

AI 应用开发面经仪表盘

牛客近 6 个月 AI 应用开发 / Agent / 大模型应用相关面经整理

面经样本数：32 | 提取题目数：464 | 覆盖公司数：16

构建时间：2026-04-13 01:37:33

Top 公司：字节(7)、淘天(4)、快手(4)、腾讯(3)、蚂蚁(2)、阿里国际(2)、小红书(1)、阿里云(1)、高德(1)、京东(1)

牛客近6个月AI应用开发面经汇总

- 生成时间：2026-04-12 17:17:22
- 统计窗口：2025-10-11 ~ 2026-04-12
- 搜索关键词数：53
- 牛客搜索候选帖数（去重后）：119
- 过滤后实际面经帖数：32
- 最近3个月面经帖数：32

引用：说明：这里的“全量”指基于牛客站内搜索+多组岗位关键词扩展后，当前可检索到并落在近6个月窗口内的相关面经帖去重结果。

搜索关键词

- AI应用开发、AI应用开发 面经、AI应用开发 一面、AI应用开发 二面、AI应用开发 实习、AI应用平台开发、AI应用平台开发 面经、AI应用服务端开发、AI应用服务端开发 面经、AI应用后端开发、AI应用后端开发 面经、AI应用前端开发、AI应用工程师、AI应用工程师 面经、AI技能应用开发、AI Agent开发、AI Agent开发 面经、AI Agent 应用开发、AI Agent研发、Agent开发、Agent开发 面经、智能体开发、智能体开发 面经、智能体应用开发、智能体与大模型应用工程、智能体与大模型应用工程 面经、大模型应用开发、大模型应用开发 面经、大模型应用开发 一面、大模型应用开发 二面、大模型应用工程、大模型应用工程师、大模型数据应用开发、大模型数据应用开发 面经、大模型开发、大模型应用、小红书 AI Agent开发 一面、米哈游-AI Agent研发-暑期一面、百度ai agent开发春招一面、字节agent开发实习一面凉经、快手AI Agent开发一面（已过）、4.9-淘天agent-二面、腾讯暑期实习 ai agent开发一面、实习-快手电商-大模型数据应用开发一面、字节暑期后端agent开发一面、快手大模型应用开发算法岗三面面经、拼多多大模型应用开发一面凉经、腾讯 大模型应用开发 二面、腾讯 大模型应用开发 一面、大模型应用开发一面-美团面经、OPPO大模型应用开发一面 攒人品、字节 大模型应用开发 二面 日常实习、字节大模型应用开发 日常实习一面

公司分布

- 字节: 7
- 淘天: 4
- 快手: 4
- 腾讯: 3
- 蚂蚁: 2
- 阿里国际: 2
- 小红书: 1
- 阿里云: 1
- 高德: 1
- 京东: 1
- 阿里灵犀互娱: 1

- 金山云: 1
- 阿里淘天: 1
- 米哈游: 1
- 百度: 1
- 蔚来: 1

月份分布

- 2026-04: 31
- 2026-03: 1

总表

日期	公司	岗位归类	轮次	标题	链接	提取题目数
2026-04-12	小红书	ai agent开发	一面	小红书 AI Agent开发 一面	[原帖](https://www.nowcoder.com/discuss/872820735335485440)	11
2026-04-12	淘天	ai应用开发	二面	淘天 AI应用开发 二面	[原帖](https://www.nowcoder.com/discuss/872816572199464960)	8
2026-04-12	阿里云	ai应用开发	一面	阿里云 AI应用开发 一面	[原帖](https://www.nowcoder.com/discuss/872810222128021504)	10
2026-04-11	快手	ai应用服务端开发	二面	快手ai应用服务端开发 二面	[原帖](https://www.nowcoder.com/discuss/872512773710696448)	11
2026-04-10	蚂蚁	agent开发	未明确	蚂蚁金融Agent开发暑期面经分享	[原帖](https://www.nowcoder.com/feed/main/detail/1f90384e8ff54172b65b50d853b00da0)	9
2026-04-10	高德	ai应用开发	一面	高德 AI应用开发 一面	[原帖](https://www.nowcoder.com/discuss/872229606252675072)	13
2026-04-10	字节	ai应用开发	未明确	字节 剪映 ai应用开发	[原帖](https://www.nowcoder.com/feed/main/detail/2e53f0ec451a4e71b2e3418252071ba4)	22
2026-04-10	腾讯	ai应用开发	未明确	腾讯 AI应用开发面经	[原帖](https://www.nowcoder.com/feed/main/detail/acd3e53eecbc405296335edae4d5cf88)	14
2026-04-10	京东	agent开发	二面	27暑期实习-京东Agent开发二面分享	[原帖](https://www.nowcoder.com/feed/main/detail/6af1c8611e434bbe8ba9cfa2bd0eaf18)	9
2026-04-10	阿里灵犀互娱	ai应用开发	二面	阿里灵犀互娱AI应用开发暑期二面	[原帖](https://www.nowcoder.com/feed/main/detail/3dbec664147348d685f3e2b8b50288f8)	9
2026-04-10	快手	ai agent开发	二面	快手 AI agent开发二面分享 1h	[原帖](https://www.nowcoder.com/feed/main/detail/afa98a916503444aa708fcfffb263d38)	9
2026-04-10	金山云	ai应用平台开发	一面	27届-日常实习-金山云-AI应用平台开发-一面	[原帖](https://www.nowcoder.com/discuss/872055401926164480)	11
2026-04-10	阿里国际	ai应用开发	一面	阿里国际AI应用开发暑期一面 1h	[原帖](https://www.nowcoder.com/feed/main/detail/580e5a35eddb47e0851952165004f233)	26
2026-04-10	淘天	ai应用开发	一面	淘天AI应用开发 agent岗一面 好难	[原帖](https://www.nowcoder.com/feed/main/detail/a0306a045d594b02b63fb0654d517901)	22
2026-04-09	阿里淘天	ai应用开发	一面	阿里淘天 AI应用开发 暑期实习一面	[原帖](https://www.nowcoder.com/feed/main/detail/39627b379e8e46ce9cb3a3b8eae46959)	9

| 2026-04-09 | 米哈游 | ai agent研发 | 一面 | 米哈游-AI Agent研发-暑期一面 | [原帖](https://www.nowcoder.com/feed/main/detail/a1566acc2660477dbf54b58885513d6c) | 14 |

| 2026-04-09 | 阿里国际 | ai应用开发 | 二面 | 阿里国际AI应用开发二面 | [原帖](https://www.nowcoder.com/feed/main/detail/c2772ead209549a0bb6e54f2df205a31) | 6 |

| 2026-04-09 | 百度 | ai agent开发 | 一面 | 百度ai agent开发春招一面 | [原帖](https://www.nowcoder.com/feed/main/detail/c807140b75bd4cf3bf4166660676db5d) | 21 |

| 2026-04-09 | 字节 | agent开发 | 一面 | 字节agent开发实习一面凉经 | [原帖](https://www.nowcoder.com/feed/main/detail/69b9cd21f1244d3cb19499f249228b50) | 28 |

| 2026-04-09 | 快手 | ai agent开发 | 一面 | 快手AI Agent开发一面(已过) | [原帖](https://www.nowcoder.com/feed/main/detail/7ce89f19368b46da853c718f2ae2f53c) | 30 |

| 2026-04-09 | 腾讯 | ai应用开发 | 一面 | 腾讯日常实习一面-AI应用开发 1h | [原帖](https://www.nowcoder.com/feed/main/detail/37e847fa267b4c3898b38e2952b5286e) | 14 |

| 2026-04-09 | 淘天 | ai agent开发 | 二面 | 4.9-淘天agent-二面 | [原帖](https://www.nowcoder.com/discuss/871774510138023936) | 9 |

| 2026-04-09 | 字节 | agent开发 | 一面 | 字节Agent开发一面90min凉经 | [原帖](https://www.nowcoder.com/feed/main/detail/91c5394e57c14927841d7a86bfe427c2) | 21 |

| 2026-04-09 | 腾讯 | ai agent开发 | 一面 | 腾讯暑期实习 ai agent开发一面 | [原帖](https://www.nowcoder.com/feed/main/detail/497e14617f01453db84dfbc264e5751e) | 7 |

| 2026-04-09 | 蔚来 | ai应用开发 | 一面 | 蔚来 AI应用开发 暑期一面分享 | [原帖](https://www.nowcoder.com/feed/main/detail/5af31c8afe254a00a69a6c0141e6908a) | 9 |

| 2026-04-08 | 快手 | 大模型数据应用开发 | 一面 | 实习-快手电商-大模型数据应用开发一面 | [原帖](https://www.nowcoder.com/discuss/871506086656761856) | 13 |

| 2026-04-08 | 字节 | ai应用开发 | 二面 | AI应用开发日常实习二面-字节 | [原帖](https://www.nowcoder.com/feed/main/detail/0607316a2492400896408c95e10951e0) | 14 |

| 2026-04-08 | 字节 | ai应用开发 | 未明确 | 字节AI应用开发实习面经分享 | [原帖](https://www.nowcoder.com/feed/main/detail/76448605da6e4aee99a394da83d1718f) | 21 |

| 2026-04-05 | 蚂蚁 | ai应用开发 | 二面 | 蚂蚁ai应用开发实习二面 | [原帖](https://www.nowcoder.com/feed/main/detail/718c5a06a65645d7a462a56b457cb35e) | 8 |

| 2026-04-03 | 淘天 | ai应用开发 | 一面 | 淘天-ai应用开发-一面 | [原帖](https://www.nowcoder.com/feed/main/detail/076a8d1acaae4795a28d4a5370872bd6) | 20 |

| 2026-04-02 | 字节 | ai应用开发 | 一面 | 字节 ai应用开发 | [原帖](https://www.nowcoder.com/feed/main/detail/c5947496fc29440883611173478dc257) | 17 |

| 2026-03-21 | 字节 | 大模型应用开发 | 二面 | 字节 大模型应用开发 二面 日常实习 | [原帖](https://www.nowcoder.com/feed/main/detail/30b6f37830414df5a32d851f2004df7a) | 19 |

逐条整理

2026-04-12 | 小红书 AI Agent开发 一面

- 公司：小红书
- 岗位归类：ai agent开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/discuss/872820735335485440>
- 命中搜索词：AI Agent开发、AI Agent开发 面经、小红书 AI Agent开发 一面、快手AI Agent开发一面(已过)

- 提炼摘要：1. 自我介绍；2. 介绍你写的这个项目；3. 为什么要做多 Agent，而不是把所有能力都塞进一个 Agent；4. MCP、Function Calling、Skills、Agent 分别是什么，它们之间是什么；5. Agent 的规划、执行、反思三段式链路怎么设计；6. Agent 的记忆一般怎么分层，为什么不能只靠聊天历史

- 提取到的面试题/考点：

- 1. 自我介绍
- 2. 介绍你写的这个项目
- 3. 为什么要做多 Agent，而不是把所有能力都塞进一个 Agent
- 4. MCP、Function Calling、Skills、Agent 分别是什么，它们之间是什么关系
- 5. Agent 的规划、执行、反思三段式链路怎么设计
- 6. Agent 的记忆一般怎么分层，为什么不能只靠聊天历史
- 7. RAG 可以怎么分类，Agentic RAG 和传统 RAG 差别在哪
- 8. RAG 项目里怎么做召回闭环，才能让系统真的越用越准
- 9. HyDE 的原理是什么，什么时候有效，什么时候会害人
- 0. IVF、PQ、IVF-PQ 分别在做什么，为什么它们能把向量检索做快
- 1. 向量索引有哪些典型类别，分别适合什么场景

2026-04-12 | 淘天 AI应用开发 二面

- 公司：淘天
- 岗位归类：ai应用开发
- 轮次：二面
- 来源：<https://www.nowcoder.com/discuss/872816572199464960>

- 命中搜索词：AI应用开发、AI应用开发 面经、AI应用服务端开发
面经、大模型应用开发、大模型应用开发

二面、大模型开发、大模型数据应用开发、拼多多大模型应用开发一面凉经、智能体开发、智能体开发
面经、腾讯大模型应用开发 一面

- 提炼摘要：1. Agent 和传统 workflow 的本质区别是什么，什么时候该选 Agent，什么时候不该；2. 智能体开发范式有哪些，分别适合什么场景；3. 智能体的记忆机制应该怎么设计，为什么“全都记住”反而是坏事；4. 短期记忆和长期记忆应该如何精准提取，分别存什么内容更合理；5. LangChain 和 LangGraph 如果要选，你更看重什么维度；6. 设计智能体时，稳定性和安全性应该优先在哪些层面做约束

- 提取到的面试题/考点：

- 1. Agent 和传统 workflow 的本质区别是什么，什么时候该选 Agent，什么时候不该选
- 2. 智能体开发范式有哪些，分别适合什么场景
- 3. 智能体的记忆机制应该怎么设计，为什么“全都记住”反而是坏事
- 4. 短期记忆和长期记忆应该如何精准提取，分别存什么内容更合理
- 5. LangChain 和 LangGraph 如果要选，你更看重什么维度
- 6. 设计智能体时，稳定性和安全性应该优先在哪些层面做约束
- 7. AI 应用里的在线实验和传统 AB 测试有什么差别，为什么更难做

- 8. MCP 和 CLI 工具接入在工程上分别解决什么问题，它们不是一回事

2026-04-12 | 阿里云 AI应用开发 一面

- 公司：阿里云
- 岗位归类：ai应用开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/discuss/872810222128021504>

• 命中搜索词：大模型应用开发、大模型应用开发

二面、大模型开发、大模型数据应用开发、拼多多大模型应用开发一面凉经、智能体开发、智能体开发面经、腾讯大模型应用开发一面

• 提炼摘要：1. 做一下自我介绍；2. 你对网络安全方向怎么看，为什么 AI 应用在这个方向上既有价值又有风险；3. 介绍一个你做的这个项目，重点讲架构和你负责的部分；4. 为什么要用 RAG，而不是直接让模型回答安全分析问题；5. RAG 的效果你怎么评估，不能只说看用户反馈；6. 幻觉问题你一般怎么处理，尤其是安全场景下的高风险幻觉

• 提取到的面试题/考点：

- 1. 做一下自我介绍
- 2. 你对网络安全方向怎么看，为什么 AI 应用在这个方向上既有价值又有风险
- 3. 介绍一个你做的这个项目，重点讲架构和你负责的部分
- 4. 为什么要用 RAG，而不是直接让模型回答安全分析问题
- 5. RAG 的效果你怎么评估，不能只说看用户反馈
- 6. 幻觉问题你一般怎么处理，尤其是安全场景下的高风险幻觉
- 7. 多智能体项目的目的是什么，什么时候真的需要多智能体
- 8. 你做完多智能体系统之后，最大的体会是什么
- 9. 前端为什么不直接做一个普通聊天页，而是要选更适合流式交互的方案
- 0. 平时你怎么利用大模型开发项目，而不是停留在写点 demo

2026-04-11 | 快手ai应用服务端开发 二面

- 公司：快手
- 岗位归类：ai应用服务端开发
- 轮次：二面
- 来源：<https://www.nowcoder.com/discuss/872512773710696448>

• 命中搜索词：AI应用开发、AI应用开发 面经、AI应用服务端开发 面经

• 提炼摘要：4.9面，没后续了。因为我得到6月才能到，面试官说还有两个月，中间会陆续招人，要是到6月份还有名额再；啥时候能来实习。能实习多久？；了解redis集群吗？了解redis底层的存储逻辑吗？；redis的性能单点能够有多少？集群能够有多少？；你的这个agent项目用了spring Ai Ailibaba，为什么用这个，他与langchain；你用spring Ai Ailibaba的时候感觉有哪些比较好的点？哪些不太方便的点？

