科研课堂开题报告

——基于生成式人工智能的课件自动生成

指导老师：葛声

学生：李昊霖、张之夏、罗浩宇

1. 研究背景及意义

近年来，人工智能技术飞速发展，特别是深度学习模型的演进极大地拓展了其应用边界。在2017年，Google团队提出的Transformer架构彻底革新了序列建模方式，为自然语言处理（NLP）领域的技术进步奠定了基础。随后，大规模预训练模型如BERT、GPT等相继推出，依托庞大的数据量和强大的计算资源，生成式人工智能（Generative AI）逐步从实验室技术走向实际应用，成为提升生产力的潜力工具。

尤其是2020年OpenAI发布GPT-3以及后续的ChatGPT 3.5版本后，生成式人工智能在社会中的应用逐渐普及。这类大模型在翻译、对话、内容生成等领域表现出高度的灵活性和创新潜力，并在国内外引发了新一轮的研究热潮。目前的行业发展大致分为两个方向：一是通过提高模型的训练方法与推理效率来增强生成效果；二是将生成式人工智能应用于各类实际场景之中，以发挥其技术潜力。许多团队聚焦于将生成式大模型用于教育、金融、医疗等领域，并已推出了一些初步应用。

然而，在教育场景中，当前AI生成课件的效果仍不理想，存在生成内容质量不稳定、信息逻辑性较差、视觉设计欠佳等问题。课件生成作为教师日常教学的重要组成部分，直接影响教学质量和课堂效果。为了减轻教师的负担、优化教学流程，生成式人工智能在教育领域的进一步应用研究显得尤为重要。

1. 研究现状
2. 大模型技术发展
   1. 递归神经网络

递归神经网络（RNN）和LSTM理论上能捕捉长距离依赖，但因梯度消失、难以并行等问题，难以扩展。尽管ELMo模型通过双向LSTM在多个NLP任务上达到SOTA，RNN和LSTM最终被Transformer取代，后者更擅长长依赖处理和并行化。

* 1. Transformer

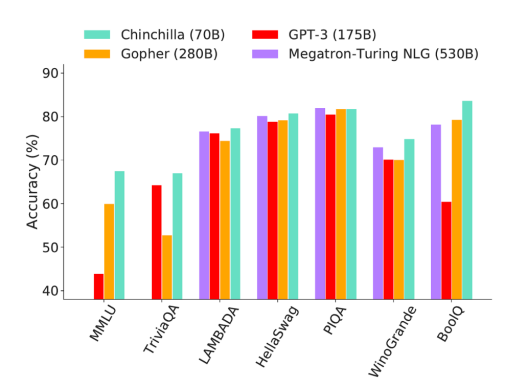
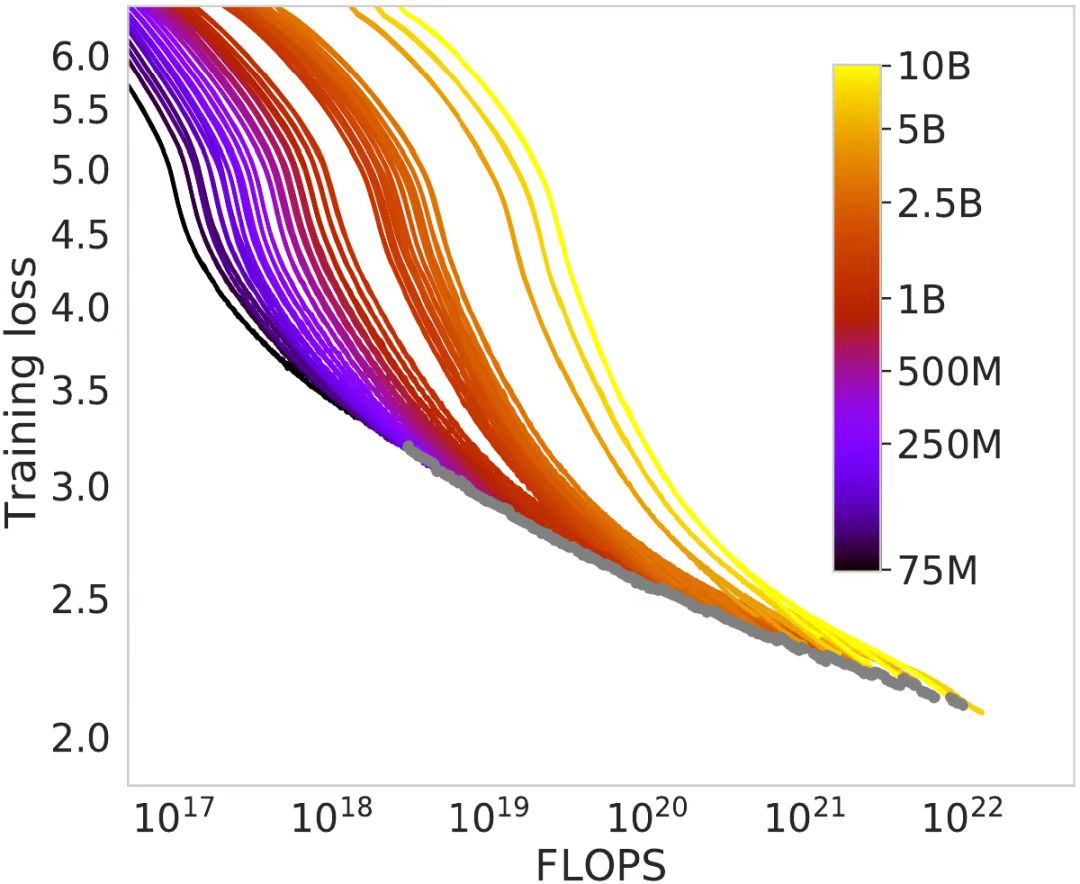
Transformer模型由Google于2017年提出，通过注意力机制解决了RNN/LSTM在长距离依赖和并行化方面的不足。其核心创新是自注意力机制，使每个token无需距离限制地关联其它token，从而消除了长距离依赖问题。同时，Transformer不依赖序列计算，能够并行处理所有输入位置，大大提升了训练速度。该模型的推出推动了大规模预训练模型（如BERT、GPT）的发展，标志了深度学习从序列建模向注意力机制的转变。

