1. AML中期汇报

在2024年秋季的课程前半段，同学们基于 ChatGPT、ChatGLM等先进的语言模型和多模态技术，各小组选取不同的方向，经过几周的努力，完成了一系列创新而有趣的项目提案。这些项目涵盖了语言模型与推理、图像与视频处理、推荐系统优化、教育与医疗应用，以及商业与其他实际应用等多个领域，展示了先进技术在各个领域的广泛应用和巨大潜力。

2024版本：东昱晓、唐杰、杜晋华、郑问迪、赖瀚宇、（共建同学请在此补足）



# 1. 语言模型与推理

## 1. LLM中的细粒度奖励建模：基于RL、PRM和增强记忆的高级推理方法

英文标题：Fine-Grained Reward Modeling in LLMs: An RL, PRM, and Memory-Augmented Approach for Advanced Reasoning

作者：任挽澜,刘家祥

摘要：本研究提出了一种用于大型语言模型 （LLM） 对齐的高级框架，集成了强化学习 （RL）、过程奖励模型 （PRM） 和动态记忆机制。与仅限于基本奖励标准（例如“有用性”和“毒性”）的传统 RLHF 方法不同，我们的模型结合了细粒度的评估指标，如“上下文连贯性”和“逻辑一致性”。通过利用另一个 LLM 进行自动反馈并实现门控内存系统，该模型可以有效地适应多步骤任务。这种架构提供了更高的可扩展性和准确性，减少了对手动标记的依赖并提高了推理质量。

## 2. 机器学习研究代理的系统化思路改进

英文标题：Systematic Idea Refinement for Machine Learning Research Agents

作者：刘子君,高骋,朱晗希

摘要：该项目旨在增强大型语言模型 （LLM） 支持的代理的机器学习 （ML） 研究能力，而不仅仅是基本代码生成。虽然最近的进展已经证明使用单代理系统在各种 ML 任务中生成代码，但这些方法通常只专注于提高代码有效性。他们缺乏为给定问题探索不同方法的能力，这限制了他们的适应性和性能。为了解决这一差距，拟议的项目开发了一个多智能体框架，通过自动提案、反馈集成和推理时间扩展系统地完善研究思路指南。通过整合来自 LLM 判断、代码生成过程和实验结果的多级反馈，这种方法使代理能够探索更广泛的解决方案路径，类似于人类研究人员。该框架在代码生成代理上是即插即用的，并将在共计75个kaggle经死啊上进行实验。预期结果是通过对方法论思想的全面探索，提高机器学习研究代理的理解和性能。

## 3. 利用基于LLM的多代理协作增强具身代理在人员密集环境中的基于文本任务的推理能力

英文标题：Leveraging LLM-based Multi-Agent Collaboration to Enhance Embodied Agents’ Reasoning Capabilities for Solving Text-based Tasks in Human-populated Environments

作者：孙楠,施铖铭,董毓文

摘要：该提案探讨了利用基于 LLM 的多智能体协作来增强具身智能体的推理能力的推理框架的设计。通过提高它们在复杂的人类环境中对基于文本的指令的理解和执行，该系统旨在改善机器人的动态推理、与人类的互动和任务完成。拟议的框架将使机器人能够自主处理任务，同时在需要时有效地寻求人工协助，确保以最少的干预完成任务。

## 4. 像人类一样自我反思，可编辑的LLM（E-LLM）是你所需要的

英文标题：Self-reflection like Humans, Editable-LLM (E-LLM) is All You Need

作者：吴彦辰,徐港新,邹冬晨

摘要：我们创新性地设计了一个 Editable-LLM，它可以不断实时反映和修改生成的内容，就像人类的反思过程一样。更准确地说，我们在传统的生成式大模型的基础上增加了一个检查机制，实现了对生成的文本进行增删改检查的操作。监督信号由仿真修改完成后的文本质量分数提供。这个想法的灵感来自强化学习领域的应用，例如 AlphaGo 决定根据作为监督信号的模拟来确定当前最佳决策。我们将使用强化学习的思想来指导我们改进传统的大型语言模型。