• 提取到的面试题/考点：

• 4.9面，没后续了。因为我得到6月才能到，面试官说还有两个月，中间会陆续招人，要是到6月份还有名额再捞我。

- 啥时候能来实习。能实习多久？
- 了解redis集群吗？了解redis底层的存储逻辑吗？
- redis的性能单点能够有多少？集群能够有多少？
- 你的这个agent项目用了spring Ai Ailibaba，为什么用这个，他与langchain4j、spring ai有什么区别？选型的时候怎么考虑的？
- 你用spring Ai Ailibaba的时候感觉有哪些比较好的点？哪些不太方便的点？
- 你了解agent编排的有向无环图吗？你用了吗？为什么没用？
- 你用过龙虾吗？对里面的技术有啥了解吗？
- 我看你的黑马点评项目用的rabbitmq，为啥选这个？这个对比其他mq有什么特别的优势？（我说当时只学了这一个，所以就用了，实际上rocketmq应该会更好）
- 手撕合并k个升序链表（前几天pdd一面刚写完）
- 转正率如何？看表现，一般都能转

2026-04-10 | 蚂蚁金融Agent开发暑期面经分享

- 公司：蚂蚁
- 岗位归类：agent开发
- 轮次：未明确
- 来源：<https://www.nowcoder.com/feed/main/detail/1f90384e8ff54172b65b50d853b00da0>
- 命中搜索词：AI Agent开发、AI Agent开发 面经、小红书 AI Agent开发 一面、快手AI Agent开发一面（已过）
- 提炼摘要：1.项目拷打；2.上传文档时，是一个一个上传还是批量上传；3.文档chunk完有多少id；4.做完rag后llm能理解图片里的内容吗，还是通过加入上下文token后靠文本理解；5.多路召回和重排序是怎么提高检索成功率的；6.怎么解决幻觉问题
- 提取到的面试题/考点：
 - 1.项目拷打
 - 2.上传文档时，是一个一个上传还是批量上传
 - 3.文档chunk完有多少id
 - 4.做完rag后llm能理解图片里的内容吗，还是通过加入上下文token后靠文本理解
 - 5.多路召回和重排序是怎么提高检索成功率的
 - 6.怎么解决幻觉问题
 - 7.介绍一下对temperature的理解
 - 8.介绍一下langchain和langgraph
 - 9.算法题口头说一下1200w的数据里，怎么找到中位数

2026-04-10 | 高德 AI应用开发 一面

- 公司：高德

- 岗位归类：ai应用开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/discuss/872229606252675072>
- 命中搜索词：AI应用开发、AI应用开发 面经、AI应用服务端开发 面经
- 提炼摘要：1. 做一下自我介绍，重点讲你在 AI 编码和跨端工程结合上的经历；2. Flutter 的核心原理是什么，如果不从 Widget 开始讲，你会怎么讲；3. Flutter 和 React Native、原生开发最大的区别，不要只说性能；4. Flutter 的 UI 为什么能回到原生页面上，这个过程本质上发生了什么；5. 如果文字内容变化导致宽度发生变化，Flutter 的重排到底是怎么发生的；6. Flutter 里 build、layout、paint、composite 这几段如果线上卡
- 提取到的面试题/考点：
 - 1. 做一下自我介绍，重点讲你在 AI 编码和跨端工程结合上的经历
 - 2. Flutter 的核心原理是什么，如果不从 Widget 开始讲，你会怎么讲
 - 3. Flutter 和 React Native、原生开发最大的区别，不要只说性能
 - 4. Flutter 的 UI 为什么能回到原生页面上，这个过程本质上发生了什么
 - 5. 如果文字内容变化导致宽度发生变化，Flutter 的重排到底是怎么发生的
 - 6. Flutter 里 build、layout、paint、composite 这几段如果线上卡顿，你怎么判断卡在哪一段
 - 7. Java 里的强引用、软引用、弱引用、虚引用在工程里怎么理解，别只背定义
 - 8. 讲一下你做过的 Agent Skill，不要从业务价值讲，直接讲技术拆分
 - 9. 团队里常用哪些开发工具比较多，如果从 AI 协作方式去分类，你会怎么讲
 - 0. 你是什么时候开始用 Claude Code 这一类工具的，真正改变你的是什么
 - 1. 用过哪些模型，模型选择不是看排行榜的话，你一般怎么选
 - 2. 对商业项目来说，开发者过去习惯手动掌控每个细节，到了 AI 提供生产力的时代，你怎么看这个变化
 - 3. 你用 AI 写代码时一般怎么拆解任务，才能让结果比较稳定

2026-04-10 | 字节 剪映 ai应用开发

- 公司：字节
- 岗位归类：ai应用开发
- 轮次：未明确
- 来源：<https://www.nowcoder.com/feed/main/detail/2e53f0ec451a4e71b2e3418252071ba4>
- 命中搜索词：AI应用开发、AI应用开发 面经、AI应用服务端开发 面经
- 提炼摘要：1. 你这里的意图判断是怎么做的？直接交给模型做，还是这个用户的查询query做了一些其他处理。；2. 为什么要设计这三层记忆架构，是一个通用的设计还是参考了什么论文。；3. Agent 的评估是怎么做的；4. 有没有评估过你的记忆架构，在没参考openclaw之前和之后的指标有什么变化，成功率有对比吗？；5. 你之前向量检索加redis缓存是怎么做的？redis键和值是什么？；6. 新的记忆加到文件里是直接append吗？覆盖？还是改写？
- 提取到的面试题/考点：
 - 1. 你这里的意图判断是怎么做的？直接交给模型做，还是这个用户的查询query做了一些其他处理。

- 2. 为什么要设计这三层记忆架构，是一个通用的设计还是参考了什么论文。
- 3. Agent 的评估是怎么做的
- 4. 有没有评估过你的记忆架构，在没参考openclaw之前和之后的指标有什么变化，成功率有对比吗？
- 5. 你之前向量检索加redis缓存是怎么做的？redis键和值是什么？
- 6. 新的记忆加到文件里是直接append吗？覆盖？还是改写？
- 7. 改的方式 模型改吗？模型调用editor去改文件是吗？
- 8. 你觉得 redis+milvus向量存储 和mysql+文件系统+向量检索，他们的优劣如何？优点缺点分别是什么。
- 9. 其实我比较感兴趣的是他们这两种方式的查询检索召回 效率的对比。
- 0. 对于每一个用户，在系统中怎么保存这些不同用户的文档？
- 1. 其实就是模型通过路径去读这些文件是吧，那如果两个人登了同一个账号，在同一时刻，这个文件会被两个不同的用户使用，在这个情况下，你们的系统如何处理？怎么解决这个问题？
- 1. 你说你的文档准确率达到90%，具体做的改变是什么。
- 2. 那为什么医院的问答项目里，不采用同样的实现
- 3. 你说pdf表格密集，那假如用户就是问一个表格里，第二行是2025年，第三行是2024年，具体问2025年的值是多少，你的系统能做到吗？
- 4. skill是自己写的吗，互相之间可以联动吗？
- 5. 为什么不用skill-creator
- 6. 你的数据集是怎么收集的，然后是怎么评估的呢？打分是谁在打分？人在打分，还是模型裁判员去打分？如果有很好用的模型裁判员，为什么不用裁判员模型
- 7. 介绍一下KV cache
- 8. 讲一下transfromer架构。
- 算法题
- 1. 字符串数字相减
- 2. 字符串最长连续序列

2026-04-10 | 腾讯 AI应用开发面经

- 公司：腾讯
- 岗位归类：ai应用开发
- 轮次：未明确
- 来源：<https://www.nowcoder.com/feed/main/detail/acd3e53eecbc405296335edae4d5cf88>
- 命中搜索词：AI Agent开发、AI Agent开发 面经、AI应用开发、AI应用开发 面经、AI应用服务端开发面经、小红书 AI Agent开发 一面、快手AI Agent开发一面（已过）
- 提炼摘要：1. 记忆模块存在哪种介质上；2. 跨会话通过压缩保留关键症状，这一块详细介绍；3. FSM状态机分哪个部分，怎么实现；4. 为什么选择ReAct；5. 用户今天问A，三天后问B，怎么关联两次提问，记忆怎么跟用户维度绑定。（session级视角、没；6. 多 Agent 框架怎么设计，为什么这么划分 Agent
- 提取到的面试题/考点：
- 1. 记忆模块存在哪种介质上

- 2. 跨会话通过压缩保留关键症状，这一块详细介绍
- 3. FSM状态机分哪个部分，怎么实现
- 4. 为什么选择ReAct
- 5. 用户今天问A，三天后问B，怎么关联两次提问，记忆怎么跟用户维度绑定。（session级视角、没考虑并发和多次咨询隔
- 6. 多 Agent 框架怎么设计，为什么这么划分 Agent
- 7. skill是什么 Auto-coder skill怎么工作。
- 8. Skill 的核心价值是什么？为什么会有这个概念？
- 9. 可插拔 RAG、一键配置切换、是不是热更新？（热更新--运行时不停机动态替换）
- 0. 工厂模式的优点是什么？
- 1. 本地部署要考虑什么？配置管理放在哪里？
- 2. langchain和langgraph的区别？为什么有了langchain还要langgraph
- 分别解决什么问题？为什么有了langchain还要langgraph
- 算法题:

2026-04-10 | 27暑期实习-京东Agent开发二面分享

- 公司：京东
- 岗位归类：agent开发
- 轮次：二面
- 来源：<https://www.nowcoder.com/feed/main/detail/6af1c8611e434bbe8ba9cfa2bd0eaf18>
- 命中搜索词：AI Agent开发、AI Agent开发 面经、大模型应用开发、大模型应用开发 二面、大模型开发、大模型数据应用开发、小红书 AI Agent开发 一面、快手AI Agent开发一面（已过）、拼多多大模型应用开发一面凉经、智能体开发、智能体开发 面经、腾讯大模型应用开发 一面
- 提炼摘要：1.项目拷打；2.介绍一下 AI 智能体的 workflow 以及 RAG 知识库是怎么设计的？；3.意图识别用的是哪个大模型？如何判断和测试意图识别结果的准确率？；4.Redis 做语义向量的相似度检索，内部用的是什么算法？；5.在使用 RAG 的程中有没有出现大模型幻觉？如何设计兜底方案？；6.JMeter 压测的具体参数是怎么设置的？
- 提取到的面试题/考点：
- 1.项目拷打
- 2.介绍一下 AI 智能体的 workflow 以及 RAG 知识库是怎么设计的？
- 3.意图识别用的是哪个大模型？如何判断和测试意图识别结果的准确率？
- 4.Redis 做语义向量的相似度检索，内部用的是什么算法？
- 5.在使用 RAG 的程中有没有出现大模型幻觉？如何设计兜底方案？
- 6.JMeter 压测的具体参数是怎么设置的？
- 7.Redis 是单线程的，为什么它还能支持高并发？
- 8.在实际业务场景中，你怎么判断什么时候需要加缓存？为什么不能只用 MySQL？
- 9.除了导购，你认为在电商领域（如京东）还有哪些场景可以用到 Agent？

2026-04-10 | 阿里灵犀互娱AI应用开发暑期二面

- 公司：阿里灵犀互娱
- 岗位归类：ai应用开发
- 轮次：二面
- 来源：<https://www.nowcoder.com/feed/main/detail/3dbec664147348d685f3e2b8b50288f8>
- 命中搜索词：AI应用开发、AI应用开发 面经、AI应用服务端开发 面经
- 提炼摘要：1.你如何使用ai coding的，怎么保证能按照你的要求完成代码；2.设计一个游戏智能客服，整机架构；3.意图识别小模型如何训练，数据集怎么构造；4.什么时候采用单agent，什么时候采用多agent，原因；5.如何打通agent对多平台数据源的访问；6.如果用户在同一个对话中每一句都是独立问题，如何进行上下文管理
- 提取到的面试题/考点：
 - 1.你如何使用ai coding的，怎么保证能按照你的要求完成代码
 - 2.设计一个游戏智能客服，整机架构
 - 3.意图识别小模型如何训练，数据集怎么构造
 - 4.什么时候采用单agent，什么时候采用多agent，原因
 - 5.如何打通agent对多平台数据源的访问
 - 6.如果用户在同一个对话中每一句都是独立问题，如何进行上下文管理
 - 7.设计一个游戏助手agent，记忆系统怎么设计
 - 8.设计一个自动挂机agent，与传统挂机脚本有什么优势
 - 9.如何定义游戏角色状态

2026-04-10 | 快手 AI agent开发二面分享 1h

- 公司：快手
- 岗位归类：ai agent开发
- 轮次：二面
- 来源：<https://www.nowcoder.com/feed/main/detail/afa98a916503444aa708fcfffb263d38>
- 命中搜索词：AI Agent开发、AI Agent开发 面经、小红书 AI Agent开发 一面、快手AI Agent开发一面（已过）
- 提炼摘要：1.你的AI项目的亮点是什么？；2.上下文压缩算法的触发条件；3.长任务(执行几天，步骤很多)在你的项目中是怎么做的？；4.有了解claudeCode的Hook机制功能知道嘛?解释一下；5.装饰器设计模式和代码设计模式；6.看过Springboot/Spring的源码嘛?里面用到了哪些设计模式
- 提取到的面试题/考点：
 - 1.你的AI项目的亮点是什么?
 - 2.上下文压缩算法的触发条件
 - 3.长任务(执行几天，步骤很多)在你的项目中是怎么做的?
 - 4.有了解claudeCode的Hook机制功能知道嘛?解释一下

- 5.装饰器设计模式和代码设计模式
- 6.看过Springboot/Spring的源码嘛?里面用到了哪些设计模式
- 7.mysql的锁有哪些?
- 8.讲一下乐观锁和悲观锁
- 手撕:一道原创题,大致题意:给定一个链表"head={4,2,8}和一棵二叉树 root=Rn{1,4,4,1,2,null,2,null,null,nul1,2,8,2,1,null,null,null,null,8,null,null}(貌似是这样的)判断一下这棵树中存不存在一条路径和链表是一样的。

2026-04-10 | 27届-日常实习-金山云-AI应用平台开发-一面

- 公司：金山云
- 岗位归类：ai应用平台开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/discuss/872055401926164480>
- 命中搜索词：AI应用开发、AI应用开发 面经、AI应用服务端开发 面经
- 提炼摘要：4.9-半个小时左右；1、自我介绍；2、慢SQL优化分析，解决了哪些问题，优化思路是怎么样的？；3、针对Kafka消息丢失采取了哪些措施？如何保障消息不重复消费？(只回答了消费端幂等，没了解过生产；4、Cache aside模式的优劣是什么？；5、项目中的RAG定位和 workflows 是什么样的？
- 提取到的面试题/考点：
- 4.9-半个小时左右
- 1、自我介绍
- 2、慢SQL优化分析，解决了哪些问题，优化思路是怎么样的？
- 3、针对Kafka消息丢失采取了哪些措施？如何保障消息不重复消费？(只回答了消费端幂等，没了解过生产端的幂等，但是好像是存在的)
- 4、Cache aside模式的优劣是什么？
- 5、项目中的RAG定位和 workflows 是什么样的？
- 6、RAG如何实现语义分块的？RAG评估用了哪些方法和指标，结果如何？
- 7、聊AI Coding，使用过AI Coding写代码吗？有使用AI Coding进行项目迭代吗？
- 8、AICoding在需求清晰和需求变更两类场景下，各自的实践方式和遇到的主要挑战是什么？
- 9、Skills，主Agent调用子Agent的话通信怎么做？回答了A2A协议，但是面试官说这个其实用的不多。
- 0、反问:具体部门业务。

2026-04-10 | 阿里国际AI应用开发暑期一面 1h

- 公司：阿里国际
- 岗位归类：ai应用开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/feed/main/detail/580e5a35eddb47e0851952165004f233>
- 命中搜索词：AI应用开发、AI应用开发 面经、AI应用服务端开发 面经

- 提炼摘要：1.你现在主要使用的开发语言有哪些?；2.有使用过Java吗，对Java的掌握程度怎么样?；3.Python里面的init_(和_new_)有什么区别?；4.了解带有下划线的_new_方法吗?；5.Python里面传参是传值还是传引用?；6.如果对参数传入一个字符串类型，如何在方法内部修改它并把这个字符串的值传递出来?

