SOFA研发测试工具链
—智能化带来质变

蚂蚁集团高级技术专家　周路

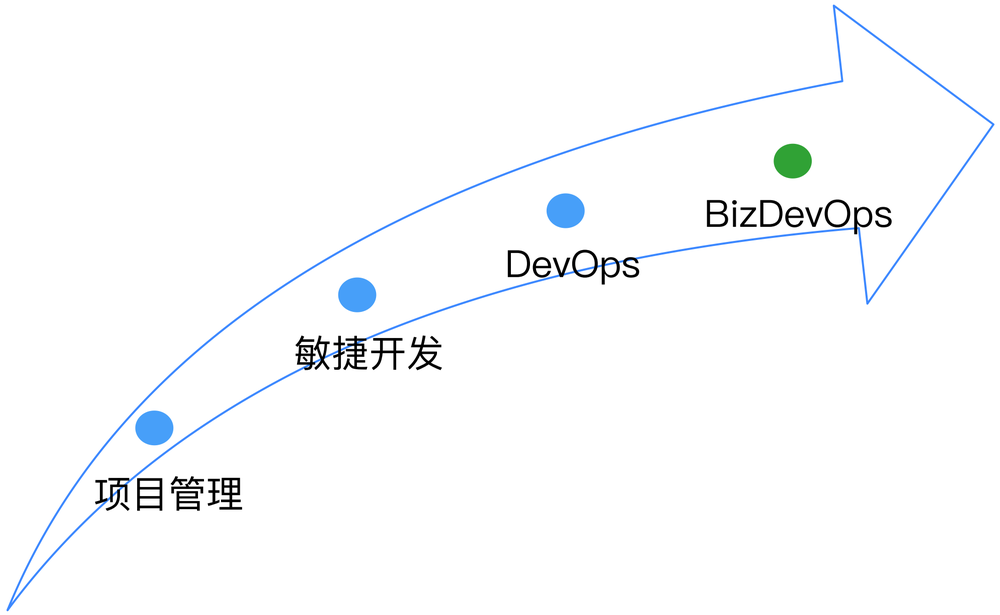
Agenda

1. **研发效能挑战和趋势**
2. 大模型加持研发效能实践
3. 总结与展望

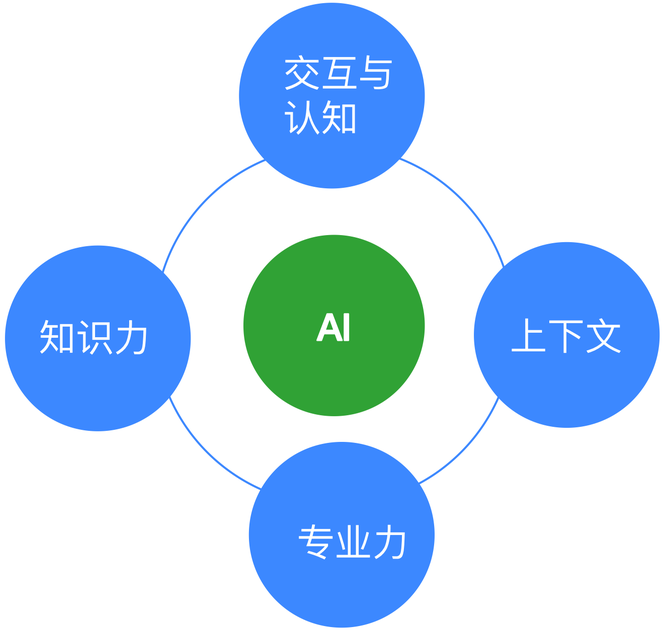
研发效能体系演进

BizDevOps落地难，新一轮AI热潮有望打通业技一体最后一公里

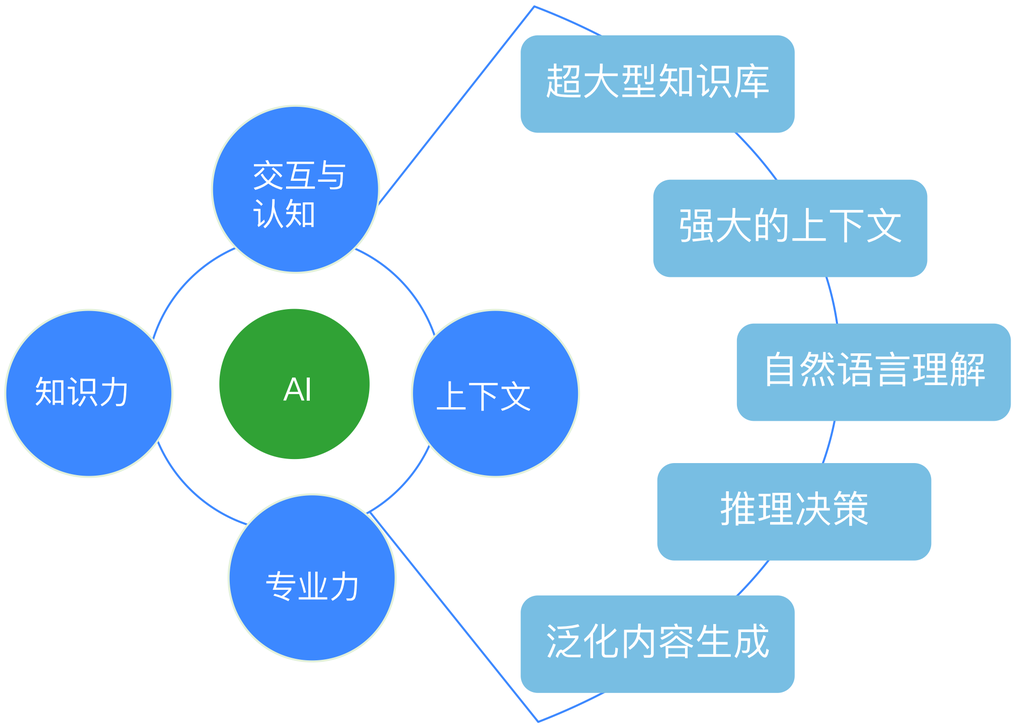
数字化研发效能工程



数智化研发效能工程



LLM为研发效能带来哪些机遇



