

摘要：随着大语言模型的蓬勃发展，人工智能（AI）技术正经历一场深刻的民主化变革。大模型及其配套工具显著降低了技术门槛，为非技术背景人员参与 AI 创新开辟了新途径。本文以“图书馆员的 AI 应用自主开发”为主线，系统探讨了图书馆员如何在这一技术变革中从观望走向实践。从机遇与挑战的认知启程，到技术民主化带来的探索可能，再到必要工具与技能的准备，文章全面梳理了图书馆员迈向 AI 开发的历程。基于实践经验，本文提出了面向图书馆员的 AI 开发 ESSE 方法论（Explore 探索、Specify 详述、Scrutinize 评估、Enhance 强化），为零基础图书馆员提供系统化的 AI 开发指南。“图书馆开放 AI-Demo 工作坊”的案例实践表明，图书馆员能够充分利用其专业知识优势，结合 AI 工具与提示词工程，开发出满足实际需求的创新解决方案。这一探索不仅有助于解决图书馆业务中的长尾需求，更展现了技术民主化时代图书馆员角色的新价值。

关键词：技术民主化 低代码 大语言模型 应用开发 AI 应用

0 引言

近年来，人工智能（Artificial Intelligence, AI）领域经历了前所未有的发展，其中大语言模型（Large Language Model, LLM，下称“大模型”）的崛起尤为引人注目。这一进展可追溯至 2017 年 Transformer 架构的提出。Transformer 凭借其高效的并行处理能力和对长距离依赖关系的有效捕捉，彻底改变了自然语言处理（Natural Language Processing, NLP）的范式，为人工智能的快速发展奠定了基础。2022 年末发布的 ChatGPT 则彻底引爆了这一轮的 AI 革命。

Makridakis 曾预测人工智能将对社会和生活的各个方面产生革命性影响。得益于大模型强大的语言理解与生成能力，以及广泛的适用性，如今这一预测正在快速成为现实。大模型的这些特性源于其庞大的参数规模和海量的训练数据。随着模型规模的不断扩大，大模型不仅展现出了强大的核心语言能力，还呈现出了令人惊叹的“涌现”能力，如跨领域知识整合、上下文理解、推理和创造性问题解决等。这些能力使得大模型能够在没有专门训练的情况下，很好地适应不同下游任务，展现出它强大的泛化能力。此外，大模型融合语言、视觉、听觉等多模态信息，通过对齐预训练和指令微调，实现多模态感知与统一表示；也可集成知识图谱、搜索引擎、代码执行、工具调用等技术，或者与小模型融合，从而实现优势互补，性能上实现“1+1>2”的效果，更加提升了大模型的功能丰富性和性能优越性。得益于大模型强大的自然语言理解和代码生成能力，以及其标准化、低门槛的特点，AI 技术开发正在经历前所未有的民主化进程。非专业技术人员，也能够通过与大模型的自然语言交互，参与到 AI 应用开发中来。这种技术民主化不仅为图书馆智能化发展提供了新的可能，也为探索人工智能技术在特定领域的落地提供了独特的实践者视角。

1 图书馆员的 AI 应用开发：从观望到实践

1.1 启程：图书馆员面对 AI 的机遇与挑战

人工智能技术在图书馆领域的应用已有数十年的历史，其影响力和重要性随着技术的发展而不断增强。早在 1984 年，美国图书馆学界就已认识到人工智能技术

对于图书馆事业可能带来的深远影响。20 世纪 90 年代，美国国家医学图书馆率先测试了人工智能专家系统在辅助规范控制方面的应用，开启了人工智能技术在图书馆具体业务中的实践探索。此后，专家系统、自然语言处理、机器学习、深度学习等多种人工智能技术在图书馆的各个方面得到了广泛应用，涵盖了信息检索、编目、分类、采访、参考咨询和自动化等多个领域。

近年来，以大模型为代表的生成式人工智能技术的突破性发展，引发了跨行业的广泛关注和讨论。国际图联（IFLA）、美国研究图书馆协会（ARL）、云瀚联盟等专业协会纷纷发表报告，提醒图书馆界认识到大模型技术对图书馆未来发展的重要性。这些报告不仅分析了大模型技术可能带来的机遇和挑战，也为图书馆如何在这一技术变革中保持竞争力提供了战略性建议。

在机遇方面，大模型技术为图书馆带来了多维度的发展可能性。研究表明，大模型提升了信息处理能力，促进了算法与技术的优化、信息资源的升级以及信息组织与检索的创新。生成式 AI 和 GPT 类技术正在颠覆传统的信息管理与传播模式，推动信息管理与传播领域的智能化进程，使得图书馆能够提供更加个性化和智能化的服务。在智慧图书馆建设方面，大模型技术提升了服务流程的智能化水平、优化了资源管理、促进了基础设施的现代化，并助力图书馆员工技能增强。同时，AIGC 技术延展了图书馆知识服务的深度，推动了知识服务维度的提升和场景的升级，增强了知识服务的亮度和新颖性。

然而，图书馆在面对这些机遇的同时，也需要应对诸多挑战。尽管图书馆员对 AI 技术持谨慎乐观态度，但在技术应用层面仍面临诸多障碍，特别是在将这些技术无缝整合到现有工作流程中的过程。同时，知识产权保护、内容真实性验证和用户隐私安全等问题需要建立相应的风险治理框架。此外，图书馆员需要不断提升自身的 AI 素养和技能，以适应新技术带来的变革。在资源整合方面，如何有效整合传统图书馆资源与 AI 技术，实现优势互补，也是一个需要深入探索的问题。