* 1. Scaling Law

OpenAI在2020年提出的缩放定律（Scaling Law）表明，语言模型的性能主要取决于计算量、模型参数量和训练数据量，而与模型的具体结构关系不大。他们指出，大模型在样本效率上优于小模型，能以较少的训练数据和步骤达到高性能。依据这一规律，OpenAI通过增加GPT模型参数，构建了具有涌现能力的大模型。尽管近年来AI的发展遵循缩放定律，但对于未来，有人认为不断扩大模型规模能实现AGI，也有人认为该定律会像摩尔定律一样终有极限。

* 1. Chinchilla

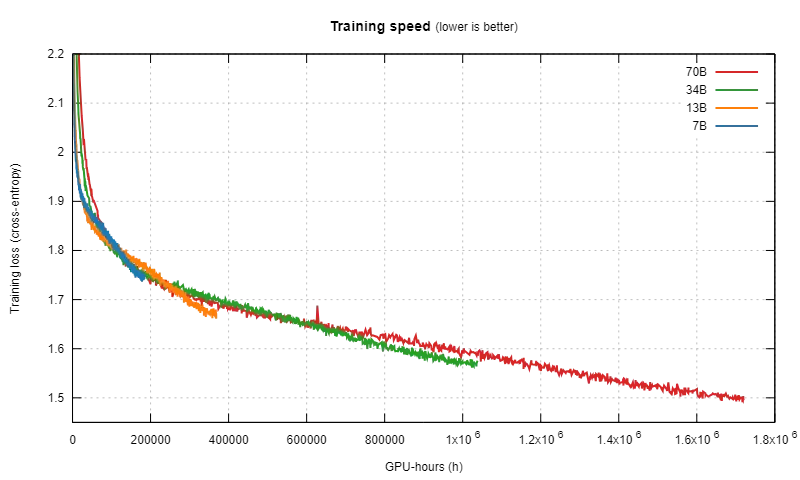
2022年，Google DeepMind在论文中提出，当前大模型通常训练不足，模型规模和训练token数量应成比例扩展。通过计算，他们发现Gopher模型的最佳规模应为现有的1/4，并在4倍的tokens上训练。为验证此假设，他们构建了70B参数的Chinchilla模型，并在1.4万亿tokens上训练，结果Chinchilla在各项测试中均优于Gopher。该研究完善了Scaling Law，提出的最优训练方法被称为Chinchilla Scaling Law。

