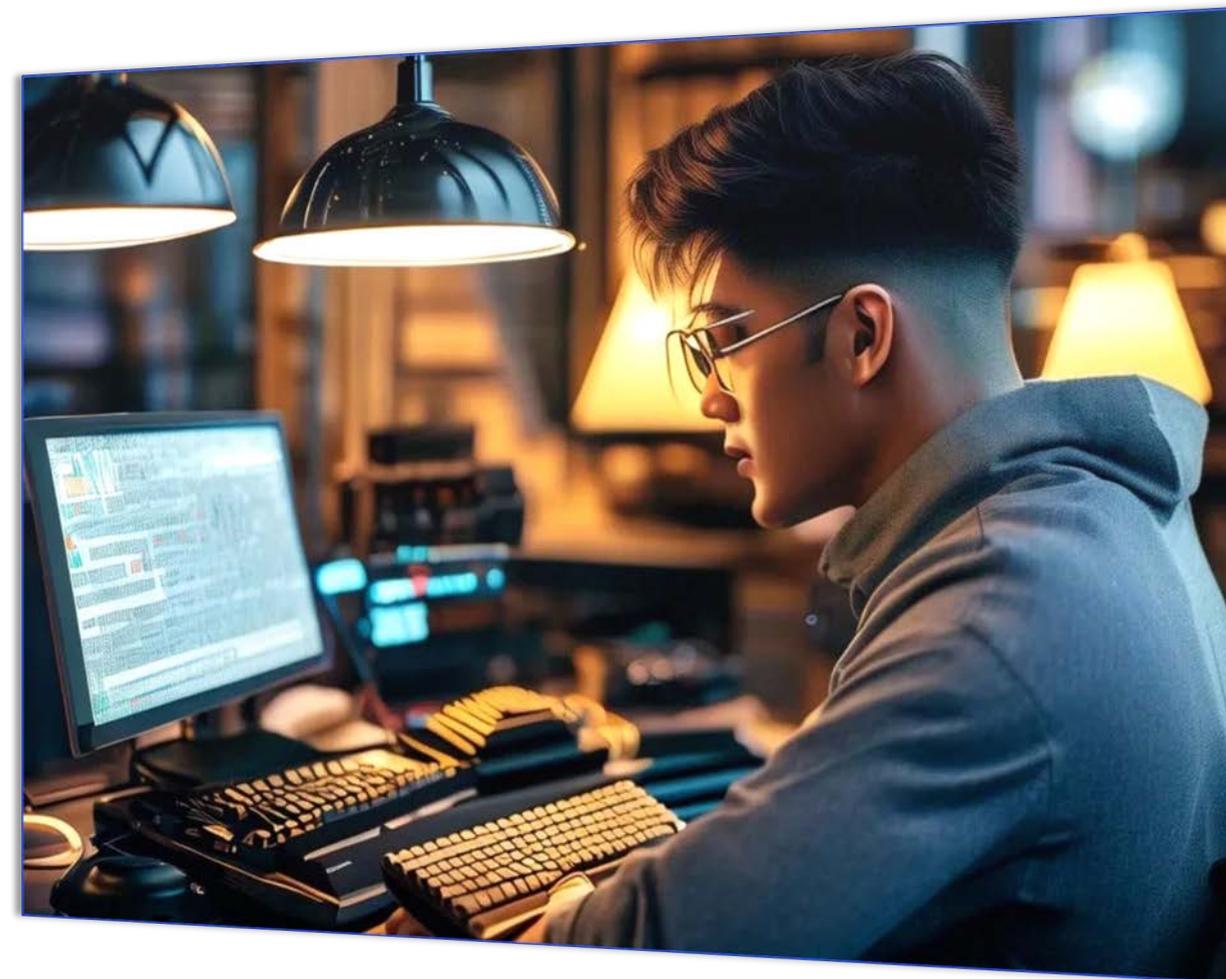
→ 面向
Github编程



→ 面向
ChatGPT编程



→ 面向
微信编程



有了 Copilot(s) 我们还是需要它们, **为什么?**

它们的表现, **你满意吗?**

还记得这些时刻吗?

当学生做论文

你的课题是基于某个已发表的开源项目进行改进，但是怎么也捋不清楚这个开源项目代码的逻辑。而且隐隐觉得跟论文写的不太对得上?

升职为架构师

公司的新产品，是基于开源项目改?
还是完全自研?
需要迅速读懂几个典型开源项目，做出选型决定

公司/部门的技术主管

乙方的代码如何验收?
是否真正掌握?
一年前的项目要做二期，有的同事已经离职了，是继续开发还是推倒重来?

