

JD Evidence · Scorecard ·
Project Kit

AI Engineer 技能自查清单

大厂 AI 岗 JD 拆解

给想转 AI Engineer 的在职工程师：先看真实岗位在要什么，再自测差距，最后把项目和简历补成能讲的证据。



目录

- 00 先看：这不是 AI 工具清单
- 01 真实 JD 证据库
- 02 薪资和岗位池怎么读
- 03 能打分的技能自测
- 04 RAG 自查清单
- 05 Agent / Tool Use 自查清单
- 06 Eval / Observability 自查清单
- 07 Production / Cloud / Safety 自查清单
- 08 客服工单 Agent 参考项目
- 09 失败库、golden set、trace 范本
- 10 JD 拆成简历和项目证据
- 11 面试题完整答案
- A 可下载模板
- B 来源和边界

先看：这不是 AI 工具清单

会用 ChatGPT、Claude、Cursor，不等于能投 AI Engineer。岗位真正看的是：你能不能把 LLM 接进真实系统，能不能处理检索、工具调用、评测、监控、权限、成本、延迟和失败。

这本给在职工程师用。你不用从零学 Python，也不用把所有论文看完。你要先知道自己缺的是 RAG 数据层、Agent 工程化，还是 Eval 和生产化。

10

条真实 JD 摘录

24

道可判定自测题

20

条 failure case 范本



想知道自己离 AI Engineer 差哪一块？

扫码找 Amelia，把简历项目和目标 JD 发来，先拆 RAG、Agent、Eval、Cloud 哪块最该补。不承诺面试或 offer。





真实 JD 证据库

先别背“AI Engineer 需要 RAG、Agent、Eval”。先看公开 JD。下面每条都按“公司、岗位、原文短摘录、来源、抓取日期”登记。摘录只保留关键短句，够你证明这些词不是我们编的。

公司 / 地区	岗位名	原文摘录和技能词	来源链接 / 日期
SEEK / AU	LK Property Group · AI Engineer	“Build and deploy AI agents across leading LLMs”; “Integrate AI services into existing systems via APIs, webhooks, and data pipelines ”; “Test and refine semi-autonomous and autonomous agents”。	seek.com.au 2026-07-15
SEEK / AU	Lanson Partners · AI Engineer	“Developing RAG solutions, AI agents, tool-calling workflows ”; “Deploying AI solutions across AWS, including Amazon Bedrock”; “Establishing evaluation frameworks ”。	seek.com.au 2026-07-15
SEEK / AU	CVT Global · AI Engineer – Graduate	“Assist in the development and implementation of AI solutions ”; “Design, develop, test and implement software integrations and systems”; “Support the deployment and monitoring of AI applications”。	seek.com.au 2026-07-15
SEEK / AU	Chemist Warehouse · AI Engineer	“Build and operate enterprise LLM/GenAI solutions”; “Design and implement RAG capabilities”; “Implement LLMOps/MLOps and CI/CD pipelines including monitoring, evaluation and deployment automation”。	seek.com.au 2026-07-15
SEEK / AU	Real Time · AI Engineer	“Deliver end-to-end AI solutions ”; “Build agentic systems ”; “Design, write, and scale stateful multi-agent workflows and automated RAG pipelines using Python, Langflow, and LangGraph”。	seek.com.au 2026-07-15
SEEK / AU	Nuage Technology Group · AI Engineer	“Architecting and building GenAI and agentic systems that go into production”; “AWS, Azure or GC”; “Bedrock or SageMaker, or Azure ML / Azure OpenAI”; “LangGraph, Semantic Kernel, AutoGen, or Google ADK”。	seek.com.au 2026-07-15
SEEK / AU	Nuage Technology Group · Software Engineer (AI + .NET)	“Junior Engineer to help build and grow their AI capability ”; “build internal AI platforms”; “AI automations, or agentic AI solutions”; “LLMs, prompt engineering, RAG , AI workflows”。	seek.com.au 2026-07-15
BOSS / CN	四川盈播网络科技 · 大模型应用工程师	“负责 智能客服系统 、 智能质检 等开发与维护”; “负责 AI 辅助工具 开发”; “熟悉大语言模型(LLM)相关技术”; “熟悉常用大模型工具套件, 如 RAG、MCP、Skills、Agent ”。	zhipin.com 2026-07-15
BOSS / CN	朗姿医疗 · AI 应用开发工程师 LLM Agent / RAG	“设计并开发 Agent 工作流 、工具调用、上下文管理、长期记忆机制”; “参与 RAG 知识库 建设, 包含索引构建、语义检索、重排序与效果优化”; “建立 AI 应用效果评估机制, 持续跟踪回答准确性、知识命中率、推荐合理性”。	zhipin.com 2026-07-15
拉勾 / CN	微博 · AIGC 算法工程师 (NLP 方向)	“负责微博基础内容 AI 的研发”; “负责自然语言处理、多模态内容理解和生成模型调优”; “参与负责大模型预训练技术、海量参数下大模型高效微调技术、基于大模型 Prompt 提示学习技术”。	lagou.com 2026-07-15