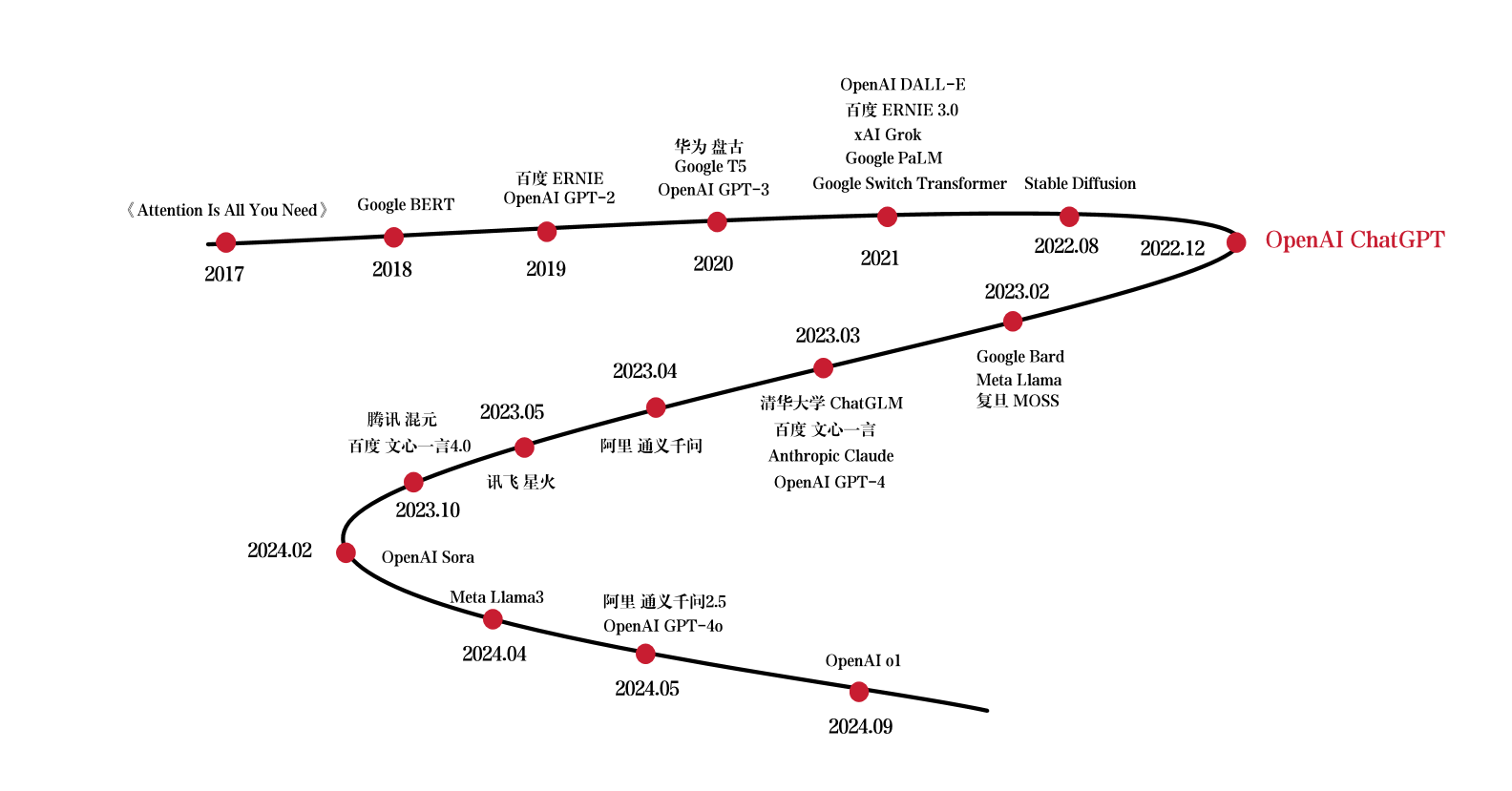


* 1. Meta Llama

在2023年7月的文章《Chinchilla’s Death》中，Thaddée Tyl分析了OpenAI和DeepMind的缩放规律，提出在充足的计算资源和数据支持下，较小的模型通过长时间训练也能超过大模型的性能。同年初，Meta的研究验证了这一观点，他们训练了四个不同大小的模型，并延长了每个模型的训练时间。

