

阿里·灵犀互娱·面试模拟手册

岗位：AI产品经理

部门：灵犀互娱-新业务中心

面试时间：2026-04-10 17:00（明天下午5点，字节面试结束2小时后）

形式：钉钉视频面试（会议号：403 292 4831）

面试官画像推测：灵犀互娱新业务中心的产品负责人/AI技术负责人，关注AI落地能力+产品商业化+数据驱动

⚠️ 明天三场面试全天作战

时间	面试	心态模式	平台
11:00	网易天下	🎨 体验设计师	Chrome浏览器
15:00	字节跳动	🔧 技术产品经理	飞书
17:00	阿里灵犀互娱	🗣️ AI产品经理（最硬核）	钉钉

这场是三场里最“正统AI产品经理”的——JD明确要求LLM/多模态/Agent/RAG转化为产品功能、Workflow编排、Prompt设计、SQL/Python数据监控。你的全部经历在这场都能直接用，不需要“翻译”。

一、岗位-候选人匹配分析

JD核心需求 → 你的对应优势

JD要求	你的匹配点	面试中怎么讲
挖掘AI落地创新机会	腾讯：从甲方“沉浸式社交”模糊诉求中提炼5个AI产品机会	"我在腾讯做AI商业化时，从一个模糊的‘沉浸式社交’需求中结构化提炼出E2E语音、声纹安全、UGC语音包、AI NPC、离线互动5个产品机会点"
AI应用方案设计 (Workflow/Prompt)	腾讯QA-Agent (RAG架构) + 米哈游NPC (RAG记忆+多Agent)	"我设计过两套RAG工作流——QA-Agent的知识库匹配→自动回复→分类→入库飞轮，以及NPC的短期记忆+长期记忆双层检索架构"
LLM/多模态/Agent/RAG转化为产品	E2E语音交互方案(多模态) + NPC多Agent方案 + QA-Agent(RAG) + HERO论文(LLM可信度)	"这四个技术方向我都有实际的产品化经验或深度研究"
产品评估体系+数据监控	PUBGM专项看板+双周指标复盘+工单质量审查	"我在腾讯搭建过面向头部产品的数据看板，建立了双周指标复盘机制"

JD要求	你的匹配点	面试中怎么讲
SQL/Python数据处理	技术背景：Python是主力语言，CVPR/ECCV论文全用Python实现	"Python是我的主力研发语言，数据处理和分析是基本功"
与算法工程师协作	推E2E方案时拉通算法团队技术评估+论文研究本身就是算法工作	"我自己就是从算法出来的，能和算法团队用同一套语言沟通，不会出现'产品提了一个算法做不到的需求'"

灵犀互娱背景速查

- **灵犀互娱**：阿里巴巴旗下游戏品牌（原三七互娱合作业务），代表产品有《三国志·战略版》《三国志幻想大陆》等SLG/卡牌游戏
- **新业务中心**：探索新方向，AI是重点赛道——大概率在探索**AI在游戏中的应用**或**AI驱动的新业务形态**
- **你的定位**：他们需要的是“能把LLM/多模态/Agent等AI能力落地为游戏产品功能的人”——你是**完美候选人**

二、自我介绍（阿里版，90秒）

面试官好，我叫张清安，中山大学计算机技术硕士在读。

我的核心优势是**既有AI算法的深度理解，又有把AI能力转化为产品功能的完整实践。**

算法层面：我以一作身份在CVPR 2026中稿，研究3D资产的逆渲染生成；ECCV 2026二作在投，做多模态大模型的幻觉抑制。对LLM、多模态、扩散模型这些技术方向的原理和边界都有第一手的理解。

产品层面：我在腾讯IEG基础技术产品部实习，做了三件事，每件都涉及AI能力的产品化。第一，推动端到端情感语音交互方案替代传统ASR-LLM-TTS串行架构——这是一个完整的AI方案从需求挖掘到商业化交付的闭环。第二，设计并推进QA-Agent自动化工具——基于RAG架构实现知识库匹配、自动回复、智能分类的工作流编排。第三，搭建数据看板和指标复盘机制，用数据驱动产品优化。之前还在米哈游做过基于RAG+多Agent架构的NPC智能体交互设计。

我看到灵犀互娱新业务中心的AI产品经理岗位要求——LLM、多模态、Agent、RAG的产品化落地、Workflow编排、数据监控——这些我都有对应的实践和研究。希望能为团队带来价值。

