大模型最新研究和技术探讨

# 1. 总结

大语言模型（Large Language Model，简称LLM） ，是一类基于深度学习的参数量极其庞大的模型，具备较强的文本理解与生成能力。这类模型主要由多层Transformer 神经网络架构组成，并可进一步扩展至多模态任务，如代码生成、图像分析、语音识别和视频处理等。通过近年来的研究，LLM已逐渐形成了一套完善的训练范式，即“预训练–微调–人类偏好对齐”（Pretrain-Finetune-Alignment），并在多种下游任务中展现出接近或超越人类平均水平的性能。这种逐步提升的能力使得大模型成为人工智能（AI）领域中极具潜力的技术之一。

对于下一阶段大模型的研究和技术发展，可能从数据、训练、架构和应用四个关键维度展开。

**数据层面**：当前的大语言模型主要通过海量数据的编码表示，结合深度神经网络进行训练。与传统机器学习范式不同，LLM更加强调上下文间的语义一致性和跨文档的关联性学习，而不强依赖于事先明确定义的数据结构。这种“大一统”的学习方式虽然效果显著，但仍然存在一些潜在问题：一是**数据编码的离散性与信息损失** ：现有大模型使用离散化的方式表达数据，这种表示方法虽然训练效率较高，但在信息层面可能导致部分细节丢失。未来研究中，采用连续化建模（如基于函数分布的数值表达或几何表示）或许能更好地捕捉数据的深层信息，从而提升模型性能和泛化能力。二是**动态探索式学习的潜力** ：尽管上下文学习在现阶段占据主导地位，但它并非唯一有效的数据学习机制。在下一步研究中，可以结合强化学习的方法（如基于强化博弈的探索策略），实现动态的、面向任务的“主动学习”。这样不仅能够进一步提升模型对复杂任务的适应能力，还能使其从现有数据中挖掘更多潜在信息。

**训练层面**：当前大模型的训练范式通常经过全面优化，其核心目标是通过大规模数据的多轮迭代，逐步提升模型的表现能力，同时避免过拟合和偏置。然而，这套训练方法也暴露出以下问题：一是**数据存量的限制** ：训练大模型所需的数据无疑是海量的，而在全球范围内优质数据（经过清理和去重）资源是有限的。一旦现有数据资源被充分利用甚至“耗尽”，训练可能趋于瓶颈。对此，可以探索在线数据生成、数据扩充和数据增强技术，例如利用生成对抗网络（GAN）或自监督预训练生成高质量样本，持续支撑大模型的训练需求。二是**训练范式的灵活性** ：当前大部分大模型的预训练仅依赖于无监督任务，例如通过自回归语言建模（autoregressive language modeling）进行目标优化。但这种单一过程并不灵活，无法在训练初期有效结合其他机制（如对抗训练或博弈论）。未来，一种方向是将强化学习、对抗学习等技术融入训练全流程，从而在预训练中实现效率提升，甚至弱化“人类偏好对齐”对微调阶段的依赖。这不仅能够降低对人工干预的需求，还可以提升模型的自主学习能力。

**架构层面**：目前大语言模型的主流架构以多层Transformer为核心 ，其依赖于注意力机制实现高效的自回归学习。然而，这种架构仍然有改进空间，尤其是与人类大脑构造和工作方式的类比还十分有限。因此，未来研究重点可能集中在以下两方面：一是**基础架构的创新** ：Transformer架构虽然强大，但并不一定是实现通用智能的最优解。对于类脑科学的进一步研究可能启发更抽象、更泛化的神经网络架构。例如，灵感来自脑神经回路的“动态网络”或基于概率理论的计算架构，或许能够实现更高效的推理和决策能力。二是**多模态信息处理能力的提升** ：当前的大语言模型正逐渐向多模态统一模型靠拢，理想状态下，一个模型应具备处理多模态输入（语言、图像、视频等）的能力，并能够在理解后实现多模态输出（如语言生成与图像生成的联动）。然而，这种全场景、端到端的多模态学习机制仍然面临诸多挑战，例如模态间的信息对齐与共享以及计算效率问题。

**应用层面**：在实际应用中，大语言模型作为基础模型（Base Model）的能力主要通过检索增强生成（RAG）和智能体（Agent）的方式落地，广泛应用于教育、艺术创作、代码生成、数据分析等领域。然而，要充分发挥大模型的潜力，其应用设计需回答以下三个核心问题：一是**目标用户是谁**？应用是否能够清晰定义目标人群，以制定特定场景的模型需求？二是**是否具备使用必要性**？该场景是否实际需要大模型的基础能力，是否存在可替代或更经济的解决方案？三是**是否实现增益**？应用场景与大模型的结合是否能显著提升效率或性能，进而创造更大的价值？只有能够清楚地回答这三个问题的应用，才更有可能在实际市场中取得成功。然而，目前许多基于大模型的应用尚未解决上述问题，其设计更多是为了展示技术能力，而非应对明确的市场需求。