Meta的实验表明，初期小模型的loss下降速度较快，符合Chinchilla规律，但随着训练持续，较大模型逐渐超过小模型。然而在足够长的训练后，小模型进入线性下降阶段，其loss稳步下降并最终再次超过大模型。这一现象为模型规模与训练时间之间的关系提供了新的视角。



1. 国内外大模型发展

目前的大语言模型发展呈现出百花齐放的局面，各个公司推出的模型在不同能力上各具特色：

1. OpenAI的GPT系列：从GPT-3到GPT-4不断升级，成为高性能任务处理的代表。免费用户可以使用GPT-4o，但频率有限，而o1-preview仅对plus会员开放。

2. Anthropic的Claude：专注于上下文处理，2024年发布的Claude 3.5 Sonnet在代码和视觉能力上提升显著。

3. Google的Gemini：以超长上下文处理能力著称，支持多达10M token的上下文处理。

4. Meta的Llama：2024年发布的Llama 3（8B、70B、405B参数）在多个基准测试中超越Gemini和Claude，并持续扩展多语言和多模态能力。

5. xAI的Grok：闭源大模型，凭借最新版本v2.0加入顶尖大模型行列，仅限x premium会员使用。

6. 阿里云的通义千问：版本2.5支持多模态生成，开源版本为Qwen系列，增强了PPT生成等功能。

7. 智谱清言：由清华KEG和智谱AI合作研发，支持实时多模态功能，增强了图片和视频的理解与生成。

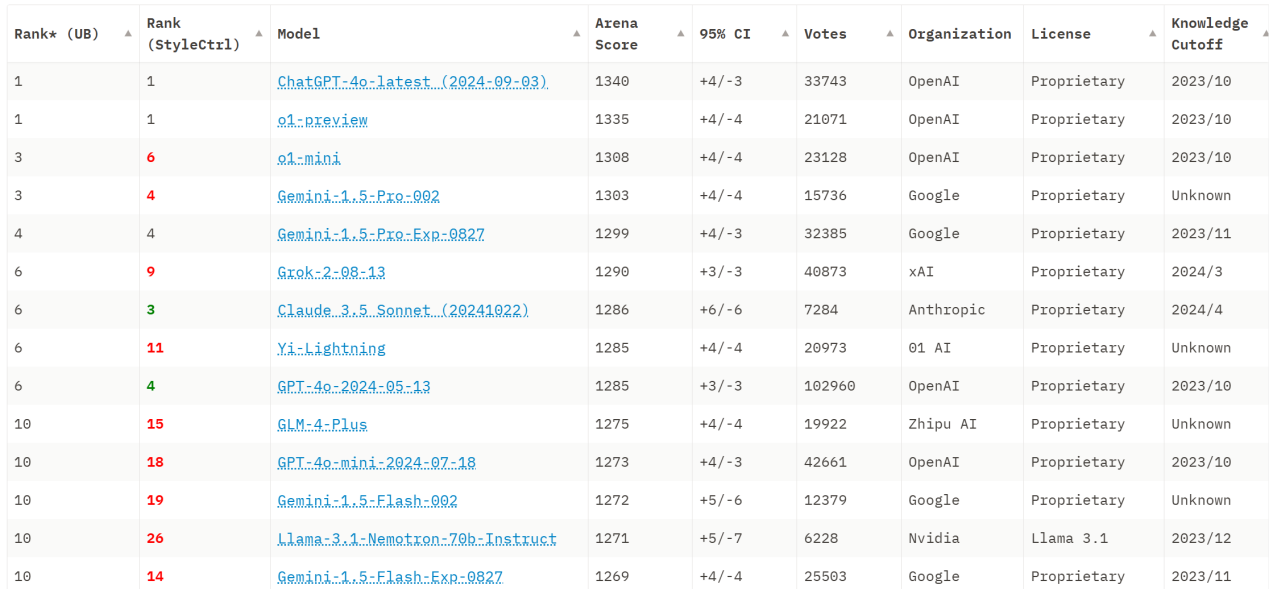
8. 百度的文心一言：文心大模型4.0在逻辑和记忆能力上显著提升，官方称综合能力媲美GPT-4。