.....

毕业了刚加入一个团队

熟悉公司的“祖传代码”成为了职业生涯的第一个挑战，看也看不懂，改也不敢改~

开始带一个团队

每天要处理各种会议、流程，还要清楚掌握团队开发工作具体情况。组会听完汇报总还觉得有些地方没搞清楚，想看看大家写的代码但时间又不够用。

作为一个CIO

我们系统里以前不可避免地引入了很多舶来的开源代码，是否都能做到可控?
我们的已经投入的研发成本得到的代码，是否能够保值增值?
还是随着人员流动存在较高损耗?

项目级代码分析的难点

1

大模型上下文长度：
10万行代码 ≈
1M+ tokens

2

代码-自然语言理解的问题：
Embedding计算
向量距离不适用

3

增量处理：代码可能每天都在变（增、改、删）

4

跨项目：前后端、微服务.....

5

跨语言：Python、java、js ...（所谓“主流”编程语言，有几十种）

6

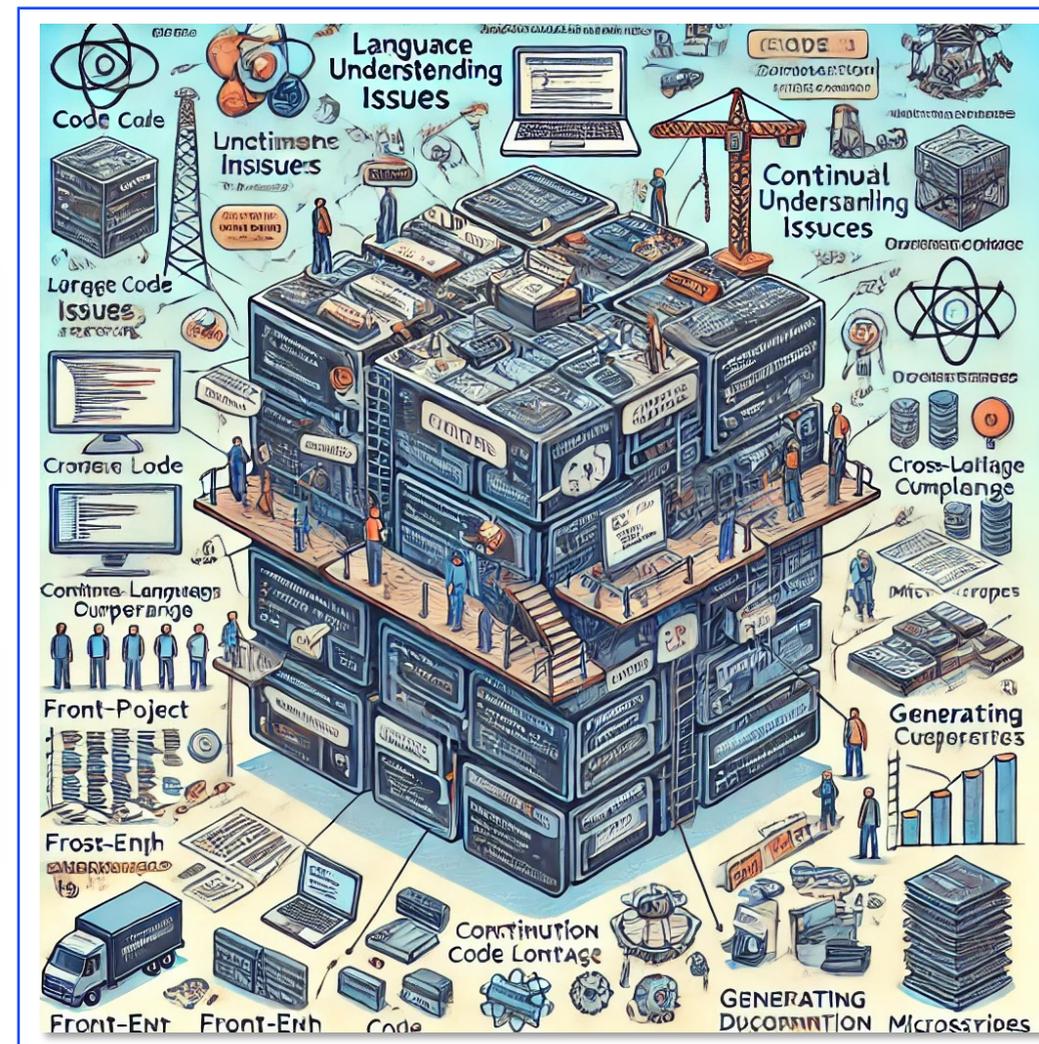
代码和文档的结合

7

从代码生成UML图

8

指定“版本”
(Branch、Release、Tag、Commit)

