

不转到大模型不改名

博主介绍 & 笔记说明

作者:	不转到大模型不改名
平台:	小红书 + B站
方向:	大模型开发
文档版本:	2026年2月

1. 个人介绍

小红书 + B站博主，双985学历，校招毕业后进入国内一线大厂，目前就职于外企一线大厂。主要开发语言是C++。

通过自学，从0AI基础出发，自学掌握了Agent、RAG等技术，成功拿到了6个大模型的offer，其中包括：

- 京东（算法岗）
- 千问C端
- 网龙
- SAP（世界500强外企）
- 平安证券
- 华林证券

2. 笔记介绍

不转到大模型不改名在自学的过程中，把所有自学过程中遇到的问题，总结在了文档中。目前文档已经有12万字。

文档设计的理念是：针对于0基础学员，以应用方向为主，算法方向为辅助，帮助0基础的同学快速转行到大模型。

笔记设计的最大的理念是：突出重点，讲清楚每个知识点为什么要考，怎么复习，要复习多深入。

笔记内容包括：

2.1 面试真题（20+公司）

包含完整的岗位JD（帮助大家有针对性复习）、问题解析、参考资料、个人反思、视频讲解。

2.2 八股内容

涵盖Agent、RAG、模型基础、微调、推理部署等内容。最重要的思想是完全根据面试内容，针对性总结八股——面试常考的就总结的深入，不常考的就总结的少，突出重点和思路，不想让转行的人陷入“感觉什么都要学，不知道学多深”的困境。

2.3 自研RAG项目

自研了一个RAG项目，包含完整的项目技术开发文档和代码。项目不光总结该项目常见的面试问题和解析、简历如何写、项目设计的代码和技术详解。更重要的是总结讲解自己写项目的思路，不光提供项目，更重要提供写项目的思路，学会以后也能轻松扩展。同样该内容包含视频讲解。

2.4 参考资料

将自学过程中遇到的好的参考资料、视频，总结在文档中。让大家做到只管照着笔记学就可以。

2.5 持续更新

从文档上线目前已经更新了2个多月了。博主每天下班就在整理笔记，做到和大家一起学习。过完年也会更新更多面经、算法相关内容，持续更新，共同进步，但从未涨价。

笔记目前在小红书链接售卖，价格199元。

图片1：设计思路

前言

▼ 笔记/路线设计思路

大家好啊，我希望设计的这份资料是：自学入行大模型应用开发的最全的记录。

为什么是自学路线？

我是从0开始自学的，自学了200多天，面试了2个月，拿到6个Offer（奇克千问，京东（算法岗），SAP（世界500强外企），网龙，平安证券，华林证券），所有信息小红书可查。

这份资料是按照我自己学习自学路线，我希望把我完整的自学路线，心得，笔记，面试重点，都分享给大家，让自学的人少走弯路。

笔记包括：详细的大模型八股，知识点参考资料（视频，文章），讲解视频，项目实战（源代码，设计思路，讲解视频）。项目实战是可以直接写在简历上的。

我的所有笔记的来源，是可以我的小红书笔记中看到的，都是我根据面试总结出来原创的，而不是随便搬运网上，这个真实可查。

为什么是最全记录？

因为我在转行路上，所以的心得，笔记，复习资料以及所有的面试记录都记录在这里面。这也是我自己学习的资料，我在面试中任何发现面试出现的问题，我都会补充在里面，分析出题频率，复习思路，面试的最真实完整的记录，还包括岗位介绍，面试反思。从0开始到上岸，这份笔记记录了我所有的学习资料和心得感受。这份笔记会更新到我真正入职新公司，真正转行完成。到目前为止我也在一直不断地更新这份笔记，一旦发现问题，就会补充迭代，这份文档会越来越完整。但到目前为止，其实内容已经非常详细和完整，对于一般的面试都足够了。

适用人群

自学进入这个行业的社招同学和校招同学。大家完全可以先按照这个路线和笔记去学习，我后面更多更新的是一些更高级，更深入的进阶技巧，支持获得更高薪资的岗位的一些技巧和知识。并且整个笔记我都想告诉你的，哪部分最重要，这部分为什么要学，这部分怎么学，除了知识部分，我更想告诉你的是学习的思路和方法。