三、面试核心问答模拟

Phase 1: AI产品能力考察（核心环节，预计占50%+）

Q1: 你对大模型（LLM）在游戏场景中的应用怎么看？能落地的方向有哪些？

我从**已验证可落地**和**高潜力待验证**两个维度来说：

已验证可落地的方向：

方向	场景	产品形态	我的相关经验
智能NPC对话	开放世界/MMO的NPC具备记忆和个性	RAG记忆+人设 Prompt+多Agent	米哈游NPC交互 (老板娘方案)
AI语音交互	组队语音中的实时翻译、AI队友语音	E2E/ASR+LLM+TTS链路	腾讯GVoice AI语音助手
智能客服/QA	游戏内问题自动解答、工单分类	RAG知识库+自动回复+飞轮	腾讯QA-Agent
AIGC美术素材	角色立绘、场景贴图、道具设计	文生图/图生图+风格一致性控制	CVPR论文 (3D资产生成)

高潜力待验证的方向：

方向	场景	关键挑战
AI剧情动态生成	根据玩家行为生成个性化支线	叙事连贯性+世界观一致性+质量把控
AI自动化测试	LLM驱动的智能测试Agent	游戏环境复杂度高，Agent的探索效率
AI辅助数值策划	用LLM分析数值平衡性、预测经济系统	需要精确数学推理，LLM目前不擅长
AI驱动的个性化运营	千人千面的活动推送和难度调节	数据隐私+用户感知到"被操控"的反感

关键判断框架：评估一个AI方向是否值得投入，我看三个条件——①技术成熟度是否过了"可用"门槛 ②对用户体验的提升是否足够大 ③商业模式是否可持续（成本可控、可计费或可变现）

Q2：请详细讲一下你做的RAG workflow，怎么设计的？

我做过两套RAG workflow，复杂度不同：

第一套：QA-Agent（腾讯，服务场景）

1	用户提问
2	↓
3	[Query理解] 提取关键信息（组件名/错误码/平台/版本号）
4	↓
5	[向量检索] 在历史工单知识库中做语义检索 → 返回Top-K相关工单
6	↓
7	[相关性判断] 匹配度是否超过阈值？
8	├ 是 → [答案生成] LLM基于检索到的工单生成回复 → 用户
9	└ 否 → [智能分类] LLM判断问题类型 → 分配给对应二线开发
10	↓
11	[闭环入库] 问题解决后，新的问答对自动入库 → 下次类似问题直接命中

关键设计决策：

- 为什么用RAG而不是纯LLM？因为游戏服务场景有大量**专业术语+内部工具名+版本特定信息**，纯LLM会幻觉。RAG用知识库做事实锚定
- 知识库怎么建？历史工单+FAQ+排查手册 → 切chunk → embedding向量化 → 存向量数据库
- 怎么评估效果？三个核心指标：知识库命中率、自动回复准确率、二线工单减少比例

第二套：NPC记忆系统（米哈游，交互场景）



关键设计决策：

- 不是所有对话都存长期记忆——只有"玩家明确意图"、"情感强互动"、"超长停留时间"才触发存储，否则记忆库会膨胀
- 检索时不能只看语义相关性，还要加时间衰减权重——最近的记忆优先级更高
- Prompt工程的核心挑战：在token限制内平衡"人设描述"+"当前上下文"+"历史记忆"+"输出格式约束"

Q3：你怎么理解Prompt设计？能举例说明你是怎么做Prompt工程的？

Prompt设计的本质是用**自然语言编程**——你在定义AI的行为边界、思考路径和输出规范。

以NPC记忆系统为例，我的Prompt结构分四层：

第一层：System Prompt（人设层）

- ```

1 | 你是"老板娘"，酒馆的主人。性格：温暖、健谈、偶尔吐槽。
2 | 你对每一位常客都很关心，会记住他们说的重要事情。
3 | 你的说话风格：自然口语化，不用敬语，偶尔用反问句。

```

### 第二层：Memory Context（记忆层）

- ```

1 | [历史记忆] 这位玩家上次来时说："明天要去挑战冰龙副本"（3天前）
2 | [历史记忆] 这位玩家曾经帮你找回了丢失的酒杯（15天前）

```

第三层：Current Context（当前层）

- 1 [当前情绪] 玩家刚完成了一场苦战，语气疲惫
- 2 [环境状态] 现在是夜晚，酒馆里人不多

第四层：Output Constraint (约束层)