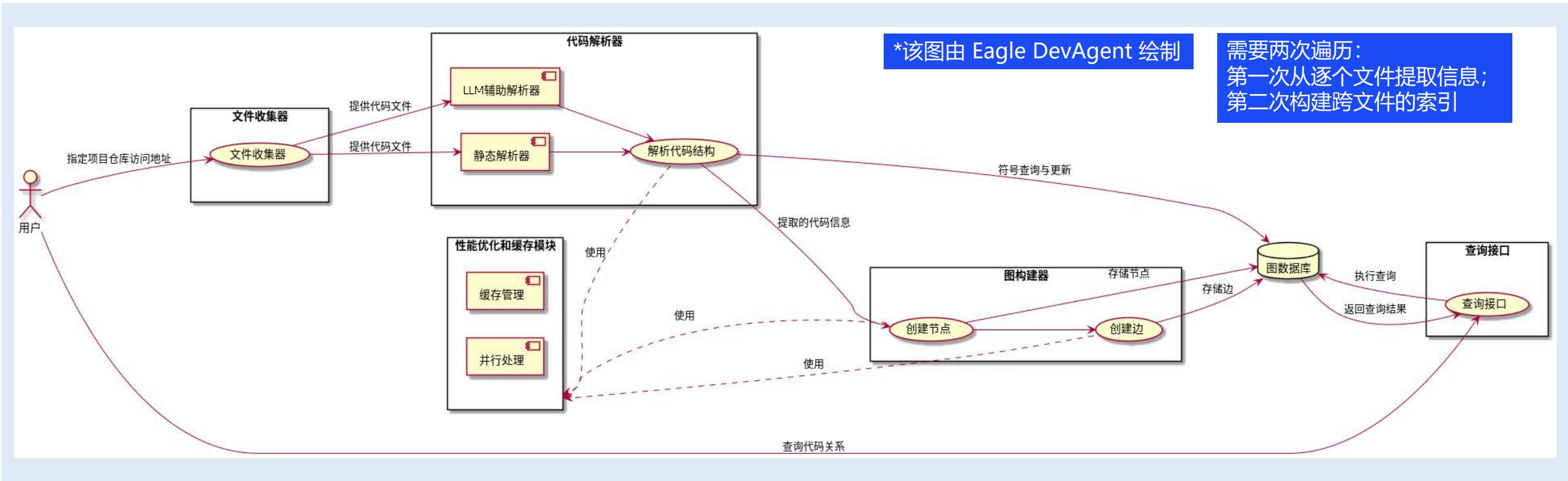


02

CodeGraphRAG

让 AI 理解项目级代码

整体架构与逻辑



图节点属性

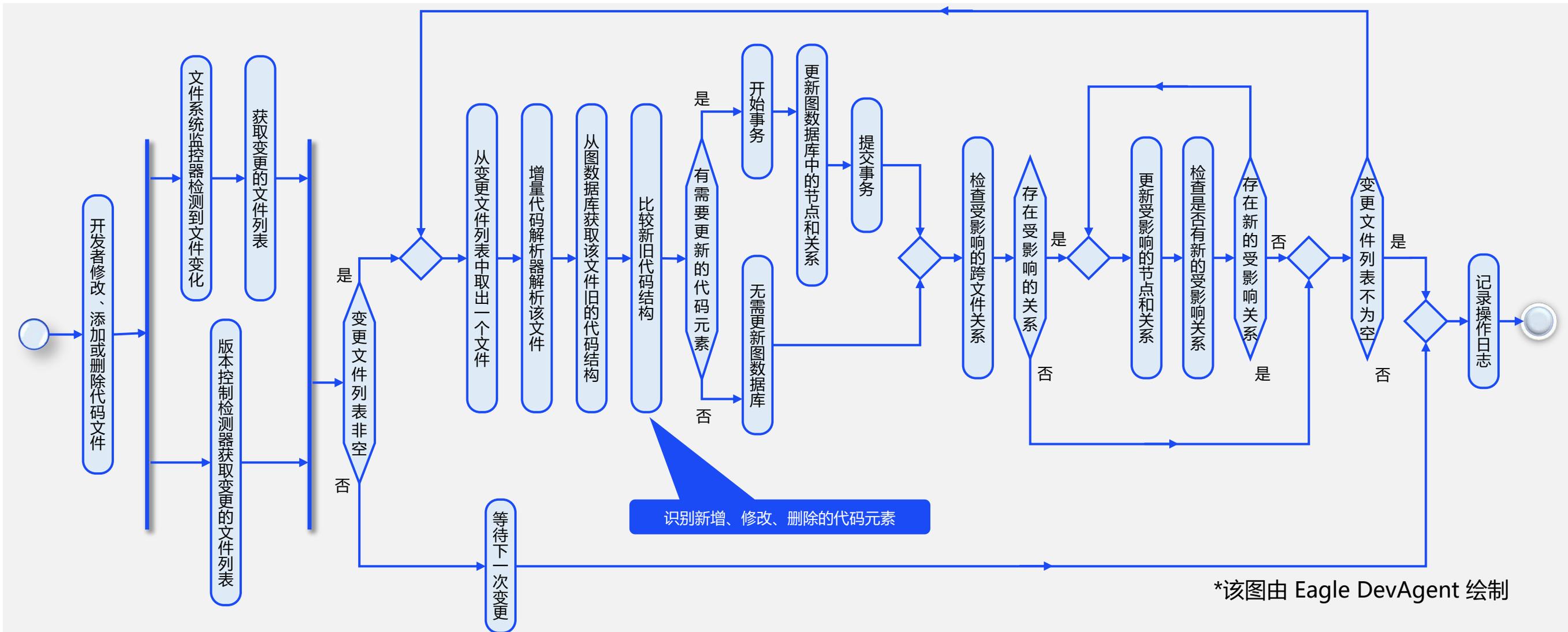
- `id`: 唯一标识符
- `name`: 名称
- `type`: 节点类型 (File、Class、Function、Var)
- `file_path`: 文件路径
- `language`: 编程语言
- `full_name`: 全限定名
- `line_number`: 行号, 便于定位
- `description`: 描述 (LLM生成)