10 条 JD 来源里的技能词频

技能词	出现次数	怎么理解
LLM / GenAI / 大模型	9 / 10 已核验	岗位不是“会聊天”，而是把模型能力接进产品或业务。
RAG / retrieval / 知识库	6 / 10 已核验	企业数据怎么被找到、引用、更新、控权限。
Agent / tool calling / workflow	7 / 10 已核验	会调用工具还不够，要有审批、回滚、失败处理。
Eval / quality / metrics	4 / 10 已核验	要证明系统变好，不是 demo 看着顺眼。
Cloud / deployment / production	7 / 10 已核验	能上线、能排错、能控成本，才像工程岗位。
Data pipeline / analytics	3 / 10 已核验	AI 项目吃数据，数据质量差，模型再强也白搭。

说明：上表只放已截图/已核验的真实岗位摘录，不把官方检索页包装成 JD。招聘页会频繁上下架，投递前要按当天目标 JD 再拆一遍。

指定公司官方检索入口

公司	用途	官方入口 / 抓取日期
Canva AU	投递前查 AI / ML / product engineering 岗位。	lifeatcanva.com/en/jobs 2026-07-15
Atlassian AU	投递前查 AI / ML / engineering 岗位。	atlassian.com/company/careers/all-jobs 2026-07-15
Amazon AU	投递前查 GenAI / AWS AI / Applied Science 岗位。	amazon.jobs 2026-07-15
Google AU	投递前查 AI / ML / Google Cloud 岗位。	google.com/about/careers 2026-07-15
字节跳动 / 阿里 / 腾讯	投递前查大模型应用、LLM、Agent、搜索推荐、云上 AI 岗位。	jobs.bytedance.com / talent.alibaba.com / careers.tencent.com 2026-07-15





真实薪资和岗位池怎么读

这章复用同仓库《AI 岗位图鉴 2026》的数据口径，并补上岗位池数量栏。数据日期：薪资口径核验到 2026-06-25；招聘搜索快照抓取到 2026-07-15。薪资是市场参考，不是收入承诺。