9. 月之暗面Kimi：支持200万字无损上下文输入，具备较强的逻辑推理和翻译能力，表现优异。

这些模型在不同维度上持续创新，为大模型领域的发展带来了新的突破。

1. 大模型排行（LMSYS Chatbot Arena Leaderboard）

LMSYS Chatbot Arena 是一个众包的开放平台，专门用于评估大型语言模型（LLM）。用户可以向两个匿名 AI 聊天机器人提出任何问题，然后投票给最佳答案，目前已经收集了超过 100 万次人类对模型的两两比较，并通过 Bradley-Terry 模型对LLM进行排名，最终将评分以 Elo 等级显示。通过这套评测系统给出的排行，更能代表用户的实际体验。



1. 研究内容
2. 调研方法
   1. 大语言模型调研

好的PPT绝不是只有华丽的外表，内容充实有趣的PPT才是人们所希望看到的。所以在调研时，我们选择先调研主流大模型的文本分析能力。

基于《链接》（艾伯特-拉斯洛·巴拉巴西）的第六链的内容，对国内外主流大模型进行多维度测试。

* + 1. 文本阅读能力对比

列出主流大模型支持的文件格式（如TXT、PDF、DOCX等），并分析不同格式对读取效果的影响。研究每个模型对输入文件大小的限制，并测试在不同大小文件上的表现（如短文与长文）。测试模型对输入文本的准确读取率，包括是否正确解析标点符号和格式。

* + 1. 文本摘要能力对比

比较模型生成的摘要长度（短摘要与长摘要）以及对关键信息的保留程度。

采用人工评估的方式对生成摘要的质量进行量化分析。测试同一文本在不同模型上的摘要差异，分析信息的多样性和重要性。

* + 1. 文本理解能力对比

基于文本内容设计多种类型的问题（如细节性、推理性、综合性问题）。记录各模型的回答正确率，并分析错误原因。测试模型在处理复杂问题（如需要推理或多步逻辑的问题）时的表现。

* + 1. 文本输出能力

比较模型支持的输出格式（如文本、表格、图像、代码等）。测试模型在不同上下文和主题下的输出变化。提供相同输入在不同模型下的输出样本，分析输出风格和结构的差异。

* + 1. 多知识点层次结构关联

测试模型从文本中提炼出知识点的准确性和完整性。要求模型生成知识点的层次结构（如思维导图、树状图等）。分析模型在关联不同知识点时的能力，如何展示它们之间的关系。

* + 1. 知识点扩写能力对比

设定扩写的标准（如字数增加50%或100%），比较模型在扩写时的信息保留情况。评估扩写内容的创造性和逻辑性，是否能有效扩展原有信息。比较模型对相同知识点的扩写结果，分析其多样性和深度。

* + 1. 图片生成能力对比

评估生成图片的质量和图文关联度。测试模型在不同主题下生成图片的能力。分析同一主题下不同模型生成的图片样式和内容差异。

* 1. AI PPT工具调研

这一部分调研专注于比较各个AI生成PPT工具的易用性，测试这些工具能否根据使用者的想法生成PPT。

* + 1. 输入输出方式对比

比较不同工具支持的输入方式（如文本输入、语音输入、文件导入等），评估其易用性和灵活性。测试生成PPT的输出格式（如PPTX、PDF等），以及兼容性和易分享性。

* + 1. 自定义功能对比

比较不同工具提供的主题和模板种类，评估其灵活性和美观度。测试用户在生成过程中能否自定义内容（如标题、段落、图片位置等）。分析工具是否允许用户调整字体、颜色、图形样式等，满足不同用户需求。

* + 1. 图片生成能力对比

比较各工具生成的图片质量，包括清晰度、色彩和主题相关性。测试工具在生成图片时的多样性（如插图、图表、照片等）。分析用户是否可以对生成的图片进行自定义修改（如裁剪、调整等）。

* + 1. 排版美观度对比

测试各工具的自动排版能力，评估生成的PPT在信息布局和视觉效果上的美观度。比较生成内容的可读性，包括字体大小、行间距、段落结构等。分析不同幻灯片之间的视觉一致性，是否保持整体风格统一。