边类型

- ****CONTAINS****
 - 表示层次结构
- ****EXTENDS****
 - 表示继承或接口实现关系, 属性`kind`区分
- ****DEPENDS_ON****
 - 表示文件、模块、类型、函数之间的依赖关系
- ****CALLS****
 - 表示函数或方法之间的调用关系

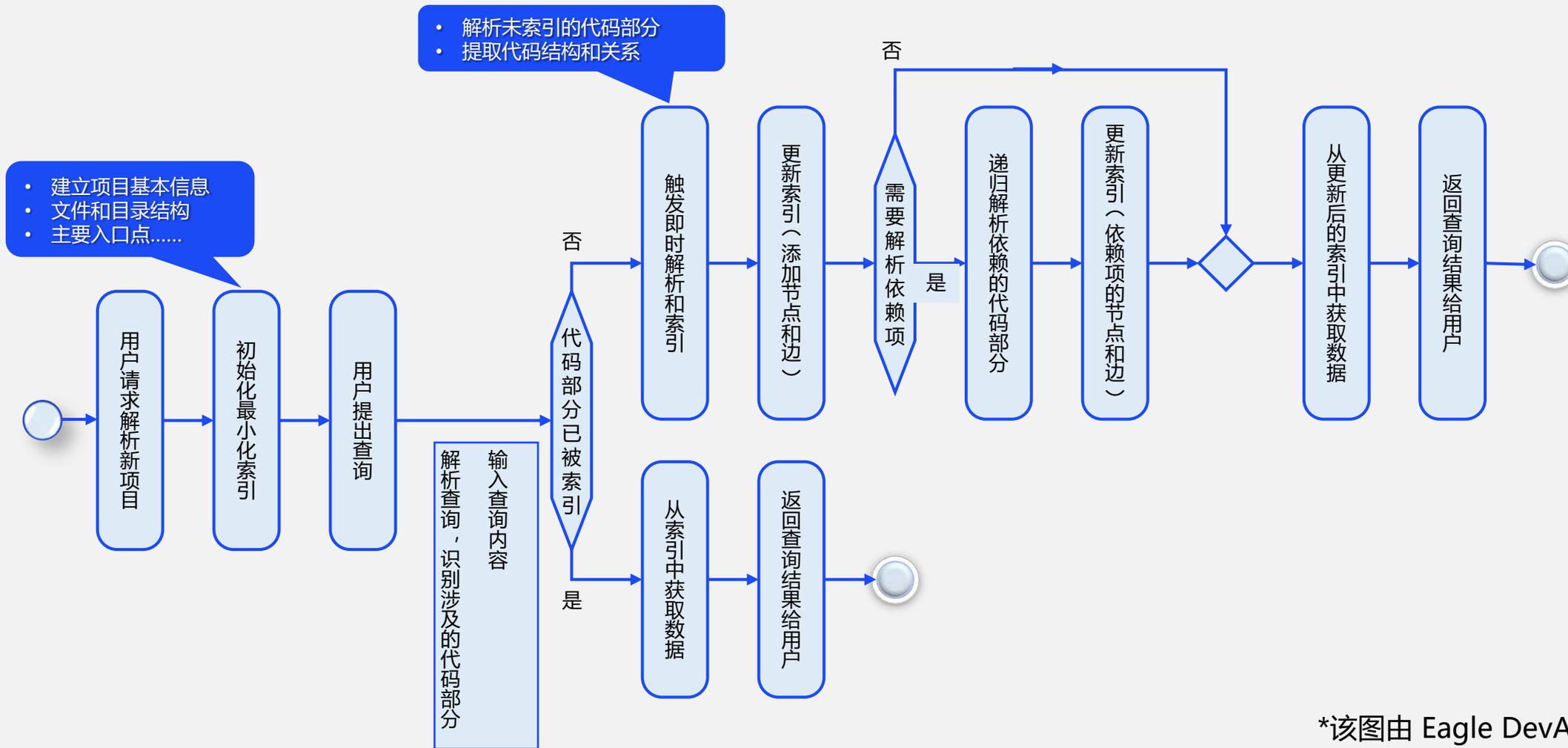
Incremental Indexing

对修改的文件重新构建索引，并递归修改索引图中的相关部分。



▶ Lazy Indexing

延迟索引
以最快的速度
响应用户提问



*该图由 Eagle DevAgent 绘制

03

Eagle DevAgent及应用



注：视频的部分片段进行了加速

Eagle DevAgent

https://github.com/o...
 https://github.com/o...
 渲染如下UML： @st...
 https://github.com/al...
 一个系统设计文档如...
 我们在写一个系统设...
 我们在写一个系统设...
 https://github.com/...
 https://github.com/...

Eagle DevAgent
 扫码免费领不限量使用机会/进群交流

会员用户 | 截止至：2025-10-31

代码问题，就问Eagle!

Eagle是一个Repo级代码分析AI，它支持通过自然语言对程序代码进行搜索，以及在整个代码库上下文中的代码解读，解决新项目上手和代码维护中的复杂问题。请把你需要解读的项目的GitHub链接及相关问题告诉我，或者以zip包的形式上传给我。（如果文件数量较多，Eagle对Github Repo的处理效果可能明显优于zip包形式）
[Eagle用户使用帮助文档>>](#)

https://github.com/WordPress/WordPress的5.0和6.0版本中对于编辑器功能有了什么升级
 分析 https://github.com/vesoft-inc/nebula 中 Raft 算法的具体实现?