- 提取到的面试题/考点：

- 1.你现在主要使用的开发语言有哪些?
- 2.有使用过Java吗，对Java的掌握程度怎么样?
- 3.Python里面的init_(和_new_)有什么区别?
- 4.了解带有下划线的_new_方法吗?
- 5.Python里面传参是传值还是传引用?
- 6.如果对参数传入一个字符串类型，如何在方法内部修改它并把这个字符串的值传递出来?
- 7.any和or有什么区别?
- 8.Python里面有哪些元素最终判断为 False?
- 9.Python里面的copy(浅拷贝)跟deepcopy(深拷贝)有什么区别?
- 0.Python里面的垃圾回收机制是怎么样的?
- 1.怎么去提高一个Python文件的运行效率?在编译层面还可以怎么做?
- 2.Python里面会出现内存泄漏吗?
- 3.怎么理解内存泄漏?全局变量算内存泄漏吗?
- 4.如果把列表(List)本身作为一、二级缓存来使用，算内存泄漏吗?
- 5.在操作系统方面，进程跟线程有什么区别?
- 6.引入线程主要是为了解决什么样的问题?
- 7.从浏览器里面输入一个URL，到整个页面的展示，中间的过程是怎样的?
- 8.DNS解析用的是TCP还是UDP?
- 9.TCP是几次握手、几次挥手?
- 0.四次挥手有可能会变成三次吗(出现三个报文的情况)?
- 1.什么是对称加密跟非对称加密?
- 2.HTTP请求中GET和POST有什么区别?
- 3.GET的长度限制是哪里限制的?是客户端还是服务端的限制?
- 4.有用过哪些AI相关的编程工具?
- 5.你觉得这些AI工具会代替掉现有的程序员吗?
- 6.AI在哪些方面会比程序员要强?

2026-04-10 | 淘天AI应用开发 agent岗一面 好难

- 公司：淘天
- 岗位归类：ai应用开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/feed/main/detail/a0306a045d594b02b63fb0654d517901>

- 命中搜索词：AI Agent开发、AI Agent开发 面经、AI应用开发、AI应用开发 面经、AI应用服务端开发面经、大模型应用开发、大模型应用开发 二面、大模型开发、大模型数据应用开发、小红书 AI Agent开发一面、快手AI Agent开发一面（已过）、拼多多大模型应用开发一面凉经、智能体开发、智能体开发面经、腾讯 大模型应用开发 一面

- 提炼摘要：1.项目拷打；2.你在这个 RAG 系统优化里面，在多阶段 RAG 系统中采用了 BM25 和向量混合检索，然后这；3.RAG支持 PDF 扫描件和 OCR，然后还有表格结构化的提取，然后在这过程中有没有遇到什么技术；4.识别准确率怎么样？；5.这个多智能体系统设计里面，然后它的这个 State 管理和Checkpoint分配机制的具体实现；6.怎么样实现 State 全局管理？

- 提取到的面试题/考点：

- 1.项目拷打

- 2.你在这个 RAG 系统优化里面，在多阶段 RAG 系统中采用了 BM25 和向量混合检索，然后这个是怎么去设计两者的结合逻辑的？然后混合策略的话具体是如何去提升检索效果的？

- 3.RAG支持 PDF 扫描件和 OCR，然后还有表格结构化的提取，然后在这过程中有没有遇到什么技术难点？

- 4.识别准确率怎么样？

- 5.这个多智能体系统设计里面，然后它的这个 State 管理和Checkpoint分配机制的具体实现方式是什么？怎么去解决对话执行中的状态竞争问题呢？

- 6.怎么样实现 State 全局管理？

- 7.将Choice

接口封装为MCP工具的时候，怎么去设计一个标准化接口？然后遇到有没有遇到过一些兼容性的挑战？

- 8.举了项目里的一个例子，问我出参入参是怎么去定义的？

- 9.大模型在调用这个工具的时候，比如说有 MCP 之前，它的调用的处理流程是什么样子的？

- 0.或者说 MCP 它有哪些缺点或者挑战呢？

- 1.提供的这个 MCP 的结果它是流式的吗？

- 2.这个多agent项目是主子agent的项目吗？

- 3.其中一个功能，然后它的 token 就是一次会话 token 大概有多少？有没有超过上限？

- 4.模型用的哪个？我答Qwen，问我Qwen具体哪个版本

- 5.在子任务过程中啊，如果它的某个子任务失败，比如说数据获取为空，它的这个整个 workflows 是怎么去重试或者是降级处理的？

- 6.在实现这个流式输出实现的时候，比如说后端用了 FastAPI 和 SSE 来实现中间结果的实时流式输出。然后在这个 Langchain 这种基于图的状态机框架中，是怎么捕获每个 node 的执行结果，然后推送到前端的？

- 7.LangGraph 和Langchain 为什么选择了 LangGraph 没有选择简单的那个 Langchain 呢？

- 8.Checkpoint 的持久化

- 9.对话之后重新连接的话，是怎么能够恢复到之前的那个状态呢？

- 0.对话持久化的话，是存储到哪里的？是存储到内存里面，还是存到硬盘上面去的？

- 1.关于 RAG 的，向量数据库在选择索引的时候是用了哪种向量数据库？为什么？

- 2.这里面你提到了 RRF 重排序，然后有没有引入什么模型进行精排？

2026-04-09 | 阿里淘天 AI应用开发 暑期实习一面

- 公司：阿里淘天
- 岗位归类：ai应用开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/feed/main/detail/39627b379e8e46ce9cb3a3b8eae46959>
- 命中搜索词：AI应用开发、AI应用开发 面经、AI应用服务端开发 面经
- 提炼摘要：1. 目前你的每月AI消耗的token量是多少？花费是多少？；2. RAG系统中，超过了模型上下文窗口怎么处理？；3. RAG相比于目前的各种coding agent，他的优势在哪里？；4. Multi-Agent系统和单Agent系统的优缺点？什么时候选用？；5. pipeline中如果其中一个agent出错了，产生错误输出，下游agent该如何处理？；6. 介绍ReAct、Plan、Reflection这些不同agent实现范式的区别，还有使用场景的

- 提取到的面试题/考点：

- 1. 目前你的每月AI消耗的token量是多少？花费是多少？
- 2. RAG系统中，超过了模型上下文窗口怎么处理？
- 3. RAG相比于目前的各种coding agent，他的优势在哪里？
- 4. Multi-Agent系统和单Agent系统的优缺点？什么时候选用？
- 5. pipeline中如果其中一个agent出错了，产生错误输出，下游agent该如何处理？
- 6. 介绍ReAct、Plan、Reflection这些不同agent实现范式的区别，还有使用场景的理解。
- 7. 什么样才是一个好的agent？从工程角度来说，做一个好的agent需要关注哪些方面？从这个Context Engineering的思路，如何去做好一个生产级的agent？
- 9. 你有没有尝试用AI解决工作/生活/学习中的一些问题？改变了一些重复性劳动的方式？
- 0. 算法:AI coding题，要求实现一个工单处理系统。

2026-04-09 | 米哈游-AI Agent研发-暑期一面

- 公司：米哈游
- 岗位归类：ai agent研发
- 轮次：一面
- 来源：<https://www.nowcoder.com/feed/main/detail/a1566acc2660477dbf54b58885513d6c>
- 命中搜索词：AI Agent开发、AI Agent开发 面经、小红书 AI Agent开发 一面、快手AI Agent开发一面（已过）、米哈游-AI Agent研发-暑期一面
- 提炼摘要：1.自我介绍；2.项目；3.问SFT；4.attention和self-attention；5.为什么attention比rnn好；6.workflow和agent区别
- 提取到的面试题/考点：
- 1.自我介绍
- 2.项目
- 3.问SFT
- 4.attention和self-attention
- 5.为什么attention比rnn好
- 6.workflow和agent区别

- 7.记忆机制的设计？如果都放在一起怎么办
- 8.mcp
- 9.few-shot有没有用？
- 0.agent怎么保证有用？
- 1.沙箱了解吗？
- 2.你会怎么去测试ai写的代码？
- 3.代码实现LRU
- 4.反问:做什么？有什么建议？面试几轮？

2026-04-09 | 阿里国际AI应用开发二面

- 公司：阿里国际
- 岗位归类：ai应用开发
- 轮次：二面
- 来源：<https://www.nowcoder.com/feed/main/detail/c2772ead209549a0bb6e54f2df205a31>
- 命中搜索词：AI应用开发、AI应用开发 面经、AI应用服务端开发 面经
- 提炼摘要：1.深挖实习内容；2.mcp是如何内置sop流程的；3.如何量化工作的成效；4.为什么不在上一家公司转正；5.你未来想做什么业务；6.你参加实习后最大的提升是什么
- 提取到的面试题/考点：
 - 1.深挖实习内容
 - 2.mcp是如何内置sop流程的
 - 3.如何量化工作的成效
 - 4.为什么不在上一家公司转正
 - 5.你未来想做什么业务
 - 6.你参加实习后最大的提升是什么

2026-04-09 | 百度ai agent开发春招一面

- 公司：百度
- 岗位归类：ai agent开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/feed/main/detail/c807140b75bd4cf3bf4166660676db5d>
- 命中搜索词：AI Agent开发、AI Agent开发 面经、小红书 AI Agent开发 一面、快手AI Agent开发一面（已过）、百度ai agent开发春招一面
- 提炼摘要：简单讲一个你最有代表性的项目，你在其中解决的最关键问题是什么？；提示词模板是怎么设计与迭代的？你如何判断一个模板真的变好了？；你们的 Agent 是单 Agent 还是多 Agent？为什么这么设计？有没有考虑过替代方案？；Agent 的任务是怎么拆分的？拆分粒度是怎么决定的？；上下文是如何构建的？你们怎么避免上下文过长或信息污染？；如果上下文窗口不够，你会优先保留哪些信息？为什么？
- 提取到的面试题/考点：

- 简单讲一个你最有代表性的项目，你在其中解决的最关键问题是什么？
- 提示词模板是怎么设计与迭代的？你如何判断一个模板真的变好了？
- 你们的 Agent 是单 Agent 还是多 Agent？为什么这么设计？有没有考虑过替代方案？
- Agent 的任务是怎么拆分的？拆分粒度是怎么决定的？
- 上下文是如何构建的？你们怎么避免上下文过长或信息污染？
- 如果上下文窗口不够，你会优先保留哪些信息？为什么？
- 做代码理解时，AST、调用关系这些信息是怎么用起来的？
- 单测生成里，哪些代码不适合生成单测？你们是如何识别并过滤的？
- 覆盖率高但测试质量很差，你见过吗？你们是如何解决的？
- mock 在单测里什么时候是必须的？什么时候反而会带来问题？
- 如果一个函数同时依赖数据库和 RPC，你怎么让模型生成的单测稳定运行？
- 你们如何评估生成单测的质量？除了覆盖率，还有哪些指标有效？
- LLM 的输入到底是什么？模型真正看到的内容包含哪些？
- self-attention 的核心作用是什么？为什么要拆成 Q、K、V？
- 为什么 attention 能建模长距离关系？
- 为什么需要 multi-head？attention 为什么可以看成动态加权？
- 同一个 token 的 Q、K、V 为什么不相同？
- attention 复杂度很高，当上下文特别长时，你会怎么优化？
- 模型产生幻觉的常见原因是什么？工程上有哪些方法可以降低幻觉？
- Python 有多线程吗？GIL 是干嘛的？什么时候多线程才是有用的？
- #Agent面试会问什么？#

2026-04-09 | 字节agent开发实习一面凉经

- 公司：字节
- 岗位归类：agent开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/feed/main/detail/69b9cd21f1244d3cb19499f249228b50>
- 命中搜索词：AI Agent开发、AI Agent开发

面经、字节agent开发实习一面凉经、字节暑期后端agent开发一面、小红书 AI Agent开发 一面、快手AI Agent开发一面（已过）

• 提炼摘要：1. 为什么选择 Spring AI 框架？；2. Spring AI 框架的主要优势是什么？；3. 如何设计多模型支持架构？；4. 多租户环境下模型切换是否支持热更新？切换是否相互独立？；5. Agent 的记忆、工具调用、知识库检索等关键组件是如何实现的？；6. 系统中的 Agent 包含哪些具体状态？

- 提取到的面试题/考点：
- 1. 为什么选择 Spring AI 框架？
- 2. Spring AI 框架的主要优势是什么？
- 3. 如何设计多模型支持架构？

- 4. 多租户环境下模型切换是否支持热更新？切换是否相互独立？
- 5. Agent 的记忆、工具调用、知识库检索等关键组件是如何实现的？
- 6. 系统中的 Agent 包含哪些具体状态？
- 7. 后端 Agent 是否支持多租户同时调用？
- 8. Session 和 User ID 是如何绑定的？
- 9. Agent 工具调用的完整业务流程是怎样的？
- 0. 长期记忆和短期记忆在设计上有什么区别？本质上有什么区别？
- 1. SSE 在前后端是如何交互的？
- 2. 后端以什么数据格式推送流式信息？
- 3. Agent 发生“工具调用”时，SSE 推送的事件结构中通常包含哪些字段？
- 4. MCP的交互流程是怎样的？
- 5. Agent 如何与 MCP Server 连接通信？
- 6. RAG 的主要模式和主要工作流程是怎样的？
- 7. Agent 一般在什么阶段去查询向量知识库？通过什么方式去查询？
- 8. 查询知识库的工具函数，其标准输入和输出是什么？
- 9. Go 语言的 GMP 调度模型是怎样的？
- 0. GMP 模型中的 Machine (M) 代表什么？
- 1. Processor (P) 上的本地队列长度通常是多少？
- 2. 系统如何实现图像识别等多模态功能？
- 3. 为什么不直接使用多模态大模型？
- 4. 在没有前端界面的情况下，本地图片是如何传到后端并进行识别的？
- 5. 对其他主流 AI 框架（如 LangChain）有了解吗？
- 6. 你最熟悉的主力编程语言是什么？
- 7. 手撕：两两交换链表中的节点
- #Agent面试会问什么？#

2026-04-09 | 快手AI Agent开发一面（已过）

- 公司：快手
- 岗位归类：ai agent开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/feed/main/detail/7ce89f19368b46da853c718f2ae2f53c>
- 命中搜索词：AI Agent开发、AI Agent开发 面经、小红书 AI Agent开发 一面、快手AI Agent开发一面（已过）

• 提炼摘要：1、为什么引入父子索引？；2、为什么在检索阶段引入BM25？；(追问)BM25和向量检索是怎样组合的？比例是如何设置的？；(追问)整体检索流程是怎样的？从query 到最终上下文的完整流程是什么？；(追问)检索阶段有没有做rerank？使用的是什方式？；1、rerank后一般返回几个块？

- 提取到的面试题/考点：
- 1、为什么引入父子索引?
- 2、为什么在检索阶段引入BM25?
- (追问)BM25和向量检索是怎样组合的?比例是如何设置的?
- (追问)整体检索流程是怎样的?从query到最终上下文的完整流程是什么?
- (追问)检索阶段有没有做rerank?使用的是什 么方式?
- 1、rerank后一般返回几个块?
- (追问)为什么选择这个数量?有没有做过验证?
- 2、rerank后的topK截断是怎么做的?
- (追问)为什么是这个值?有没有尝试过其他策略?
- (追问)如果上下文长度不够或过长,你是怎么处理的?
- 1、讲一下上下文工程是怎么设计的。
- (追问)上下文拼接的结构是怎样的?
- (追问)如何避免上下文过长导致模型性能下降?
- 2、记忆机制是怎么做的?
- (追问)短期记忆和长期记忆是如何区分和存储的?
- (追问)记忆更新策略是什么?
- 1、Function Calling 是怎么设计的?
- 2、Agent的任务规划是怎么做的?
- (追问)规划是由模型完成还是通过规则实现?
- (追问)多工具调用时如何决定调用顺序?
- (追问)如果工具调用失败如何处理?
- 1、Prompt注入攻击如何防御?
- (追问)有没有做输入过滤或规则校验?
- 2、工具调用的安全控制是怎么实现的?
- (追问)如何限制模型调用敏感接口?
- (追问)是否有权限控制或白名单机制?
- 1、讲一下分布式令牌桶限流。
- 2、漏桶算法是什么?
- 3、滑动窗口算法是怎么实现的?
- (追问)如果使用滑动窗口,结构体中会包含哪些字段?