面对这些机遇与挑战，一个有效的策略是从小规模、可逆的任务入手，如资源分类与标注等。这样，图书馆员能够在可控和灵活的环境中逐步适应 AI 技术，体验其优势，为更广泛的应用奠定基础。通过这种渐进式方法，图书馆员可以在保持专业判断的同时，充分利用 AI 技术提升工作效率和服务质量。

正如研究者们指出的，只有通过实践者的视角，我们才能更加充分理解和掌握大模型的技术，更好地评估大模型在特定领域的实用性，进而更敏锐地捕捉到这些模型在实际应用中可能存在的伦理问题和社会影响。这也正是本文探索图书馆员 AI 应用自主开发的意义所在。从 AI 使用民主化的角度出发，探讨非技术出身的图书馆员如何通过“实践式学习”，借助大模型的语义理解和代码生成能力，深度参与 AI 技术创新过程。通过赋予图书馆员参与 AI 应用开发的能力，我们可以确保这些应用更好地满足图书馆和读者的实际需求，同时也为图书馆员的职业发展开辟新的途径。

1.2 探索：技术民主化带来的可能性

在人工智能和大模型技术快速发展的今天，技术民主化（**Democratizing Technology**）为图书馆员等非技术背景的专业人士带来了前所未有的机遇。技术民主化的核心在于使技术相关的知识、工具和方法能够被更广泛人群获取、使用 and 参与决策。这种趋势正在重塑传统的技术开发模式，使得先进技术变得更加普及、易于获取和应用，让更广泛的群体能够参与技术开发和使用。

1.2.1 人工智能的使用民主化

传统的人工智能开发依赖跨学科的专业知识，涵盖数据科学、机器学习和软件工程等领域，还依赖于大规模的硬件和软件基础设施投资。这些因素共同导致了所谓的“计算鸿沟”（**Compute Divide**），限制了 AI 技术的普及和应用范围。然而，随着大模型技术的发展，特别是开源模型的广泛可用，AI 技术的参与度和可及性正在显著提升。正如 Mark Riedl 所指出的，AI 民主化意味着任何人都能创建和使用 AI 系统，而无需深厚的数学或计算机科学背景。这一民主化趋势还引入了 AI 应用开发的新范式，使得无论是技术专家还是非技术背景的个人，都能更轻松地参与其中。

对于开发者来说，大模型技术不仅使现有产品中的 AI 功能集成变得更加简单和高效，还在重塑软件开发领域的整体格局。早在 2017 年，GitHub CEO Chris Wanstrath 就预言，未来的编码将完全由机器取代，当前的技术发展正在印证这一预见。以 ChatGPT、Claude-3.5、DeepSeek-Coder 等为代表的大模型在代码生成、补全、解释等多个编程环节表现卓越。借助这些先进的 AI 工具，开发人员逐步摆脱传统编程方式的束缚，能够更快速地构建和推出最小可行产品（**Minimum Viable Products, MVPs**），从而高效地验证市场需求并进行产品迭代。这种开发模式的转变不仅提升了开发效率，还使开发者能够将更多精力投入创新和解决复杂问题上，而非陷入烦琐的编码细节中。

对于非技术背景的图书馆员来说，这种技术民主化趋势提供了积极参与技术开发的机会。设计师、教育工作者和图书馆员等各种背景的专业人士，现在都可以成为 AI 应用开发的重要贡献者。这种多元化的参与不仅打破了传统技术开发的壁垒，还为 AI 解决方案的设计和 implementation 带来了新的视角和洞察。

1.2.2 MaaS 平台的技术民主化

随着大模型技术的不断发展，模型即服务（**Model as a Service, MaaS**）平台正成为图书馆员等非技术人员进入 AI 领域的关键基础设施。MaaS 平台提供了一种可扩展且易于访问的解决方案，使得使用者无需拥有昂贵的基础设施和深厚的模型训练知识即可利用大模型能力。得益于此，MaaS 平台正在成为 AI 利用民主化的关键催化剂。

首先，MaaS 平台降低了技术门槛。通过提供预训练模型和简化的 API 调用集成过程，开发者无需从头开始训练模型或深入了解模型的内部工作原理。即使没有 AI 专业背景的开发者也可以通过简单的 API 调用集成使用先进的 AI 模型，从而将更多的 AI 能力转化为实际应用。

其次，MaaS 平台简化了模型开发流程。传统的 AI 模型开发需要大量时间和资源进行数据准备、模型选择、训练和优化。MaaS 平台提供以大模型为核心的数据处理、模型托管、训练、调优、推理部署和智能应用开发等全周期管理工具，大大降低了模型开发的时间和资源需求，使开发者能够专注于应用逻辑和创新，从而加快产品开发和上市速度。

此外，MaaS 平台提高了成本效益。建设和维护人工智能基础设施需要在硬件和人力专业知识方面投入大量资金。通过采用“即用即付”的定价模式，小型企业和个人开发者也能够负担得起使用高质量的 AI 模型。这样，资源有限的用户也可以参与到 AI 应用浪潮中。

MaaS 平台还促进了知识共享和协作。这些平台通常拥有一个开发者社区，这些社区通常由模型贡献者和用户组成，是交流见解、经验和最佳实践的中心。这种合作环境不仅提高了个人的专业知识，还推动了集体进步，让更多人了解人工智能知识。