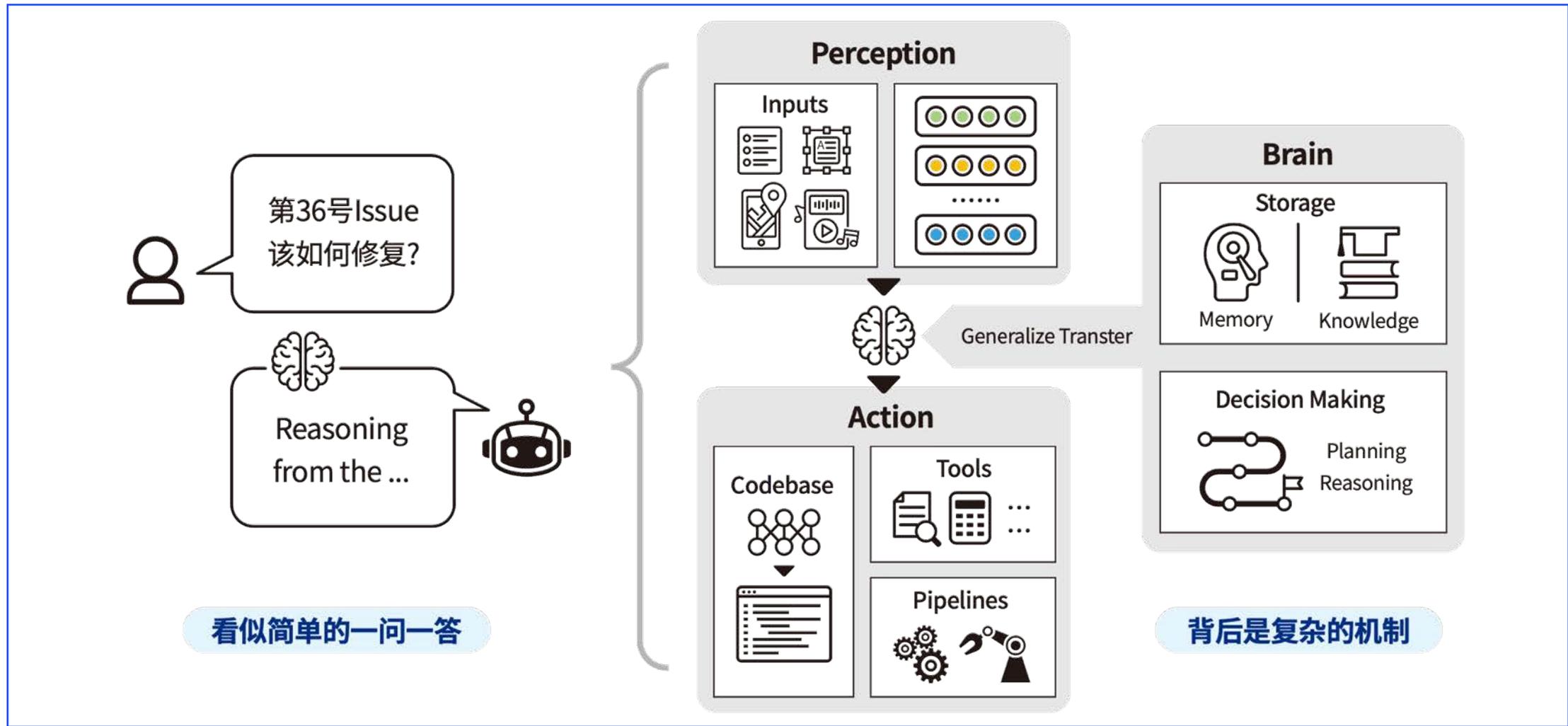
请点击此处上传文件

请在此输入您需要解读的内容

v1.2.0 京ICP备2024078129号-1 Copyright © 2024 北京兆人智能技术有限公司 - 人工智能生成内容仅供参考，使用前请校验内容的正确性。

▶ 代码问题，就问 Eagle !

并不是简单把 RAG 结果返回给用户。
而是通过智能体推动一个分析过程，以解决较为复杂的问题。



Eagle DevAgent 的功能