2026-04-09 | 腾讯日常实习一面-AI应用开发 1h

- 公司：腾讯
- 岗位归类：ai应用开发
- 轮次：一面

- 来源：<https://www.nowcoder.com/feed/main/detail/37e847fa267b4c3898b38e2952b5286e>
- 命中搜索词：AI应用开发、AI应用开发 面经、AI应用服务端开发 面经
- 提炼摘要：手撕:LRU Cache；1.短期记忆的存储结构；2.生成摘要是怎么存储的？；3.短期记忆是针对会话级的，那跨对话的呢？；4.摘要压缩的时候，是怎么限制摘要长度呢？压缩到原来的一半还是让大模型自由发挥？；5.长期记忆的触发时机是怎么样的？如果只是主动录入+召回和rag有什么区别？
- 提取到的面试题/考点：
 - 手撕:LRU Cache
 - 1.短期记忆的存储结构
 - 2.生成摘要是怎么存储的？
 - 3.短期记忆是针对会话级的，那跨对话的呢？
 - 4.摘要压缩的时候，是怎么限制摘要长度呢？压缩到原来的一半还是让大模型自由发挥？
 - 5.长期记忆的触发时机是怎么样的？如果只是主动录入+召回和rag有什么区别？
 - 6.两种检索方式是怎么融合的呢？
 - 7.查询重写是怎么实现的？是怎么评估重写效果？
- 1.对于Linux有了解吗？如果Linux的磁盘满了或者CPU高负载需要用什么命令定位？
- 2.假如说有一个服务部署在Linux上面，上游调用失败，怎么去排查失败原因
- 3.讲一讲计算机网络的IO多路复用？epoll的实现是怎么样的
- 4.讲一讲进程和线程的区别？线程之间的通讯方式有哪些？
- 5.MySql中执行查询语句的过程，是怎么找到对应行的？描述一下中间过程
- 6.如果输入的是二级索引，是一个什么过程呢？

2026-04-09 | 4.9-淘天agent-二面

- 公司：淘天
- 岗位归类：ai agent开发
- 轮次：二面
- 来源：<https://www.nowcoder.com/discuss/871774510138023936>
- 命中搜索词：AI Agent开发、AI Agent开发 面经、小红书 AI Agent开发 一面、快手AI Agent开发一面（已过）
- 提炼摘要： 面试时间：60min； 面试问题：；自我介绍；聊聊对agent的认识；有没有用过open claw；skills了解吗
- 提取到的面试题/考点：
 - 面试时间：60min
 - 面试问题：
 - 自我介绍
 - 聊聊对agent的认识
 - 有没有用过open claw
 - skills了解吗

- 你刚才想到了用子agent的思路，在什么情况下会这么做，有什么好处
- skills和子agent有什么区别，什么场景下选择不同的解决方法（我回答的是看整体的任务复杂度，问题太复杂的话用子agent，同时开发复杂度会高）
- 问了个java开放题，我的机器每隔一段时间就会full gc你有什么解决方式 或排查方式吗

2026-04-09 | 字节Agent开发一面90min凉经

- 公司：字节
- 岗位归类：agent开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/feed/main/detail/91c5394e57c14927841d7a86bfe427c2>
- 命中搜索词：字节agent开发实习一面凉经、字节暑期后端agent开发一面
- 提炼摘要：1.自我介绍；2.提示词模板是怎么构建的？；3.单Agent还是多Agent的？子Agent任务是什么？；4.分支覆盖率是怎么统计的？原理有没有了解过？代码插桩具体是怎么实现的？；5.对于代码解析有没有前置分析？有效性判断怎么实现的？未来让你来优化这些指标你会怎么设计？；6.有没有思考过哪些代码会让模型生成的代码准确度和覆盖率降低？这些用AST和LSP都生成不了单测的代码
- 提取到的面试题/考点：
 - 1.自我介绍
 - 2.提示词模板是怎么构建的？
 - 3.单Agent还是多Agent的？子Agent任务是什么？
 - 4.分支覆盖率是怎么统计的？原理有没有了解过？代码插桩具体是怎么实现的？
 - 5.对于代码解析有没有前置分析？有效性判断怎么实现的？未来让你来优化这些指标你会怎么设计？
 - 6.有没有思考过哪些代码会让模型生成的代码准确度和覆盖率降低？这些用AST和LSP都生成不了单测的代码如何过滤？
 - 7.mock是怎么实现的？
 - 8.多维度的查询改写是什么？改写的时候遇到需要用户参与补充信息时怎么设计？技术上怎么实现？
 - 9.并行化意图识别是什么？为什么要并行化？如何实现的？
 - 0.讲一下项目里召回的流程
 - 1.对于上下文工程有什么经验吗？有没有做过to do list？为什么这样会让模型更聚焦，怎么实现的？
 - 2.skills的原理有没有了解过？怎么实现的？
 - 3.LLM的底层原理有没有了解？输入给模型的是什么？
 - 4.self attention怎么实现的有了了解吗？为什么要分成QKV，怎么计算的？同一个token向量是一样的吗？
 - 5.python有没有多线程？为什么要有GIL，有没有了解过？多线程的锁都有什么？lock和Rlock有什么区别？
 - 6.信号量的底层是怎么实现的？
 - 7.cpp编译链接流程有没有了解？编译分哪些步骤？
- 手撕：
 - 8.岛屿最大面积
 - 9.反问

- #Agent面试会问什么？#

2026-04-09 | 腾讯暑期实习 ai agent开发一面

- 公司：腾讯
- 岗位归类：ai agent开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/feed/main/detail/497e14617f01453db84dfbc264e5751e>
- 命中搜索词：AI Agent开发、AI Agent开发 面经、小红书 AI Agent开发 一面、快手AI Agent开发一面（已过）
- 提炼摘要：（3）了解过抓包吗？介绍一下tcpdump？；面试很难，还是要多多练习，感觉自己说不明白业务逻辑；（1）实习拷打；（2）介绍一下垃圾回收算法。；（4）介绍一下分布式系统的一致性哈希算法。；（5）介绍一下HTTPS四次握手。
- 提取到的面试题/考点：
 - （3）了解过抓包吗？介绍一下tcpdump？
 - 面试很难，还是要多多练习，感觉自己说不明白业务逻辑
 - （1）实习拷打
 - （2）介绍一下垃圾回收算法。
 - （4）介绍一下分布式系统的一致性哈希算法。
 - （5）介绍一下HTTPS四次握手。
 - （6）算法题：实现两个字符串类型的大数相加

2026-04-09 | 蔚来 AI应用开发 暑期一面分享

- 公司：蔚来
- 岗位归类：ai应用开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/feed/main/detail/5af31c8afe254a00a69a6c0141e6908a>
- 命中搜索词：AI应用开发、AI应用开发 面经、AI应用服务端开发 面经
- 提炼摘要：1. SSE怎么实现的？；2. RabbitMQ怎么保证时序性？；3. 介绍一下IOC和AOP；4. 介绍一下Bean的生命周期。；5. Redis的持久化策略。；6. 说说Mysql的引擎。
- 提取到的面试题/考点：
 - 1. SSE怎么实现的？
 - 2. RabbitMQ怎么保证时序性？
 - 3. 介绍一下IOC和AOP
 - 4. 介绍一下Bean的生命周期。
 - 5. Redis的持久化策略。
 - 6. 说说Mysql的引擎。
 - 7. 场景题:a,b,c的联合索引，只查a,c，会走联合索引吗？

- 8. 介绍一下JVM内存结构。
- 9. 手撕:欢乐数 (LeetCode 202 , ACM风格)

2026-04-08 | 实习-快手电商-大模型数据应用开发一面

- 公司：快手
- 岗位归类：大模型数据应用开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/discuss/871506086656761856>
- 命中搜索词：实习-快手电商-大模型数据应用开发一面
- 提炼摘要：介绍netty的多路复用机制？介绍一下epoll是什么？；（简历）介绍一下双token机制？refresh token的安全性这么考虑的？；（简历）会话机制怎么校验当前客户端是他自己？；（简历）布隆过滤器的业务场景是什么？他是为了做什么的？高并发的流量是由什么决定的？；介绍一下synchronized和ReentrantLock的区别？；介绍一下synchronized的底层实现？Monitor是用来做什么的？
- 提取到的面试题/考点：
- 介绍netty的多路复用机制？介绍一下epoll是什么？
- （简历）介绍一下双token机制？refresh token的安全性这么考虑的？
- （简历）会话机制怎么校验当前客户端是他自己？
- （简历）布隆过滤器的业务场景是什么？他是为了做什么的？高并发的流量是由什么决定的？
- 介绍一下synchronized和ReentrantLock的区别？
- 介绍一下synchronized的底层实现？Monitor是用来做什么的？
- 介绍一下ReentrantLock的底层？
- 线程池的构造参数？
- 介绍一下Spring的AOP？
- 有没有了解过JDK动态代理为什么是基于接口的？
- 介绍一下Spring的IoC？
- 讲讲IoC的循环依赖问题和解决？
- 算法，维护二叉树next指针，用常量级空间

2026-04-08 | AI应用开发日常实习二面-字节

- 公司：字节
- 岗位归类：ai应用开发
- 轮次：二面
- 来源：<https://www.nowcoder.com/feed/main/detail/0607316a2492400896408c95e10951e0>
- 命中搜索词：AI应用开发、AI应用开发 面经、AI应用服务端开发 面经
- 提炼摘要：1. 问关于ReAct的知识，那几个模式有什么差别；2. 给了一个场景，就是设计一个全自动化的agent进行AI漫剧创作，你会怎么设计？；3. 为什么要引入父子索引？BM25引入的原因是什么，比例是怎么设置的，整个具体流程是怎样的，有没有；4. 如果做了rerank，rerank之后返回几个块？有没有做过一些验证来确保效果？；5.

rerank之后的topk截断是怎么实现的？为什么选这个k值，有没有考虑过其他方案？；6. 讲一下上下文工程，还有记忆功能是怎么实现的。

- 提取到的面试题/考点：
- 1. 问关于ReAct的知识，那几个模式有什么差别
- 2. 给了一个场景，就是设计一个全自动化的agent进行AI漫剧创作，你会怎么设计？
- 3. 为什么要引入父子索引？BM25引入的原因是什么，比例是怎么设置的，整个具体流程是怎样的，有没有做rerank操作？
- 4. 如果做了rerank，rerank之后返回几个块？有没有做过一些验证来确保效果？
- 5. rerank之后的topk截断是怎么实现的？为什么选这个k值，有没有考虑过其他方案？
- 6. 讲一下上下文工程，还有记忆功能是怎么实现的。
- 7. 讲一下LRU的原理和实现。
- 8. 讲一下你项目agent框架
- 9. langchain langgraph了解多少 你在你的项目中有使用吗
- 0. 你做的agent响应这么快 你是否有工程优化 是什么
- 1. chunk怎么切的 除了你这么切 还有别的办法吗
- 2. 如果用滑动窗口 如何去重？
- 3. 准确率提升是如何量化的
- 4. 如果你切chunk的时候 第一个段落和第十个段落有联系你怎么做

2026-04-08 | 字节AI应用开发实习面经分享

- 公司：字节
- 岗位归类：ai应用开发
- 轮次：未明确
- 来源：<https://www.nowcoder.com/feed/main/detail/76448605da6e4aee99a394da83d1718f>
- 命中搜索词：AI应用开发、AI应用开发 面经、AI应用服务端开发 面经
- 提炼摘要：1. AI agent 与function call有什么区别。；2. AI Coding用过什么。；3. AI Coding和AI Agent如何一起使用，即如何用AI Coding开发一个AI Ag；4. 有没有试过用langchain和langgraph做AI Agent的开发？有什么感受？；5. 对云熟不熟？有哪些产品？相关的实践有哪些？；6. 有哪些推理引擎？大模型推理有哪些指标？
- 提取到的面试题/考点：
- 1. AI agent 与function call有什么区别。
- 2. AI Coding用过什么。
- 3. AI Coding和AI Agent如何一起使用，即如何用AI Coding开发一个AI Agent
- 4. 有没有试过用langchain和langgraph做AI Agent的开发？有什么感受？
- 5. 对云熟不熟？有哪些产品？相关的实践有哪些？
- 6. 有哪些推理引擎？大模型推理有哪些指标？
- 7. 为什么要用PD分离？业界方案是怎么样的？mooncake的KV Cache是否有外挂？

- 8. 你如何看待AI Agent的发展现状？未来又是什么看法？
- 9. langchain的主要在这个chain 怎么理解这个chain
- 0. 提示词怎么设计的 如何设计
- 1. rag怎么构建的
- 2. 模型微调qwen
- 2.5-3B 用llamafactory 怎么做的
- 3. mcp mcp是什么
- 4. 工具调用失败 怎么办
- 5. 想要按照结构输出怎么办？
- 6. 多agent协作 如何做 设置 workflow
- 7. 介绍一下你的项目
- 8. 你提到的coze, coze有两版, 一个是拖拽式的编排, 新的一版是直接使用agent?, 你使用的是哪一种? 你是怎么设计节点的?
- 9. 你编写了哪些MCP工具, 介绍一下
- 0. 上下文管理是怎么做的, 如何进行记忆

2026-04-05 | 蚂蚁ai应用开发实习二面

- 公司：蚂蚁
- 岗位归类：ai应用开发
- 轮次：二面
- 来源：<https://www.nowcoder.com/feed/main/detail/718c5a06a65645d7a462a56b457cb35e>
- 命中搜索词：大模型应用开发、大模型应用开发

二面、大模型开发、大模型数据应用开发、拼多多大模型应用开发一面凉经、智能体开发、智能体开发面经、腾讯大模型应用开发一面

- 提炼摘要：1.实习拷打；2.简历上项目ai coding的占比有多少；3.自己ai coding的使用范式/习惯；4.项目从0到1，以及1-100开发时有什么区别？；5.function call, map, skills. 智能体演进过程；6.从安全的角度去考虑，用户与Agent，Agent与Agent，以及Agent与后端交互这些链路
- 提取到的面试题/考点：
 - 1.实习拷打
 - 2.简历上项目ai coding的占比有多少
 - 3.自己ai coding的使用范式/习惯
 - 4.项目从0到1，以及1-100开发时有什么区别？
 - 5.function call, map, skills. 智能体演进过程
 - 6.从安全的角度去考虑，用户与Agent，Agent与Agent，以及Agent与后端交互这些链路中存在哪些安全问题
 - 7.mcp与命令行shell脚本实现一些功能有什么区别
 - 8.AI时代对于传统后端开发有哪些影响？是不是程序员就不用懂后端相关的技术了？

2026-04-03 | 淘天-ai应用开发-一面

- 公司：淘天
- 岗位归类：ai应用开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/feed/main/detail/076a8d1acaae4795a28d4a5370872bd6>

• 命中搜索词：大模型应用开发、大模型应用开发

二面、大模型开发、大模型数据应用开发、拼多多大模型应用开发一面凉经、智能体开发、智能体开发面经、腾讯大模型应用开发一面

• 提炼摘要：1. 性能优化演进、；面对 1~2 亿条msg的突发消息洪峰（MQ 压力极大），除了单机单线程处理，还有什么更好的架构级解；2. RabbitMQ vs RocketMQ 技术选型；这两种 MQ 分别适用于什么样的业务场景？；RocketMQ 的 ACK 机制和消息重发机制是怎样的？；如何在这两种 MQ 中实现顺序消费？