通过上述方式，MaaS 平台使 AI 技术不再仅限于有大量资源和专业知识的大型组织，而是让更多用户能获得、负担得起并适应人工智能，从而将人工智能技术覆盖到更广泛的群体和领域。

1.3 准备：AI 开发工具与知识储备

1.3.1 大模型基础知识与应用技术

在开始 AI 应用开发之前，图书馆员需要深入理解大模型的基础知识及其能力边界。这不仅包括对不同模型的特性和适用场景的认识，还应充分了解它们的局限性。虽然 MaaS 平台为下游行业用户降低了大模型的应用门槛，但大模型在语言生成、知识利用和复杂推理等基础能力上仍存在一些典型问题，例如幻觉、知识时效性差、专业化生成能力较弱以及推理不一致等问题。熟悉这些特性有助于在开发过程中合理利用大模型的能力，避免过度依赖或错误使用。

为了提高大模型应用的稳定性与一致性，业界发展出多种技术方案。其中，提示词工程可以优化与模型的交互方式，检索生成增强（Retrieval Augmented Generation, RAG）能够提升知识时效性和准确度，智能体技术有助于增强模型的推理能力，而模型微调则可以提升在特定领域的生成能力。这些技术方案的综合运用，能够有效克服大模型的固有局限。

在知识储备方面，持续积累并构建个人的 AI 知识库同样重要。“通往 AGI 之路”、“Learn Prompting”、“提示词图书馆”等公开知识库提供了基础性的知识资源。DeepLearning.AI、LangChain、MicroSoft 等提供了系统化的学习课程。LlamaIndex、Cohere、Grop 等项目则提供了丰富的 AI 应用开发案例。这些资源的积累和整合将为后续的 AI 开发工作奠定坚实基础。

1.3.2 开源框架与工具生态

对于没有技术背景的非专业人士，大模型行业应用技术方案都有一些可快速上手的低代码工具和开源框架。这些工具和框架大多可以在代码托管平台上找到，其

中 GitHub 作为最大的开源社区平台尤为重要。正如 Burkhardt 指出，GitHub 是当今编码文化的中心，在 AI 民主化进程中扮演着重要角色。

在开发框架方面，LangChain、LlamaIndex、CrewAI 多个大模型应用开发框架提供了丰富的组件和接口，简化了复杂功能的实现。LangChain 提供了从开发到部署的全套解决方案，包含 langchain-core 等开源工具。CrewAI 专注于基于角色的多智能体系统的开发。LlamaIndex 则提供了数据连接器、检索引擎等工具，支持创建基于用户数据的定制化 AI 应用。

而近年来低代码开发范式的快速发展也为 AI 应用提供了新的可能。传统低代码工具的功能和灵活性仍受限于预制组件和有限的代码定制能力。Martin 等人的研究强调了大模型与低代码平台结合的重要性，形成了基于大模型的低代码工具。如引入了 LangChain、LLM Models 等 AI 节点的传统自动化平台 n8n，以及 Dify、FastGPT、RAGFlow 等完全基于大模型的新一代 AI 低代码开发工具。这些工具不仅继承了传统低代码平台的易用性，还利用了大模型的强大能力，提供了更智能的自动化支持和更高级的自然语言交互，使用户能够通过简单配置实现复杂的业务流程与 AI 功能。

表 1 是笔者整理的部分大模型 AI 应用开发的开源框架与低代码工具，包括大模型 API 服务商与管理工具、提示词、智能体、RAG、文档处理等。

表1 AI应用开发框架、平台与低代码工具

名称	技术/服务类型	是否开源	简要说明	链接
Agenta	提示词工程	Y	开源的LLM开发者平台，支持prompt工程和管理、评估、人类标注以及应用部署的全流程	https://github.com/Agenta-AI/agenta
PromptPerfect	提示词工程	—	提示词优化工具	https://promptperfect.jina.ai/
Huggingface	开源模型MaaS	—	HuggingFace提供开源模型的体验、推理、训练、部署和应用的平台	https://huggingface.co/
Ollama	开源模型管理工具	—	Ollama是一个简单易用的本地大语言模型运行框架，支持多平台部署和OpenAI接口兼容	https://ollama.com/
OneAPI	API管理与分发	Y	支持国内外众多大模型API管理，并统一为 OpenAI 格式调用	https://github.com/songquanpeng/one-api
LiteLLM	API管理与分发	Y	litellm是一个 Python SDK和代理服务器，支持使用OpenAI格式调用100多个大型语言模型（LLM）API	https://github.com/BerriAI/litellm
LangChain	AI应用开发框架	Y	LangChain 是一个开源框架，提供了众多模块化组件，简化大模型应用开发过程	https://github.com/langchain-ai/langchain
LlamaIndex	AI应用开发框架	Y	LlamaIndex是一个针对大模型应用的数据框架，旨在帮助用户将私有数据与大模型集成，提供数据连接器、结构化数据工具以及高级检索和查询接口等模块	https://github.com/run-llama/llama_index
RAGFlow	RAG	Y	基于深度文档理解的端到端开源RAG引擎	https://github.com/infiniflow/ragflow
Dify	RAG、Agent	Y	大模型应用开发平台。提供可视化界面结合了工作流编排、RAG管道、智能体、模型管理等	https://github.com/langgenius/dify
n8n	低代码自动化平台	Y	n8n是一种开源的工作流自动化工具，通过可视化界面创建和管理复杂的自动化任务。它允许用户将不同的应用程序和服务连接起来，以实现数据传输、任务自动化和流程优化	https://github.com/n8n-io/n8n
MinerU	文档解析	Y	MinerU是一个开源高质量数据提取工具，支持PDF/网页/多格式电子书转换为机器可读格式（例如，markdown, JSON）的数据格式	https://github.com/opendatalab/MinerU
OmniParse	多媒体文档解析	Y	OmniParse是一个用于处理各种格式数据的平台，它能够将文档、多媒体文件和网页内容解析成结构化数据，以便更好地与大模型框架兼容	https://github.com/adithya-s-k/omniparse
firecrawl	网络文档解析	Y	firecrawl能够将整个网站转换为适用于大模型的标记语言或结构化数据，提供强大的网页抓取、爬虫和数据提取功能	https://github.com/mendableai/firecrawl
PaddleOCR	OCR	Y	PaddleOCR是一套丰富实用的OCR工具库，提供数据注释和合成工具，支持模型培训和部署	https://github.com/PaddlePaddle/PaddleOCR
MindSearch	AI搜索	Y	MindSearch使用大模型来驱动一个多智能体框架，以构建高级Web搜索引擎	https://github.com/InternLM/MindSearch