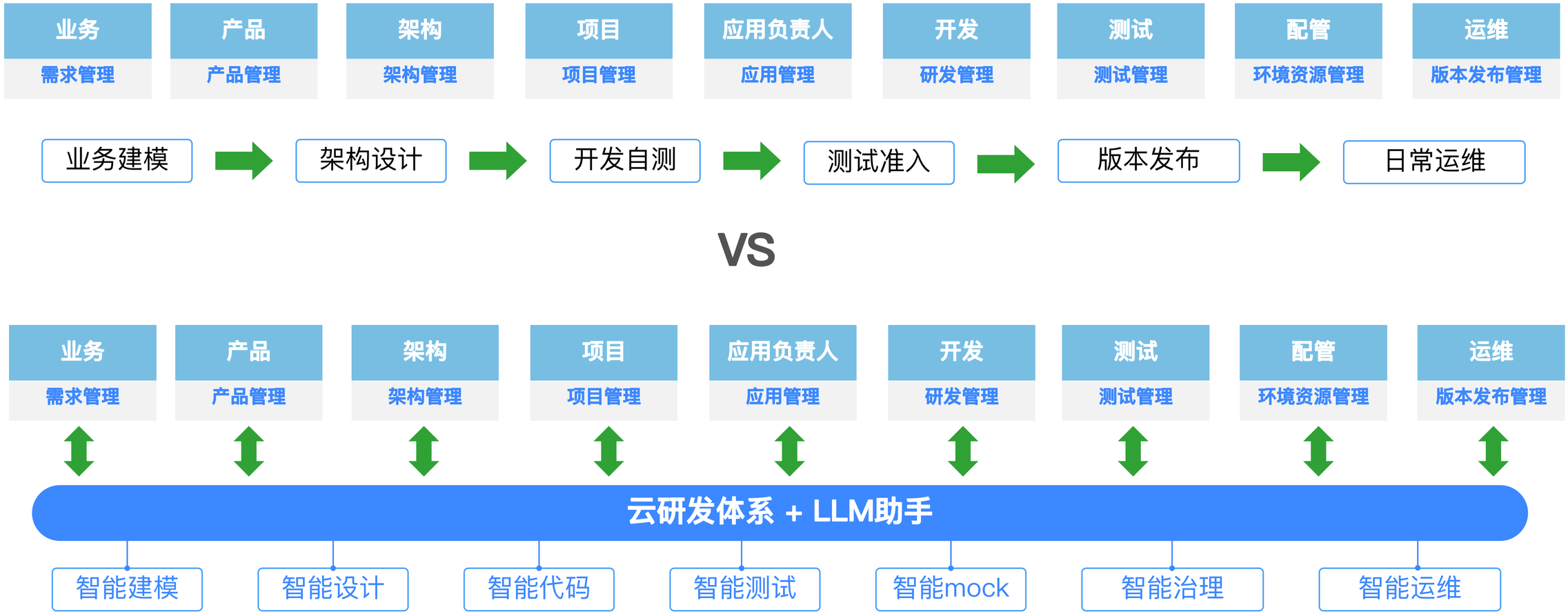
理解需求
- 需求结构化
- 内容总结
- 知识问答

标准研发
- 标准溯源
- 规范检索
- 架构巡检

提效编码
- 代码补全
- SQL生成
- 低码审查

提升质量
- 生成用例
- BUG定位/修复
- 质量门禁

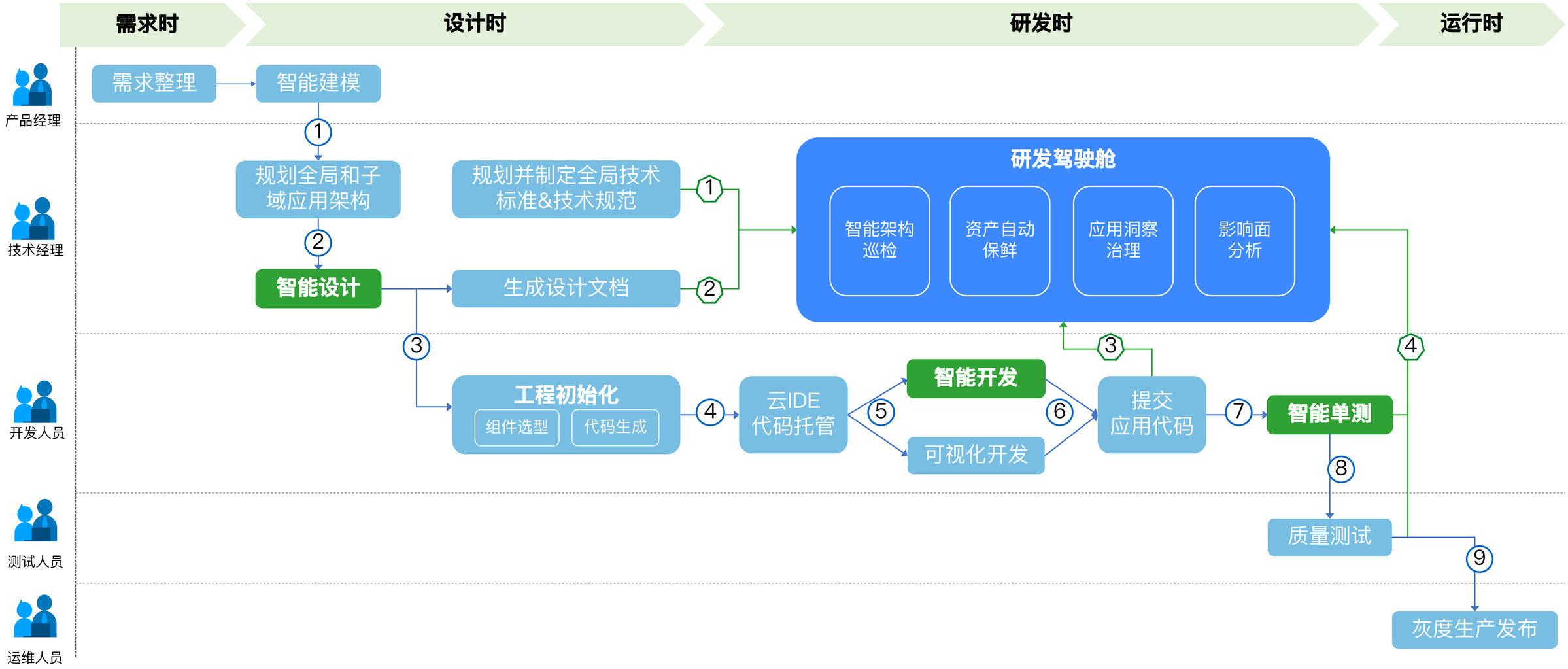
AI驱动软件交付模式变革



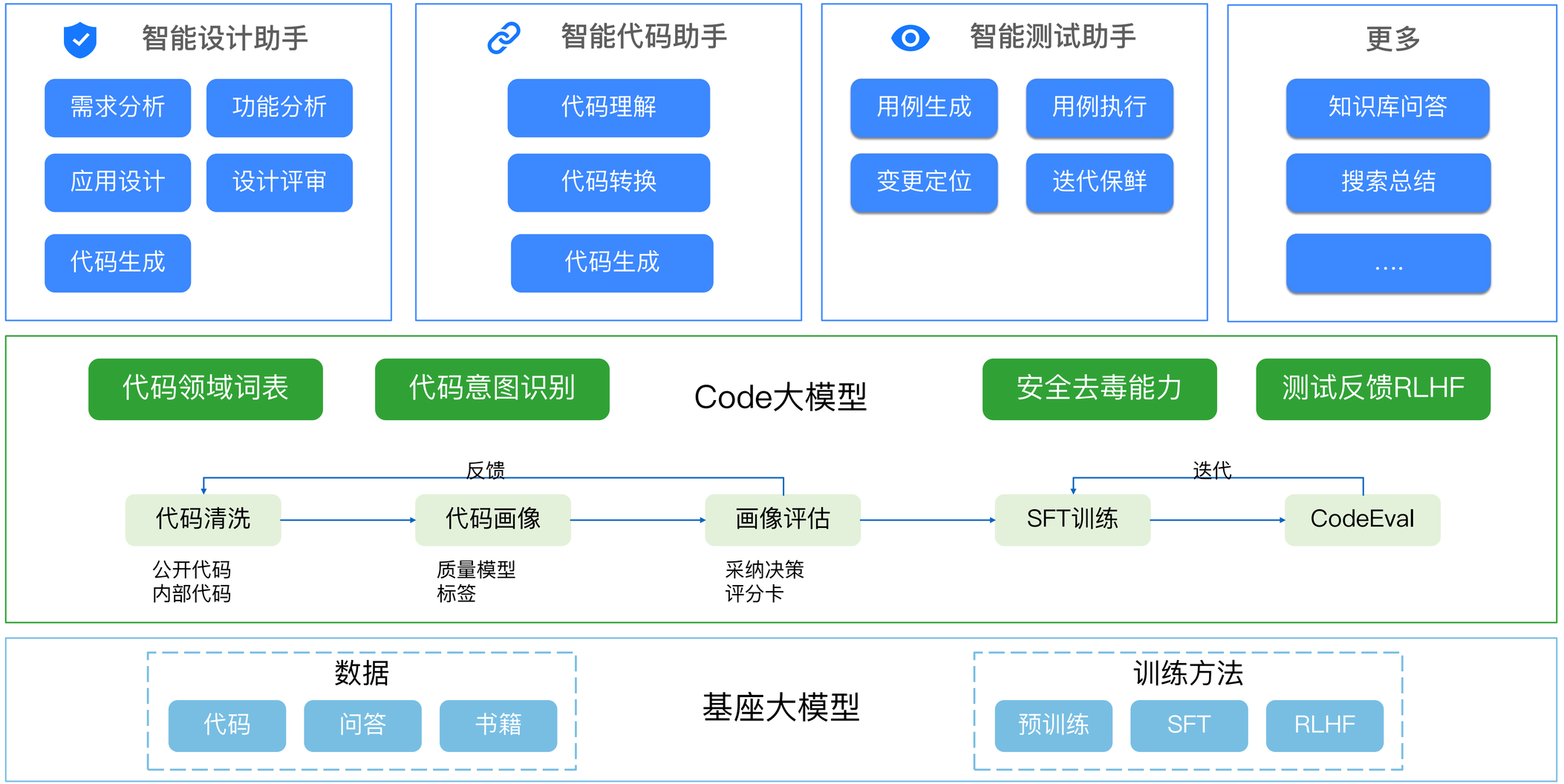
Agenda

1. 研发效能挑战和趋势
2. **大模型加持研发效能实践**
3. 总结与展望

蚂蚁研发效能实践 AI + BizDevOps



Code大模型


智能设计助手
需求分析
功能分析
应用设计
设计评审
代码生成
智能代码助手
代码理解
代码转换
代码生成
智能测试助手
用例生成
用例执行
变更定位
迭代保鲜
更多
知识库问答
搜索总结
....
代码领域词表
代码意图识别
Code大模型
安全去毒能力
测试反馈RLHF
反馈
迭代
代码清洗
代码画像
画像评估
SFT训练
CodeEval
公开代码
内部代码
质量模型
标签
采纳决策
评分卡
数据
代码
问答
书籍
基座大模型
训练方法
预训练
SFT
RLHF