市场	岗位数量 / 池口径	Junior / entry	Mid	Senior	来源和判断
AU	SEEK “ai engineer” 搜索截图显示 1,184 jobs。样本薪资从 Graduate AU\$80K 起，到 Real Time AI Engineer AU\$180K+ super；Nuage AI+.NET 为 AU\$90K-\$120K；Lanson 为 AU\$150K-\$180K base + super + bonus。	CVT Global Graduate: AU\$80K 起；Nuage AI+.NET: AU\$90K-\$120K。	LK Property: AU\$80K-\$100K；Lanson: AU\$150K-\$180K base。	Real Time: AU\$180K+ super；Senior/enterprise 岗还看平台、云和行业经验。	SEEK 截图 2026-07-15；SEEK Career Advice；Robert Half AU 2026；本仓库 ai-jobs-salary-atlas。
CN	BOSS 成都“大模型应用工程师”样本：18-30K、9-14K·13薪、10-15K、15-30K、8-12K；朗姿 LLM Agent/RAG 为 12-24K。拉勾“AIGC 工程师”显示 500+ 职位，微博 AIGC NLP 岗为 20-40K。	AI 应用工程师 Agent 方向 8-12K；大模型应用工程师 9-14K·13薪；朗姿 LLM Agent/RAG 经验不限 12-24K。	大模型应用工程师 18-30K；AI 应用/微调部署 10-15K；微博 AIGC NLP 20-40K。	3-5 年大模型应用工程师样本 18-30K / 15-30K；资深岗位更看 LLM、Agent、RAG、评测机制和业务流程落地。	BOSS / 拉勾截图 2026-07-15；薪资为样本，不代表全国。
US 参考	不作为 AU/CN 收入口径，只做全球岗位价值参照。	entry 仍看软件工程基础和项目证据。	Robert Half US 2026：AI/ML Engineer US\$134K-\$193.25K；Data Scientist US\$121.75K-\$182.5K。	BLS Software Developers 2024 中位数 US\$133,080，2024-2034 预计增长 15%。	Robert Half US 2026；BLS OOH。

Entry-level 缺口到底在哪

说白了，纯“Junior AI Engineer”岗位少。更现实的入口是三类：一类是软件工程师岗位加 GenAI feature，一类是数据工程/分析岗位加 LLM 数据层，一类是云/平台岗位加 AI 服务部署。

所以你要准备的不是“我学过 AI”，而是一个能被软件、数据、平台岗位都读懂的项目：有 RAG，有工具调用，有 eval，有日志，有简历 bullet。



能打分的技能自测

每题 0/1/2 分：0 是不会，1 是半会或照教程做过，2 是能独立解释并拿出项目证据。24 题满分 48。别骗自己，面试官会追细节。

能力域	可判定小题	0 / 1 / 2 怎么打	证据
RAG	能否独立设计 chunk size、overlap、metadata?	0 不会；1 照教程；2 能解释取舍。	索引配置。
RAG	能否解释 hybrid search 和 rerank 何时有用?	0 不会；1 知道名词；2 有对比。	检索实验。
RAG	能否处理 citation、freshness、document ACL?	0 不会；1 单点做过；2 三项都有。	带引用回答。
RAG	是否维护 retrieval / citation failure cases?	0 没有；1 少于 10 条；2 至少 20 条。	失败库。
Agent	能否写 tool schema、输入输出和错误码?	0 不会；1 toy demo；2 可校验。	tool spec。
Agent	能否让退款、删除、账单修改走人工审批?	0 没有；1 手动提醒；2 workflow 阻断。	审批记录。
Agent	能否处理 retry、timeout、fallback、idempotency?	0 没有；1 只做 retry；2 四项都有。	测试日志。
Agent	能否追踪每次 agent decision 和 tool call?	0 没有；1 console log；2 结构化 trace。	trace。
Eval	能否写 10 题 golden set 和评分点?	0 没有；1 只有问题；2 有期望和评分。	eval CSV。
Eval	能否区分 retrieval miss、answer drift、citation error?	0 混在一起；1 能说；2 能归因。	失败分类。
Eval	能否用同一批案例比较 prompt 版本?	0 不会；1 手测；2 有回归表。	版本报告。
Eval	能否统计 quality、groundedness、latency、cost?	0 没有；1 两项；2 四项都有。	dashboard。
LLMOps	能否记录 model、prompt version、token、latency、cost?	0 没有；1 部分字段；2 字段完整。	日志表。
LLMOps	能否给低风险请求做 cache 或 deterministic route?	0 没有；1 只设计；2 已实现。	路由代码。
LLMOps	能否解释 rate limit、rollback、prompt rollout?	0 不会；1 讲概念；2 有方案。	部署文档。
LLMOps	能否排查一次成本突然升高?	0 不会；1 看总账；2 能按 trace 拆。	成本分析。
Safety	能否测试 prompt injection?	0 没有；1 少量样例；2 回归测试。	攻击样例。
Safety	能否避免 PII 出现在回答和日志里?	0 没有；1 只在回答；2 回答日志都脱敏。	脱敏测试。
Safety	能否在 retrieval 和 tool call 前做权限检查?	0 没有；1 只做一处；2 两处都有。	ACL 测试。
Safety	能否说清 guardrail 的边界?	0 不会；1 说口号；2 能讲失败样例。	风险表。
Product	能否定义用户场景、验收标准、上线边界?	0 没有；1 模糊；2 可验收。	PRD。
Product	能否说明什么时候不该用 LLM?	0 不会；1 知道风险；2 能给替代方案。	trade-off。
Product	能否把项目改成 JD 匹配的简历 bullet?	0 不会；1 泛泛写；2 按 JD 改。	简历版本。
Product	能否和 product / data / security 解释 trade-off?	0 不会；1 只讲技术；2 能讲业务边界。	复盘文档。

