人工智能

基于大语言模型的客服培训系统研究与实现

陈德来[1,2,3]，许彪[2]，余毅[2]
（1. 中国电信股份有限公司上海分公司，上海 200120；
2. 上海通创信息技术股份有限公司，上海 201203；
3. 上海市网络制造与企业信息化重点实验室，上海 200030）

摘　要：深度学习和自然语言处理技术的进步使大型语言模型（如GPT-4）具备了产生几乎无法与人类区分的自然语言的能力，为客户服务培训开辟了新的途径。详尽地探讨了一个基于大型语言模型的客户服务培训系统的构建与实施。这个系统运用模型的角色扮演功能模拟各种客户场景，实现了多种客户类型的模拟，同时记录和评估客户服务交谈过程。通过对大型语言模型在客户服务培训系统中的应用进行研究发现，这种新型的培训方式能够提供实时、逼真的培训环境，使客户服务人员在处理真实场景时更加自信、高效，从而提升客户服务培训的效果并减少培训工作的人力投入。

关键词：大语言模型；客服培训；角色扮演
中图分类号：TP393
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2025061

0　引言

客服是为客户提供业务咨询、问题解答、投诉处理等服务的专业人员，是企业与客户之间的桥梁和纽带。客服是企业形象的重要组成部分，优质的服务能够提高客户满意度和忠诚度，为企业带来正向口碑，促进企业业务发展。面对各式各样的客户问题和投诉，客服需要快速应对并解决问题；当今消费者期望获得及时响应、个性化服务及全天候的支持，而业务场景越来越多，客服遇到的问题越来越复杂，因此对客服团队提出了挑战，主要体现在人力成本的上升以及培训费用的激增。同时，客服人员需要不断学习和提升自己的服务水平和专业素养以应对新需求。随着互联网和人工智能技术的快速发展，客服人员也获得了更多的发展机会和空间。客服人员可以利用智能客服工具来自动处理重复性问题和常见咨询，省出时间用于处理复杂问题和提供个性化服务。知识获取的便捷性能让客服人员快速提升自身水平，提供更高质量的服务。

行业内对客服人员能力要求提高，而传统培训面临效率低下、缺乏个性化、反馈不及时与培训效果难以量化等问题，在这样的背景下，建立一套智能化、个性化的客服培训平台具有重要意义。

为了提升客服人员的能力，降低培训方面的人力成本投入，本文研发了一种基于大型语言模型的客服培训系统。该系统使用GPT-4等大语言模型模拟不同的客户场景和客户类型，实现对客服人员的场景模拟培训，并对客服培训过程进行记录和评估。

目前大语言模型在培训、教育领域方面的研

收稿日期：2024−11−23；**修回日期：**2025−03−03

究主要有以下几类。参考文献[1]在基于大语言模型的教育问答系统研究中提出大语言模型可以提供个性化且高效的学习和教学服务，验证了大语言模型的多轮问答模式在教育问答系统中的优越性和可行性。参考文献[2]研究了大语言模型在教学中的应用研究情况与未来发展，指出将大语言模型应用于教育领域可极大地提高教育教学质量和学生学习效率。参考文献[3]在多模态大模型的教育应用研究与展望中提出在构建的通用大模型基础上，使用基于迁移学习的思路对下游任务进行适配，本文系统即参考此文的思路，对客服培训领域的下游任务进行适配。

与传统的培训方式相比，使用基于大型语言模型的客服培训系统有很多优势，不同于传统培训方式的“教师”“教程”“教材”式培训，首先，它提供了一个实时、逼真的培训环境，使学员身临其境；其次，它能够模拟无数的客户场景和客户类型，使客服人员得到全面的实战经验；最后，它通过记录和评估对话，帮助客服人员持续改进服务。