- 1 回复长度不超过50字。自然引用记忆，不要生硬。
- 2 如果记忆中有相关事件可以提及，用"对了..."或"话说..."开头。
- 3 如果没有相关记忆，就按当前情境自然回应。

Prompt设计中的三个常见坑：

1. **指令冲突**：人设说"偶尔吐槽"但约束说"始终友好" → 模型不知道听谁的 → 需要明确优先级
2. **记忆注入过多** → token超限+噪声淹没信号 → 需要做Top-K截断+相关性阈值
3. **输出格式不稳定** → 需要Few-shot示例固定格式，或者加后处理层做格式校验

Q4：你怎么评估一个AI产品功能的效果？

我的评估体系分三层：

第一层：技术指标 (算法侧看)

- 模型准确率/召回率/F1
- 推理延迟 (P50/P99)
- 知识库命中率 (RAG场景)
- 幻觉率 (生成场景)

第二层：产品指标 (产品侧看)

- 任务完成率 (用户达成目标的比例)
- 用户满意度 (NPS/评分/反馈情绪)
- 使用频率和留存 (用户是否持续用这个功能)
- 降级触发率 (需要回退到非AI方案的比例)

第三层：业务指标 (业务侧看)

- 效率提升 (自动处理比例 vs 人工处理)
- 成本节约 (API调用费 vs 人工成本)
- 收入贡献 (如果是付费功能的话)

实际案例：我在腾讯做QA-Agent评估时的指标体系：

- 知识库命中率 → 衡量RAG检索质量
- 自动回复采纳率 → 衡量回复有没有用
- 二线工单减少比例 → 衡量对人工的替代效果
- 新工单入库率 → 衡量飞轮效应是否运转
- 用户二次反馈率 → 衡量"一次解决率"

关键认知：技术指标好不等于产品效果好。比如知识库命中率95%，但自动回复的用户采纳率只有60% → 说明检索到了对的知识，但生成的回复不够"人话" → 需要优化Prompt的回复风格而非RAG检索。

Phase 2: 项目经历深挖

Q5: 你做的端到端情感语音方案，完整链路是什么样的？

这个项目的完整商业化链路可以分为四步：

Step 1: 需求挖掘

- 甲方是一个星际题材的SOC+MMO大作，核心社交场景是"酒馆对话"
- 甲方的诉求是"沉浸式社交"——但这是一个模糊概念
- 我做的是把这个模糊诉求**结构化拆解**为5个可执行的产品机会：E2E情感交互、声纹身份验证、UGC语音包、AI NPC语音、离线互动语音

Step 2: 方案设计

- 核心决策：E2E vs 串行？我做了四维对比分析（情感保真度/时延/多模态/可控性），结论是主推E2E+串行作为降级兜底
- 商业化分层：免费层（基础魔音变声+默认声线包）→ 付费层（云端魔音+定制音色+UGC语音包分成+声纹增值）
- 安全模块：声纹验证作为独立可计费的增值功能

Step 3: 技术评估

- 拉通内部算法团队做可行性评估——E2E模型的延迟、算力需求、多语言支持能力
- 和算法团队的分歧处理：算法要更长验证周期，产品有交付时间窗口 → 分层交付策略（确定的先交、不确定的PoC验证）

Step 4: 交付落地

- 定义验收标准和灰度方案
- 设计降级策略（E2E置信度低时自动切串行）
- 建立效果监控指标

Q6: 你怎么做用户调研和需求挖掘的？

以AI商业化项目为例：

信息来源分三路：

1. **甲方直接沟通**：项目对接会，记录甲方的原始诉求和优先级排序
2. **竞品分析**：调研GPT-4o语音模式、竞品游戏的语音社交方案，提取可借鉴点
3. **内部数据**：分析GVoice现有客户的使用数据，找出哪些场景用户使用最频繁/抱怨最多

从原始信息到产品机会的转化方法：

1. 甲方说"我要沉浸式社交" → 这是一个**情感需求**，不是功能需求
2. 拆解：什么场景下需要沉浸感？ → 酒馆对话/组队副本/开放世界随机遭遇

3. 每个场景下"沉浸感"的缺失原因是什么? → 语音没有情感/NPC千篇一律/没有社交身份感
4. 每个原因对应什么技术方案? → E2E语音/RAG记忆NPC/声纹身份
5. 每个方案的可行性和优先级? → 按技术成熟度×需求紧迫度排序

这个过程本质上是**结构化的需求拆解能力**——也是AI产品经理最核心的能力之一。

Phase 3: 技术理解深度

Q7: 你怎么理解RAG和纯LLM的区别? 什么时候该用RAG?