Code大模型应用场景

智能设计助手

智能设计助手可以快速创作高质量领域模型和流程设计

需求分析
- 微服务应用
- 功能描述
- 业务规则

功能分析
- 接口分析
- 领域模型分析
- 数据库表分析

应用设计
- 接口调整
- 出入参调整
- 领域模型调整
- 数据库表调整

设计评审
- UML转换
- 领域模型
- 流程状态

代码生成
- 代码
- 注释

智能代码助手

智能代码助手可以高质量理解代码、生成代码，编码自测一气呵成

代码理解
- 代码注释
- 代码解释
- 代码翻译

代码生成
- NL2SQL
- NL生成代码
- 代码搜索

代码转换
- 代码补全
- 代码修复
- 用例生成

智能测试助手

智能测试助手可以自动化生成用例、并辅助软件测试工作

用例生成
- 单测用例
- 多粒度精准用例
- Assert自动生成

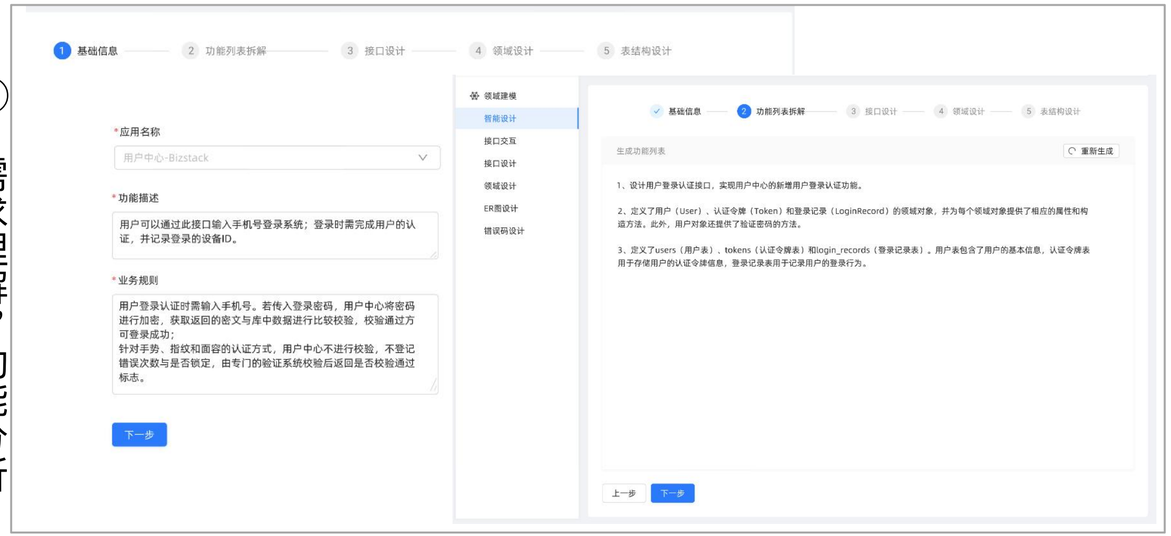
用例执行
- Runing agent
- 多模执行

变更定位
- 失败用例聚合
- 变更行定位
- 推荐修复

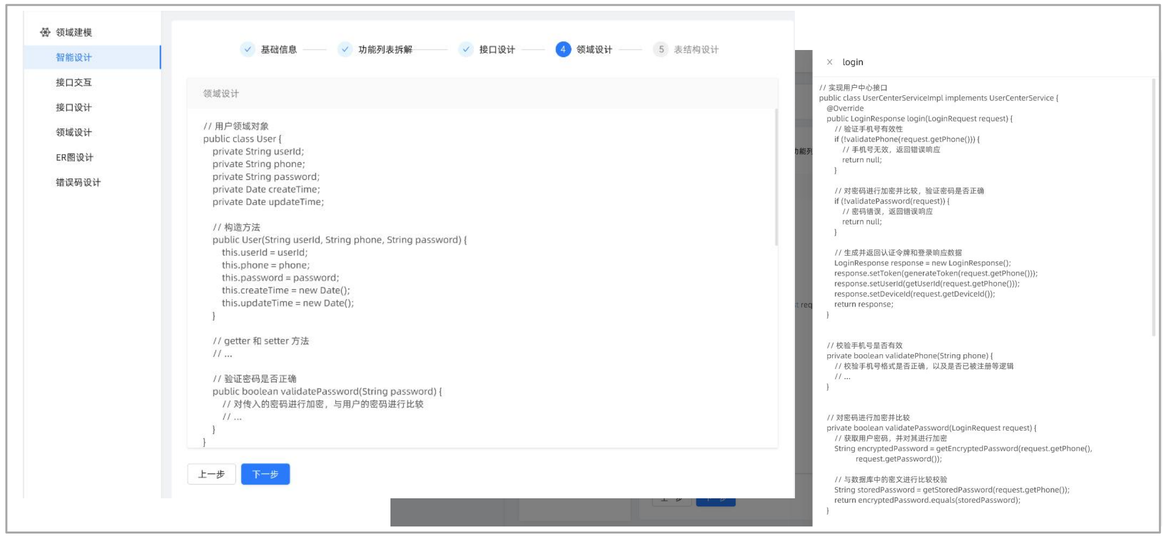
迭代保鲜
- 存量用例汰换
- 存量用例生成
- 用例回归分析

智能设计助手演示

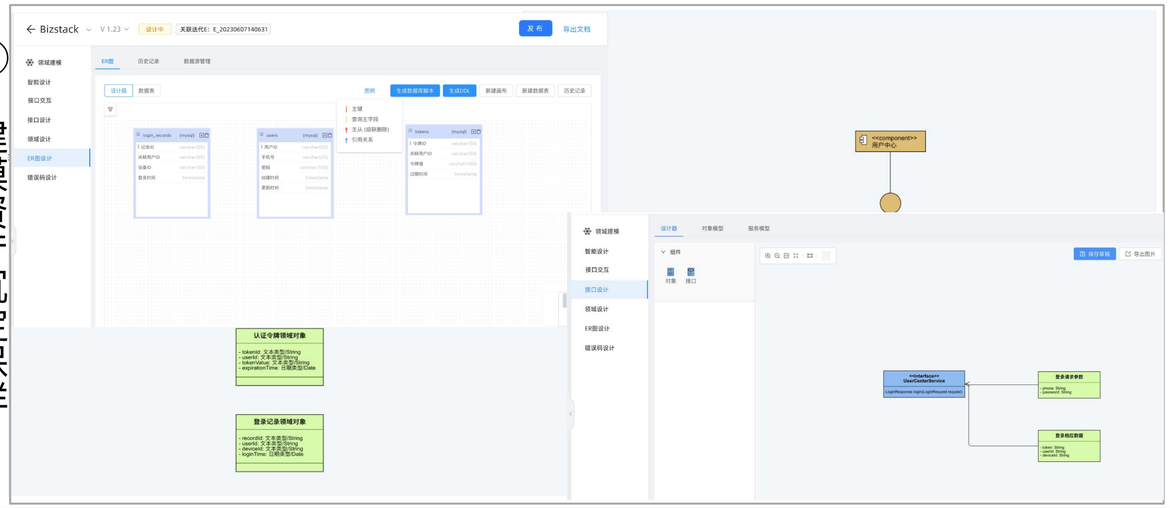
① 需求理解，功能分析