## 5. Mini-Omni：语言模型可以听、说，同时在流媒体中思考

英文标题：Mini-Omni: Language Models Can Hear, Talk While Thinking in Streaming

作者：谢之非

摘要：我们创新性地设计了一个 Editable-LLM，它可以不断实时反映和修改生成的内容，就像人类的反思过程一样。更准确地说，我们在传统的生成式大模型的基础上增加了一个检查机制，实现了对生成的文本进行增删改检查的操作。监督信号由仿真修改完成后的文本质量分数提供。这个想法的灵感来自强化学习领域的应用，例如 AlphaGo 决定根据作为监督信号的模拟来确定当前最佳决策。我们将使用强化学习的思想来指导我们改进传统的大型语言模型。

# 2. 图像、视频与多模态模型

## 1. 场景剪辑长视频用于更好的理解

英文标题：Scene-Clipping Long Video For Better Understanding

作者：赵子钰,王锦,肖锦松

摘要：近年来，在各种平台上视频内容量不断增加的推动下，对有效长视频理解的需求激增。然而，主要为短视频剪辑设计的现有模型难以捕捉长视频中固有的复杂时空动态。为了应对这一挑战，我们提出了一种新颖的场景剪辑长视频 LLM，它根据场景分布动态分割视频，无需预先指定剪辑数量，保证了语义的一致性。我们的方法将视频分割成片段，使用预先训练的图像编码器提取帧表示，并采用基于熵的场景剪辑算法通过 Video-Qformer 生成剪辑嵌入，同时结合时间位置信息。这种方法使 LLM 能够全面理解长视频的时空内容，为增强视频摘要、问答和交互式视频分析中的应用铺平了道路。我们将采用这种方法在各种基准上进行训练，并与不同的基线进行比较以验证其有效性。

## 2. 预测材料特性的Transformer模型

英文标题：Transformer Models for Predicting Material Properties

作者：张开维,于竣丞,李昊楠

摘要：图形神经网络 （GNN） 已广泛用于预测材料特性。然而，目前的 GNN 模型无法分配每个原子的贡献。同时，GNN 在捕获全局信息方面表现不佳，导致对某些属性的预测不令人满意。广泛使用的变压器可能适用于晶体，可用于解决这两个问题。我们建议训练一个用于预测晶体特性的 transformer，然后分析习得的注意力以分配贡献并演示模型的全局视图。如果成功，可以进行进一步的讨论。在相关工作方面，我们介绍了关于所提出的模型、注意力机制的使用、常见模型变体以及网络应用的研究。

## 3. 通过双编码器增强医学图像字幕生成

英文标题：Enhanced medical image captioning via dual encoder

作者：胥广洁,高杨驰

摘要：本研究的目标是通过提高模型在医学图像中捕获细粒度细节的能力来增强医学图像标题，从而生成更准确和具有临床意义的描述。我们计划使用双图像编码器模型，使用 CLIP 编码器提取一般图像信息，使用 SAM 模型（使用 MAE 编码器）来捕获更精细的细节。然后，这些编码器通过 Q-Former 集成，以连接到大型语言模型，以实现有效的文本生成。我们预计这种方法将提高医学图像描述的质量，捕捉广泛和具体的图像特征，最终以更详细和相关的输出使诊断过程受益。

## 4. VideoAlign：用于评估文本与生成视频对齐性的综合模型

英文标题：VideoAlign: A Comprehensive Model for Evaluating Alignment Between Text and Generated Videos

作者：杨苑明,刘晓倩,陈健

摘要：文本到视频生成模型最近取得了重大进展，但在实现与人类偏好的一致性方面仍然存在挑战。生成的视频往往与相应的文本描述缺乏可靠的一致性，人工评估既费力又昂贵。本研究提出了一种全面的解决方案来解决这些对齐问题。我们将介绍 VideoAlign，这是一种端到端的奖励模型，旨在自动评估视频生成模型的指令跟踪能力。

## 5. 大型多模态模型的高效推理

英文标题：Efficient Inference for Large Multimodal Models

作者：杨超群,张远达,黄茜瑛

摘要：大型多模态模型 （LMM） 在视觉指令调优方面取得了显著的成功，但由于大型语言模型 （LLM） 主干的自回归解码，它们的推理非常耗时。 推测解码 （SD） 已被证明可以通过 draft-then-verify 模式实现无损自回归解码加速。 在这项工作中，我们探讨了推测解码的应用，以提高 LMM 的推理效率，特别关注在多模态任务的上下文中利用 LMM 处理过程中产生的有用信息，例如视觉嵌入、隐藏状态和键值 （KV） 缓存。同时，我们尝试在这种情况下开发目标模型和草图模型之间的对齐技术，以尽可能提高加速效果。我们预计加速比将达到 2 倍以上。 代码和模型将在不久的将来发布。