总体而言，大模型是机器学习和深度学习技术发展的里程碑式成果。不仅如此，它还代表了人类在探索通用人工智能（Artificial General Intelligence，简称AGI）过程中迈出的一大步。然而，大模型是否成为实现AGI的必经之路，或其表现是否足以作为通向强人工智能（Strong AI）甚至超人工智能（Superintelligence）的桥梁，这些问题尚需时间去进一步验证。未来，随着学术界与工业界对大模型技术的持续探索，更先进、更高效的智能系统必将涌现。

机器智能的曙光终将驱散未知的迷雾，而人类不断创新的脚步，亦将在这一领域书写新的篇章！

# 2. 大纲

涵盖了大模型训练的优化策略、数据处理的新趋势、多模态生成模型的创新、与人类思维一致性的探讨，以及在教育等实际应用中的前沿进展。这些探讨为与会者提供了丰富的科研启示和实践思路，有助于推动AI领域的进一步创新与发展。如下是具体内容：

1. 数据处理与特征挖掘：数据的处理方法迭代、tokenizer的存废
   1. 数据特征挖掘发展过程：
      1. 单独的数据，是点：线性——非线性——概率图
      2. 数据关联，加边：CNN、RNN
      3. 数据拓扑，加图：GNN
      4. 未来信息更本质的关联如何构建，或者说下一个信息增量是什么，导致熵降低
   2. Tokenizer的存废
      1. 今天的范式是离散化
      2. 未来的范式感觉可能是级联的离散化，比如层级式的
      3. 完全去除离散化，其指数的计算空间是目前无法处理的
      4. 理论上我们对数据的处理不希望只用具体的值，而是用分布，分布是更加连续的表示
2. 模型训练与优化：多模态场景的增量训练、预测问题的短期策略、预训练范式与RL结合
   1. 多模态场景的增量训练：目前中心是避免过拟合，通过混合数据配平防止其他能力掉落
   2. 预测问题：
      1. 只需要过拟合一个临时的模型，然后临时性使用
      2. 因为更长的预测是无法达到的，当你的技术突破后，总有人会很快跟上的，对抗
      3. 本质上是一个熵减的过程，数据会被越来越用到，用到后就无效了
      4. 扩展出一个研究点：可以做一个auto LM，每天都会基于过去快速变化，然后模型只在当天有效
   3. 预训练范式：
      1. 当前的范式相较固定，比如数据配比
      2. 但也有可以扩展的点：比如把RL博弈论的相关内容放到Pre-train中——把合成数据用于预训练中
      3. 类似于人类一样：
         1. 任务导向性的，找到相关的数据后，筛选出最核心的数据用于任务
   4. 模型的可解释性
      1. 传统的方式是通过可视化的方式
      2. 考虑大模型和小模型是否需要协作
         1. 考虑到大模型本身的信息密度
         2. 考虑通用大模型的天花板
3. 模型架构：人类一致性、以及架构的变化
   1. 大模型与人类思维和行为的一致性和相互模仿
      1. 张钹院士提到脑科学有两种：一种是仿生，一种是类脑计算
         1. 仿生的本质是为了理解人类自己
         2. 类脑计算是希望实现类似于脑的计算方式
      2. 未来对意识的探索
         1. 一种可能的途径是：找到最简单的组成部分，然后搭建起来复杂的系统
   2. Transformer等底层架构的提出方式
      1. 是一个自然演化的产物
      2. 从数据的关联，到把数据压缩到一个固定的向量，再到优化数据到信息的表示，比如注意力机制
      3. 计算机学家的出发方式通常是找到一种更容易计算的方法
   3. 多模态模型的scaling law
      1. 计算量、数据量和参数量的tradeoff
      2. CV之前的模型主要在CNN，后来又产生了Stable diffusion等，但仍然没有统一
      3. 多模态的理解和生成不像文字一样可以统一
         1. 目前不能让视频生成太过跟随指令，容易出现坍塌现象，所以看起来不是特别跟随指令
         2. 正在尝试多模态的统一生成，端到端的一个模型
      4. 对齐工作其实没有做得很好，对齐工作的本质是标数据
4. 应用与未来方向：音乐生成、数字人、检索增强、智能体
   1. 音乐生成
      1. 复杂的语音，比如不止包括时间维度，而且在空间维度，可以参考图像生成这样的二维生成任务
      2. 与文字生成的离散生成不一样，是一个连续生成的场景，考虑类stable diffusion扩散模型
   2. 数字人和端到端的视频生成视频区别
      1. 数字人的可控性非常强
      2. 大部分更精细的生成效果都是通过比如进一步超分辨率生成实现的
   3. RAG检索增强生成与LLM的区别
      1. 直观感受是RAG无法实现自由度检索，所以会存在通过生成检索来优化RAG效果
      2. RAG的优势在于对于实时性场景的效果较好，以及对于未见知识的回应效果较好，而且适用于个人数据或者领域数据
   4. Agent
      1. 个性化教育 & 标准化教育
         1. 人类会有两种方式，一种是碰撞式交互，另一种是独立性思考
         2. 具体的功能点可能包括：prompt的自动生成、理解用户意图
         3. 解决实际问题
         4. 向下兼容服务，比如向更低层级的用户提供服务
         5. 问题在于特别个性化的工作的访问量比较少
         6. 引入激励机制，可以让用户更愿意使用，提供用户成就感 / 情绪价值
         7. 总得来说，需要陪伴、助手 和 记忆
         8. 卡尔加里大学和 Adobe研究院的论文提出了一种新可能：用AI系统将物理课本上的图表变成动画，演示物理原理的过程，从而让抽象的物理直观化
         9. 其中一种创业方式是做：学习可交互，比如开源的3b1b/manim，具体用于比如进行编程教学
      2. 全息通讯
      3. 数据分析相关的工作
         1. 关键在于谁会用，用与不用
         2. 考虑会不会被基座大模型秒杀