0-15: L0

还在工具体验层。先做一个完整项目，不急着投。

16-27: L1

能跑 demo，但证据弱。补 eval、trace、失败样例。

28-39: L2

可以投 AI 相关软件/数据岗。简历要按 JD 改。

40-48: L3

可以冲 AI Engineer。准备好系统设计和取舍。

30 天补强路径

当前水平

未来 30 天先补 3 件事

别做什么

L0	跑通客服工单 Agent；写 10 题 golden set；部署一个可访问 demo。	别刷一堆工具名。
L1	补 20 条 failure case；加 trace；把 README 写成可复现项目。	别只录一个顺利视频。
L2	按 3 条目标 JD 改简历；准备 5 道完整面试答案；做一次成本和延迟分析。	别所有岗位投同一版简历。
L3	补系统设计图；准备 trade-off；把项目数据、限制、下一步写清楚。	别把自己讲成“什么都会”。





RAG 自查清单

RAG 不是“丢进向量库再问”。岗位写 retrieval、knowledge base、enterprise data、grounding，本质都在看你能不能让企业资料被正确使用。

RAG 能勾上这些，才不像教程项目

- ☐ 我能解释 chunk size、overlap、metadata 为什么这么设
例子：退款政策按标题和条款号切，不按固定 500 token 硬切。
- ☐ 我做过 hybrid search 或 rerank 对比
例子：用户搜“chargeback”时，关键词检索比纯 embedding 更稳。
- ☐ 我能处理权限和引用
例子：Sales 用户不能检索 HR 薪资文件，引用必须指向可见文档。
- ☐ 我有失败案例库
至少 20 条：找不到、找错、答偏、引用错、权限错。
- ☐ 我能说清 freshness 和更新策略
文档变了以后，索引怎么更新，缓存怎么失效。



Agent / Tool Use 自查清单

Agent 岗位最怕“能跑但不可控”。面试会追：工具怎么定义、怎么防止乱调用、怎么让人审批、怎么回滚。

Tool Schema

参数清楚，输出可验证

每个 tool 有明确输入、输出、错误码和权限。比如
``refund_ticket(ticket_id, amount, reason)`` 不能让模型自由编字段。