为什么要从应用入手？

对于转行或者想进入这个赛道的同学，大模型应用相比大模型算法/Infra，大模型应用是最好入手的。大模型算法/Infra 技术难度相对更大，需要你对深度学习/机器学习 有比较深入的了解，同时也需要比较高的硬件资源来支持你完成相关的实验和项目，同时很多大厂在算法这一块对背景的要求同样更高，倾向于有论文/有竞赛的背景。所以，如果你想入行，从算法/Infra上入行，明显难度更高，从进入这个行业的角度来说，从应用侧切入是更更好的选择？我的项目是面向于应用的，应用的笔记非常详细。但是同时，针对于应用的岗位经常也会问算法，我的整套学习路线包含了一些算法：微调/模型推理/压缩（量化）的知识。对于大模型应用岗位来说，这个足够了。这样的设计让大家既有广度，又能抓住重点。

路线/笔记的设计思路

1.提供实战项目（可写简历）

我在做一个RAG-Agent项目，这个项目也是我自己用来学习，准备写在简历的。

我用了方法帮助大家理解，掌握这个项目：提供详细的项目设计文档（看目录的项目实战），提供讲解视频，

图片2：项目介绍

项目实战

1. 项目介绍

朋友们，最近一直在调研这个项目，今天正式给大家介绍一下：

1.背景：

其实虽然自学转行AI应用开发，也得到了一些Offer，但之前项目都是以Agent为主，Rag理论知识丰富，但实践甚少，比如面试官经常问这种分块，召回等技巧，RAG评估，这些基本都复习了，但是其实还是停留在理论部分。我的Agent项目其实并没有特别深入做这些步骤。所以我也想着借着这个机会，自己再做一个项目，把这些面试最常考RAG细节真正做一遍，同时我也真正想要把这个项目确实能够有一些实际的产出，解决工作具体问题，能写在简历上。当我在技术选型，确定项目时，我发现我的选型项目完全可以给大家参考。因为有太多小伙伴一直问我项目如何做，怎么找，之前我更多的给的是思路，所以我希望刚好通过这个机会，帮助大家一起做一个都能写在简历的项目。我做这件事核心动机目标如下：

1. 将所有RAG面试问题揉入项目，深入做一遍，从实践中掌握RAG各流程，掌握RAG的整个步骤。
2. 项目要可以和工作结合，能够确实解决某个业务需求，这样可以写在简历上。
3. 从文档，代码，视频讲解，思路设计等所有心得汇总，帮助大家一起实现这两个目标。
4. 代码分阶段交付，每个commit做一小件事，你可以随着一个commit感受整个项目完成的过程。
5. 配有讲解视频。关键地方，项目设计思路，代码配置等配有讲解视频。
6. 可以结合Client端，完成Agent的调用，让它既是Agent，也是RAG项目。

+

2.项目介绍：

阶段一：

以RAG为核心的基础架构，将RAG封装成一个MCP - Server。这个Server上保存了本地处理过的RAG文本，任何MCP - Client端连接了这个Server后，可以通过自己手写的逻辑（分块，检索，召回），返回相关文本。至此是实现了个通用的RAG框架，完成了RAG所有的关键步骤。

这里的代码我会完整公布，如果你不想继续做，你只用做数据准备，使用成熟的MCP Client（Cursor, Github Copilot）去连接你的server，就可以实现一个针对于特定文档/业务的Agent。

通过阶段一，你会基本完成了所有面试常考的RAG问题，你能弄透，对RAG也算入门了。并且你要真的不想扩展，你确实可以直接拿来写在简历上。

阶段二：

实现自己的业务逻辑和扩展。

如果你想做的更好，那么你就结合你的业务场景去优化，把效果调到最好，去自己写一个Agent连接这个RAG Server处理相关逻辑。去调整分块策略，召回策略，模型，数据库等，适配你自己的业务，结合阶段一完成的RAG系统的评测去提高整体性能，真正是一个可以根据你的业务场景，找到最合适的文本块，产生最准确的回答的项目。这个过程中你会有更深入体会，而且可以用一些Agent开发的技巧，来把Agent的内容也集成到一个项目中。

这块我可能会结合我的工作需求去扩展，这块代码不一定能展示，不过我会展示这个思路，怎么扩展，怎么和自己业务结合，这个可以大家自己去琢磨，讨论。