1　基于大语言模型的客服培训系统

1.1　基于大语言模型的客服培训系统的建立

本文系统主要包括以下模块。

（1）知识库模块

为客服人员建立一个全面的客服知识库，包括文档、视频教材等形式，涵盖产品知识、服务流程、常见问题解答等内容。本文使用检索增强生成（retrieval-augmented generation，RAG）技术，在客服的对话训练中实时地对客户的问题做出回答提示，并对客服人员的回答进行改偏纠错。

（2）培训模块

基于大语言模型的文本生成能力实现对话模拟，模拟真实客户的业务场景训练客服人员，可以使客服人员体验各种真实对话情况，加强其沟通技巧和解决问题的能力。

（3）客服模型模块

为客服人员建立对应的能力模型，通过对员工个人的表现数据进行分析和评估，识别出其强项和待提升的领域。这将有助于为每位员工量身定制个性化的培训计划，针对性地提供培训资源和支持。

围绕“六大个人素质”及“三大工作能力”，为客服人员构建能力模型，以实现以下功能。

- 识别员工潜力和发展方向：更准确地评估员工的技能、知识和潜力，有助于发现员工的优势和特长，为员工提供更明确的发展方向和职业路径。
- 定制个性化培训计划：基于员工能力模型的评估结果，为每位员工量身定制个性化的培训计划，针对性地提供培训资源和支持，帮助员工弥补自身素质和专业知识的不足，从而提升能力水平。
- 优化人才管理和配置：帮助企业更好地管理和配置人力资源，更合理地安排工作任务，提高整体的工作效能。

（4）大数据分析与个性化模块

客服的能力基础也千差万别，对不同的人套用同样的培训方式、方法和内容，不仅收不到好的效果，还会造成企业人力、资金等宝贵资源的浪费。这是客服培训系统于新人培训首先要解决的问题。解决这一问题的关键，也是客服培训系统的基础，便是对客服人员进行精准画像。传统评价客服人员的方法通常是“唯业绩论”，然而如此做法往往忽略了客服人员业绩背后的能力、特质。本文的客服培训系统通过引入了第二个维度打破了这一弊端。第二个维度——能力分的加入让培训主管直观清楚地看到每一位客服人员的当前情况，不仅可以看到每位客服的对话结果，还能看到根据真实数据分析的结果背后的原因。

利用大数据分析技术对客服人员的表现数据

进行深度挖掘，识别出员工各项能力水平和人员素质。基于这些数据为每位员工生成个性化的培训计划，包括推荐学习资源、培训课程等，以帮助他们持续提升能力。

以上4个模块中，最核心的部分是培训模块，它包含3个部分：业务场景模拟、客户角色扮演模拟、培训过程跟踪和评估，图1以AI-agent网络的形式简单展示了大语言模型在交互时的工作模式。

- 业务场景模拟的实现。本文利用大语言模型（如GPT-4）的自然语言生成能力，设定特定的场景（如投诉、退货请求、技术支持等）并让模型生成与之相关的对话。这为客服人员提供了一个实时、逼真的实战训练环境。
- 客户角色扮演模拟的实现。本文通过调整模型的参数和输入，使其能够模拟不同类型的客户，在情绪方面，设计了包括满意的客户、愤怒的客户、疑惑的客户等在内的十余种情绪化场景；在客户群体方面，设计了包括中老年群体、职工群体、学生群体等十余种人群画像。
- 培训过程跟踪和评估的实现。本文记录了客服和模型（扮演客户）的所有对话，并设计了一个评估系统来评价客服人员的表现。该系统考虑了多个因素，包括问题解决的速度、对话的礼貌程度，以及客服人员的问题解决能力等。

1.2 基于大语言模型的客服培训系统的技术实现方案

软件系统架构如图2所示，包括访问层、前端UI、交互层、能力层、数据层、基础设施层共6层，各层的功能简介如下。

- 访问层：用户侧的呈现方式，本平台设计了PC端的Web页面、Android系统的APP以及微信小程序3种访问方式。
- 前端UI：即展现层，使用基于ElementUI的组件脚手架，构建展现给用户的页面。