1.3.3 技术与业务的融合实践

正如图 1 所示，在积累了一定的大模型基础知识以及 AI 应用的案例与工具库后，我们还需要深入了解图书馆业务场景和流程。只有全面掌握业务流程，对业务的输入、输出、中间处理过程以及相关规则进行详尽分析，才能精准识别出可以通过 AI 优化的环节。通过这种理论与实践的结合，图书馆员才能真正开启富有成效的 AI 应用自主开发。

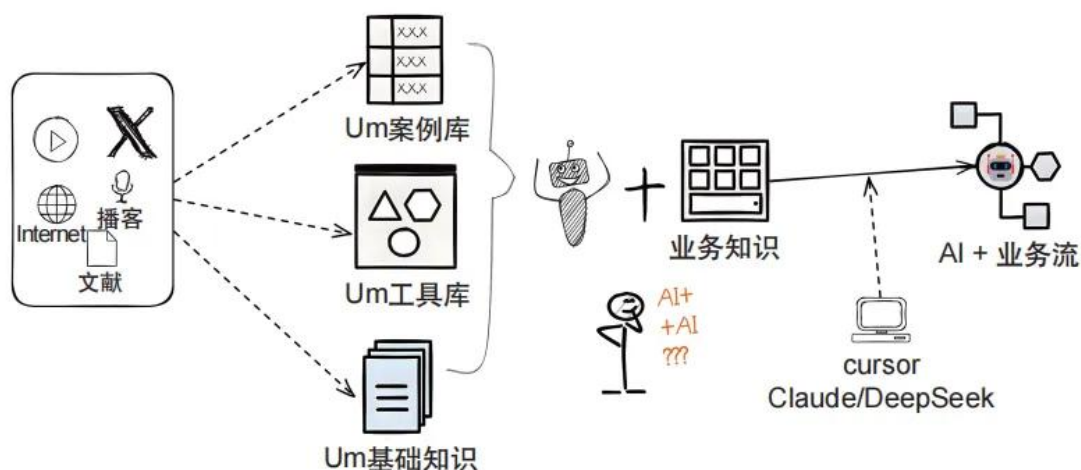


图1 技术与业务融合的AI demo开发实践

2 图书馆员的 AI 开发方法论：ESSE 框架

为了帮助完全没有编程基础的图书馆员进行 AI 应用开发，本文提出了如图 2 所示的 ESSE 方法论，即 Explore（探索）、Specify（详述）、Scrutinize（评估）和 Enhance（强化）。这个方法论旨在系统化地指导从需求分析到应用优化的整个开发过程，充分利用大模型的对话能力，将复杂的技术过程简化为直观的对话交互。