• 提取到的面试题/考点：

- 1. 性能优化演进、
- 面对 1~2 亿条msg的突发消息洪峰（MQ 压力极大），除了单机单线程处理，还有什么更好的架构级解决方案？
- 2. RabbitMQ vs RocketMQ 技术选型
- 这两种 MQ 分别适用于什么样的业务场景？
- RocketMQ 的 ACK 机制和消息重发机制是怎样的？
- 如何在这两种 MQ 中实现顺序消费？
- 如何支持分布式事务消息？
- 1. RAG的痛点与优化
- 如何解决 RAG 检索出来的知识忠实度不够、回答有偏差的问题？
- 对于长文档（如 200~300M 包含图片的文档），Chunk（文本块）的切分策略是怎样的？
- 大模型上下文窗口变大后（如从 256K 提升至 1M），如何解决“中间遗忘 (Lost in the middle)”的问题？
- 2. Agent 编排与大模型框架
- 在简单的问答任务和复杂的配置任务中，分别采用了哪种 Agent 架构？
- MAS 多智能体架构在处理复杂任务时有什么明显的缺点？
- 为什么选择 Spring AI 框架？在实际使用中发现它有什么缺陷？
- 1. Agent 评测体系 (Eval)
- 你们通过哪些指标来评测 Agent 的表现？
- 如何定义和评判一个回答是否是“高分回答”？
- 2. 真实大厂高容错场景设计
- 问题要求: 在此背景下，如何升级你的系统架构？如何进行自动化的投放与评测闭环？

2026-04-02 | 字节 ai应用开发

- 公司：字节

- 岗位归类：ai应用开发
- 轮次：一面
- 来源：<https://www.nowcoder.com/feed/main/detail/c5947496fc29440883611173478dc257>
- 命中搜索词：AI应用开发、AI应用开发 面经、AI应用服务端开发 面经
- 提炼摘要：手撕:丑数；自我介绍；DAG编排怎么做的；每个agent用的啥；langgraph怎么编排的；评估怎么做的
- 提取到的面试题/考点：
- 手撕:丑数
- 自我介绍
- DAG编排怎么做的
- 每个agent用的啥
- langgraph怎么编排的
- 评估怎么做的
- 数据集怎么建的
- 语义分块
- 微调和强化学习应用场景
- HNSW怎么构建
- 多文件怎么处理
- 益智题：一个八升水杯，一个五升水杯，怎么得到四升水
- 第二个面试官提问：
- threadlocal
- langchain的checkpoint
- agent harness
- 第三个项目中的感知调度怎么做的

2026-03-21 | 字节 大模型应用开发 二面 日常实习

- 公司：字节
- 岗位归类：大模型应用开发
- 轮次：二面
- 来源：<https://www.nowcoder.com/feed/main/detail/30b6f37830414df5a32d851f2004df7a>
- 命中搜索词：字节 大模型应用开发 二面 日常实习、字节大模型应用开发 日常实习一面
- 提炼摘要：1.实习介绍；2.拷打第一个项目；3.拷打第二个项目；4.在向量化之前，为什么要对长文档进行切片？如果不切片会有什么后果？；5.切片时设置重叠区域的作用是什么？这个比例你通常怎么来确定？；6.讲一下稠密向量与稀疏向量的区别，分别适合处理什么样的搜索需求？
- 提取到的面试题/考点：
- 1.实习介绍

- 2.拷打第一个项目
- 3.拷打第二个项目
- 4.在向量化之前，为什么要对长文档进行切片？如果不切片会有什么后果？
- 5.切片时设置重叠区域的作用是什么？这个比例你通常怎么来确定？
- 6.讲一下稠密向量与稀疏向量的区别，分别适合处理什么样的搜索需求？
- 7.向量库检索出的Top-K结果，如果K值设置得过大，对后续的生成质量有哪些负面影响？
- 8.余弦相似度和欧氏距离在衡量文本相似性时，各自的优缺点是什么？
- 9.为什么在初筛召回之后，还要加一个Rerank模型？它能解决向量搜索的哪些局限？
- 0.如果文档发生了局部更新，如何通过增量索引来避免全量重新向量化？
- 1.在RAG的生成阶段，如何在Prompt中设定边界条件来防止模型在没搜到内容时产生幻觉？
- 2.了解HyDE吗？介绍一下原理，它在处理模糊提问时有哪些优势？
- 3.随着超长上下文模型的出现，你认为传统RAG架构的必要性是否降低了？
- 4.你了解哪些大模型推理框架？SGLang相比vLLM的PagedAttention在推理延迟上有哪些优势？
- 5.调用大模型API时，为什么要使用asyncio异步编程？它在处理高并发请求时有何优势？
- 6.针对大规模PDF解析这种任务，你选择多线程还是多进程？
-
- 7.如何确保Agent返回的结果是标准的JSON格式？如果模型输出中有多余的说明文字，你在后端如何提取？
- 8.场景题:对于RAG，如果检索到了针对同一故障的两份手册，内容相互冲突，请你设计一套逻辑，让模型能够识别冲突并优先选择时效性更高的信息？
- 9.手撕:第k大元素

排除项

- 2026-04-12 | Agent开发 | 更像讨论/求助帖，非稳定面经 | <https://www.nowcoder.com/feed/main/detail/437dadeddf6041ca9ea7575e0b890434>
- 2026-04-11 | Agent岗位爆发，作为后端开发要不要无脑转？ | 明显是求助/讨论帖 | <https://www.nowcoder.com/discuss/872566113119322112>
- 2026-04-10 | 网易互娱 C++开发实习一面(暑期) | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/872225421591404544>
- 2026-04-10 | 4.10-阿里云agent-一面 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/872120976190816256>
- 2026-04-10 | 蚂蚁27暑期智能体大模型应用一面 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/8d817a84839949ed85d2bc7fe5f8e9bf>
- 2026-04-09 | 蚂蚁智能体ai开发一面 暑期实习 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/ce29d4456f5a4474a52fd21f9d9cf741>
- 2026-04-09 | agent开发外包面试常见问题 | 明显是汇总/攻略帖 | <https://www.nowcoder.com/feed/main/detail/4a2bfe1cc9d34a3a90b419840e73e5d9>
- 2026-04-09 | 竞技世界C++开发一面 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/871804904774393856>

- 2026-04-09 | Agent 小厂面经 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/d2812ee9853e41dc9d4885948143fd79>
- 2026-04-09 | 现在走ai应用开发行么 | 明显是求助/讨论帖 | <https://www.nowcoder.com/feed/main/detail/5b2ecf12435f4d3f9580e73cb12924d9>
- 2026-04-09 | 测试和开发怎么选 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/f1dee05b86134e329e164eb76a251e6f>
- 2026-04-09 | 阿里 AI Agent 开发一面 (附答题思路) | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/871729066020257792>
- 2026-04-09 | 大模型Agent面试全攻略 (附答题思路) | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/871718560224112640>
- 2026-04-08 | 27届-暑期实习-字节国际广告CRM-一面凉经 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/871521961380241408>
- 2026-04-08 | 腾讯后台开发二面凉经-内容服务部 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/03b223fc497b42cfad05b1f3c4017b78>
- 2026-04-08 | Agent开发八股合集真实面经总结版 | 明显是汇总/攻略帖 | <https://www.nowcoder.com/feed/main/detail/76b321bffc5e460fb316813352d8d950>
- 2026-04-07 | 高德地图-汽车业务-C++开发-二面面经 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/871039265591877632>
- 2026-04-07 | 评论赏金任务来啦 0410 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/870969335479554048>
- 2026-04-04 | 美团AI软件开发一面-暑期实习 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/0c3300eba2ff4ae1a265571fb371469e>
- 2026-04-03 | 【温氏股份】急招算法、人工智能方向应届生，快速安排面试~ | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/438ad97af0db4dfbb20a01471fe27b82>
- 2026-04-02 | 27届-AI应用开发 | 缺少面试特征词 | <https://www.nowcoder.com/feed/main/detail/df88b45946f843799e3603675520efca>
- 2026-03-29 | 字节测试开发-数据平台部门一面凉经 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/867788123156008960>
- 2026-03-28 | 后端还是AI应用开发 | 明显是求助/讨论帖 | <https://www.nowcoder.com/feed/main/detail/53dc9eea05a4aff845cde193fe9cacf>
- 2026-03-27 | 评论赏金任务来啦 04月3日 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/867039674815045632>
- 2026-03-26 | 走传统后端开发还是应该直接ai应用开发 | 明显是求助/讨论帖 | <https://www.nowcoder.com/feed/main/detail/8852c115be4648f9ab465f3f478a1606>
- 2026-03-19 | 评论赏金任务来啦 0326 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/864104469254975488>
- 2026-03-14 | 字节跳动-实习面经- AI开发 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/6c7120c856564da0b42cf9092908d203>
- 2026-03-12 | 评论赏金任务来啦 0313 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/861564783961571328>

- 2026-03-06 | 字节跳动AI开发一面 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/f24537d842214021b8e617f10482ede8>
- 2026-02-28 | 今日评论激励赏金任务来啦！02月28日任务 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/857225787529818112>
- 2026-02-26 | 今日评论激励赏金任务来啦！02月26日任务 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/856496290396401664>
- 2026-02-25 | 评论激励赏金任务来啦！02月25日任务 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/856125466250342400>
- 2026-02-14 | 今日评论激励赏金任务来啦！02月14日任务 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/852130727087861760>
- 2026-02-13 | 今日评论激励赏金任务来啦！02月13日任务 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/851791558532665344>
- 2026-02-12 | 今日评论激励赏金任务来啦！02月12日任务 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/851421206564630528>
- 2026-02-10 | 今日评论激励赏金任务来啦！02月11日任务 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/850692157953355776>
- 2026-02-06 | 今日评论激励赏金任务来啦！02月06日任务 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/849249984636596224>
- 2026-02-05 | 今日评论激励赏金任务来啦！02月05日任务 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/848879871160745984>
- 2026-02-04 | 今日评论激励赏金任务来啦！02月04日任务 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/848520474282766336>
- 2026-01-30 | 评论激励赏金任务来啦！01月30日任务 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/846719823630331904>
- 2026-01-28 | 评论激励赏金任务来啦！01月28日任务 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/846025564459708416>
- 2026-01-19 | 1月面试的大厂日常实习记录贴 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/842882899173994496>
- 2026-01-07 | 评论激励赏金任务来啦！01月22日任务 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/838418769830506496>
- 2025-12-20 | 给27届家人找实习的经验分享 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/831928503183355904>
- 2025-12-17 | 评论赏金任务来啦
12.26 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/830766628437512192>
- 2025-12-15 | 字节大模型算法实习一面凉经 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/b70dc597f95f4b8394d97ff3bd88b858>
- 2025-11-19 | 27届菜鸡小硕 后端找实习记录 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/461164f75fde43cd87f409ac31033915>
- 2025-11-15 | 啊啊啊啊怎么周六给我offer，电话都没有直接甩个邮件，佬们快来帮我选一下 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/20dce7fd9d964b4191b6153b6b67f5cd>

- 2025-11-11 | 签了三方，我的秋招结束了！面不动了！ | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/817714902344220672>
- 2025-11-05 | 评论赏金任务来啦
11.11 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/815611758369845248>
- 2025-08-27 | 计算机日常答疑，一起寻找问题的最优解 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/790328901208707072>
- 2025-08-19 | 评论赏金任务来啦0929 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/787285884310650880>
- 2025-07-19 | 回忆录：后端鼠鼠苦苦哀求日常实习 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/776102677414539264>
- 2025-04-15 | 暑期实习 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/f8905f97a26b49128ddcbec824f5d082>
- 2025-04-14 | 大模型 0 实习经验暑期实习投递记录 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/741384896458817536>
- 2025-04-12 | 记录一下面试经历 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/b3f1d3070c414d07ab7bc4b0995b5edf>
- 2025-04-09 | 记录一下实习/秋招 投递情况 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/feed/main/detail/298cc12bdf51458dac9ee6e8381b4291>
- 2025-03-31 | 同程旅行 全栈 春招 凉经 一面
SK精选 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/736323632040779776>
- 2025-02-21 | 2025 2 19 Java 面试题（美团
快手） | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/722397808745033728>
- 2024-03-01 | 阿里云暑期实习提前批一面（技术服务平台）|讲解 | 0301 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/593104748941684736>
- 2024-02-29 | 【快手】电商Java日常实习一面 | 讲解 | 0229 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/592736631489146880>
- 2024-02-18 | 面试：百度一面，吓尿了 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/588738046242177024>
- 2023-07-03 | 谈谈这次招聘季的经验和教训 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/505378429559173120>
- 2023-05-11 | 牛客广告产品再上新！搜索品牌专区迎战秋招季 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/486210903277432832>
- 2023-03-31 | 字节-测试开发-日常实习-三面面经（已发offer） | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/471324815715069952>
- 2022-09-06 | 顺利上岸字节本地化后端，但也真的很不顺利 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/396740931939053568>
- 2022-06-19 | 数据分析暑期实习面经总结 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/367868926699556864>
- 2022-03-26 | 23届前端实习 腾讯、美团、字节、蚂蚁
面经 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353159419414388736>

- 2021-10-02 | 150+场面试 拿到了20+offer【22双非本科求职路】 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353158660908064768>
- 2021-09-30 | 秋招上岸，个人历程和面经总结 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353158653773553664>
- 2021-04-22 | 实习面经汇总：腾讯、阿里、字节 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353157855148711936>
- 2021-04-21 | 字节跳动Android日常实习一二三面面经(附部分回答) | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353157849465430016>
- 2021-03-11 | 初识产品经理校招 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353148704649584640>
- 2020-11-13 | 2020秋招人力岗面试题 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353157328121831424>
- 2020-09-29 | 记录一下自己艰苦的2020秋招---累觉不爱 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353157099884584960>
- 2020-09-16 | 2021秋招数据分析个人总结-持续更新 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353156983446511616>
- 2020-05-08 | 牛客匿名发布功能说明 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353156414929575936>
- 2020-04-04 | 3-4月面经汇总-字节跳动，美团，腾讯，阿里算法岗 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353156268942630912>
- 2019-11-27 | 秋招c++面经，少量算法、java，百度、滴滴、京东、58等 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353155939832373248>
- 2019-10-29 | 一个双非渣硕的自我救赎之路 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353155830084214784>
- 2019-09-23 | 后台秋招面经总结：字节，百度，网易，腾讯，阿里等 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353155546041753600>
- 2019-09-01 | 机械狗软件求职记 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353155336729206784>
- 2019-03-19 | 大数据开发2019春招面经 | 标题/正文前段缺少岗位关键词 | <https://www.nowcoder.com/discuss/353154895723307008>
- None | index | 抓取失败: window.__INITIAL_STATE__ not found | <https://www.nowcoder.com/discuss/offer/index>
- None | 826147231207055360 | 抓取失败: window.__INITIAL_STATE__ not found | <https://www.nowcoder.com/discuss/826147231207055360>
- None | 505159fb5f874d909c0c252f3f959ea0 | 抓取失败: window.__INITIAL_STATE__ not found | <https://www.nowcoder.com/feed/main/detail/505159fb5f874d909c0c252f3f959ea0>
- None | 0733c5af82ff4bc8aaa1ddd45ae041b0 | 抓取失败: window.__INITIAL_STATE__ not found | <https://www.nowcoder.com/feed/main/detail/0733c5af82ff4bc8aaa1ddd45ae041b0>

牛客近6个月AI应用开发面经：知识点聚类

- 面经样本数：32
- 提取题目数：464

聚类总览

知识点	题目数	涉及公司数	涉及面经数
Agent架构与范式	66	13	25
RAG与检索优化	51	10	18
后端基础与工程化	45	10	17
框架、MCP与工具调用	38	9	19
记忆与上下文管理	34	10	18
算法与编码题	22	7	17
评测、幻觉与数据集	20	8	14
模型基础与推理机制	17	6	11
项目介绍与项目拷打	14	10	12
业务场景设计	11	7	9
AI Coding与协作方式	10	8	8
前端与跨端工程	7	3	3
HR与流程信息	1	1	1
其他	122	12	26