01 十万到百万行量级的全项目和跨项目程序代码理解

02 支持多语种自然语言对代码库进行快速搜索和答疑

03 从代码一键生成架构图、类图、流程图、思维导图

04 深度解读Github项目，掌握开源项目不再困难

05 广泛适用几十种主流编程语言，跨编程语言项目解读无障碍

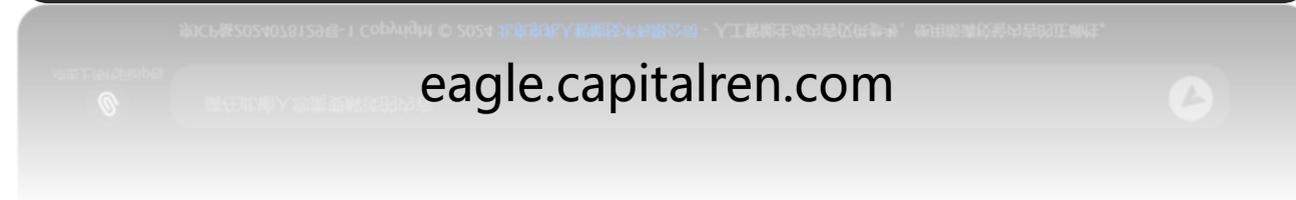
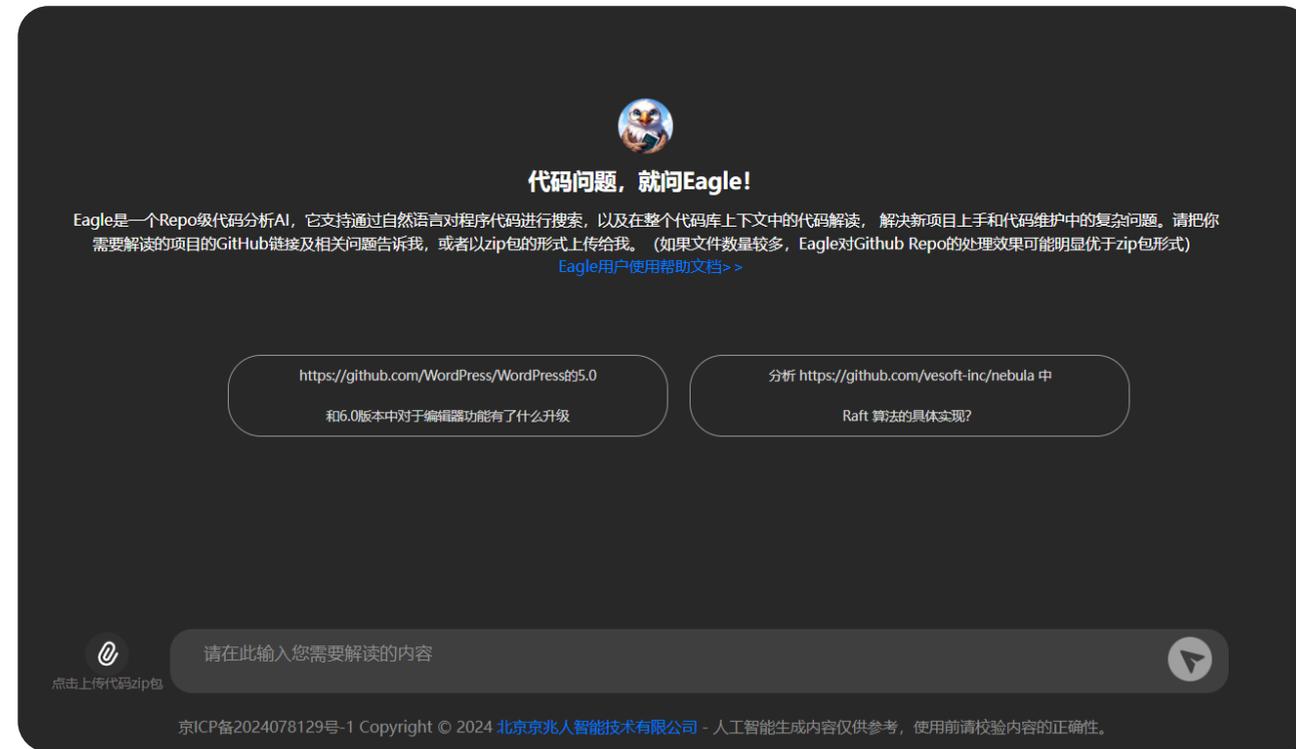
企业版功能