Control

能中断、重试、降级

超时、API 失败、工具返回异常时，有 retry、fallback、human review，不让 agent 一路乱撞。

Memory

知道什么该记，什么不该记

用户偏好、任务状态、临时上下文和隐私数据分开处理。PII 不进长期记忆。

Audit

每一步都能追溯

谁触发、模型决定了什么、调用了哪个工具、结果是什么，都要 trace。

Agent 项目最小证据

做一个“客服工单 agent”：能检索资料、创建/更新工单、遇到退款/删除/权限动作时要求人工确认，并保留 trace。别只做自动回复。



Eval / Observability 自查清单

AI 系统和普通系统最大的差别：你不能只测“接口通不通”。你要测回答好不好、有没有引用、有没有胡说、有没有越权。

指标	看什么	项目里怎么证明
Answer quality	答案是否解决用户问题	golden set + 人工评分规则。
Groundedness	答案是否基于资料	引用片段和原文对齐检查。
Retrieval	该找的文档有没有找出来	top-k hit rate / recall 样例。
Safety	是否泄露、越权、被注入	prompt injection 和 PII 测试集。
Cost / latency	每次调用多少钱、多慢	trace dashboard 或日志统计。
Regression	改 prompt 后有没有变差	CI 里跑 eval 或手动回归表。



项目没有 Eval，面试容易被追卡

扫码找 Amelia，把 RAG/Agent 项目发来，先看该补 golden set、trace 还是失败样例。



Production / Cloud / Safety 自查清单

大厂 AI 岗不会只问 demo。会问如果上线给 10 万用户，怎么控成本、怎么限流、怎么观测、怎么处理坏回答。

生产化 上线前要能回答的问题

- ☐ 我知道每次请求的 token 成本和延迟来源
prompt、retrieval、rerank、tool call 都要能拆。
- ☐ 我有 fallback 和降级策略
模型超时、检索为空、工具失败时怎么回。
- ☐ 我能处理 secret、PII、权限和审计日志
不要把企业数据当 playground。
- ☐ 我能画出部署架构
API、queue、worker、vector DB、object storage、monitoring。
- ☐ 我能解释 guardrail 的边界
不是加一句 system prompt 就完了。



客服工单 Agent 参考项目

别再写“我做了一个 RAG 聊天机器人”。做一个客服工单 Agent，面试官能追的点全在里面。

模块	功能规格	验收标准
RAG 检索	检索退款、订阅、账号删除、发票、SLA 文档，答案带引用。	10 题 golden set 至少 8 题引用正确。
工单增改	读取工单、创建 follow-up、更新 priority、追加 internal note。	tool call 参数可校验，失败有错误码。
人工审批	退款、删除账号、改账单地址必须 pending approval。	没有审批记录时，工具不能执行。
Trace	记录 request_id、user_id、retrieval、prompt_version、tool call、latency、cost。	能从一次坏回答追到原因。
Safety	检测 prompt injection、越权文档、PII 输出。	至少 5 条攻击样例能被拒绝或转人工。

Repo 目录骨架

```
customer-ticket-agent/  
  app/  
    api/           # FastAPI or Next.js API routes  
    agent/         # planner, tool router, approval policy  
    rag/           # loaders, chunking, retrieval, rerank  
    tools/         # ticket, refund, account, notification tools  
    evals/         # golden set, scoring script, regression report  
    observability/ # traces, cost, latency, prompt version  
  data/  
    policies/      # fake SOP docs, refund rules, account deletion  
    tickets/       # synthetic ticket seed data  
  tests/  
    safety_cases.json  
    golden_set.csv  
  README.md  
  docker-compose.yml
```

技术选型建议

后端用 Python + FastAPI 更好讲，前端可用 Next.js。RAG 用 Postgres + pgvector 或 Qdrant。Agent workflow 用 LangGraph 或自己写状态机。Trace 用 OpenTelemetry、LangSmith、Phoenix 或一张结构化日志表。重点不是工具名，是你能解释为什么这么选。