# 3. 优化与推荐系统

## 1. 逻辑提取：在抽象和推理语料任务中增强AI泛化能力

英文标题：Logic Extraction: Enhancing AI Generalization in Abstraction and Reasoning Corpus Tasks

作者：陈远乐,耿双越,朱少廷

摘要：抽象与推理语料库 （ARC） 基准测试是评估 AI 系统从有限示例进行概括的能力、模拟人类智能的基本方面的关键挑战。尽管 AI 取得了进步，但当前最先进的模型在 ARC 任务上的表现落后于人类，在准确性和泛化能力方面存在巨大差距。在这项研究中，我们提出了一种新的框架，通过从任务样本中提取隐含的 “内部逻辑 ”并将其用作逻辑提取模型和生成模型之间的信息瓶颈来增强人工智能泛化。这种方法在模式识别和推理中模拟人类的认知过程。我们的方法结合了数据增强和课程训练，以提高不同任务复杂度级别的模型稳健性。通过推广封装这种内部逻辑的新兴语言，我们的目标是提高模型性能，同时探索超越显式编程或自然语言描述的更抽象的推理表示。这项研究可以为 AI 推理和抽象提供关键见解，有助于在人工系统中开发更像人类的泛化能力。

## 2. 基于RoPE位置编码的对比研究：一个扩展视角

英文标题：A Comparative Study of RoPE-based Positional Encodings from A Scaling Perspective

作者：杨天行,闫梓涵,张柱

摘要：Transformer 已成为自然语言处理的主要架构，尤其是作为大型语言模型 （LLM） 的支柱。然而，它们的二次计算复杂性为长序列的高效训练带来了挑战。一种常见的策略包括对较短的序列进行预训练，然后使用较长的序列进行长度外推。位置编码在此过程中起着关键作用，旋转位置编码 （RoPE） 因其强大的性能而被广泛采用。尽管 RoPE 很实用，但当序列长度超出预训练的上下文窗口时，RoPE 会面临分布外 （OOD） 问题，这促使了各种 RoPE 变体的开发，例如 PI、ABF、NTK 和 YaRN。这些方法旨在通过缩放机制提高模型在较长序列上的性能，但目前尚不清楚哪种变体更好，或者为什么 RoPE 在 Transformer 中有效执行。这项工作研究了这些基于 RoPE 的位置编码在长期上下文场景中的比较性能和原理，并试图揭示 RoPE 成功背后的机制。最终，我们的目标是提出一种新的位置编码方法，该方法在处理扩展上下文方面超越了现有方法。

## 3. 基于大型语言模型的带宽预测算法

英文标题：A Large Language Model-based Bandwidth Prediction Algorithm

作者：陈鑫,容逸朗,蒋政

摘要：在流媒体行业，带宽预测对于保障用户体验和优化资源至关重要。在低延时直播中，实时预估网络状况，调整传输策略，减少卡顿和延迟。对于长视频，它可以帮助自适应算法智能选择比特率，平衡画质、流畅度和缓冲。在短视频中，它决定视频码率组合，实现高清无缝播放，提高用户留存率。带宽预测还通过优化转码比特率和调度来影响 CDN 分发成本和效率。准确的带宽预测可以增强决策、用户体验和技术架构，这对竞争力至关重要。由于用户需求和网络复杂性，传统算法难以应对长期带宽变化。最近，来自时间序列预测的 Transformer 和大型语言模型 （LLM） 提供了新的解决方案。Transformer 擅长特征提取和长期依赖关系建模，而 LLM 则使用大型数据集的预训练快速适应。将这些模型应用于带宽预测可以大大提高准确性、泛化和实时性能，从而更有效地应对带宽预测挑战。