* + 1. 再编辑功能对比

比较工具对已生成PPT的再编辑功能，是否方便用户修改内容、格式和布局。评估修改后的PPT能否方便地导出为不同格式，并保持格式一致性。

1. 关键技术

2.1 Transformer

2.1.1 位置编码

由单词嵌入和位置嵌入组成。单词嵌入如Word2Vec可将单词映射到向量空间，位置嵌入可通过特定公式计算，弥补模型处理序列数据时对单词顺序信息的缺失，在自注意力机制中起关键作用。

2.1.2 自注意力机制

包括单头和多头注意力。单头注意力通过计算查询、键、值得到注意力权重并加权求和；多头注意力并行计算多个自注意力头，捕捉不同子空间信息，提升模型表示能力。

2.1.3 编码器

由多个相同层堆叠而成，每层含自注意力和前馈神经网络，通过AddNorm层连接。AddNorm层用于残差连接和层归一化，前馈神经网络由两个线性变换和一个激活函数组成，可提取深入特征，提升序列数据建模能力。

2.1.4 解码器

每层通过AddNorm层连接，含掩码自注意力层确保生成下一词时仅依赖当前词及其之前的词，通过交叉注意力层与编码器输出交互，最后经前馈神经网络和AddNorm层输出。

2.2 GPT

2.2.1 架构特点

只利用Transformer架构中的解码器部分，专注于生成任务。解码器每层通过AddNorm层连接形成堆叠结构，其输入经掩码自注意力层、AddNorm层、前馈神经网络和AddNorm层输出。

2.2.2 原理

解码器的每一层通过 AddNorm 层连接，形成深层的堆叠结构。具体来说，输入 Y 首先经过掩码自注意力层，得到输出 YAttention，然后通过AddNorm 层得到 YAddNorm1。接着，YAddNorm1 经过前馈神经网络（FFN）和 AddNorm 层，形成该层的最终输出：







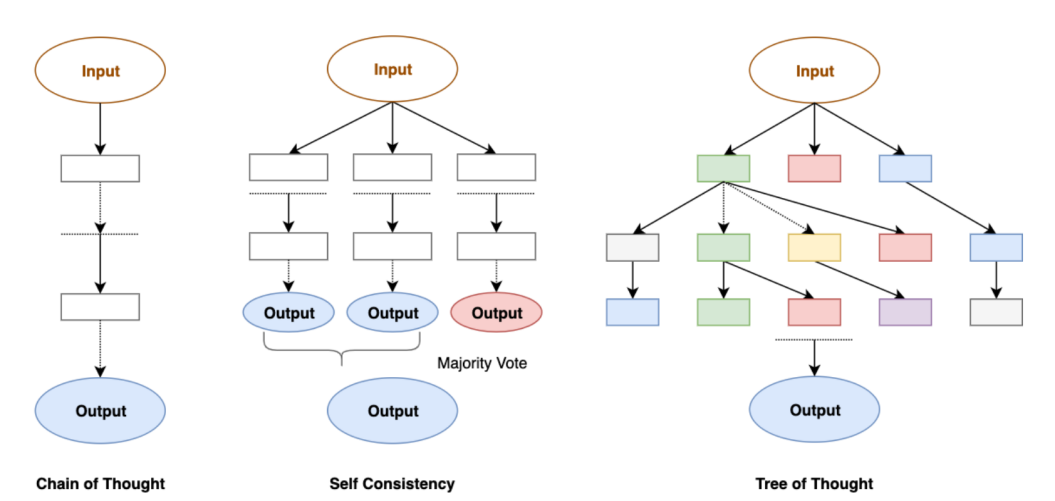
经过 N 层解码器的输入通过堆叠的方式形成最终的解码器输出。

我们发现，GPT 采用仅有解码器的架构，与传统 Transformer 的编码器结构在某些方面很相似，其解码器专注于自回归生成任务，强调自注意力机制，能够有效捕捉输入序列中词之间的依赖关系，生成更连贯的输出。与依赖交叉注意力的编码器不同，GPT 的解码器减少了对交叉注意力的依赖，提升了生成效率，同时保持上下文的一致性。这种设计不仅优化了文本生成过程，还体现了自注意力机制的核心优势。