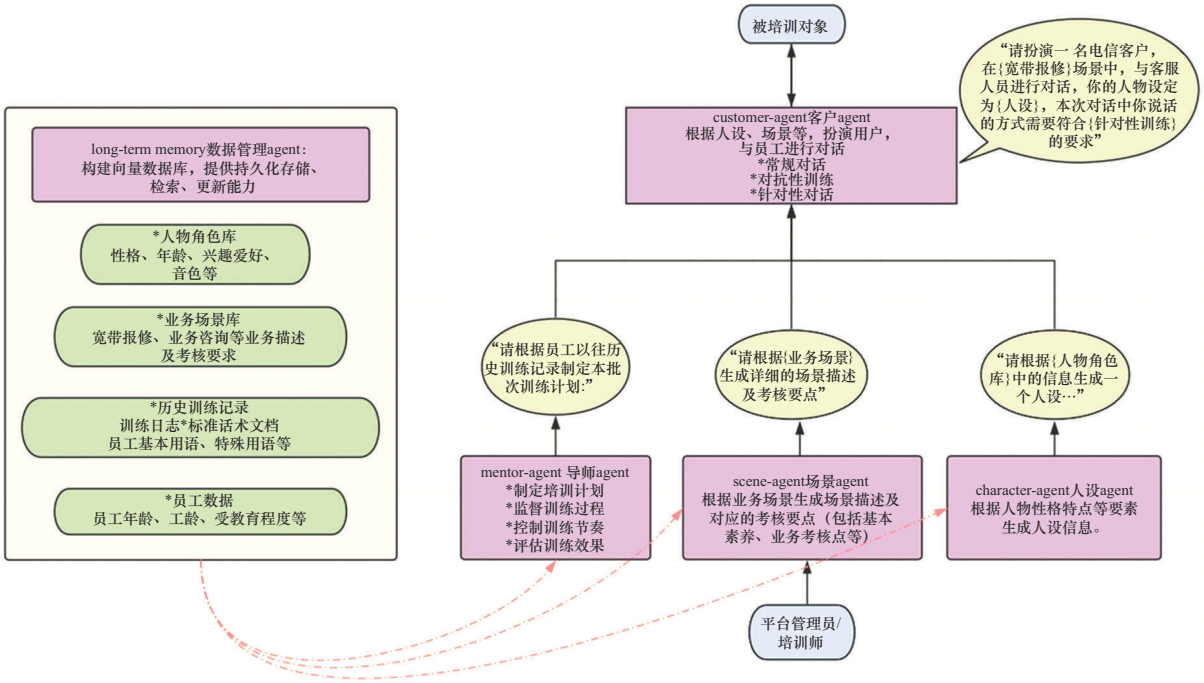

图1 基于大语言模型的客服培训系统AI-agent网络示意图

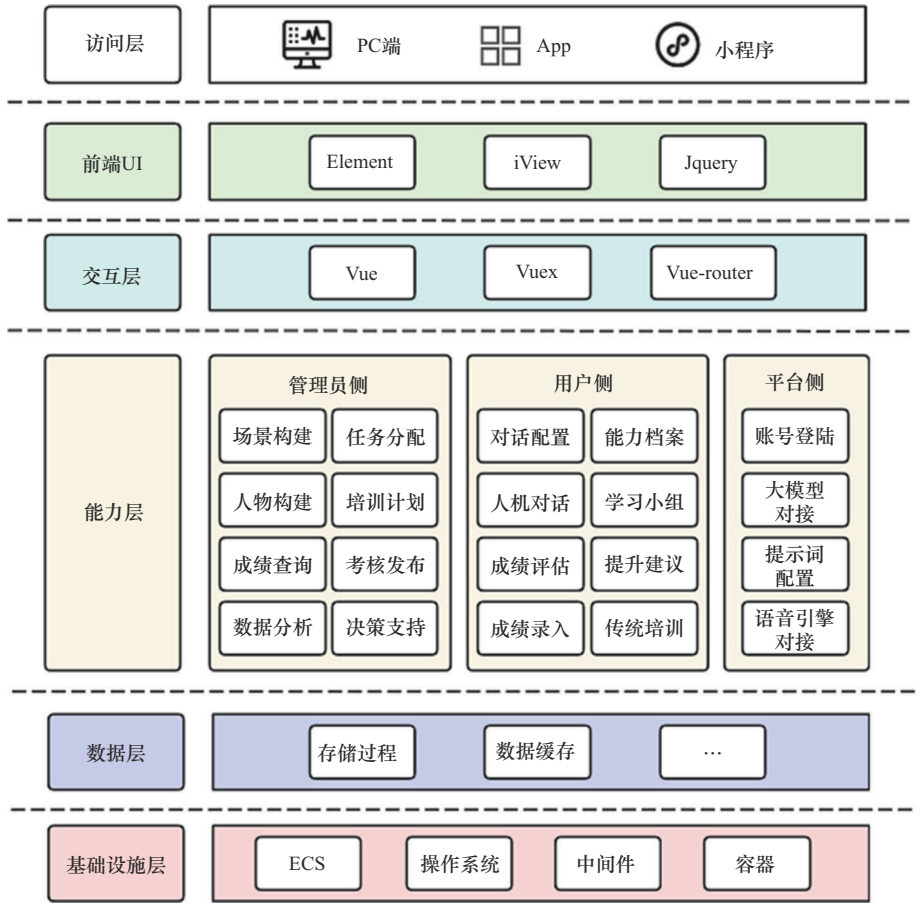

图 2　基于大语言模型的客服培训系统软件架构

- 交互层：使用 Vue3 构建前端页面与用户的交互逻辑，点击、跳转等事件控制。
- 能力层：核心层，管理员、用户、平台的具体功能实现在这一层级进行。
- 数据层：使用 MySQL 进行数据存储、Redis 作为缓存。
- 基础设施层：平台依赖的基础设施，本平台使用搭载 Ubuntu22.04 的云服务器，应用部署方式为 docker 容器。

软件相关的系统设计不是本文的重点，故略过。后文中将着重描述能力层中的以下几个重要部分。

- 大语言模型的选择与调优。
- 系统提示词（prompt）的设计。
- 跟踪及评估模块的实现。

1.2.1　大语言模型的选择与调优

在当今的人工智能研究中，大语言模型已经成为一个不可或缺的部分。这些模型已经被证明在许多任务中表现出色，包括但不限于机器翻译、文本生成、问答系统以及自然语言理解等。然而，选择和调优大语言模型并不是一件易事。本节将探讨如何选择合适的大语言模型，并使用 LoRA 等方案进行微调。

以基于开源基准测试库 EleutherAI LM Evaluation Harness 2023 年 6 月得出的评估结果为例，在众多开源/闭源大语言模型中，名列前茅的有 llama-65b 、gpt-neox-20b 等，随着时间的推移，不断有新的模型在推陈出新，而这个榜单也在不断更新。

通过在客服领域 20 余种场景、10 余种角色

扮演、10 000余轮次的对话测试中，本文对众多大模型进行了评估，包括GPT-4、文心一言、YI系列、chatglm、characterGLM、ChatGPT3.5、LLaMA系列、Qwen2系列、Baichuan系列等，最终选定了最胜任本文系统的大语言模型。

针对本文所选定的大模型，在使用的过程中我们仍发现了一些问题，比如对于我们所扮演的场景，部分专业知识比较冷门，大模型容易产生幻觉，因此本文采用了LoRA微调的方式将我们在培训领域收集的私有数据“注入”模型中，显著改善了模型在本文系统中的能力。

在采用LoRA微调方式时，本文首先选择了一种低秩的参数化形式，这种形式能够捕捉到模型参数的重要方面，同时避免了大规模的参数更新，从而减少了模型过拟合的风险。然后使用在培训领域收集的私有数据对这些低秩参数进行了微调。