逐类展开

Agent架构与范式

- 题目数：66
- 覆盖公司：京东、字节、小红书、快手、淘天、百度、米哈游、腾讯、蚂蚁、阿里云、阿里淘天、阿里灵犀互娱、高德
- 代表题目：
 - (3) #Agent面试会问什么？# —— 字节、百度
 - (1) 为什么要做多 Agent，而不是把所有能力都塞进一个 Agent —— 小红书
 - (1) Agent 的规划、执行、反思三段式链路怎么设计 —— 小红书
 - (1) RAG 可以怎么分类，Agentic RAG 和传统 RAG 差别在哪 —— 小红书
 - (1) Agent 和传统 workflow 的本质区别是什么，什么时候该选 Agent，什么时候不该选 —— 淘天
 - (1) 智能体开发范式有哪些，分别适合什么场景 —— 淘天
 - (1) 智能体的记忆机制应该怎么设计，为什么“全都记住”反而是坏事 —— 淘天
 - (1) 设计智能体时，稳定性和安全性应该优先在哪些层面做约束 —— 淘天

RAG与检索优化

- 题目数：51
- 覆盖公司：京东、字节、小红书、快手、淘天、腾讯、蔚来、蚂蚁、金山云、阿里云
- 代表题目：
 - (1) RAG 项目里怎么做召回闭环，才能让系统真的越用越准 —— 小红书
 - (1) HyDE 的原理是什么，什么时候有效，什么时候会害人 —— 小红书
 - (1) IVF、PQ、IVF-PQ 分别在做什么，为什么它们能把向量检索做快 —— 小红书
 - (1) 向量索引有哪些典型类别，分别适合什么场景 —— 小红书
 - (1) 为什么要用 RAG，而不是直接让模型回答安全分析问题 —— 阿里云
 - (1) 文档chunk完有多少id —— 蚂蚁
 - (1) 多路召回和重排序是怎么提高检索成功率的 —— 蚂蚁
 - (1) 你之前向量检索加redis缓存是怎么做的？redis键和值是什么？ —— 字节

后端基础与工程化

- 题目数：45
- 覆盖公司：京东、字节、快手、淘天、百度、腾讯、蔚来、金山云、阿里国际、高德
- 代表题目：
 - (1) 了解redis集群吗？了解redis底层的存储逻辑吗？ —— 快手
 - (1) redis的性能单点能够有多少？集群能够有多少？ —— 快手
 - (1) 我看你的黑马点评项目用的rabbitmq，为啥选这个？这个对比其他mq有什么特别的优势？（我说当时只学了那个，所以就用了，实际上rocketmq应该会更好） —— 快手
 - (1) Java 里的强引用、软引用、弱引用、虚引用在工程里怎么理解，别只背定义 —— 高德
 - (1) 本地部署要考虑什么？配置管理放在哪里？ —— 腾讯
 - (1) JMeter 压测的具体参数是怎么设置的？ —— 京东
 - (1) Redis 是单线程的，为什么它还能支持高并发？ —— 京东
 - (1) 在实际业务场景中，你怎么判断什么时候需要加缓存？为什么不能只用 MySQL？ —— 京东

框架、MCP与工具调用

- 题目数：38
- 覆盖公司：字节、小红书、快手、淘天、米哈游、腾讯、蚂蚁、金山云、阿里国际
- 代表题目：
 - (1) MCP、Function Calling、Skills、Agent 分别是什么，它们之间是什么关系 —— 小红书
 - (1) LangChain 和 LangGraph 如果要选，你更看重什么维度 —— 淘天
 - (1) MCP 和 CLI 工具接入在工程上分别解决什么问题，它们不是一回事 —— 淘天

- (1) 你的这个agent项目用了spring Ai Ailibaba，为什么用这个，他与langchain4j、spring ai有什么区别？选型的时候怎么考虑的？—— 快手
- (1) 你用spring Ai Ailibaba的时候感觉有哪些比较好的点？哪些不太方便的点？—— 快手
- (1) 介绍一下langchain和langgraph—— 蚂蚁
- (1) skill是自己写的吗，互相之间可以联动吗？—— 字节
- (1) 为什么不用skill-creator—— 字节

记忆与上下文管理

- 题目数：34
- 覆盖公司：字节、小红书、快手、淘天、百度、米哈游、腾讯、蚂蚁、阿里淘天、阿里灵犀互娱
- 代表题目：
- (1) Agent 的记忆一般怎么分层，为什么不能只靠聊天历史—— 小红书
- (1) 短期记忆和长期记忆应该如何精准提取，分别存什么内容更合理—— 淘天
- (1) 做完rag后llm能理解图片里的内容吗，还是通过加入上下文token后靠文本理解—— 蚂蚁
- (1) 为什么要设计这三层记忆架构，是一个通用的设计还是参考了什么论文—— 字节
- (1) 新的记忆加到文件里是直接append吗？覆盖？还是改写？—— 字节
- (1) 记忆模块存在哪种介质上—— 腾讯
- (1) 跨会话通过压缩保留关键症状，这一块详细介绍—— 腾讯
- (1) 用户今天问A，三天后问B，怎么关联两次提问，记忆怎么跟用户维度绑定。（session级视角、没考虑并发和多次咨询隔—— 腾讯

算法与编码题

- 题目数：22
- 覆盖公司：字节、快手、米哈游、腾讯、蔚来、蚂蚁、阿里国际
- 代表题目：
- (2) 算法题—— 字节、腾讯
- (1) 手撕合并k个升序链表（前几天pdd一面刚写完）—— 快手
- (1) 算法题口头说一下1200w的数据里，怎么找到中位数—— 蚂蚁
- (1) 字符串数字相减—— 字节
- (1) 字符串最长连续序列—— 字节
- (1) 手撕:一道原创题，大致题意:给定一个链表"head={4,2,8}"和一课二叉树 root=Rn{1,4,4,1,2,null,2,null,null,nul 1,2,8,2,1,null,null,null,null,8,null,null}(貌似是这样的)判断一下这棵树中存不存在一条路径和链表是一样的—— 快手
- (1) 如果对参数传入一个字符串类型，如何在方法内部修改它并把这个字符串的值传递出来?—— 阿里国际
- (1) 代码实现LRU—— 米哈游

评测、幻觉与数据集

- 题目数：20
- 覆盖公司：京东、字节、淘天、百度、腾讯、蚂蚁、阿里云、阿里灵犀互娱
- 代表题目：
- (1) AI 应用里的在线实验和传统 AB 测试有什么差别，为什么更难做 —— 淘天
- (1) RAG 的效果你怎么评估，不能只说看用户反馈 —— 阿里云
- (1) 幻觉问题你一般怎么处理，尤其是安全场景下的高风险幻觉 —— 阿里云
- (1) 怎么解决幻觉问题 —— 蚂蚁
- (1) 有没有评估过你的记忆架构，在没参考openclaw之前和之后的指标有什么变化，成功率有对比吗？ —— 字节
- (1) 你说你的文档准确率达到90%，具体做的改变是什么 —— 字节
- (1) 你的数据集是怎么收集的，然后是怎么评估的呢？打分是谁在打分？人在打分，还是模型裁判员去打分？如果有很好用的模型裁判员，为什么不用裁判员模型 —— 字节
- (1) 意图识别用的是哪个大模型？如何判断和测试意图识别结果的准确率？ —— 京东

模型基础与推理机制

- 题目数：17
- 覆盖公司：字节、百度、米哈游、蚂蚁、阿里云、高德
- 代表题目：
- (1) 平时你怎么利用大模型开发项目，而不是停留在写点 demo —— 阿里云
- (1) 介绍一下对temperature的理解 —— 蚂蚁
- (1) 用过哪些模型，模型选择不是看排行榜的话，你一般怎么选 —— 高德
- (1) 介绍一下KV cache —— 字节
- (1) attention和self-attention —— 米哈游
- (1) 为什么attention比rnn好 —— 米哈游
- (1) LLM 的输入到底是什么？模型真正看到的内容包含哪些？ —— 百度
- (1) self-attention 的核心作用是什么？为什么要拆成 Q、K、V？ —— 百度

项目介绍与项目拷打

- 题目数：14
- 覆盖公司：京东、字节、小红书、快手、淘天、米哈游、蚂蚁、金山云、阿里云、高德
- 代表题目：
- (6) 自我介绍 —— 字节、小红书、淘天、米哈游、金山云
- (3) 项目拷打 —— 京东、淘天、蚂蚁
- (1) 介绍你写的这个项目 —— 小红书
- (1) 做一下自我介绍 —— 阿里云
- (1) 介绍一个你做的这个项目，重点讲架构和你负责的部分 —— 阿里云

- (1) 做一下自我介绍，重点讲你在 AI 编码和跨端工程结合上的经历 —— 高德
- (1) 你的AI项目的亮点是什么? —— 快手

业务场景设计

- 题目数：11
- 覆盖公司：京东、字节、快手、腾讯、阿里云、阿里国际、阿里灵犀互娱
- 代表题目：
 - (1) 你对网络安全方向怎么看，为什么 AI 应用在这个方向上既有价值又有风险 —— 阿里云
 - (1) 那为什么医院的问答项目里，不采用同样的实现 —— 字节
 - (1) 除了导购，你认为在电商领域（如京东）还有哪些场景可以用到 Agent？ —— 京东
 - (1) 设计一个游戏智能客服，整机架构 —— 阿里灵犀互娱
 - (1) 设计一个游戏助手agent，记忆系统怎么设计 —— 阿里灵犀互娱
 - (1) 如何定义游戏角色状态 —— 阿里灵犀互娱
 - (1) 你未来想做什么业务 —— 阿里国际
 - (1) 工具调用的安全控制是怎么实现的? —— 快手

AI Coding与协作方式

- 题目数：10
- 覆盖公司：字节、快手、蚂蚁、金山云、阿里国际、阿里淘天、阿里灵犀互娱、高德
- 代表题目：
 - (1) 你是什么时候开始用 Claude Code 这一类工具的，真正改变你的是什么 —— 高德
 - (1) 你用 AI 写代码时一般怎么拆解任务，才能让结果比较稳定 —— 高德
 - (1) 你如何使用ai coding的，怎么保证能按照你的要求完成代码 —— 阿里灵犀互娱
 - (1) 有了解claudeCode的Hook机制功能知道嘛?解释一下 —— 快手
 - (1) 聊AI Coding，使用过AI Coding写代码吗？有使用AI Coding进行项目迭代吗？ —— 金山云
 - (1) 有用过哪些AI相关的编程工具? —— 阿里国际
 - (1) 算法:AI coding题，要求实现一个工单处理系统 —— 阿里淘天
 - (1) AI Coding用过什么 —— 字节

前端与跨端工程

- 题目数：7
- 覆盖公司：字节、阿里云、高德
- 代表题目：
 - (1) 前端为什么不直接做一个普通聊天页，而是要选更适合流式交互的方案 —— 阿里云
 - (1) Flutter 的核心原理是什么，如果不从 Widget 开始讲，你会怎么讲 —— 高德
 - (1) Flutter 和 React Native、原生开发最大的区别，不要只说性能 —— 高德

- (1) Flutter 的 UI 为什么能回到原生页面上，这个过程本质上发生了什么 —— 高德
- (1) 如果文字内容变化导致宽度发生变化，Flutter 的重排到底是怎么发生的 —— 高德
- (1) Flutter 里 build、layout、paint、composite 这几段如果线上卡顿，你怎么判断卡在哪一段 —— 高德
- (1) 在没有前端界面的情况下，本地图片是如何传到后端并进行识别的？ —— 字节

HR与流程信息

- 题目数：1
- 覆盖公司：快手
- 代表题目：
- (1) 转正率如何？看表现，一般都能转 —— 快手

其他

- 题目数：122
- 覆盖公司：字节、快手、淘天、百度、米哈游、腾讯、蔚来、蚂蚁、金山云、阿里国际、阿里淘天、高德
- 代表题目：
- (1) 上传文档时，是一个一个上传还是批量上传 —— 蚂蚁
- (1) 团队里常用哪些开发工具比较多，如果从 AI 协作方式去分类，你会怎么讲 —— 高德
- (1) 对商业项目来说，开发者过去习惯手动掌控每个细节，到了 AI 提供生产力的时代，你怎么看这个变化 —— 高德
- (1) 你这里的意图判断是怎么做的？直接交给模型做，还是这个用户的查询query做了一些其他处理 —— 字节
- (1) 改的方式 模型改吗？模型调用editor去改文件是吗？ —— 字节
- (1) 对于每一个用户，在系统中怎么保存这些不同用户的文档？ —— 字节
- (1) 其实就是模型通过路径去读这些文件是吧，那如果两个人登了同一个账号，在同一时刻，这个文件会被两个不同的用户使用，在这个情况下，你们的系统如何处理？怎么解决这个问题？ —— 字节
- (1) 讲一下transformer架构 —— 字节

牛客近6个月AI应用开发面经：高频题题库与答案框架

- 面经样本数：32
- 高频主题数：14

高频题总览

高频题主题	命中题数	涉及公司数	涉及面经数
检索/召回/重排/向量索引优化	38	10	16
RAG/Agent 评测与幻觉治理	27	10	17
算法题/手撕题	26	9	19
MCP / Function Calling / Skills / Tool Calling	24	10	15
记忆系统设计	22	7	12
LangChain / LangGraph / 编排框架选型	20	5	10
后端基础：Redis/缓存/MQ/高并发	17	7	9
场景设计题（游戏/电商/安全/客服）	15	7	10
项目介绍/项目拷打	14	10	12
模型基础：Transformer/KV Cache/Temperature	13	4	8
AI Coding / Claude Code 实践	12	9	9
Agent 与 workflow/单多 Agent 选型	10	7	9
RAG 系统设计	5	4	4
意图识别与 Query 理解	5	4	5

题库与答案框架

检索/召回/重排/向量索引优化

- 命中题数：38
- 涉及公司：京东、字节、小红书、快手、淘天、腾讯、蔚来、蚂蚁、金山云、高德
- 高频问法：
 - (1) RAG 项目里怎么做召回闭环，才能让系统真的越用越准
 - (1) HyDE 的原理是什么，什么时候有效，什么时候会害人
 - (1) IVF、PQ、IVF-PQ 分别在做什么，为什么它们能把向量检索做快
 - (1) 向量索引有哪些典型类别，分别适合什么场景
 - (1) 文档chunk完有多少id
 - (1) 多路召回和重排序是怎么提高检索成功率的
- 答案框架：

1. 先拆问题：召回质量、排序质量、检索延迟、索引成本。

- 回答常见优化：BM25 + 向量混检、query rewrite、HyDE、rerank。
- 索引层说明：HNSW 适合高召回低延迟，IVF/PQ 更偏大规模压缩。
- 针对 PDF/OCR/表格，强调结构化抽取和版面信息保留。
- 结尾说评估方式：Recall@K、MRR、NDCG、端到端正确率。

RAG/Agent 评测与幻觉治理

- 命中题数：27
- 涉及公司：京东、字节、小红书、淘天、百度、腾讯、蚂蚁、金山云、阿里云、阿里灵犀互娱
- 高频问法：
 - (1) RAG 项目里怎么做召回闭环，才能让系统真的越用越准
 - (1) AI 应用里的在线实验和传统 AB 测试有什么差别，为什么更难做
 - (1) RAG 的效果你怎么评估，不能只说看用户反馈
 - (1) 幻觉问题你一般怎么处理，尤其是安全场景下的高风险幻觉
 - (1) 多路召回和重排序是怎么提高检索成功率的
 - (1) 怎么解决幻觉问题
- 答案框架：