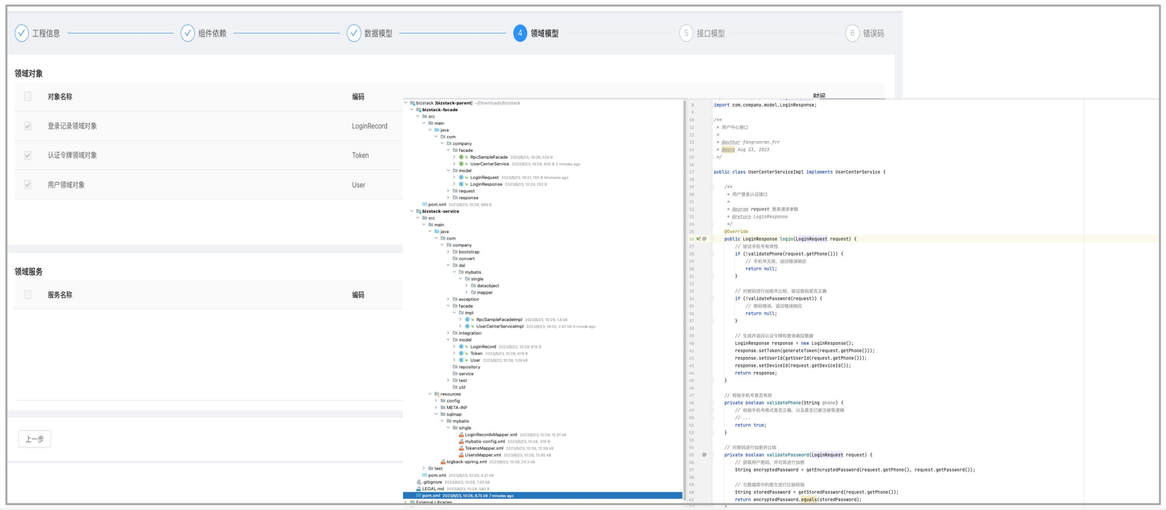
② 功能微调，应用设计



③ 建模资产沉淀保鲜

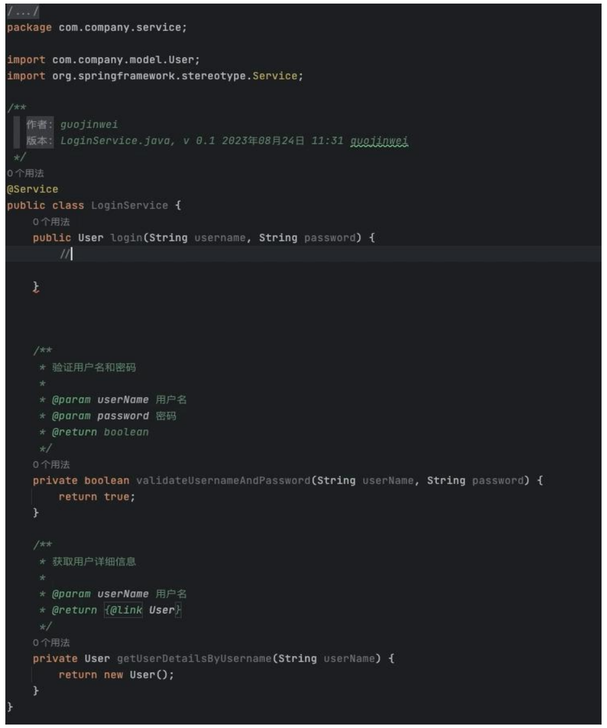


④ 设计文档及代码产出



智能代码助手演示

代码补全/续写

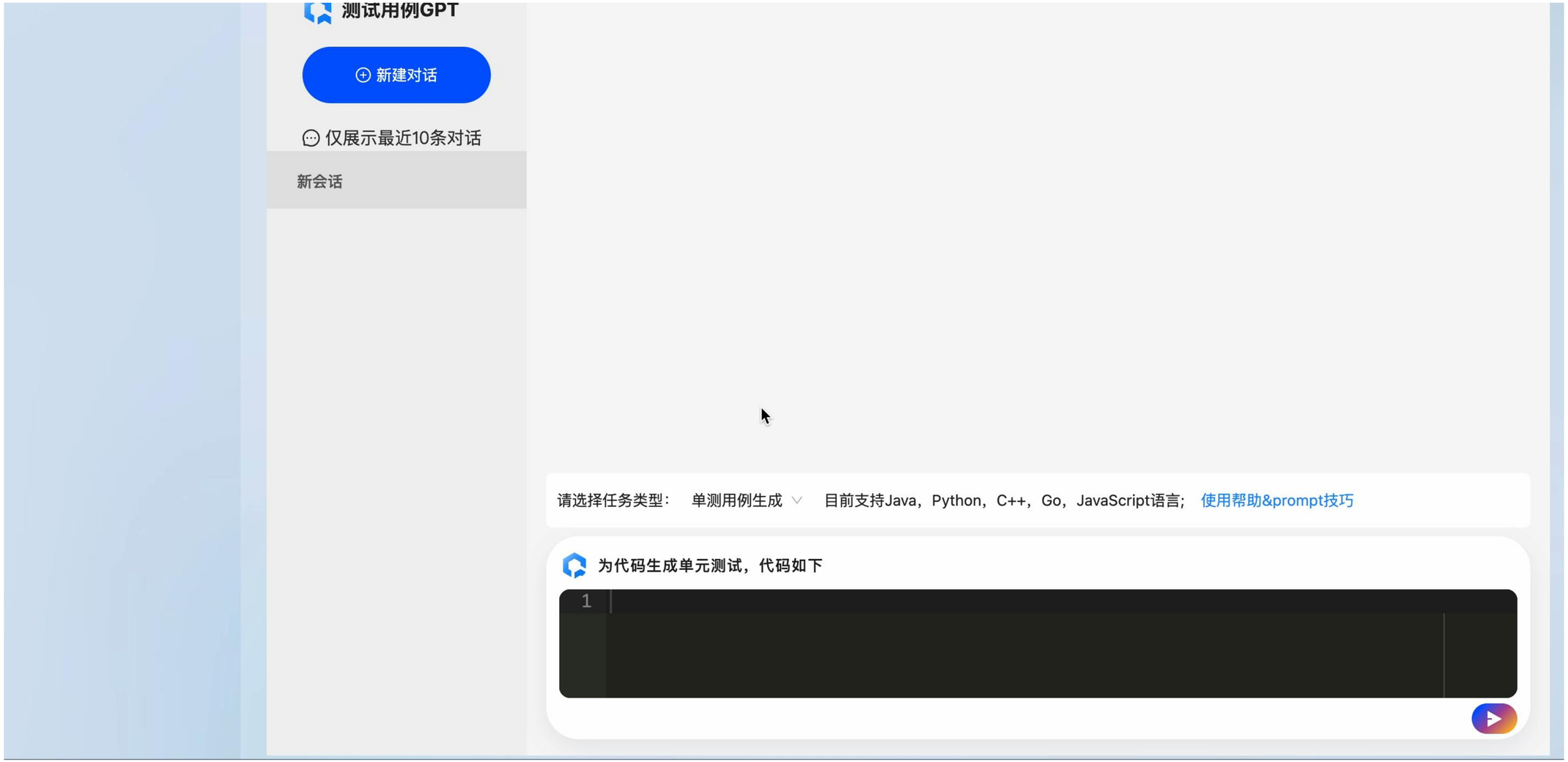


注释生成代码

添加注释

代码对话

智能测试演示


测试用例GPT
新建对话
仅展示最近10条对话
新会话
请选择任务类型：
单测用例生成
目前支持Java，Python，C++，Go，JavaScript语言；
使用帮助&prompt技巧
为代码生成单元测试，代码如下
1