核心区别:

维度	纯LLM	RAG
知识来源	训练数据中的"记忆"	外挂知识库的实时检索
更新频率	需要重新训练/微调	只需更新知识库文档
幻觉风险	高 (没学过就瞎编)	低 (有事实锚定)
适用场景	通用对话、创意生成	专业领域问答、事实性要求高的场景
成本	推理成本	推理成本+检索成本+知识库维护成本

什么时候用RAG? 三个判断条件:

1. **知识有"保质期"**——游戏版本每周更新、客服FAQ每天变化 → 不可能每次都微调模型 → 用RAG
2. **需要事实准确性**——客服答疑不能幻觉、NPC不能"记错"玩家的历史 → 需要知识锚定 → 用RAG
3. **知识有权限边界**——不同用户看到不同的知识 (A公会的历史 vs B公会的历史) → 需要检索时过滤 → 用RAG

RAG的常见坑 (我实际遇到过的):

- **检索质量是天花板**——如果检索到的chunk不相关, 后面LLM再强也没用。核心优化点在embedding模型选择和chunk切分策略
- **上下文窗口争夺**——System Prompt + 检索结果 + 用户输入 + 历史对话 全部要塞进有限的token窗口 → 需要做优先级和截断策略
- **知识库质量管理**——garbage in garbage out, 不能只管入库不管质量审查

Q8: 你对多模态大模型的理解? 结合你的ECCV论文说说。

多模态大模型的核心突破是**打通了不同模态信息之间的语义理解**——文字、图像、音频、视频不再是孤立的处理管道, 而是在一个统一的语义空间里理解和生成。

但关键问题是可信度。这也是我ECCV论文研究的方向。

我们发现多模态模型有一个致命弱点——**Confidence Trap**: 当视觉信息不清晰时, 模型不是说"我不确定", 而是依赖语言先验**极度自信地说错误**。更危险的是, 传统的不确定性检测方法 (看熵值高不高) 完全失效——因为错误答案的熵值反而很低。

对产品的启示：

1. **不能盲信模型的置信度**——产品设计时必须要有独立于模型置信度的验证机制
2. **需要降级策略**——多模态能力适合做"增强"而非"依赖", 关键场景必须有非AI兜底
3. **评估体系必须分区间**——不能只看平均指标, 要特别关注"高置信低准确率"区间的表现

我们的HERO方案是从RL训练层面解决这个问题——动态熵感知惩罚让模型"知道自己不知道"。低熵区间准确率从≈50%恢复至>92%。

Phase 4: 产品设计思维

Q9: 如果让你为灵犀互娱的一款SLG游戏设计一个AI功能, 你会怎么做?

灵犀互娱的代表产品是《三国志·战略版》这类SLG。我从SLG玩家的核心痛点出发来设计。

SLG玩家的三大痛点：

1. **策略学习门槛高**——新手面对复杂的兵种克制、地形加成、武将搭配一脸懵
2. **时间成本极高**——SLG需要持续在线管理资源和军队, 对休闲玩家不友好
3. **社交协调成本大**——联盟作战需要大量沟通和协调, 信息散落在各个聊天群

我会选"AI战略参谋"作为切入方向：

核心功能设计：

功能模块	描述	技术方案
局势分析	基于当前地图状态, 自动生成局势报告 (敌我力量对比/资源分布/威胁预警)	多模态理解 (地图截图+数据) → LLM生成结构化报告
策略建议	"你的北方防线薄弱, 建议调兵2000+加固关隘"	RAG检索历史战例+LLM推理当前态势
联盟协调助手	自动汇总联盟成员的战力/位置/状态, 生成作战计划	Agent workflow: 信息收集→分析→生成计划→发送通知
新手引导	根据新手的当前状况给出个性化的发展路线建议	RAG (攻略知识库) + LLM个性化生成

为什么选这个方向：

1. **解决真实痛点**：SLG信息过载是最大流失原因之一
2. **技术可行**：RAG+LLM方案成熟, 不需要等技术突破
3. **商业模式清晰**：基础分析免费 (提升留存) → 高级策略付费 (变现)
4. **不影响核心体验**：是"辅助决策"不是"代替决策", 不会破坏SLG的策略乐趣

验证步骤：

1. 先做最小MVP——只做"局势分析"这一个功能, 给内测玩家试用
2. 核心指标：生成准确率、用户使用频率、7日留存对比
3. 如果验证OK, 逐步扩展到策略建议和联盟协调

Phase 5: 文化匹配 & 软素质

Q10: 你怎么和算法工程师协作? 有没有遇到过冲突?