2.3 Prompt工程

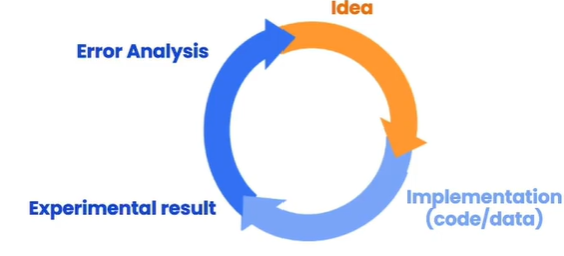
2.3.1 链式提示技术

链式提示技术（Chain-of-Thought Prompting）是一种用于大语言模型的提示优化方法，旨在通过引导模型进行分步推理和逐步思考，来生成更准确、合理的答案。与传统的直接输出结果的提示不同，链式提示通过让模型逐步解释自己的推理过程，从而增强其逻辑能力和解决复杂任务的能力。链式提示技术的原理图如下：



2.3.2 提示工程迭代技术

大语言模型提示工程迭代技术旨在通过不断优化提示词的设计和结构，提高模型的生成效果和质量。其原理图如下：



提示工程经历了多个迭代阶段，从最初的固定模板设计到引入上下文信息与示例引导，再到动态调整和自适应个性化，显著提升了生成质量和用户体验。这一过程使模型更有效地理解用户意图，能够更灵活地适应不同的应用场景。一个提示工程的迭代一般包括四个步骤：1.初始提示的设计；2.生成结果评估；3.识别问题（通过2找到生成内容中的不足之处）；4.提示优化。不断重复四个步骤，对提示词进行不断地优化最终达到理想的效果。

1. 技术路线

3.1 内容规划

首先根据输入的主题和要求，利用Transformer架构强大的建模能力，规划出课件的大致结构。调用现有的大模型，通过其Transformer编码器对输入序列的特征进行提取，通过多层的自注意力和前馈神经网络处理，确定课件内容的框架。

3.2 大纲生成

在确定课件结构后，需要为每一页幻灯片生成标题。这可以通过解码器部分的掩码自注意力机制来实现。掩码机制确保在生成当前幻灯片标题时，只依赖于之前规划的结构信息和已生成的标题信息，保证标题的逻辑性和连贯性。

3.3 文本生成

对于每一页幻灯片的详细内容，再次利用解码器的能力。根据幻灯片标题和之前规划的上下文信息，生成详细的讲解文本。

3.4 图表生成

一些先进的大模型可能会尝试根据生成的文本内容，自动生成相关的图表或示例。这可能需要额外的模块或算法，例如采用Diffusion模型基于文本内容生成与之相关的图片。从Transformer架构的角度来看，其生成的文本内容可以为图表和示例的生成提供指导。

3.5 布局规划

在生成PPT的整体流程中，后续可以结合其他算法或规则来规划PPT的布局。例如，根据幻灯片内容的重要性和逻辑关系，确定文字、图表的位置和大小比例等，使课件更加美观。

1. 实施方案

4.1 方案设计

通过对现有AI课件工具生成PPT的原理研究发现，其通常是先对输入文本进行理解以生成大纲，再基于大纲生成内容。然而，经过调研对比，这些工具在大纲内容总结能力方面不如常规大模型。因此，本方案采用先利用常规大模型生成高质量大纲，再将大纲输入AI课件工具生成PPT的方法，以期望达到更好的效果。