06 企业研发环境接入 (Gitlab、BitBucket、Jira)

07 代码质量效率评估

08 Issue 分析与修复建议

09 融合企业的开发文档和知识库



典型场景一：新项目接手

当团队接手一个二次开发项目，或者开发者进入一个新的项目时，帮助开发者快速分析项目的代码结构和逻辑，减少阅读文档和梳理代码的时间，迅速融入新项目，实现高效的开发对接，降低工作中错误几率。



典型场景二：国外开源代码自主可控

对开源项目进行深度分析，帮助企业自主掌控开源代码的应用和二次开发。通过Eagle，企业和开发者能够全面了解开源项目的代码逻辑，确保在关键场景下实现自主可控。



典型场景三：研发质量和进度风险预警

通过对代码的智能审查，识别代码中的潜在问题，发出**质量警示**，给出改进建议。Eagle还可以对不同技术点和人员的**研发进度**进行评估，识别项目进展中的关键瓶颈。帮助研发主管有效预测研发风险并做出预防措施。



典型场景四：软件代码审计与验收

可用于支持各种高要求的代码审计与验收场景，包括**融资尽调**、**代码资产化**、软件委托开发项目的**验收及获奖评审**等。Eagle协助评审专家进行深度代码分析，评估代码的**先进性**、**原创性**和**潜在风险**，提供详尽的代码评估报告。