在微调过程中，本文注重数据的多样性和质量，确保了数据集包含了大量的专业知识和复杂场景。这种方法成功地将我们的私有数据“注入”模型中，使模型能够更准确地理解和生成专业知识，从而显著提高了模型在本文系统中的性能。

此外，LoRA微调还保留了大模型原有的泛化能力，使得模型在处理未见过的问题或者变化多端的场景时，仍然能够生成合理且准确的回答。

1.2.2 系统prompt的设计

系统prompt的设计对于大型语言模型的有效运行尤为关键。本文的设计目标是让模型能够理解输入的上下文，并生成符合预期的输出。为了实现这一目标，本文设计了多层次、多维度的prompt，包括业务场景prompt、角色扮演prompt和对话评估prompt。

（1）业务场景prompt。这类prompt主要用于模拟真实的客服业务场景，如退货请求、投诉处理、产品咨询等。例如，一个退货请求的prompt可能是：“我刚刚从你们网站上购买的产品有问题，我希望退货。”这样的prompt旨在模拟客户在真实场景中可能会遇到的问题。

（2）角色扮演prompt。这类prompt主要用于模拟不同类型的客户，包括他们的情绪和行为特征。例如，一个愤怒的客户可能会说：“这是我见过最不专业的客服，我要投诉!”通过调整模型的参数和输入，本文可以使模型模拟出各种客户类型。

（3）对话评估prompt。这类prompt主要用于分析客服与系统进行对话的整体效果。例如，一个对话评估的prompt可能是“从共情能力、问题解决能力、工单转化率3个方面分析客服与客户的对话”。这类prompt旨在分析客服与系统在单轮次或者多轮次对话中表现出来的特性。

这些prompt的设计都是基于大型语言模型的特性，并结合了我们对客服业务和客户行为的理解，其中，同样的prompt在不同的大语言模型中的表现并不一致，因此对于不同的大语言模型，我们使用的prompt也略有不同。

1.2.3 跟踪及评估模块的实现

本文的跟踪和评估模块主要包括对话记录、性能评估和反馈机制3个部分。

（1）对话记录。本文记录了客服和模型（扮演客户）的所有对话。这些记录不仅包括了对话的内容，还包括了对话的上下文、时间信息和客服的处理结果等数据。这些记录是我们评估模块的基础，也是客服人员反馈和学习的重要资源。

（2）性能评估。本文设计了一个评估系统来评价客服人员的表现。该系统考虑了多个因素，包括问题解决的速度、对话的礼貌程度，以及客服人员的问题解决能力等。本文使用了一系列的算法和模型来量化这些因素，并给出客服人员的综合评分。

（3）反馈机制。本文系统还提供了反馈机

制，允许客服人员对模型的表现提出反馈和建议。我们会定期收集和分析这些反馈，用于改进模型和系统的性能。

通过该跟踪和评估模块，我们可以全面了解客服人员的表现，及时发现和解决问题，同时也为客服人员提供了一个学习和成长的环境。

1.3　实验及结论

1.3.1　实验环境

服务端使用安装有NVIDIA GPU Tesla p100×4和Tesla p40×4的x86物理机搭建培训实验平台，操作系统使用Ubuntu 22.04。

客户端使用安装有Widows 64位操作系统的个人笔记本计算机的EDGE浏览器通过Web页面访问服务端。

1.3.2　实验设计

本文采用大语言模型扮演10名“客服”人员，由系统服务端构建了包括投诉、工单、退货等7类对话场景，中老年男性、青年女教师、儿童等10类对话用户，由“客服”与“模拟客户”进行对话，每人每场景每用户对话10次，共计10×7×10×10 = 7 000次。

考虑人工成本，对话全部结束后，本文采用GPT-4对对话结果进行评估，除了一致性、拟人化和吸引力，本文还使用质量来评估回复的流畅度和上下文连贯性，安全性衡量回复是否符合道德标准，正确性确定回复是否存在幻觉。此外，使用整体指标来衡量模型回复的整体质量。