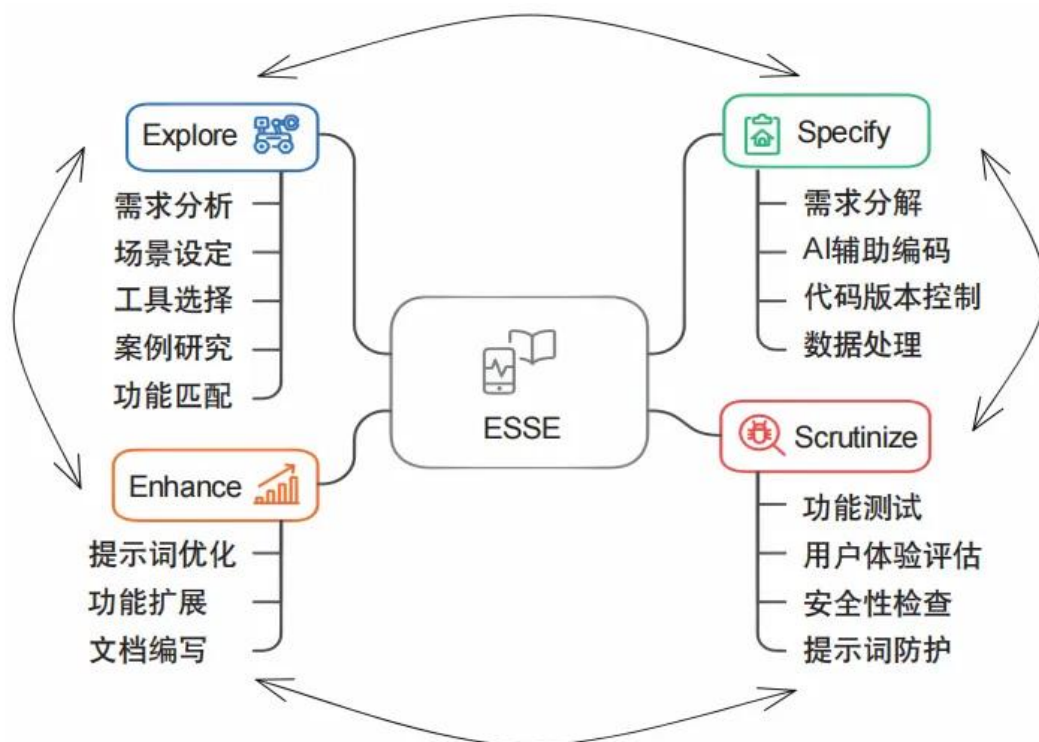


图2 零基础AI应用开发方法论ESSE

2.1 Explore（探索）：需求分析与场景设计

在探索阶段，图书馆员需要进行需求分析、场景设定、工具选择、案例研究和功能匹配。首先，明确图书馆服务中的痛点和需要优化的业务，定义具体的问题陈述和目标。其次，描述目标用户群体和使用场景。然后，评估不同的 AI 对话平台和相关工具，选择适合项目需求和技能水平的工具。最后，搜索分析类似的 AI 应用案例，学习成功案例的关键特征，并将需求与 AI 工具的功能进行匹配。

2.2 Specify（详述）：借助大模型进行开发

详述阶段的核心是利用大模型强大的语义理解和代码生成能力，通过交互式对话引导大模型生成项目代码。这一过程涉及多个关键步骤，包括需求分解、辅助 AI 编码以及版本控制。

首要任务是需求分解。这一步骤旨在将复杂的项目需求拆分为一系列小型、可管理的任务单元。对每个任务单元，创建明确的描述和具体的目标。这种方法不仅有助于简化开发过程，还能提高大模型对需求的理解和执行效率。随后，进入以 AI 为主导的代码编写阶段。这个过程从预先构建的工具和案例库中选择适当的代码示例开始。这些示例与实际项目需求一同提交给大模型，作为初始参考点。然后，与 AI 模型进行持续的对话反馈，逐步完善代码质量，实现预定功能。在这个过程中，需要遵循一些有效的策略。

其中一个关键原则是每次仅提交单一任务单元。这种方法建立在需求分解的基础之上，确保大模型能够精准理解并回应特定需求，避免因多任务并行导致的混淆和错误。另一个与之密切相关的关键策略是版本控制和代码备份。在开发每个新的任务单元之前，对当前可运行的代码进行备份，并为新增功能和需求添加详细注释。这种做法不仅符合软件工程的最佳实践，还为 AI 辅助开发提供了额外的安全网。它允许开发者在 AI 模型产生错误或偏离预期时，迅速回滚到先前的稳定版本，从而确保开发过程的连续性和代码质量的稳定性。

而当遇到大模型反复出现同类错误时，开发者可以采取多种应对措施。一种方法是尝试使用不同的模型，以获得新的视角和解决方案。另一种有效策略是使用思维链提示词，详细描述开发者的问题解决逻辑，引导 AI 模型按照特定思路修复代码。

大模型驱动的应用开发呈现出独特的双重性质。它们不仅作为独立的应用程序存在，还同时充当构建其他应用程序的基础组件。因此，当最终开发出的应用包含了由大模型驱动的功能模块，还需特别考虑因大模型自身的不稳定可能导致的输出错误。因此，在设计基于大模型的 AI 应用时，必须纳入额外的防护措施，以增强系统的安全性和可靠性。这包括输入预处理、提示词防护、输出后处理、错误处理与重试机制等。通过应用代码中嵌入这些安全策略，可以有效地管理大模型的输出不确定性，提高应用的稳定性和可靠性。

2.3 Scrutinize（评估）：应用测试与优化

评估阶段包括功能测试、用户体验评估和安全性检查。为每个功能创建测试用例并执行全面的功能测试，记录并分类发现的问题。进行小规模用户测试，收集用户反馈和建议，评估应用的易用性和直观性。同时，审查代码中的潜在安全漏洞，确保用户数据的保护和隐私，并实施必要的安全措施。特别需要注意的是，当最终开发出的应用包含了由大模型驱动的功能模块时，还需考虑因大模型自身的不稳定可能导致的输出错误。因此，在设计基于大模型的 AI 应用时，必须纳入额外的防护措施，以增强系统的安全性和可靠性。这包括输入预处理、提示词防护、输出后处理、错误处理与重试机制等。通过应用代码中嵌入这些安全策略，可以有效地管理大模型的输出不确定性，提高应用的稳定性和可靠性。

2.4 Enhance（强化）：迭代改进与完善

在强化阶段，重点是优化提示词、功能扩展和文档编写。根据测试结果调整提示词，优化 AI 助手的语言风格和回应方式，创建一个高效提示词的模板库。基于用户反馈添加新功能，改进现有功能的性能和可用性，探索与其他系统的集成可能性。最后，编写详细的用户手册和技术文档，记录开发过程中的经验和教训。