04

未来展望

Eagle Workspace

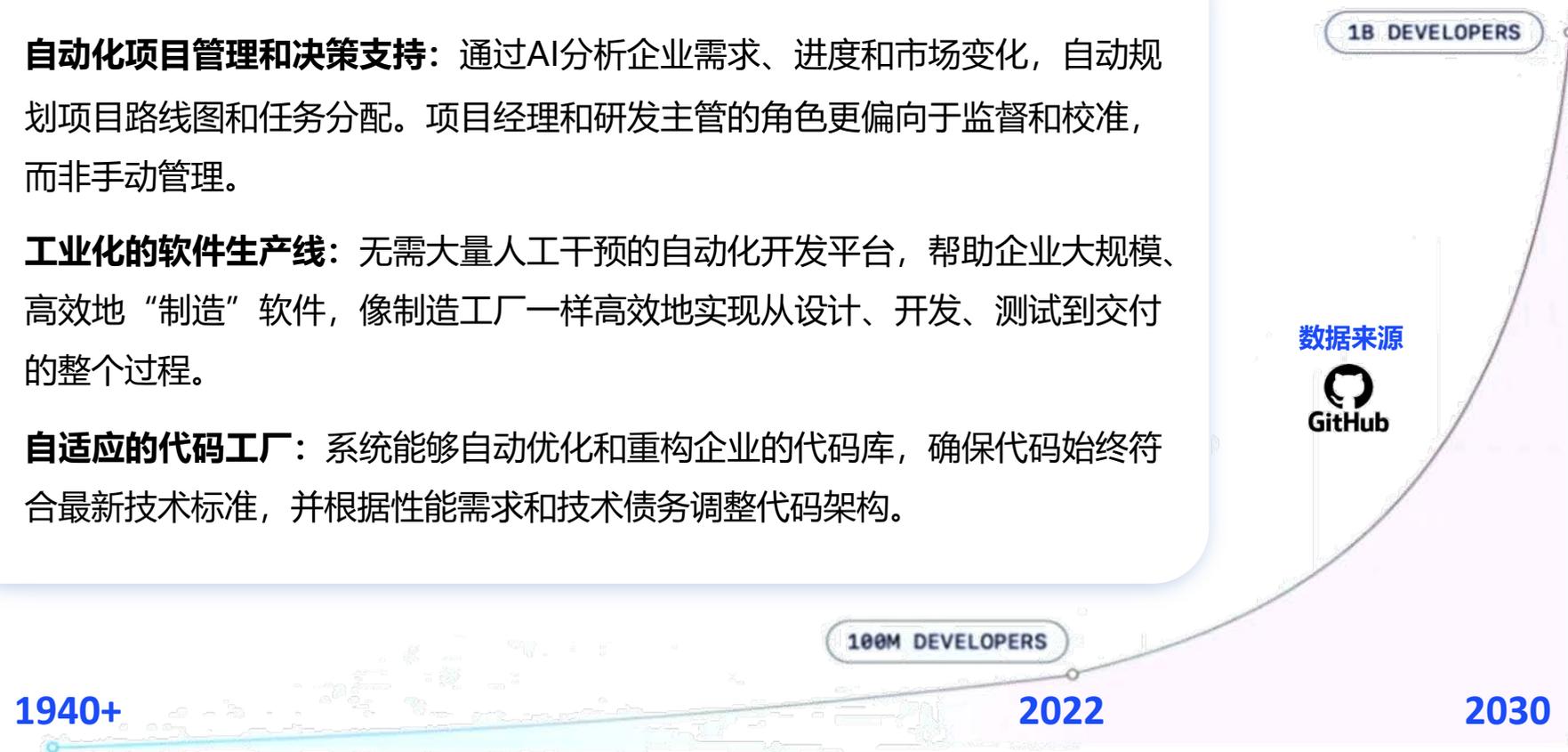
真正使开发者摆脱代码的束缚，专注于更高层次的创造性工作，让**人人都能成为开发者**。

- **思维驱动开发**：通过自然语言提示和思维导图，AI能够帮开发者实现从概念构思到代码实现的全过程。
- **实时反馈和自动优化**：不仅能生成代码，还能根据实时需求自动检测和优化代码，无需开发者手动调整性能和质量问题。
- **代码封装与高层次设计**：开发者可以专注于产品设计和提出问题，AI负责代码实现部分，降低对代码细节的关注。

Eagle Factory

面向**企业**和研发团队的AI驱动的**全自动软件制造平台**：

- **自动化项目管理和决策支持**：通过AI分析企业需求、进度和市场变化，自动规划项目路线图和任务分配。项目经理和研发主管的角色更偏向于监督和校准，而非手动管理。
- **工业化的软件生产线**：无需大量人工干预的自动化开发平台，帮助企业大规模、高效地“制造”软件，像制造工厂一样高效地实现从设计、开发、测试到交付的整个过程。
- **自适应的代码工厂**：系统能够自动优化和重构企业的代码库，确保代码始终符合最新技术标准，并根据性能需求和技术债务调整代码架构。



NiDD AI+研发数字峰会 **5**
AI+ Development Digital summit 第5届



代码问题，就问 Eagle!

科技生态圈峰会 + 深度研习

—1000+ 技术团队的选择



敦煌站

K+思考周®研习社

时间: 2025.08.29-30



上海站

K+金融专场

时间: 2025.10.17-18



香港站

K+思考周®研习社

时间: 2025.11.25-26



K+峰会详情



上海站

AI+研发数字峰会

时间: 2025.05.17-18



北京站

AI+研发数字峰会

时间: 2025.08.08-09



深圳站

AI+研发数字峰会

时间: 2025.11.28-29



AIDD峰会详情



利用AI技术深化计算机对现实世界的理解
推动研发进入智能化时代