## 4. RLOJF：通过在线评测反馈增强LLM在奥林匹克编程中的表现

英文标题：RLOJF: Enhancing LLMs in Olympiad Programming with Online Judge Feedback

作者：徐家骏,吴垒,汪子涵

摘要：大型语言模型 （LLM） 在编程任务中取得了重大成功，尤其是在 LeetCode 等面向访谈的平台上表现出色。然而，我们观察到这些模型在奥林匹克级别比赛中的更复杂的问题上仍然表现不佳。这种性能差距主要源于奥林匹克竞赛编程问题所需的深度数学推理、复杂的算法思维和多样化的解决方案策略。为了解决这个问题，我们提出了一种新的方法：带有在线裁判反馈的强化学习 （RLOJF）。这种方法模拟了真实编程环境中的迭代过程，允许模型根据在线裁判 （OJ） 系统提供的分数和错误消息动态调整其输出。RLOJF 旨在提高模型生成的代码的正确性和效率，发展其使用自动反馈迭代优化代码的能力，并增强其在复杂编程任务中的推理和解决问题的能力。我们的研究贡献包括：为复杂的编程任务提出一种新的强化学习框架，设计一种利用 OJ 反馈的训练方法，对大量复杂的编程问题进行广泛的实验以验证该方法的有效性。

## 5. Syn-QL：用于文本到SQL的合成数据优化偏好

英文标题：Syn-QL: Preferrec Optimization with Synthetic Data for Text-to-SQL

作者：樊陆,胡瑞麟,陈奕哲

摘要：本白皮书解决了在文本到 SQL 任务中提高开源大型语言模型 （LLM） 性能的挑战，其中自然语言查询被转换为 SQL 语句以进行数据库交互。尽管开源 LLM 具有可访问性和成本效益，但在准确性方面落后于闭源模型。为了弥合这一差距，我们引入了 Syn-QL，这是一个利用合成数据生成和自我训练技术迭代微调模型的框架。我们的方法采用双模型方法，将 SQL Writer 和 SQL Verifier 配对，通过反复优化来提高 SQL 输出的质量。实验结果表明，在已建立的基准测试（包括 Spider 和 BIRD）上，性能得到了显著提升，这凸显了 Syn-QL 在使开源 LLM 在文本到 SQL 任务中更具竞争力的潜力。

## 6. 通过 LLM 生成的数据增强多领域推荐

英文标题：Enhancing Multi-Domain Recommendations via LLM-Generated Data

作者：江楚萌,罗开荣,潘至璇

摘要：由于日益严重的信息过载问题，推荐系统 （RS） 的应用在各种互联网平台上盛行，因为它们可以为每个用户提供个性化的项目。近年来，随着同应用短视频推荐、文章推荐、产品推荐等专业领域的应用越来越多，多领域推荐受到了广泛关注。多域推荐可以同时利用来自不同领域的知识，缓解数据稀疏问题，并允许单个模型跨多个域提出建议，从而降低部署成本。然而，不同域之间的历史数据大小各不相同。某些区域的数据可能比其他区域多得多，即丰富或冷启动方案。数据大小的这种差异可能会导致模型训练期间的某些限制。例如，在冷启动场景中，对特定领域参数的学习可能不足，而对领域共享参数的学习可能以丰富的场景为主。以前的工作 主要通过对模型进行细致的结构设计来解决这些问题。在本文中，我们从数据的角度来解决这个问题。LLM 的出现使生成虚拟用户和监控项数据成为可能。此外，LLM 凭借其广泛的世界知识和出色的理解能力，在冷启动场景中展示了令人印象深刻的推荐能力。在这种情况下，我们利用 LLM 来模拟冷启动场景中的用户，并在从现有的多域历史交互中学习后合成更充分的正样本。通过精心设计的数据过滤和降噪策略，可以提高多域模型的推荐质量。此外，通过推荐系统的视角，我们可以更深入地了解现有 LLM 的合成数据。

# 4. 教育、科学与医疗

## 1. AI赋能的智能教育：基于LLM的问题生成

英文标题：AI-empowered Intelligent Education: Question Generation based on LLMs

作者：路宜达,郑克寒,武文静

摘要：该项目体现了人工智能在解决现实世界教育挑战方面的实际应用。通过大型语言模型实现个性化学习，我们的目标是弥合课堂教学和个别学生需求之间的差距。如果成功，这种方法可能会产生深远的影响，彻底改变学生与教育内容的互动方式。个性化的 AI 驱动练习可以让学生按照自己的节奏工作，获得即时反馈，并加强他们对困难主题的理解。

## 3. MathLLaMA：一种专门用于数学推理和问题解决的语言模型

英文标题：MathLLaMA: A Specialized Language Model for Mathematical Reasoning and Problem-Solving