ESSE 方法充分利用了大模型的对话能力，将复杂的技术过程简化为直观的对话交互。在具体操作中，探索、详述、评估和强化四个步骤常常需要反复迭代，以不断完善应用效果。你可以用如下提问开始你的 AI 编程：“请根据提供的示例代码，编写一段符合需求的 Python 代码。我的具体需求如下。”在遇到错误时，可请求修复：“这段代码运行错误提示如下，请进行修复。”生成代码后，可要求 AI 进行代码审查：“请检查这段代码，确保没有语法错误和逻辑错误”；并添加代

码注释：“请在代码中添加清晰、简洁、易懂的代码注释，以提高代码可读性和可维护性”。在文档编写时，可以询问：“请根据代码内容，编写详细的用户部署和使用手册”。这种对话式的开发方式，让技术开发变得更加平易近人，即使他们没有编程基础，也能在大模型的辅助下顺利完成整个开发流程。

3 实践案例：图书馆开放 AI-Demo 工作坊

3.1 工作坊概述

“图书馆开放 AI-Demo 工作坊”是一个开源的 AI 应用实践项目，以“最小计算、落地实践”为导向，利用开源技术、框架和低代码 AI 开发工具，开发一系列可应用于图书馆实际业务的 AI Demos。通过分享可落地的 AI 应用案例，帮助图书馆员学习大模型 AI 应用开发的基础知识，提升 AI 意识和思维，同时探索大模型在图书馆业务场景中的应用潜力。部分已开发项目见表 2。

表2 图书馆开放AI-Demo工作坊部分项目介绍

项目名	工具框架	技术方案	项目描述
参考咨询助手	模型：GLM、01.YI、Grop 低代码平台：FastGPT、coze 框架：crewAI、LangChain 其他：Google book API、网络爬虫	提示词工程、RAG、Agent	自动识别分类读者问题，调用不同的API接口或代码函数生成回复。现阶段分类处理了：读者投诉、常规业务咨询、数据库咨询、馆藏图书咨询、学术文献检索咨询
图书馆讲座视频增强描述	框架：LangChain 模型：GLM、Paraformer	提示词工程、Agent	首先将视频转为音频，再利用Paraformer模型将音频转录为文本。依据转录的文本，再利用大模型对转录文本进行总结、元数据提取与小红书宣传文案撰写
读者荐购信息整理	框架：LangChain、streamlit 模型：GLM 其他：百度OCR API	提示词工程	将来自读者的手写单和网络表单截图、Excel文本，通过OCR和大模型，对荐购书目、读者信息按要求格式化整理后保存到荐购信息表中
特藏资源（节目单）实体提取	模型：DeepSeek、Doubao 其他：百度OCR API	提示词工程、Agent	对节目单图片进行OCR，然后利用大模型从OCR文本中按照实体词表规则，分步骤对多实体进行结构化抽取，存储为JSON数据
AI实体编辑器	框架：streamlit 模型：GLM、SiliconFlow/Qwen-2 低代码平台：Dify 其他：上海图书馆开放数据API	提示词工程、RAG	对大模型抽取的JSON数据，先预处理后导入本地sqlite数据库。然后在前端页面中加载、查看、编辑、保存。编辑页面嵌入了开放数据API，大模型API，知识库API

3.2 demo 案例分析

使用 ESSE 方法，AI-Demo 工作坊实现了对上海图书馆的读者荐购信息的结构化信息提取。上海图书馆的读者荐购信息来源多样，包括手写信息单、电话荐购、在线表单和政务平台等。这些信息最终以书写拍照图片、截图图片和 Excel 表格等形式汇总到文献采访老师手中。目前的处理流程是人工从这些信源中提取题名、ISBN/ISSN、责任者、出版社、出版时间、版本等文献基本信息，以及姓名、电话、

邮箱、提交时间等读者信息。这项业务属于典型的结构化信息抽取，而这正是大模型所擅长的任务。

生成式信息抽取（**Generative Information Extraction, GIE**）是利用大模型生成结构化信息的技术，它通过数据增强、提示词设计、零样本和少样本学习、约束解码生成的技术引导模型识别和组织文本中的实体、关系和事件，以实现更灵活和通用的信息抽取能力。少样本提示词在生成性信息抽取中显示出了极高的成本效益。通过精心设计的提示词，大模型能够在几乎没有标注数据的情况下，利用其内在的语言理解能力有效地进行结构化信息的抽取。这种方法极大地降低了对大规模标注数据的依赖，同时保持了较高的抽取精度和泛化能力。因此，综合实施成本与个人能力，读者荐购信息整理的 **Aldemo** 可以只利用少样本提示词工程进行结构化信息的整理。

读者荐购信息结构化整理 **Demo** 的处理流程包括文本识别、结构化信息提取、提取信息的后处理。这部分主要体现了 **ESSE** 方法论中的“探索”和“详述”阶段。在“探索”阶段，我们明确了读者荐购信息处理的痛点，选择了结构化信息抽取作为解决通过分享可落的 **AI** 应用案例，帮助图书馆员学习方案，并评估了少样本提示词工程的适用性。在“详述”阶段，我们利用大模型进行代码编写，通过预构建的提示词模板和少样本示例，引导模型生成结构化输出。

首先，文本识别模块利用百度 **OCR API** 实现对手写拍照图片或截屏图片中文字的精确识别，或者直接读取 **Excel** 表格中的文本数据，生成初始文本。