数智化研发驾驶舱

看

架构大盘和指标体系
- 应用增长曲线
- 服务变更曲线
- 应用依赖热度
- 架构规范命中

业务能力地图
- 业务能力列表
- 调用链路关系
- 代码知识图谱
- 应用运营指标

应用健康分
- 应用健康度
- 技术债务评分
- 变更影响占比
- 版本变更元数据

诊

变更影响面分析
- 代码变更监听
- 变更因子提取
- 影响面分析报告

应用依赖分析
- 发布jar依赖列表
- 各版本依赖占比
- 第三方依赖列表

架构评估分析
- 应用体量维度
- 耦合冗余分析
- 服务依赖分析
- 规范健康分析
- 全局应用图谱

服务血缘管理
- 核心服务检测
- 服务依赖关系
- 服务闭环检测
- 不良资产检测
- 服务血缘图谱

防

架构元数据
- 技术模型
- 领域模型
- 数据模型

架构规范
- 全局规范
- 信息标准
- 数据冠标

架构巡检
- 巡检规则
- 巡检调度
- 巡检报告

程序分析
- 代码库扫描
- 源码分析（AST/BCEL/反射/正则匹配）
- 解析调用分析（callgraph）
- 代码大数据

★架构大盘和指标体系

能力：全局拓扑、服务目录、应用健康分
要点：架构资产结构化、资产保鲜、开箱即用

★技术影响面分析

能力：链路追踪、变更分析、安全诊断、依赖拆解
要点：代码扫描、特征分析，自动埋点、运行时数据

★微服务智能扫描巡检

能力：全局规范定义、常态化巡检
要点：无侵入，可防可控

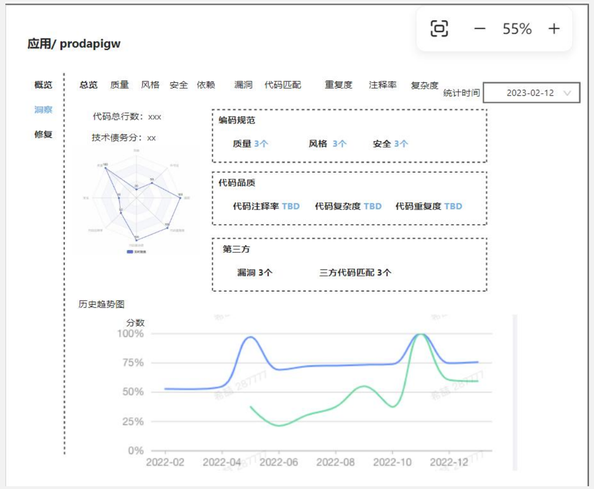
数智化研发驾驶舱演示

业务场景

业务监控

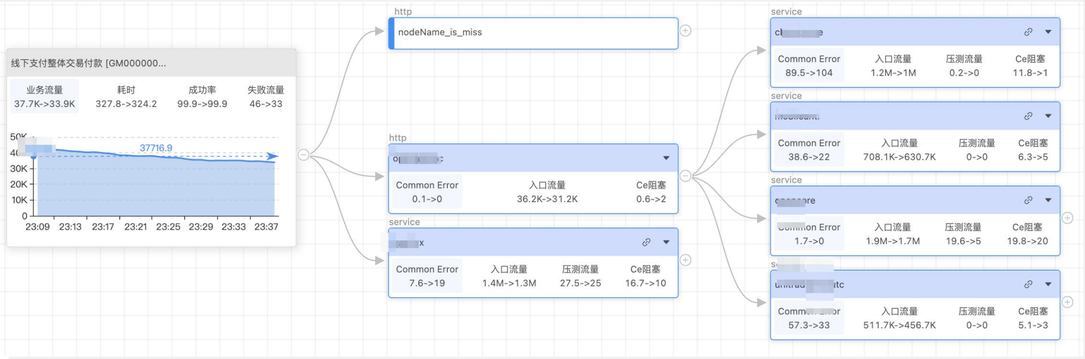
代码智能扫描引擎

应用洞察评分



全局拓扑

变更影响面分析



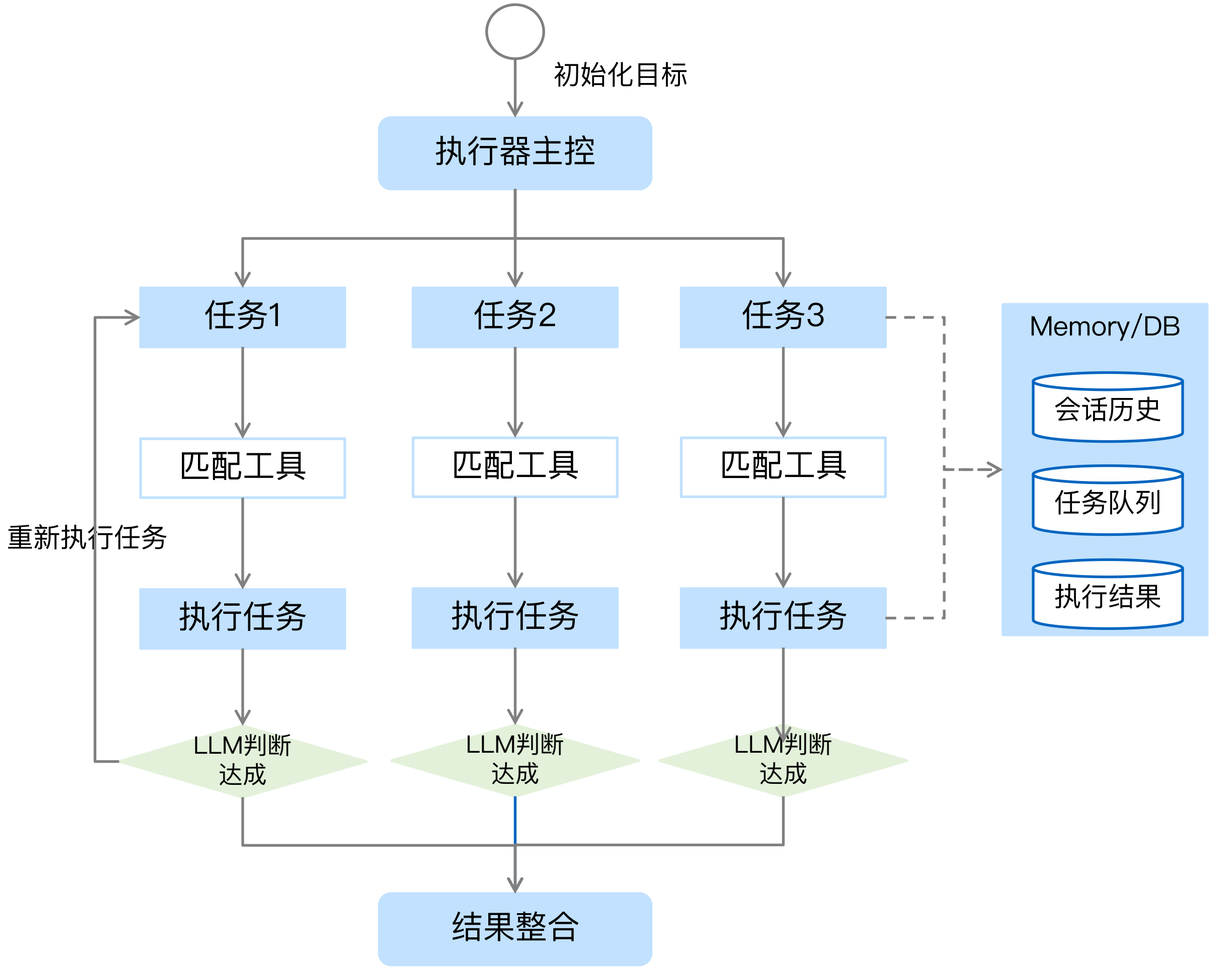
链路追踪

SOFA智能研发整体技术架构

关键设计 — LLM执行器

智能设计面临的问题：