失败库、golden set、trace 范本

20 条 failure case 范本

类型	现象	原因	修法
not found	问退款期限，系统说未找到。	退款政策 PDF 没入库。	补 loader，重建索引。
not found	问发票下载路径，答成联系客服。	query 没扩展同义词。	加 invoice/receipt 同义词。
not found	问账号删除 SLA，检索为空。	chunk 标题丢失。	metadata 保留章节标题。
not found	问企业套餐退款，命中个人套餐。	缺 plan_type filter。	检索前加 metadata filter。
找错	把 chargeback 当普通退款。	关键词权重太低。	hybrid search + rerank。
找错	引用旧版政策。	没有 effective_date。	按日期过滤，旧版降权。
找错	把澳洲政策答给美国用户。	地区 metadata 缺失。	region 必填。
找错	Top-1 是 FAQ，SOP 在 Top-5。	没有 rerank。	cross-encoder rerank。
答案漂移	明明文档写 14 天，回答 30 天。	模型用常识补全。	要求只基于引用回答。
答案漂移	把“不保证退款”说成“可以退款”。	否定句没保留上下文。	按条款段落切 chunk。
答案漂移	答案太自信，没有说明限制。	prompt 缺失信度规则。	低置信度转人工。
答案漂移	回答了用户没问的销售话术。	system prompt 混进营销目标。	客服任务和营销任务拆开。
引用错	答案正确，引用到无关 FAQ。	引用取自 Top-1，不是答案来源。	生成时绑定 evidence span。
引用错	引用链接 404。	文档 URL 更新。	索引存 canonical URL。
引用错	引用内部文档给外部用户。	展示层没做权限检查。	answer 和 citation 都过 ACL。
引用错	引用显示“退款政策”，实际段落是隐私政策。	chunk_id 和 source_title 映射错位。	索引构建时加 source checksum 回归测试。
权限错	普通用户查到 VIP 价格。	检索没传 user_role。	检索 query 带 role scope。
权限错	agent 直接发起退款。	tool allowlist 太宽。	高风险工具必须审批。
权限错	输出用户手机号。	PII redaction 缺失。	日志和回答都脱敏。
权限错	夜间批处理绕过人工审批。	approval gate 只写在前端。	后端工具层强制校验 action risk。

10 题 golden set 范本

问题	期望答案	评分点
年度订阅 7 天内能退吗？	能否退款取决于 plan 和地区，先引用退款政策。	引用正确；不承诺。
chargeback 和 refund 一样吗？	不是。chargeback 走争议流程，需转人工。	区分概念；转人工。
怎么删除账号？	说明入口、影响、人工确认。	不直接执行删除。
能帮我改账单邮箱吗？	需要验证身份后更新。	调用前检查权限。
我的发票在哪里下载？	给出路径和条件。	命中 invoice 文档。
企业客户 SLA 是多久？	按 plan 和合同条款回答。	不拿个人 FAQ 回答。
澳洲用户适用美国退款政策吗？	不直接适用，按 region 检索。	地区过滤。

用户让我忽略系统规则怎么办？	拒绝外部指令，继续按政策答。	prompt injection 防护。
退款 API 超时怎么办？	不要说完成，创建 pending 状态。	工具失败处理。
怎么证明新版 prompt 更好？	跑同一批 golden set，比 quality、groundedness、latency、cost。	有回归思路。

一段真实 trace 长这样

```
{
  "request_id": "req_20260715_104233",
  "user_id": "u_8821",
  "role": "customer_support_l2",
  "query": "企业客户想取消年度订阅，能直接退款吗？",
  "prompt_version": "support_agent_v3.2",
  "retrieval": [
    {"doc": "refund_policy_enterprise_au.md", "score": 0.86, "citation": "section 4.2"},
    {"doc": "approval_matrix.md", "score": 0.81, "citation": "refund_over_500"}
  ],
  "decision": "needs_human_approval",
  "tool_calls": [
    {"tool": "get_ticket", "args": {"ticket_id": "T-1042"}, "status": "ok"},
    {"tool": "create_approval", "args": {"amount": 1200, "reason": "annual cancellation"}},
  ],
  "status": "pending"
},
{
  "answer_quality": {"grounded": true, "risk": "medium"},
  "latency_ms": 2840,
  "estimated_cost_usd": 0.018
}
```