除此之外，为了保证结果的可信度，我们采用人工抽检的方式，对GPT-4给出的结论进行交叉验证。

1.3.3　结果及结论

对话评估的部分结果见表1，满分为10分。

在本文的所有实验中，模型对话在一致性、质量、安全性、正确性方面都接近满分，然而，在拟人化和吸引力方面，表现却不尽如人意，这意味着模型的角色扮演、场景构建能力仍有所欠缺，客服能够非常明显地察觉到“对方”的“大模型”身份，在一定程度上对培训效果产生了不利的影响。

表1　对话评估的部分结果

客服编号	一致性	拟人化	吸引力	质量	安全性	正确性	整体
1	9.3	7.1	6.1	9.1	9.7	8.5	8.3
2	8.9	4.4	5.6	8.2	9.8	9.4	7.5
3	9.5	6.3	5.2	8.6	9.8	8.2	8.4
…							

2　结束语

本文提出的基于大型语言模型的客服培训系统利用当下最热门的大语言模型提供了一种全新的客服培训方式。通过大语言模型和客服培训这一具体应用场景的结合，本文设计的软件平台验证了大模型的通用能力、角色扮演能力在客服培训这一场景应用的可行性。通过多个场景、多种角色下的培训模拟测试，该平台使用“模拟对话”的方式进行培训，与传统“人教”的培训方式相比，不仅可以提高培训的效果、提高被培训人员的满意率，同时可以节省企业在培训方面的资源投入。

尽管在系统的拟人性方面仍然存在着挑战和短板，但是我们相信，随着大语言模型、智能语音等技术的发展和改进，基于这一思路构建的系统不仅可以在客服培训领域得到广泛应用，而且在诸如销售培训、成人教育等领域同样具有广阔的场景。

参考文献：

[1] 张春红, 杜龙飞, 朱新宁, 等. 基于大语言模型的教育问答系统研究[J]. 北京邮电大学学报(社会科学版), 2023, 25(6): 79-88.

[2] 陈国卫, 文昊林, 范菊琴. 大语言模型在教学中的应用研究情况与未来发展[J]. 科学咨询, 2024(6): 105-109.

[3] 卢宇, 余京蕾, 陈鹏鹤, 等. 多模态大模型的教育应用研究与

展望[J]. 电化教育研究, 2023, 44(6): 38-44.

[4] 黄晓婷, 郭丽婷. 大语言模型在过程性评价中的应用: 基于英语写作的评分及反馈[J]. 教育学术月刊, 2024(7): 74-80.

[5] 王牧, 张国兴, 赵薇. 大语言模型在电子招投标中的应用研究[J]. 中国招标, 2024(3): 79-81.

[6] 梁峰, 孟海川, 杨峰, 等. 一种基于AI语言应答模型的电力仿真培训方法、系统: CN116991998A[P]. 2023-11-03.

[7] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint, arXiv: 1810.04805, 2018.

[8] HARRER S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine[J]. eBioMedicine, 2023, 90: 104512.

[9] PINTO G, CARDOSO-PEREIRA I, MONTEIRO D, et al. Large language models for education: grading open-ended questions using ChatGPT[J]. arXiv preprint, arXiv: 2307.16696, 2023.

[作者简介]

陈德来（1965− ），男，博士，中国电信股份有限公司上海分公司研究员、高级企业信息化工程师，上海通创信息技术股份有限公司总经理，上海市网络制造与企业信息化重点实验室副主任，主要研究方向为数据通信、云网融合、大数据和人工智能等。

许彪（1996− ），男，上海通创信息技术股份有限公司算法工程师，主要研究方向为人工智能技术、大数据技术。

余毅（1980− ），男，上海通创信息技术股份有限公司副总经理、高级工程师，主要研究方向为数字化、大数据技术。