4.2 生成大纲

4.2.1 明确输入文本

确定需要生成PPT的主题相关文本内容。收集相关的介绍性文本、应用案例、发展趋势等资料，并整理成一个较为详细的输入文本。

4.2.2 设计prompt

根据输入文本和期望生成的大纲方向，设计合适的prompt，采用prompt工程的核心思想，例如迭代、给出示例等方法，逐步引导大模型输出最为合理的结果。

4.2.3 调用大模型

选择合适的、总结和生成能力较强的大模型，如GPT系列或其他性能良好的语言模型。将设计好的prompt和输入文本按照模型要求的格式输入模型，获取生成的大纲内容。

4.3生成PPT

4.3.1 选择工具

根据实际需求和工具的功能特点，选择一款适合的AI课件工具。一些工具可能在图表生成方面有优势，而另一些可能在文本排版上表现更好。

4.3.2 输入大纲

将由大模型生成的大纲内容准确无误地输入到选定的AI课件工具中。确保大纲的格式和内容符合工具的要求，避免出现错误导致生成不理想的PPT。

4.3.3 生成PPT

启动AI课件工具的生成功能，让其根据输入的大纲生成PPT。在生成过程中，工具会基于自身的算法和模型，对大纲中的每个项目进行内容填充，包括生成文本、可能的图表等。

4.4 对PPT内容进行修改

4.4.1 内容审核

对生成的PPT内容进行仔细审核。检查文本内容是否准确、完整地表达了大纲中的主题思想，是否存在逻辑错误或信息遗漏、重复，是否出现大模型的幻觉问题。

4.4.2 图表检查

如果PPT中包含图表，检查图表是否清晰、准确地反映了文本内容中的相关信息。检查图表的坐标轴标注是否正确，图表的信息是否真实可靠，数据是否与文本描述相符。

4.4.3 排版调整

根据PPT的整体美观度和可读性，对内容的排版进行调整。可以调整文字的大小、颜色和字体，合理安排图表和文本的位置，确保页面布局平衡、美观。

4.4.4 语言优化

对PPT中的文本语言进行优化，使其更加通顺、简洁、专业。避免使用过于口语化或复杂难懂的语句，提高PPT的质量和专业性。

1. 研究方法
2. 明确目标

通过问卷或访谈收集用户对大模型和AI PPT工具的使用体验、反馈和需求。明确研究目标。

1. 分析当前大模型与AI生成PPT工具的问题所在

查阅相关研究和报道，通过实际调研，了解大模型和AI PPT工具的现状、优势和不足，识别常见的问题并分析其成因，例如生成内容的准确性、可用性和用户友好性。

1. 提出解决方案

根据前一步的分析，明确具体问题（如生成内容质量低、操作复杂等）。提出针对性的解决方案，例如改进算法、增加自定义功能等。同时要考虑技术实现的可行性和所需资源，确保方案在技术上可行。

1. 评估并调整方案

邀请相关专业人员对提出的方案进行评估，获取专业意见。在小范围内实施方案，收集用户反馈，观察效果。通过数据分析评估方案的有效性，根据结果调整方案细节。

1. 实际开发

根据调整后的方案，开始实际开发或实施工作。在实施过程中定期检查进展，确保按计划进行，并及时解决出现的问题。详细记录实施过程中的关键步骤、决策和变化，为后续总结提供依据。

1. 反思总结

总结实施后的效果，与最初设定的目标进行对比。分析在研究和实施过程中遇到的问题及解决方案，提炼经验教训。提出未来改进的方向和可能的研究课题，确保研究成果的持续发展。

1. 工作计划及进度安排

第9周：完成开题报告

第10周：在之前调研结果的基础上，进一步优化生成ppt的方法，将制作的流程整理出来，形成一套完整的、可操作性强的ppt自动生成指南，并制作《链接》一书各章节的ppt。其中应该包括使用的大语言模型(应为国内且可免费使用)，已经成熟的prompt，对大模型生成文本的修改方式(要求适用范围广且清楚)，AI生成ppt工具，对生成ppt的修改方式(要求同文本修改)。

第11周：完成中期review，对过去成果进行总结讨论，同时思考之后的具体的工作内容。

第12-15周：基于整理的ppt自动生成指南，利用大模型API、脚本等工具，应用github上的开源项目，不断迭代完善，实现书籍生成ppt全过程自动化。

第16周：完成结题报告。

第17周：进行成果展示。

1. 预期成果

科研课堂最终的成果包括：

1．一套完整的、可操作性强的ppt自动生成指南。

2．对生成的指南进行全过程自动化。

3．制作过程中的调研材料，成果展示ppt等过程性材料。

1. 参考资料
2. Vaswani A. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017.
3. Kaplan J, McCandlish S, Henighan T, et al. Scaling laws for neural language models[J]. arXiv preprint arXiv:2001.08361, 2020.
4. Hoffmann J, Borgeaud S, Mensch A, et al. Training compute-optimal large language models[J]. arXiv preprint arXiv:2203.15556, 2022.
5. Tyl, T. (2023, July 23). Chinchilla’s Death. Espadrine blog.
6. Touvron, H., et al. (2023). LLaMA: Open and efficient foundation language models. Meta AI. Correspondence: {htouvron, thibautlav, gizacard, egrave, glample}@meta.com.