然后，这些文本作为用户输入，被传递给预先构建的提示词模板。提示词模板包含系统提示词、少样本样例和用户输入三个部分。接着，完整的提示词被发送给大模型，模型根据提示生成结构化的输出信息。

最后，提取信息的后处理则利用规则对大模型提取信息进行校验与纠正，并格式化保存到 **Excel** 中。

其中核心的结构化信息提取使用智谱的 **glm-4-air API**，并结合少样本提示词实现灵活的信息抽取与结构化。代码示例如下，样本示例涵盖了典型的输入场景和输出格式，这些代表性的样本示例与系统提示词和用户输入一起构成完整的提示。

```
llm=ChatZhipuAI (  
  
model_name=llm_config.get ( "model_name" ,"default-model" ) ,  
  
temperature=llm_config.get ( "temperature" ,0) ,  
  
api_key=llm_config.get ( "api_key" ) ,  
  
)
```

examples=[

{

“input”：“希望图书馆能不能多采购或者收录台版 日版的漫画和轻小说 现在图书馆的轻小说和漫画数量非常少 如彻夜之歌 中二病也要谈恋爱阿尼呀等等。范闲，邮箱，woshi，提交时间：2023 年 3 月 1 日”，“output”：“{ ‘name’：‘范闲’， ‘phone’：None， ‘email’：None， ‘submission_time’：‘2023-3-1’， ‘books’：[{ ‘title’：‘彻夜之歌’， ‘author’：None， ‘ISBN’：None， ‘publisher’：None， ‘publication_time’：None， ‘edition’：None}， { ‘title’：‘中二病也要谈恋爱’， ‘author’：None， ‘ISBN’：None， ‘publisher’：None， ‘publication_time’：None， ‘edition’：None}， { ‘title’：‘阿尼呀’， ‘author’：None， ‘ISBN’：None， ‘publisher’：None， ‘publication_time’：None， ‘edition’：None}]}”

},

{

“input”：“流水编号 202442722138128 录入时间 2024-05-1108:08:02 受理日期 2024-05-

11 投诉人 姓名：黄先生、电话：19100000035 (/)投诉渠道 12345 涉及城市 中国/上海/上海中国/上海/上海 图书借阅市民建议：粤语教程及广府文化图书的种类稀少，对读者构成明显不便。建议补充：一、《实用粤语播音主持 语言基础教程（第 2 版）》(9787504385581) 二、《新时空粤语上册》(9787566811585) 三、轻松说粤语 (978-7-5100-8765-3) 四、‘偷听广州’：实用粤语口语 (9787519207717) 五、事实与理由 《大话广府下册》(9787218152899) 。”，

“output”：“{ ‘name’：‘黄先生’， ‘phone’：‘19100000035’， ‘email’：None， ‘submission_time’：‘2024-05-11’， ‘books’：[{ ‘title’：‘实用粤语播音主持语言基础教程（第 2 版）’， ‘author’：None， ‘ISBN’：‘9787504385581’， ‘publisher’：None， ‘publication_time’：None， ‘edition’：‘第 2 版’}， { ‘title’：‘新时空粤语上册’， ‘author’：None， ‘ISBN’：‘9787566811585’， ‘publisher’：None， ‘publication_time’：None， ‘edition’：None}， { ‘title’：‘轻松说粤语’， ‘author’：None， ‘ISBN’：‘978-7-5100-8765-3’， ‘publisher’：None， ‘publication_time’：None， ‘edition’：None}， { ‘title’：‘“偷听广州”：实用粤语口语’， ‘author’：None， ‘ISBN’：‘9787519207717’， ‘publisher’：None， ‘publication_time’：None， ‘edition’：None}， { ‘title’：‘大话广府下册’， ‘author’：None， ‘ISBN’：‘9787218152899’， ‘publisher’：None， ‘publication_time’：None， ‘edition’：None}]}”

},

{

“input”：“ome ind 上海图书馆外文图书征订读者推荐登记表 F 2023 年 12 月 13 日 独立学者职务 读者姓名 刘先生 学历 职称 单位名称 单位地址 单位邮编 电话或手机 13000000988 E-mail 单位经济类型 公有经济 私营经济 外商投资经济 其他 学科研究领域 社会科学 自然科学 生物、医药 农业 工业技术 其他 推荐征订的外文图书：书名 illiberal America: A HISTORY 版次 著者 stevesHahn 出版社 Wiw. Norton ISBN 出版年 OfficiallyMarch 19 2024 信息来源 pabliher’s Webpage”，

“output”：“{ ‘name’：‘刘先生’，‘phone’：‘13000000988’，‘email’：None，‘submission_time’：‘2023-12-13’，‘books’：[{ ‘title’：‘illiberal America: A HISTORY’，‘author’：‘stevesHahn’，‘ISBN’：None，‘publisher’：‘Wiw. Norton’，‘publication_time’：‘March 19 2024’，‘edition’：None}]}

},

{

“input”：“““## 提交信息

提交时间=2024-03-30 00:00:00

读者信息=读者姓名：丁某某 (女)，联系电话：13000000028，xianfan@163.com

咨询内容

读者 3/29 反映问题：读者为《绑架=Kidnap》（ISBN：978-7-5321-8833-8，索书号：I247.7/1261-8，作者丁某翼），反映该本作品在馆藏书目查询系统中“多了一个英文译名”，且没有图书封面，和另一本著作《小灰》（索书号：I247.7/1261-7）在系统中不太一样。读者需要回电答复 今天读者又来电希望尽快添加封面，提议可在豆瓣网找到封面图片，望尽快处理答复。”“““，