1. 先分离检索评测和生成评测，不把问题混成一个指标。

- 检索看 Recall@K / MRR / NDCG；生成看正确率、引用一致性、任务成功率。
- 数据集来源要说清：真实日志、人工构造、失败样本回流。
- 幻觉治理从三层答：输入约束、检索增强、输出校验/拒答。
- 最好补线上监控：抽样复核、回流闭环、AB 实验。

算法题/手撕题

- 命中题数：26
- 涉及公司：京东、字节、快手、米哈游、腾讯、蔚来、蚂蚁、阿里国际、阿里淘天
- 高频问法：
 - (2) 算法题
 - (1) 手撕合并k个升序链表（前几天pdd一面刚写完）
 - (1) 算法题口头说一下1200w的数据里，怎么找到中位数
 - (1) 字符串数字相减
 - (1) 字符串最长连续序列
 - (1) Redis 做语义向量的相似度检索，内部用的是什么算法？
- 答案框架：

1. 先确认题意和输入输出边界。

- 说思路时先给复杂度，再落到数据结构选择。
- 编码时注意鲁棒性、边界条件、变量命名。

- 写完主动做样例验证。
- 如果时间不够，先给可行解再优化。

MCP / Function Calling / Skills / Tool Calling

- 命中题数：24
- 涉及公司：字节、小红书、快手、淘天、米哈游、腾讯、蚂蚁、金山云、阿里国际、高德
- 高频问法：
 - (1) MCP、Function Calling、Skills、Agent 分别是什么，它们之间是什么关系
 - (1) MCP 和 CLI 工具接入在工程上分别解决什么问题，它们不是一回事
 - (1) 讲一下你做过的 Agent Skill，不要从业务价值讲，直接讲技术拆分
 - (1) skill是自己写的吗，互相之间可以联动吗？
 - (1) 为什么不用skill-creator
 - (1) skill是什么 Auto-coder skill怎么工作
- 答案框架：

1. 先逐一概念，不要混着说。

- Function Calling/Tool Calling 是模型触发工具；Skill 更像能力封装；MCP 更像工具/上下文接入协议。
- 说明它们如何协同：模型规划 -> 选择工具/skill -> 调用执行 -> 回填结果。
- 补充工程问题：权限、超时、幂等、失败重试、可观测性。
- 最后说实际价值：降低接入成本，提高复用和治理能力。

记忆系统设计

- 命中题数：22
- 涉及公司：字节、小红书、快手、淘天、米哈游、腾讯、阿里灵犀互娱
- 高频问法：
 - (1) Agent 的记忆一般怎么分层，为什么不能只靠聊天历史
 - (1) 智能体的记忆机制应该怎么设计，为什么“全都记住”反而是坏事
 - (1) 短期记忆和长期记忆应该如何精准提取，分别存什么内容更合理
 - (1) 为什么要设计这三层记忆架构，是一个通用的设计还是参考了什么论文
 - (1) 有没有评估过你的记忆架构，在没参考openclaw之前和之后的指标有什么变化，成功率有对比吗？
 - (1) 新的记忆加到文件里是直接append吗？覆盖？还是改写？
- 答案框架：

1. 分层回答：短期记忆、长期记忆、用户画像/外部知识。

- 说明写入时机：并不是“全量记忆”，而是事件触发/摘要压缩/显著性提取。
- 说明存储介质：会话缓存、数据库/向量库、文件或KV。
- 说明召回策略：按用户、会话、任务阶段、时间窗口过滤。
- 补充治理：过期淘汰、冲突解决、并发一致性、隐私隔离。

LangChain / LangGraph / 编排框架选型

- 命中题数：20
- 涉及公司：字节、快手、淘天、腾讯、蚂蚁
- 高频问法：
 - (1) LangChain 和 LangGraph 如果要选，你更看重什么维度
 - (1) 你的这个agent项目用了spring Ai Ailibaba，为什么用这个，他与langchain4j、spring ai有什么区别？选型的时候怎么考虑的？
 - (1) 你了解agent编排的有向无环图吗？你用了吗？为什么没用？
 - (1) 介绍一下langchain和langgraph
 - (1) FSM状态机分哪个部分，怎么实现
 - (1) langchain和langgraph的区别？为什么有了langchain还要langgraph
- 答案框架：

1. 先说定位：LangChain 偏快速搭链路，LangGraph 偏复杂状态机编排。

- 如果有分支、循环、重试、人工介入、checkpoint，LangGraph 更合适。
- 如果只是简单工具串联，LangChain 更轻更快。
- 选型时讲维度：开发效率、可观测性、状态管理、扩展性、维护成本。
- 最后结合自己项目说“为什么这次这么选”。

后端基础：Redis/缓存/MQ/高并发

- 命中题数：17
- 涉及公司：京东、字节、快手、淘天、蔚来、金山云、阿里国际
- 高频问法：
 - (1) 了解redis集群吗？了解redis底层的存储逻辑吗？
 - (1) redis的性能单点能够有多少？集群能够有多少？
 - (1) 我看你的黑马点评项目用的rabbitmq，为啥选这个？这个对比其他mq有什么特别的优势？（我说当时只学了this，所以就用了，实际上rocketmq应该会更好）
 - (1) 你之前向量检索加redis缓存是怎么做的？redis键和值是什么？
 - (1) 你觉得 redis+milvus向量存储 和mysql+文件系统+向量检索，他们的优劣如何？优点缺点分别是什么
 - (1) Redis 做语义向量的相似度检索，内部用的是什么算法？
- 答案框架：

1. 先回到业务问题：为什么引入缓存/MQ，而不是只背定义。

- Redis 重点答数据结构、单线程高性能原因、持久化/集群/一致性。
- MQ 重点答解耦、削峰、异步，以及顺序性、幂等、丢消息治理。
- 缓存题要讲失效策略、穿透击穿雪崩、冷热数据。
- 最好结合自己项目里实际怎么用。

场景设计题（游戏/电商/安全/客服）

- 命中题数：15
- 涉及公司：京东、字节、快手、淘天、蚂蚁、阿里云、阿里灵犀互娱
- 高频问法：
 - (1) 设计智能体时，稳定性和安全性应该优先在哪些层面做约束
 - (1) 你对网络安全方向怎么看，为什么 AI 应用在这个方向上既有价值又有风险
 - (1) 为什么要用 RAG，而不是直接让模型回答安全分析问题
 - (1) 幻觉问题你一般怎么处理，尤其是安全场景下的高风险幻觉
 - (1) 那为什么医院的问答项目里，不采用同样的实现
 - (1) 除了导购，你认为在电商领域（如京东）还有哪些场景可以用到 Agent？
- 答案框架：
 1. 先画业务目标：用户是谁，核心任务是什么，成功指标是什么。
 - 再拆系统：意图识别、检索/知识库、工具调用、状态管理、输出策略。
 - 说明为什么需要 Agent/多 Agent，以及每个角色负责什么。
 - 补稳定性与安全：权限、审计、拒答、异常兜底。
 - 最后补评测与上线：离线集、灰度、AB、人工复核。

项目介绍/项目拷打

- 命中题数：14
- 涉及公司：京东、字节、小红书、快手、淘天、米哈游、蚂蚁、金山云、阿里云、高德
- 高频问法：
 - (6) 自我介绍
 - (3) 项目拷打
 - (1) 介绍你写的这个项目
 - (1) 做一下自我介绍
 - (1) 介绍一个你做的这个项目，重点讲架构和你负责的部分
 - (1) 做一下自我介绍，重点讲你在 AI 编码和跨端工程结合上的经历
- 答案框架：
 1. 先给业务背景：用户是谁、核心痛点是什么、为什么值得用 AI/Agent 做。
 - 再讲系统方案：输入 -> 处理链路 -> 模型/检索/工具 -> 输出。
 - 明确你自己的 ownership：你主导了哪几块，做了什么关键决策。
 - 量化结果：准确率、成功率、响应时延、成本、人工替代率。
 - 最后补一段复盘：踩过什么坑、怎么权衡、如果重做会怎么改。

模型基础：Transformer/KV Cache/Temperature

- 命中题数：13
- 涉及公司：字节、百度、米哈游、蚂蚁
- 高频问法：
- (1) 介绍一下对temperature的理解
- (1) 介绍一下KV cache
- (1) attention和self-attention
- (1) 为什么attention比rnn好
- (1) self-attention的核心作用是什么？为什么要拆成 Q、K、V？
- (1) 为什么 attention 能建模长距离关系？
- 答案框架：

1. Transformer 至少讲清 self-attention、前馈层、位置编码、残差归一化。

- KV Cache 要说明它缓存什么、为什么能加速 decode、代价是什么。
- Temperature/Top-p/Top-k 回答采样多样性与稳定性的权衡。
- 如果问微调/强化学习，先分清 SFT、偏好优化、在线强化。
- 强调“懂概念 + 懂工程影响”，不要只背术语。

AI Coding / Claude Code 实践

- 命中题数：12
- 涉及公司：字节、快手、腾讯、蚂蚁、金山云、阿里国际、阿里淘天、阿里灵犀互娱、高德
- 高频问法：
- (1) 你是什么时候开始用 Claude Code 这一类工具的，真正改变你的是什么
- (1) 你用 AI 写代码时一般怎么拆解任务，才能让结果比较稳定
- (1) skill是什么 Auto-coder skill怎么工作
- (1) 你如何使用ai coding的，怎么保证能按照你的要求完成代码
- (1) 有了解claudeCode的Hook机制功能知道嘛?解释一下
- (1) 聊AI Coding，使用过AI Coding写代码吗？有使用AI Coding进行项目迭代吗？
- 答案框架：

1. 先说你怎么用：需求澄清、方案草拟、代码生成、测试补齐、文档整理。

- 强调边界：AI 适合提效，不直接代替关键设计和最终验收。
- 保证质量的办法：任务拆小、明确约束、提供上下文、单测/人工 review。
- 补充失败场景：需求模糊、跨文件重构、隐式约束多时容易跑偏。
- 最后说你自己的最佳实践和收益。

Agent 与 workflow/单多 Agent 选型

- 命中题数：10
- 涉及公司：京东、字节、小红书、淘天、米哈游、阿里淘天、阿里灵犀互娱

- 高频问法：
- (1) 为什么要做多 Agent，而不是把所有能力都塞进一个 Agent
- (1) Agent 和传统 workflow 的本质区别是什么，什么时候该选 Agent，什么时候不该选
- (1) 智能体开发范式有哪些，分别适合什么场景
- (1) 介绍一下 AI 智能体的 workflow 以及 RAG 知识库是怎么设计的？
- (1) 什么时候采用单agent，什么时候采用多agent，原因
- (1) 这个多agent项目是主子agent的项目吗？
- 答案框架：

1. 先下定义：workflow 是预定义流程，Agent 是基于上下文动态决策。

- 再讲适用场景：稳定高频任务用 workflow；开放复杂任务才考虑 Agent。
- 单 Agent 适合轻量任务，多 Agent 适合角色分工明显、工具复杂、需要并行/反思的场景。
- 补充风险：多 Agent 会带来状态同步、成本、时延、调试难度。
- 结尾给出选型原则：先 workflow，必要时再逐步 Agent 化。

RAG 系统设计

- 命中题数：5
- 涉及公司：京东、小红书、淘天、阿里云
- 高频问法：
- (1) RAG 可以怎么分类，Agentic RAG 和传统 RAG 差别在哪
- (1) RAG 项目里怎么做召回闭环，才能让系统真的越用越准
- (1) 为什么要用 RAG，而不是直接让模型回答安全分析问题
- (1) 介绍一下 AI 智能体的 workflow 以及 RAG 知识库是怎么设计的？
- (1) 你在这个 RAG 系统优化里面，在多阶段 RAG 系统中采用了 BM25 和向量混合检索，然后这个是怎么去设计两者的结合逻辑的？然后混合策略的话具体是如何去提升检索效果的？
- 答案框架：

1. 先回答为什么要用 RAG：补齐私域知识、降低幻觉、提升可控性。

- 再讲链路：文档接入 -> 清洗切块 -> 索引 -> 召回 -> 重排 -> 生成。
- 讲清关键设计：chunk 粒度、元数据、混合检索、引用回传。
- 补充失败兜底：低置信度拒答、回退模板、人工介入。
- 最后讲如何评估：检索命中率、答案正确率、业务指标。

意图识别与 Query 理解

- 命中题数：5
- 涉及公司：京东、字节、快手、阿里灵犀互娱
- 高频问法：
- (1) 你这里的意图判断是怎么做的？直接交给模型做，还是这个用户的查询query做了一些其他处理

- (1) 意图识别用的是哪个大模型？如何判断和测试意图识别结果的准确率？
- (1) 意图识别小模型如何训练，数据集怎么构造
- (1) (追问)整体检索流程是怎样的?从query到最终上下文的完整流程是什么?
- (1) 并行化意图识别是什么？为什么要并行化？如何实现的？
- 答案框架：

1. 先说明方法：规则、小模型分类、大模型判断、混合路由。

- 再说输入特征：用户 query、上下文、历史行为、业务元数据。
- 重点讲评估：准确率、召回率、混淆类目、线上误判样本。
- 强调兜底：低置信度回退到澄清问句或保守策略。
- 如果业务复杂，要补“多意图/歧义 query 如何处理”。

牛客近6个月AI应用开发面经：公司维度对比分析

• 统计公司数：16

公司总览

| 公司 | 面经数 | 高频知识点 | 风格提示 |

|---|---:|---|---|

| 字节 | 7 | Agent架构与范式、RAG与检索优化、记忆与上下文管理 | 项目细节会深挖到记忆、评测、数据集、并发隔离，常伴随模型基础或算法题。 |

| 淘天 | 4 | Agent架构与范式、框架、MCP与工具调用、评测、幻觉与数据集 | 偏 Agent 范式、记忆设计、RAG 优化和业务场景落地，也会问实验与成本。 |

| 快手 | 4 | 算法与编码题、后端基础与工程化、Agent架构与范式 | AI 应用 + 后端八股混合考，Redis/MQ/框架选型比重高。 |

| 腾讯 | 3 | 后端基础与工程化、算法与编码题、记忆与上下文管理 | 偏应用架构与工程实现，常问记忆、Skill/MCP、状态机、LangChain/LangGraph。 |

| 蚂蚁 | 2 | 框架、MCP与工具调用、项目介绍与项目拷打、RAG与检索优化 | 偏 RAG 细节、重排、幻觉、采样参数和项目追问。 |

| 阿里国际 | 2 | 后端基础与工程化、算法与编码题、AI Coding与协作方式 | 除了 AI 应用本身，还会穿插 Python/计网/操作系统等通用基础。 |

| 小红书 | 1 | 项目介绍与项目拷打、Agent架构与范式、框架、MCP与工具调用 | Agent / RAG / 向量索引问得系统，喜欢考概念边界与检索细节。 |

| 阿里云 | 1 | 项目介绍与项目拷打、业务场景设计、RAG与检索优化 | 偏场景系统设计和高风险场景治理，重视 RAG 评测与幻觉控制。 |

| 高德 | 1 | 项目介绍与项目拷打、前端与跨端工程、后端基础与工程化 | 偏跨端工程与 AI Coding 协作方式，也会问 Agent skill 拆分。 |

| 京东 | 1 | 项目介绍与项目拷打、Agent架构与范式、评测、幻觉与数据集 | 重 workflow、RAG、意图识别和电商业务场景，兼顾缓存与压测。 |

| 阿里灵犀互娱 | 1 | AI Coding与协作方式、业务场景设计、评测、幻觉与数据集 | 强业务场景题，尤其游戏客服/游戏助手/多 Agent 设计。 |

| 金山云 | 1 | 项目介绍与项目拷打、后端基础与工程化、RAG与检索优化 | 以具体项目深挖为主，建议按该公司的高频知识点准备。 |

| 阿里淘天 | 1 | 记忆与上下文管理、Agent架构与范式、AI Coding与协作方式 | 以具体项目深挖为主，建议按该公司的高频知识点准备。 |

| 米哈游 | 1 | 项目介绍与项目拷打、模型基础与推理机制、Agent架构与范式 | 偏游戏场景下的 Agent 设计、状态定义与记忆管理。 |

| 百度 | 1 | Agent架构与范式、记忆与上下文管理、评测、幻觉与数据集 | 偏 Agent 项目深挖，关注规划、记忆、工具调用和评测。 |

| 蔚来 | 1 | 后端基础与工程化、RAG与检索优化、算法与编码题 | 更偏服务端基础 + AI 应用结合。 |

逐公司分析

字节

• 面经数：7

- 风格画像：项目细节会深挖到记忆、评测、数据集、并发隔离，常伴随模型基础或算法题。
- 高频知识点：
- Agent架构与范式: 7
- RAG与检索优化: 7
- 记忆与上下文管理: 6
- 框架、MCP与工具调用: 6
- 模型基础与推理机制: 6
- 算法与编码题: 6
- 高频题型：
- RAG/Agent 评测与幻觉治理: 6
- 检索/召回/重排/向量索引优化: 6
- 算法题/手撕题: 6
- 模型基础：Transformer/KV Cache/Temperature: 5
- 记忆系统设计: 4
- MCP / Function Calling / Skills / Tool Calling: 4
- 相关面经：
- 2026-04-10 | [字节 剪映 ai应用开发](<https://www.nowcoder.com/feed/main/detail/2e53f0ec451a4e71b2e3418252071ba4>)
- 2026-04-09 | [字节agent开发实习一面凉经](<https://www.nowcoder.com/feed/main/detail/69b9cd21f1244d3cb19499f249228b50>)
- 2026-04-09 | [字节Agent开发一面90min凉经](<https://www.nowcoder.com/feed/main/detail/91c5394e57c14927841d7a86bfe427c2>)
- 2026-04-08 | [AI应用开发日常实习二面-字节](<https://www.nowcoder.com/feed/main/detail/0607316a2492400896408c95e10951e0>)
- 2026-04-08 | [字节AI应用开发实习面经分享](<https://www.nowcoder.com/feed/main/detail/76448605da6e4aee99a394da83d1718f>)
- 2026-04-02 | [字节 ai应用开发](<https://www.nowcoder.com/feed/main/detail/c5947496fc29440883611173478dc257>)
- 2026-03-21 | [字节 大模型应用开发 二面 日常实习](<https://www.nowcoder.com/feed/main/detail/30b6f37830414df5a32d851f2004df7a>)