我和算法团队协作的最大优势是**我自己就是从算法出来的**——CVPR一作、ECCV二作, 我懂算法工程师的思维方式 and 关注点。

具体协作方式:

1. **需求文档用"算法能听懂的语言"写**——不是说"体验要好", 而是说"推理延迟P99要<500ms"、"幻觉率<5%"、"检索命中率>85%"
2. **给足够的技术空间**——我定义"要什么效果", 不定义"怎么实现"。比如"NPC记忆检索相关性要高", 具体用什么embedding模型、用什么检索策略, 让算法同学决定
3. **共同定义评估标准**——上线前就对齐好"什么数据说明方案OK/需要迭代/需要回退"

冲突案例: E2E方案推进时算法团队要更长验证周期 vs 产品有交付窗口。解法是分层交付——确定的先上, 不确定的做PoC。核心原则是**找到兼顾双方关切的第三条路**。

Q11: 你为什么选择来灵犀互娱?

三个原因:

第一, 赛道。 AI+游戏是我最擅长和最热爱的交叉领域。灵犀互娱的新业务中心在探索AI的新应用方向——这意味着有机会做从0到1的产品定义, 而非在已有产品上做增量优化。

第二, 匹配度。 JD里列的每一条要求——LLM/多模态/Agent/RAG的产品化、Workflow编排、Prompt设计、数据监控——我都有对应的实践经验。这不是"我愿意学", 而是"我已经做过"。

第三, 互补性。 我的算法背景能帮助团队更准确地评估AI技术的落地边界——哪些能做、哪些做不到、哪些需要等待。在AI产品这个岗位上, **能判断"做不到什么"和"能做到什么"同样重要**。

Phase 6: 反向提问

① (**最推荐**): 新业务中心目前在AI方向上最优先探索的是什么? 是AI赋能现有游戏产品, 还是AI原生的新品类/新业务形态?

②: 团队目前的AI技术栈是什么样的? 是自研大模型能力, 还是基于开源模型/API的产品层创新?

③ (**备选**): 这个岗位在团队中的角色定位是偏"AI能力平台化"还是偏"具体产品落地"?

四、可能的陷阱题

问题	潜在意图	应对思路
"你做的更多是技术产品/toB的工作, 怎么适应toC?"	考察视角转换	米哈游经历就是toC视角+腾讯的甲方对接也是理解C端需求
"你对SLG品类了解多少?"	考察游戏品类认知	坦诚说了解程度, 然后用结构化分析能力快速展示对SLG核心机制的理解

问题	潜在意图	应对思路
"如果你设计的AI功能效果不好怎么办？"	考察失败应对	"先看数据定位是技术问题还是产品问题。技术问题调模型/Prompt/RAG；产品问题重新审视需求假设。预设好降级方案确保用户体验底线"
"你有SQL/Python数据分析的实际经验吗？"	考察数据能力	"Python是我研究的主力语言，数据处理是基本功。SQL我在搭建数据看板时有使用。具体来说..."

五、紧急速查表

自我介绍关键词

AI算法深度(CVPR+ECCV) + AI能力产品化实践(E2E语音/RAG/Agent) + 数据驱动

三段腾讯30秒版 (AI产品视角)

- E2E商业化：模糊需求→5个产品机会→E2E vs 串行四维决策→分层交付
- QA-Agent：RAG架构 workflow→知识库匹配→自动回复→飞轮效应
- 数据看板：PUBGM专项→多维分析→双周指标复盘→异常监测

米哈游30秒版 (AI产品视角)

- NPC智能体：RAG双层记忆+多Agent架构+E2E情感语音→一测验证

四大AI技术方向的产品化理解

- **LLM**：NPC对话、智能客服、文案生成 → 核心挑战是幻觉(HERO论文)
- **多模态**：语音交互、图像理解 → 核心挑战是Confidence Trap
- **Agent**：QA-Agent、多NPC协作 → 核心挑战是执行可靠性
- **RAG**：记忆系统、知识库问答 → 核心挑战是检索质量

灵犀互娱为什么选我

"我不只是'愿意学AI'的PM，而是'已经做过AI产品化'的人——LLM/多模态/Agent/RAG四个方向都有对应的实践和研究"

明天17:00，三场面试的最后一场，也是最匹配你背景的一场。全力以赴! 🙌