“output”：“{ ‘name’：‘丁某某’，‘phone’：‘13000000028’，‘email’：‘xianfan@163.com’，‘submission_time’：‘2024-03-30’，‘books’：[{ ‘title’：‘绑架=Kidnap’，‘author’：‘丁某某’，‘ISBN’：‘978-7-5321-8833-8’，‘publisher’：None，‘publication_time’：None，‘edition’：None}]}

},

]

example_prompt=ChatPromptTemplate.from_messages (

```
[
    ( "human" , "{input}" ),
    ( "ai" , "{output}" ),
]

)

few_shot_prompt=FewShotChatMessage PromptTemplate (

example_prompt=example_prompt,

examples=examples,

)

system_template="""
```

角色

你是一位行业内顶级的信息提取专家。你的主要任务是从提交给图书馆的文献采购建议的文本中提取提交者的基本信息和他们推荐的书籍信息。

任务

请按照如下的规则说明，慢慢来，一步一步分析提取信息。

技能

技能 1: 提取提交者信息

- 从文字中识别并提取出提交者信息：姓名、电话、邮箱、提交时间。

技能 2: 提取图书信息

- 从文字中识别并提取出所有图书信息，包括：题名、作者、ISBN、出版社、出版时间、版本。

输出格式

- **【重要】** 严格按照示例输出，其中提交者的基本信息（姓名、电话、邮箱、提交时间）和推荐的书籍列表信息（书名、ISBN、出版社、出版时间、版本）。

- **【注意】** 提取的邮箱、电话的格式要正确。电话是纯数字。如：13600000028。邮箱必须有 '@' 符号。如：ds-52@163.com。

限制

- **【重要】** 若某些信息无法从文本中提取，返回 null 作为结果。

- **【重要】** 只提取要求的字段信息，其他信息不要提取。

- **【重要】** 不要给出多余的解释或任何其他内容。

"""

```
final_prompt=ChatPromptTemplate.from_
messages (
[
( "system" , system_template) ,
few_shot_prompt,
( "human" , "{input}" ) ,
]
)

chain=final_prompt | llm

llm_output=chain.invoke ({ "input" : query})
```

在评估阶段，我们对生成的应用进行了小规模的功能测试。测试结果显示，对于标准格式的荐购信息，应用的信息提取准确率较高。但在处理非标准格式或包含特殊字符的内容时，准确率会有所下降。此外，为进一步提高大模型输出结果的稳定性，应用还增加了如输入预处理、输出后处理、错误处理与重试机制等。

虽然在本次 Demo 中，“强化”阶段的迭代改进尚未完全展开，但未来的工作方向已经明确。我们将根据测试结果，进一步优化提示词，提高模型在复杂场景下的抽取精度。同时，我们也将探索与更加通用的资源描述与结构化信息抽取处理管道，扩展 Demo 的功能，并编写详细的用户手册和技术文档，以便其他图书馆员能够复用和改进该方案。

4 结语

图书馆在人工智能技术的应用上仍处于起步阶段，但其潜力巨大。但图书馆员的人工智能素养和图书馆之间的技术分享是应用 AI 技术的主要挑战之一。“图书馆开放 AI-Demo 工作坊”项目致力于为图书馆员提供一个实践和分享的平台，展示在 AI 技术平权时代，非技术背景的图书馆员也能成为 AI 应用创新的生力军。

这一趋势的核心在于“最小可行产品”思维的应用。在 AI 技术快速发展的背景下，快速验证想法比追求完美更为重要。即使编程经验有限，利用开源工具和低代码平台，图书馆员也能快速构建 AI 应用原型。这种快速原型开发方法不仅降低了技术门槛，也为图书馆员提供了一个让他们能够将自身的专业知识与 AI 技术相结合的实践机会。在这个过程中，一个关键但常被忽视的环节是提示词工程。看似技术含量不高的提示词工程，实际上是连接业务需求与 AI 能力的关键桥梁。精心设计的提示词可以大幅提升 AI 应用的实用性和准确性，这正是图书馆员可以发挥专长的领域。通过将复杂的业务需求分解为可管理的小任务，并在提示词中嵌入专业知识，图书馆员可以充分发挥其业务优势，而无需深入掌握复杂的 AI 算法。通过参与式实践，图书馆员可以更好地理解 AI 的能力与边界，提升 AI 意识和思维，了解大模型行业应用的技术方案，作为业务与技术专家协调者发挥重要作用。

另一方面，图书馆的 IT 建设主要依赖外包形式，由专业开发者构建中大型业务和服务管理系统。然而，这种模式难以满足图书馆日常运营中大量存在的长尾需求。这些需求虽然小众，但种类繁多，直接影响工作效率。由于其特殊性和多样性，这些“痒点”和“痛点”往往被忽视，未能纳入系统开发的考量范围。低代码/无代码开发工具的出现，结合大模型技术，使得普通图书馆员有机会转变为“赤脚程序员”，能够自主开发小型应用，以应对专业系统未能覆盖的业务场景。

总的来说，大模型为图书馆员提供了前所未有的机会，使他们能够直接参与到 AI 应用创新中来。结合对图书馆业务的深刻理解，非技术背景的图书馆员完全有能力在这个技术平权的时代推动大模型在图书馆落地应用。这不仅将推动图书馆服务的智能化转型，也将重新定义图书馆员在大模型时代的角色和价值。