除去上述内容，讨论中也提到：

1. 研究上，算法创新和工程创新是不同的，算法创新不一定需要work，而是找到某种想法并尝试实现；工程创新更多是找到市场上比较成熟的工作，然后进行评估后使用。
2. 大部分的技术实现，最早应用于军事和交友
3. 产品侧的思维方式是：这个产品最终会有谁在用，以及用还是不用，以及是否会被基座大模型之后融合

# 3. 过程

本章内容主要由AI生成。

## 3.1 详述

讨论了信息增量、脑科学、多模态、数字人等多个领域的相关问题及应用方向。主要内容包括：

1. 信息增量与模型预测
   1. 信息增量挖掘：挖掘信息增量需关注数据本质和内在信息量，找到更能增大信息量的信息特征。
   2. 模型预测稳定性：当前模型预测存在稳定性问题，股票预测和量化交易中的预测可能是伪命题，建议构建能根据当日数据快速预测的模型。
   3. 预测范式变化：未来模型预测范式可能在信息增量、速度提升、训练机制等方面发生变化，如通过博弈方法、生成大量数据、模仿人的学习过程等。
2. 脑科学与计算机以及音乐生成
   1. 脑科学与计算的关系：探讨了仿生和借鉴人脑机制用于计算，仿生本质是理解人类，而借鉴人脑特性用于计算更具可行性，如参考简单的计算机理。
   2. 模型发展的逻辑：以机器学习模型的发展为例，其并非先参考仿生思路，而是从计算角度出发，为解决问题不断改进，如为减少信息损失出现了 encode decode 模式。
   3. 音乐生成的难题：音乐生成存在提取二维信息的困难，如不同声部、乐器和人的泛音差异，可尝试借鉴图像处理方式。
   4. 视觉模型的发展：视觉模型发展迅速，数据量和参数量对其重要，Transformer 在商用版出现后，视觉领域的追赶较快，但仍存在理解和生成结合等问题。
3. 多模态的对齐、离散化及应用场景
   1. 多模态信息对齐：多模态信息处理的首要任务是对齐，只有对齐才能有实用性，且大量优质对齐数据能使效果更好。
   2. Token 离散化问题：如何去掉 token 离散化，未来可能以像素进像素出的方式更好，但当前计算量无法实现，级联思路或许更可行。
   3. 多模态应用场景：多模态在视频中的应用场景需进一步思考，如文本到视频的转换，要考虑成本、安全审查等问题，当前存在指令跟随效果不佳等情况。
4. 数字人、语言模型及搜索
   1. 数字人控制方式：数字人通过控制坐标等数值来实现各种功能，数据收集的清晰度决定其建模精细程度，相比从文字转视频生成控制更简单。
   2. 语言模型知识处理：语言模型对于知识的压缩和记忆是基础能力，其发展可能改变传统知识处理框架和模型交互方式。
   3. 网页搜索发展方向：网页搜索从浅层向深层发展，涉及多来源数据分析和推理，RAG 在时效性问题处理上对实际用户体验有提升。
5. 模型应用与教育个性化
   1. 企业模型应用局限：企业使用模型提供数据和搭建框架，但内部不对端到端效果负责，导致使用较浅，无法深度定制任务。
   2. 模型新兴招聘形式：模型之间以新兴形式进行对话，可能用于招聘，将不同模型的知识和 IP 故事整合。
   3. 模型检索优化方式：提出让模型更多参与搜索过程，自行判断搜索结果，为其提供更多环境和上下文，而非依赖传统算法。
   4. 教育个性化趋势：未来教育会走向个性化，效果可能比传统教育更好，如为每个学生配备助教。
6. 在线教育中 AI 应用
   1. 早期上线与阶段规划：2012 年上线学堂在线，2013 年计划将AI做进去。
   2. 助手作用的思考：探讨了每个人需要的助手在在线教育中的作用。
   3. 刷题与作业问题：学校系统内考前刷题及作业完成方式。
   4. AI 校验与成本：AI 校验的需求及成本与用户使用的关系。