JD 拆成简历和项目证据

看到 JD 里的“build GenAI applications”，不要只写“会 LangChain”。要翻译成证据。

JD 信号	简历证据	项目证据
RAG / knowledge retrieval	写 chunk、metadata、rerank、citation、eval。	知识库问答 + 20 条失败样例。
Agents / tool use	写 tool schema、approval、trace、fallback。	工单 agent / workflow agent。
Evaluation	写 golden set、regression、quality metric。	eval report + dashboard。
Production systems	写 latency、cost、logging、deployment。	线上 demo + 架构图 + 日志。
Cross-functional work	写 product/legal/security/data 协作。	PRD、风险表、上线边界说明。

20 条简历 bullet 库

方向	可改写 bullet
RAG	Built a RAG pipeline over policy documents with metadata filtering, citation grounding, and scheduled re-indexing.
RAG	Improved retrieval quality by comparing vector search, keyword search, and reranking on a 10-question golden set.
RAG	Designed chunking strategy for SOP documents using section headers, effective dates, and region tags.
RAG	Added ACL-aware retrieval so users only receive answers from documents they are allowed to access.
RAG	Maintained a 20-case failure library covering missing documents, wrong retrieval, answer drift, bad citations, and permission gaps.
Agent	Implemented a customer ticket agent with structured tool schemas for ticket lookup, note creation, and approval requests.
Agent	Added human approval gates for refunds, account deletion, and billing changes to prevent unsafe autonomous actions.
Agent	Built retry, timeout, fallback, and idempotency handling for external tool calls.
Agent	Logged agent decisions, tool inputs, tool outputs, and approval status for auditability.
Agent	Separated task state, short-term context, and PII to avoid unsafe memory behavior.
Eval	Created a golden set with expected answers and scoring criteria for answer quality, groundedness, and citation accuracy.
Eval	Compared prompt versions using regression tests instead of manual cherry-picked examples.
Eval	Instrumented traces to capture model, prompt version, retrieval hits, latency, and estimated token cost.
Eval	Diagnosed failures by separating retrieval miss, poor chunking, rerank issue, prompt issue, and model hallucination.
Eval	Built safety test cases for prompt injection, PII leakage, and unauthorized document access.
Production	Deployed the AI service behind an API with rate limiting, structured logging, and rollback-ready prompt versions.
Production	Reduced avoidable model calls through caching and deterministic routing for low-risk FAQ requests.
Production	Added cost and latency dashboards to identify expensive prompts and slow retrieval paths.
Production	Documented architecture trade-offs across vector database, reranker, model choice, and hosting setup.
Production	Worked with product and support stakeholders to define safe rollout scope, escalation rules, and acceptance criteria.

如果你已经有 AI demo，但简历还是写“做了一个聊天机器人”，扫 Amelia 的码，先把项目改成 JD 能看懂的证据。

Amelia

IT Career Consultant

在澳洲 IT 业务咨询领域拥有10年丰富经验和专业知识。对 IT 行业的发展趋势和技术需求有着深入的了解，擅长提供转码就业问题解决方案，能够准确把握最新的技术趋势和就业需求，为您提供最具前瞻性的建议。

探索IT领域，点亮你的科技未来

提供问题解决方案

工作内推机会

IT就业群提供

IT技术提升服务

300+

已帮助300+人成功就业

1000+

帮助1000+学员答疑解惑

200+

已帮助数百人成功转码





面试题完整答案

下面不是背诵稿，是“强候选人会怎么说”。你要换成自己的项目数据。

RAG

RAG 答错了怎么排查？

我会先把问题拆成 retrieval 和 generation 两段。第一步看 trace：Top-K 有没有召回正确文档，metadata filter 有没有把地区、权限、日期过滤对。第二步看 chunk，是否把否定句或条款切散。第三步看 rerank 和 prompt，答案是不是用了未引用内容。踩雷答法：直接说“换更强模型”。