淘天

- 面经数：4
- 风格画像：偏 Agent 范式、记忆设计、RAG 优化和业务场景落地，也会问实验与成本。
- 高频知识点：
- Agent架构与范式: 4
- 框架、MCP与工具调用: 4
- 评测、幻觉与数据集: 3

- 记忆与上下文管理: 2
- 项目介绍与项目拷打: 2
- RAG与检索优化: 2
- 高频题型 :
- LangChain / LangGraph / 编排框架选型: 3
- RAG/Agent 评测与幻觉治理: 3
- MCP / Function Calling / Skills / Tool Calling: 3
- Agent 与 workflow/单多 Agent 选型: 2
- 场景设计题 (游戏/电商/安全/客服) : 2
- 项目介绍/项目拷打: 2
- 相关面经 :
- 2026-04-12 | [淘天 AI应用开发 二面](<https://www.nowcoder.com/discuss/872816572199464960>)
- 2026-04-10 | [淘天AI应用开发 agent岗一面好难](<https://www.nowcoder.com/feed/main/detail/a0306a045d594b02b63fb0654d517901>)
- 2026-04-09 | [4.9-淘天agent-二面](<https://www.nowcoder.com/discuss/871774510138023936>)
- 2026-04-03 | [淘天-ai应用开发-一面](<https://www.nowcoder.com/feed/main/detail/076a8d1acaae4795a28d4a5370872bd6>)

快手

- 面经数 : 4
- 风格画像 : AI 应用 + 后端八股混合考 , Redis/MQ/框架选型比重高。
- 高频知识点 :
- 算法与编码题: 4
- 后端基础与工程化: 3
- Agent架构与范式: 3
- 框架、MCP与工具调用: 2
- 记忆与上下文管理: 2
- 业务场景设计: 2
- 高频题型 :
- 算法题/手撕题: 4
- 后端基础 : Redis/缓存/MQ/高并发: 2
- 记忆系统设计: 2
- 场景设计题 (游戏/电商/安全/客服) : 2
- LangChain / LangGraph / 编排框架选型: 1
- 项目介绍/项目拷打: 1
- 相关面经 :

- 2026-04-11 | [快手ai应用服务端开发 二面](<https://www.nowcoder.com/discuss/872512773710696448>)
- 2026-04-10 | [快手 AI agent开发二面分享 1h](<https://www.nowcoder.com/feed/main/detail/afa98a916503444aa708fcfffb263d38>)
- 2026-04-09 | [快手AI Agent开发一面 (已过)](<https://www.nowcoder.com/feed/main/detail/7ce89f19368b46da853c718f2ae2f53c>)
- 2026-04-08 | [实习-快手电商-大模型数据应用开发一面](<https://www.nowcoder.com/discuss/871506086656761856>)

腾讯

- 面经数 : 3
- 风格画像 : 偏应用架构与工程实现 , 常问记忆、Skill/MCP、状态机、LangChain/LangGraph。
- 高频知识点 :
- 后端基础与工程化: 3
- 算法与编码题: 3
- 记忆与上下文管理: 2
- Agent架构与范式: 2
- RAG与检索优化: 2
- 框架、MCP与工具调用: 1
- 高频题型 :
- 算法题/手撕题: 3
- 记忆系统设计: 2
- LangChain / LangGraph / 编排框架选型: 1
- MCP / Function Calling / Skills / Tool Calling: 1
- AI Coding / Claude Code 实践: 1
- 检索/召回/重排/向量索引优化: 1
- 相关面经 :
- 2026-04-10 | [腾讯 AI应用开发面经](<https://www.nowcoder.com/feed/main/detail/acd3e53eecbc405296335edae4d5cf88>)
- 2026-04-09 | [腾讯日常实习一面-AI应用开发 1h](<https://www.nowcoder.com/feed/main/detail/37e847fa267b4c3898b38e2952b5286e>)
- 2026-04-09 | [腾讯暑期实习 ai agent开发一面](<https://www.nowcoder.com/feed/main/detail/497e14617f01453db84dfbc264e5751e>)

蚂蚁

- 面经数 : 2
- 风格画像 : 偏 RAG 细节、重排、幻觉、采样参数和项目追问。
- 高频知识点 :
- 框架、MCP与工具调用: 2

- 项目介绍与项目拷打: 1
- RAG与检索优化: 1
- 记忆与上下文管理: 1
- 评测、幻觉与数据集: 1
- 模型基础与推理机制: 1
- 高频题型 :
- 项目介绍/项目拷打: 1
- 检索/召回/重排/向量索引优化: 1
- RAG/Agent 评测与幻觉治理: 1
- 模型基础 : Transformer/KV Cache/Temperature: 1
- LangChain / LangGraph / 编排框架选型: 1
- 算法题/手撕题: 1
- 相关面经 :
- 2026-04-10 | [蚂蚁金融Agent开发暑期面经分享](<https://www.nowcoder.com/feed/main/detail/1f90384e8ff54172b65b50d853b00da0>)
- 2026-04-05 | [蚂蚁ai应用开发实习二面](<https://www.nowcoder.com/feed/main/detail/718c5a06a65645d7a462a56b457cb35e>)

阿里国际

- 面经数 : 2
- 风格画像 : 除了 AI 应用本身 , 还会穿插 Python/计网/操作系统等通用基础。
- 高频知识点 :
- 后端基础与工程化: 1
- 算法与编码题: 1
- AI Coding与协作方式: 1
- 框架、MCP与工具调用: 1
- 业务场景设计: 1
- 高频题型 :
- 算法题/手撕题: 1
- 后端基础 : Redis/缓存/MQ/高并发: 1
- AI Coding / Claude Code 实践: 1
- MCP / Function Calling / Skills / Tool Calling: 1
- 相关面经 :
- 2026-04-10 | [阿里国际AI应用开发暑期一面1h](<https://www.nowcoder.com/feed/main/detail/580e5a35eddb47e0851952165004f233>)
- 2026-04-09 | [阿里国际AI应用开发二面](<https://www.nowcoder.com/feed/main/detail/c2772ead209549a0bb6e54f2df205a31>)

小红书

- 面经数：1
- 风格画像：Agent / RAG / 向量索引问答系统，喜欢考概念边界与检索细节。
- 高频知识点：
- 项目介绍与项目拷打: 1
- Agent架构与范式: 1
- 框架、MCP与工具调用: 1
- 记忆与上下文管理: 1
- RAG与检索优化: 1
- 高频题型：
- 项目介绍/项目拷打: 1
- Agent 与 workflow/单多 Agent 选型: 1
- MCP / Function Calling / Skills / Tool Calling: 1
- 记忆系统设计: 1
- RAG 系统设计: 1
- 检索/召回/重排/向量索引优化: 1
- 相关面经：
- 2026-04-12 | [小红书 AI Agent开发 一面](<https://www.nowcoder.com/discuss/872820735335485440>)

阿里云

- 面经数：1
- 风格画像：偏场景系统设计和高风险场景治理，重视 RAG 评测与幻觉控制。
- 高频知识点：
- 项目介绍与项目拷打: 1
- 业务场景设计: 1
- RAG与检索优化: 1
- 评测、幻觉与数据集: 1
- Agent架构与范式: 1
- 前端与跨端工程: 1
- 高频题型：
- 项目介绍/项目拷打: 1
- 场景设计题（游戏/电商/安全/客服）: 1
- RAG 系统设计: 1
- RAG/Agent 评测与幻觉治理: 1
- 相关面经：

- 2026-04-12 | [阿里云 AI应用开发 一面](<https://www.nowcoder.com/discuss/872810222128021504>)

高德

- 面经数：1
- 风格画像：偏跨端工程与 AI Coding 协作方式，也会问 Agent skill 拆分。
- 高频知识点：
- 项目介绍与项目拷打: 1
- 前端与跨端工程: 1
- 后端基础与工程化: 1
- Agent架构与范式: 1
- AI Coding与协作方式: 1
- 模型基础与推理机制: 1
- 高频题型：
- 项目介绍/项目拷打: 1
- 检索/召回/重排/向量索引优化: 1
- MCP / Function Calling / Skills / Tool Calling: 1
- AI Coding / Claude Code 实践: 1
- 相关面经：
- 2026-04-10 | [高德 AI应用开发 一面](<https://www.nowcoder.com/discuss/872229606252675072>)

京东

- 面经数：1
- 风格画像：重 workflow、RAG、意图识别和电商业务场景，兼顾缓存与压测。
- 高频知识点：
- 项目介绍与项目拷打: 1
- Agent架构与范式: 1
- 评测、幻觉与数据集: 1
- RAG与检索优化: 1
- 后端基础与工程化: 1
- 业务场景设计: 1
- 高频题型：
- 项目介绍/项目拷打: 1
- Agent 与 workflow/单多 Agent 选型: 1
- RAG 系统设计: 1
- RAG/Agent 评测与幻觉治理: 1
- 意图识别与 Query 理解: 1

- 检索/召回/重排/向量索引优化: 1

- 相关面经 :

- 2026-04-10 | [27暑期实习-京东Agent开发二面分享](<https://www.nowcoder.com/feed/main/detail/6af1c8611e434bbe8ba9cfa2bd0eaf18>)

阿里灵犀互娱

- 面经数 : 1

- 风格画像 : 强业务场景题 , 尤其游戏客服/游戏助手/多 Agent 设计。

- 高频知识点 :

- AI Coding与协作方式: 1

- 业务场景设计: 1

- 评测、幻觉与数据集: 1

- Agent架构与范式: 1

- 记忆与上下文管理: 1

- 高频题型 :

- AI Coding / Claude Code 实践: 1

- 场景设计题 (游戏/电商/安全/客服) : 1

- RAG/Agent 评测与幻觉治理: 1

- 意图识别与 Query 理解: 1

- Agent 与 workflow/单多 Agent 选型: 1

- 记忆系统设计: 1

- 相关面经 :

- 2026-04-10 | [阿里灵犀互娱AI应用开发暑期二面](<https://www.nowcoder.com/feed/main/detail/3dbec664147348d685f3e2b8b50288f8>)

金山云

- 面经数 : 1

- 风格画像 : 以具体项目深挖为主 , 建议按该公司的高频知识点准备。

- 高频知识点 :

- 项目介绍与项目拷打: 1

- 后端基础与工程化: 1

- RAG与检索优化: 1

- AI Coding与协作方式: 1

- 框架、MCP与工具调用: 1

- 高频题型 :

- 项目介绍/项目拷打: 1

- 后端基础：Redis/缓存/MQ/高并发: 1
- 检索/召回/重排/向量索引优化: 1
- RAG/Agent 评测与幻觉治理: 1
- AI Coding / Claude Code 实践: 1
- MCP / Function Calling / Skills / Tool Calling: 1
- 相关面经：
- 2026-04-10 | [27届-日常实习-金山云-AI应用平台开发-一面](<https://www.nowcoder.com/discuss/872055401926164480>)

阿里淘天

- 面经数：1
- 风格画像：以具体项目深挖为主，建议按该公司的高频知识点准备。
- 高频知识点：
- 记忆与上下文管理: 1
- Agent架构与范式: 1
- AI Coding与协作方式: 1
- 高频题型：
- Agent 与 workflow/单多 Agent 选型: 1
- AI Coding / Claude Code 实践: 1
- 算法题/手撕题: 1
- 相关面经：
- 2026-04-09 | [阿里淘天 AI应用开发暑期实习一面](<https://www.nowcoder.com/feed/main/detail/39627b379e8e46ce9cb3a3b8eae46959>)

米哈游

- 面经数：1
- 风格画像：偏游戏场景下的 Agent 设计、状态定义与记忆管理。
- 高频知识点：
- 项目介绍与项目拷打: 1
- 模型基础与推理机制: 1
- Agent架构与范式: 1
- 记忆与上下文管理: 1
- 框架、MCP与工具调用: 1
- 算法与编码题: 1
- 高频题型：
- 项目介绍/项目拷打: 1

- 模型基础：Transformer/KV Cache/Temperature: 1
- Agent 与 workflow/单多 Agent 选型: 1
- 记忆系统设计: 1
- MCP / Function Calling / Skills / Tool Calling: 1
- 算法题/手撕题: 1
- 相关面经：
- 2026-04-09 | [米哈游-AI Agent研发-暑期一面](<https://www.nowcoder.com/feed/main/detail/a1566acc2660477dbf54b58885513d6c>)

百度

- 面经数：1
- 风格画像：偏 Agent 项目深挖，关注规划、记忆、工具调用和评测。
- 高频知识点：
- Agent架构与范式: 1
- 记忆与上下文管理: 1
- 评测、幻觉与数据集: 1
- 模型基础与推理机制: 1
- 后端基础与工程化: 1
- 高频题型：
- RAG/Agent 评测与幻觉治理: 1
- 模型基础：Transformer/KV Cache/Temperature: 1
- 相关面经：
- 2026-04-09 | [百度ai agent开发春招一面](<https://www.nowcoder.com/feed/main/detail/c807140b75bd4cf3bf4166660676db5d>)

蔚来

- 面经数：1
- 风格画像：更偏服务端基础 + AI 应用结合。
- 高频知识点：
- 后端基础与工程化: 1
- RAG与检索优化: 1
- 算法与编码题: 1
- 高频题型：
- 后端基础：Redis/缓存/MQ/高并发: 1
- 检索/召回/重排/向量索引优化: 1
- 算法题/手撕题: 1

- 相关面经：

- 2026-04-09 | [蔚来 AI应用开发

暑期一面分享](<https://www.nowcoder.com/feed/main/detail/5af31c8afe254a00a69a6c0141e6908a>)