作者：张宇,雷长松

摘要：由于 AI 和自然语言处理的最新进展使语言模型能够理解和生成类似人类的文本，因此数学语言处理领域仍然面临独特的挑战。数学文本通常涉及复杂的符号符号、专业术语和形式结构，因此需要为训练模型量身定制的方法才能有效地处理此类内容。为了应对这些挑战，MathLLaMA 利用 LLaMA-Factory 框架创建了一个针对各种数学任务优化的模型，范围从代数操作和微积分问题解决到离散数学和数论等更高级别的领域。本文介绍了 MathLLaMA，这是 LLaMA 模型的微调版本，专为数学问题解决和推理而设计。MathLLaMA 利用 LLaMA-Factory 框架，该框架为训练和微调 LLM 提供了一个全面的工具包。MathLLaMA 的主要目标是扩展 LLaMA 在数学领域的使用能力，使其能够理解形式数学语言、通过多步解决方案进行推理以及生成准确的数学表达式。我们的方法包括在不同的数学数据集上微调模型，并使用专门的技术来解决数学文本带来的独特挑战。

## 4. 用于大型语言模型的自验证问题强化学习框架

英文标题：A Reinforcement Learning Framework for Large Language Model on Self-Verifiable Problems

作者：李颖欣,何光磊,周九阳

摘要：对于某些类型的自我验证问题，例如编程竞赛、博弈论和数学问题，是否有可能允许大型语言模型根据结果的反馈不断尝试解决方案，而不引入任何人类先验知识，仅依靠搜索和总结历史记录来最终找到答案？然后，通过使用这些尝试的记录（无论成功与否）来强化模型，是否有可能通过持续迭代来训练一个远远超过人类解决问题能力的大型语言模型？基于这个想法，我们计划设计一个强化学习系统，应用于 NOI/ACM 等自我验证的编程领域。这将检查对于可以自我验证的问题，是否存在一种系统的方法可以通过强化学习来训练超级专家模型。

## 5. 蛋白质功能预测的强化学习框架

英文标题：A Reinforcement Learning Framework for Protein Function Prediction

作者：田仁睿,龙飞,胡晨曦

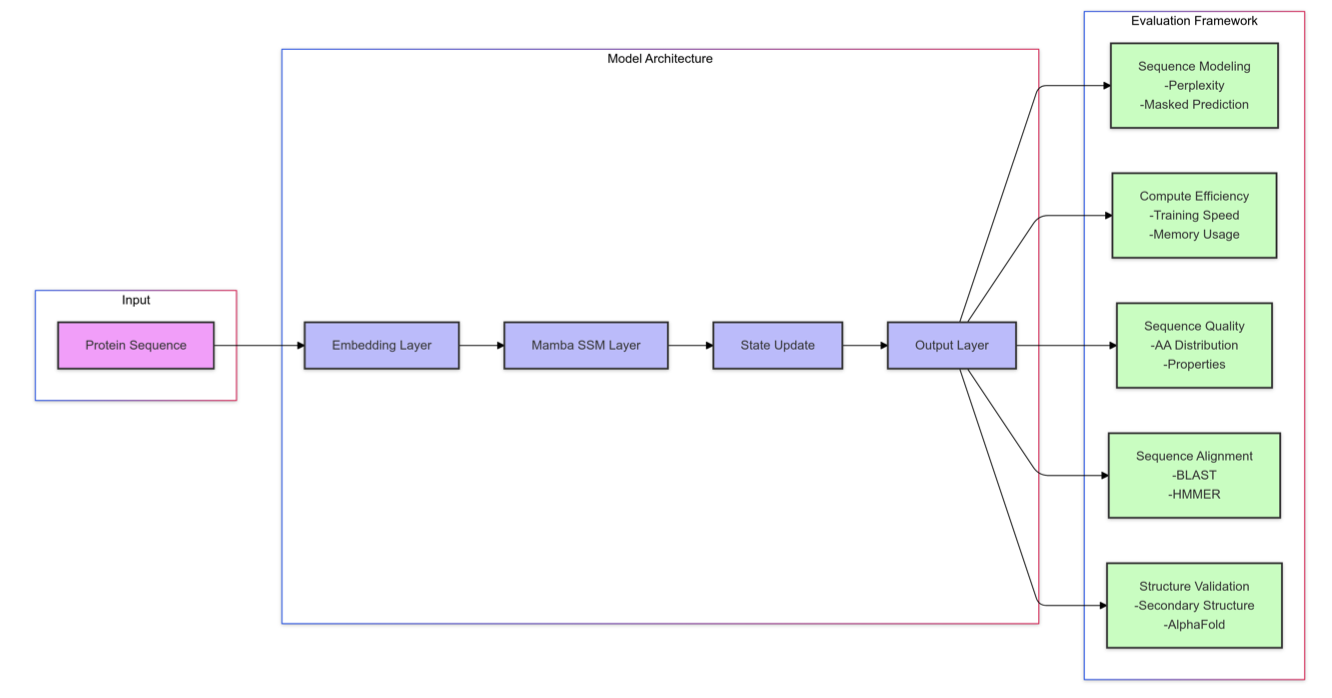
摘要：蛋白质是所有生命活动的基础。准确预测蛋白质功能不仅对于了解分子水平的生命至关重要，而且对于推进药物发现和疾病治疗中的应用也至关重要。目前的方法使用 CNNs 和 GNNs 从蛋白质序列和结构中提取信息来预测蛋白质功能，但我们发现这些方法经常存在不一致的问题。我们将开发一种新的强化学习框架来解决这个问题并实现准确的预测。