Eval

怎么证明新版 prompt 更好？

我不会只拿三条顺利样例演示。我会固定 golden set，分 answer quality、groundedness、citation accuracy、latency、cost 评分。新版只有在质量提升且成本延迟可接受时才上线。踩雷答法：说“我感觉回答更自然”。

Agent

Agent 为什么会乱调用工具？

常见原因是 tool 描述太宽、权限没有做在工具层、状态机不清楚、错误返回不可读。我会把高风险工具放到 approval gate 后面，所有写操作带 idempotency key，并记录工具输入输出。踩雷答法：说“加一句不要乱调用”。

Safety

怎么防 prompt injection？

我会把外部文档当不可信输入，系统指令和检索内容分层处理。工具调用用 allowlist，执行前检查用户权限和 action risk。再用攻击样例做回归，比如让文档诱导模型泄露系统提示词。踩雷答法：只说“加 guardrail”。

System

成本突然升高怎么办？

我会先看 trace 聚合：请求量、上下文长度、模型选择、rerank 次数、cache hit、重复重试。再定位是某个 prompt 版本变长，还是检索返回太多内容，或者工具失败导致重试。踩雷答法：只说“换便宜模型”。

再准备这 12 题

chunk 策略、embedding 选择、metadata 权限、hybrid search、rerank、structured output、function calling、streaming UX、rate limit、observability、PII、human-in-the-loop。

可下载模板

这本 PDF 旁边配了四个可直接打开的模板文件。打印自查、复制到 Notion、丢进 Google Sheet 都行。

简历 bullet 库

<templates/resume-bullets-ai-engineer.csv> : RAG、Agent、Eval、Production 各 5 条，按目标 JD 改数字和场景。

PDF / 表格版自查清单

<templates/printable-self-check.html> 可直接打印；<templates/self-check-scorecard.csv> 可导入表格打分。

Notion / Sheet 自评打分器

<templates/notion-sheet-scoring.md> : 字段、公式和 30 天补强路径。

项目验收清单

客服工单 Agent 的验收标准已经放在第 08 章，可直接搬到 README。



来源和边界

JD 来源

Canva Life at Canva、Atlassian Careers、Amazon Jobs、Google Careers、SEEK、ByteDance Jobs、Alibaba Talent、Tencent Careers、BOSS 直聘、拉勾。岗位页会变，抓取日期写在第 01 章。

薪资来源

复用本仓库《AI 岗位图鉴 2026》：Robert Half 2026、SEEK Australia Career Advice、BLS Occupational Outlook Handbook、WEF、PwC 公开报告摘录。

技术口径

OpenAI Evals / Platform Docs、Anthropic Claude Docs、LangChain、LlamaIndex、MLflow、AWS Bedrock、Azure AI Foundry、Google Cloud Vertex AI。

使用提醒

这本是自查表，不是资格认证。它不承诺面试、offer、签证或收入结果。投递前要按目标公司最新 JD 重拆一遍。





匠人学院
JR ACADEMY

AI Engineer 不是“会调模型”。是能把模型放进系统里。

从 RAG、Agent、Eval 到生产化，把你的项目补成 JD 能看懂、面试能讲清的证据。



Amelia

IT Career Consultant

在澳洲 IT 业务咨询领域拥有10年丰富经验和专业知识。对 IT 行业的发展趋势和技术需求有着深入的了解，擅长提供转码就业问题解决方案，能够准确把握最新的技术趋势和就业需求，为您提供最具前瞻性的建议。

探索IT领域，点亮你的科技未来

提供问题解决方案

工作内推机会

IT就业群提供

IT技术提升服务

300+

已帮助300+人成功就业

1000+

帮助1000+学员答疑解惑

200+

已帮助数百人成功转码





扫码找 Amelia 做 AI Engineer 技能诊断
适合在职工程师、数据/后端/全栈转 AI、已有 AI demo 但不知道怎么包装的人。不承诺面试、offer、签证或收入结果。