- 需求文档内容较多，上下文不能丢失
- 智能场景多，提示词千差万别
- 大模型token限制

解法：
- 执行器主控：
  - ✓ 具备任务串行、并行执行的能力
  - ✓ 具备单任务反复执行，并跳出的能力
  - ✓ 记录AI思考过程by日志
- Context：
  - ✓ 记录会话历史
  - ✓ 记录任务执行过程，包括执行记录、输入、输出
- 反馈收集：
  - ✓ 依靠工程或AI内部对话形式判断任务是否完成，依据反馈决定是否生成新任务，并自适应循环下一次迭代
- 防止无限循环
  - ✓ 在目标级别控制子任务上限，超过即跳出
  - ✓ 在prompt层面优化

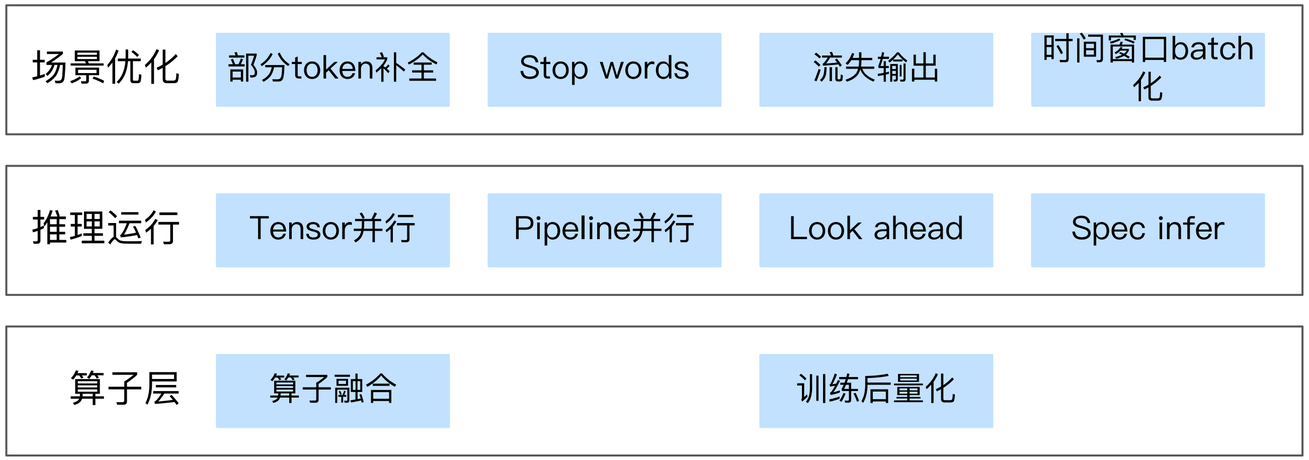


智能代码推理优化方案

挑战：

- 大模型推理耗时普遍较长，在代码补全等研发人员使用的高频场景，普遍对耗时忍耐度很低

解法



效果：

优化方案	效果
Tensor并行	• tensor并行数提高，推理的耗时显著下降，卡推理耗时提升60%~70%
int8量化	• 模型大小缩小至1/4 • 多卡并行时，模型提速至1.2～1.6倍左右，减少15%～40%左右的显存占用
部分token补全	• select optional last tokens kernel方案
流式输出	• 分隔符 • 方块字语言字符切割

Agenda

1. 研发效能挑战和趋势
2. 大模型加持研发效能实践
3. **总结与展望**

总结及展望

- AI为研发效能带来机遇和挑战，AI助手改变了软件交付模式，也引发了效能标准的新思考
- AI更有效的打通Biz-DevOps的鸿沟，让业技一体更具落地性
- 大部分企业不需要犹豫，低门槛的模型训练平台即可引入AI能力
- AI不是万能的，当前融合大于颠覆，先考虑结合现状，再考虑重构
- 对技术人员来讲，改变思路，主动迎接

THANKS

软件正在重新定义世界

Software Is Redefining The World