## 6. 蛋白质序列生成模型

英文标题：Protein Sequence Generation Model

作者：潘江, Zeeshan Zulfiqar, Ivan Iazykov

摘要：介绍了一种用于蛋白质序列生成的新方法，这是计算生物学的一个关键领域，其重点是生成具有特定功能和结构特征的序列。尽管最近在深度学习模型方面取得了进展，但在处理长序列和捕捉生物动态方面仍然存在挑战。我们建议采用 Mamba 架构，利用选择性状态空间更新实现高效的蛋白质序列生成。这种架构具有线性计算复杂性，能有效处理长程依赖关系，因此在降低资源需求的同时更适合生物动态。对 UniRef50 数据集的全面评估将证明它具有提供有竞争力性能的潜力，为计算资源有限的研究环境提供了可行的解决方案。



## 7. 可靠的药物-靶标相互作用预测的多代理框架

英文标题：A Multi-Agent Framework for Reliable Drug-Target Interaction Prediction

作者：高博文,白华君,张沛东

摘要：在本项目中，我们旨在探索如何在 DTI 预测任务中有效地利用 LLM 的强大功能。通过解决与在该领域应用 LLM 相关的挑战，我们希望开发一个可以有效利用 LLM 进行准确可靠的 DTI 预测的框架。成功解决这个问题可以通过更有效地识别有前途的候选药物来加速药物发现过程。此外，我们预计我们提出的框架可以扩展到其他科学发现领域，在这些领域中，LLM 的天真应用并非易事。

# 5. 商业与应用

## 1. ADSFT：生成模型的自适应和领域特定的微调范式

英文标题：ADSFT: Adaptive and Domain-Specific Fine-Tuning Paradigms for Generative Models

作者：王俊逸,隋元培,刘涛

摘要：在电子商务行业，为特定产品生成量身定制的高质量广告对于提高产品知名度和提高销售转化率至关重要。Amazon 产品数据集提供了一组丰富的产品信息，包括关键字和键值属性列表（kv-list），支持创建精确的广告生成任务。但是，生成的广告需要确保高生成速度以及与输入产品信息的事实一致性。因此，本研究的目标是微调 ChatGLM3-6B 以更快、更准确地生成电子商务广告，提高广告生成过程的整体质量和效率。

## 2. 基于检索增强生成的个人投资助手（RBPA）

英文标题：RBPA: Retrieval Augmented Generation based Personal Investment Assistant

作者：曾雪,李一诺

摘要：近年来，大型语言模型 （LLM） 取得了长足的进步，其功能被用于众多行业的各种应用。这些模型具有大量的参数，旨在处理复杂的任务和数据，提供增强的表达能力和预测性能。尽管 LLM 可以掌握基本的世界知识，但它们不能直接应用于动态游戏中的金融市场。金融市场的影响因素复杂，从宏观到微观，涉及的方面很广，对金融市场的分析需要建立金融领域的专业知识基础，包括基本专业知识、逻辑链知识、相关的网络知识。检索增强生成 （RAG） 是一种开创性的方法，它将大型语言模型的功能与从可靠数据库中检索信息的精度相结合。通过利用庞大的知识库，RAG 能够生成高度准确、相关和及时的响应，使其成为个人投资助理的理想技术，在这种情况下需要大量专业数据和个人数据。

## 3. 基于LLM的宝可梦对战代理

英文标题：Pokémon Battle Agent based on LLMs

作者：岑奇航,吕子涵

摘要：LLM 的快速发展导致了在交互式环境中的广泛应用，尤其是在游戏中，LLM 代理展示了令人印象深刻的决策和策略执行能力。本文重点介绍开发基于 LLM 的 Pokémon 战斗代理。我们结合了监督微调、KAG 和自洽等技术，以提高代理对上下文的理解并生成有效的战斗命令。通过机器人和人类对 Pokémon Showdown 的实验，我们将评估模型的胜率和战略绩效，旨在为开发能够处理动态和复杂环境的 LLM 代理做出贡献。

## 4. 探索用于沟通游戏的大型语言模型：谁是间谍？

英文标题：Exploring Large Language Models for Communication Games: Who is the Spy?

作者：魏晨天,许金朱,陈杰伟

摘要：随着小规模游戏和快速工程的扩展，逻辑推理和解决问题已成为我们日常生活中不可或缺的一部分。同时，大型语言模型 （LLM） 和提示工程技术的最新进展通过机器学习 （ML） 改变了游戏开发和人工智能 （AI） 领域。鉴于该研究领域的快速发展，迫切需要一个全面的概述，以概括 LLM 的现状并提示工程。因此，我们提出了实施思维链 （COT） 提示的想法，以指导 LLM 在解决小游戏的谜题时遵循推理过程。这就引出了我们最终项目的标题“谁是间谍？ 在本提案中，我们首先介绍了小规模游戏中大型语言模型 （LLM）、提示工程和思维链 （CoT） 提示的基础背景。接下来，我们对有关快速工程原则（CoT 和 ToT）的相关文献进行了回顾，这将在我们的最终报告中进一步探讨。最后，我们确定并讨论了该领域中值得进一步研究的建议方法。

## 5. 海外市场支持代理

英文标题：Agent for Oversea Marketing Support

作者：沈智杰,吴倩,沈珂宇

摘要：传统的海外营销方法需要大量的人工工作，例如阅读文件、进行采访、关注新闻以及分析有关目标国家的统计报告。在这个项目中，我们提出了一个基于代理的系统，利用大型语言模型 （LLM） 来提高国家研究的效率和智能性。

## 6. GoalAct：全球自适应动态法律多代理协作系统

英文标题：GoalAct: A Globally Adaptive Dynamic Legal Multi-agent Collaboration System

作者：陈俊杰,赵若雯,冯志远

摘要：本文提出了我们提出的基于 GLM-4 的多智能体协作系统，该系统采用一种结合全球和本地信息的策略，通过 API 调用访问相关法律数据库来提供法律服务。我们方法的优势在于将规划、反思和记忆在全球范围内和本地整合在一起，从而提高准确性和适应性。

## 7. BenCourt：基于LLM的法庭模拟的基准和框架

英文标题：BenCourt: A Benchmark and Framework for Court Simulation using LLM-based Agents

作者：张开元,李兆曦,李宣成

摘要：这是一个新框架的初步提案：BenCourt，它旨在为使用基于 LLM 的代理的法庭模拟提供一个测试和评估平台。完整版预计将于 12 月发布。Github 链接：https://github.com/Miracle-2001/BenCourt。

## 8. 使用大型语言模型进行时间因果发现：在工业系统中利用文本信息

英文标题：Temporal Causal Discovery with Large Language Models: Leveraging Textual Information in Industrial Systems

作者：孙卓凡,李青仪

摘要：该项目旨在探索一种使用大型语言模型 （LLM） 进行时间因果发现的新方法，以解决传统方法在工业系统中面临的挑战。该项目将利用 LLM 来处理和分析系统日志中的文本信息，以揭示时间因果关系，并开发一个 LLM 指导的因果发现框架，以优化发现过程并提高准确性。该框架将在真实的工业数据集上进行测试，并与现有的基线模型进行比较。预期成果包括一个强大的时间因果发现框架，可以有效识别复杂工业系统中的因果关系，以及在 AI 操作 （AIOps） 和系统维护中应用因果发现的新机会。

# 源代码