   5. 不同学段的需求：小学生、初中生等不同学段对教育方式和知识呈现的不同需求。
   6. 个性化与知识水平：强调根据不同知识水平提供个性化教育的重要性及实现方式。
   7. AI 应用的难点：包括 AI 输出内容的准确性和难以准确表达需求等。
   8. 教育中的动机问题：探学生在教育过程中的学习动机和可能的解决方式。
7. 知识交互项目
   1. 知识可交互项目目标：旨在将理工科知识以可交互形式呈现，促进知识传授，降低对老师的高要求，促进知识普及。
   2. 项目初步尝试：以编程为首个尝试方向。
   3. 项目面临的问题：成本高，如用 Unity 引擎做 40 分钟游戏需 6 个人/月；生成代码有难度。
   4. 未来知识传授设想：未来知识不应只是文字和图，应是可交互的，老师更多做组织和讨论工作。
8. 编程、全息通信和大模型
   1. 编程教育年龄：小孩十来岁左右开始学编程较为常见，且竞赛优化并非必要，重要的是培养思维和提高能力。
   2. 游戏类编程：大部分游戏类编程旨在吸引小孩玩耍，粘性较小。
   3. 全息通信效果：当前成像方面在特定场景效果不错，但在通用场景和功能上效果不佳，如复杂场景中的视频投放和会议场景中的人物抠像。
   4. 大模型可解释性：大模型准确性和可解释性的问题，从数需要据侧和参数侧入手，研究数据对模型学习的影响，以提高模型的可解释性和性能。
9. 大模型数据分析
   1. 模型通用型困惑：面临专用数据分析模型种类多、结构和边界不同，通用模型效果存疑，领域数据建立不佳。
   2. 数据计算成本高：如用数据模拟器工作，绩效要求巨大，计算成本很高。
   3. 数据对齐的复杂性：要对齐参数和结果等，过程复杂，需尽可能记住每一步。
   4. 用户需求的关键：先明确谁需要数据分析，用何种方式打动用户，避免功能被超越。
   5. 数据分析的场景：平台上有数据分析智能体，每日有几万使用量，说明存在应用场景。

## 3.2 时间流

[00:00](https://zhipu-ai.feishu.cn/minutes/obcnlu1q5du7jyoidvcyrsh4?t=0) 关于信息增量、模型预设及未来训练机制变化的探讨

主要讨论信息增量相关问题，其需要新数据，回顾人工智能研究历史中对数据特征、结构等的探索及重复过程，还谈及模型预设存在的问题，如稳定性等，提到可做auto ML等，以及对训练框架、算法、机制变化等的猜测和思考。

[13:07](https://zhipu-ai.feishu.cn/minutes/obcnlu1q5du7jyoidvcyrsh4?t=787000) 关于脑科学、计算逻辑及音乐生成等问题的讨论

强调讨论应自由发表观点，无严谨标准答案。谈及脑科学与计算，仿生本质是理解人类，借鉴人脑特性有选择性。还讨论了计算模型发展逻辑，如加边、编解码等，以及音乐生成的二维信息提取难题、视觉数据和模型情况、理解与深层结合等问题。

[29:44](https://zhipu-ai.feishu.cn/minutes/obcnlu1q5du7jyoidvcyrsh4?t=1784000) 关于多模态对齐、离散化及理解生成gap等问题的讨论

主要讨论了多模态相关内容，包括文本对齐最好有相关图片数据，提到token离散化问题及去除离散化可能是研究新点，探讨级联思路，还谈及多模态生成和理解存在gap、指令跟随存在问题以及视频端到端应用场景和成本等相关话题。

[42:00](https://zhipu-ai.feishu.cn/minutes/obcnlu1q5du7jyoidvcyrsh4?t=2520000) 关于视频内容、数字人及RAG相关问题的讨论

主要讨论了多方面内容，包括视频内容变化及安全审查相关需求，数字人的特点及视频生成问题，还谈及RAG相关，从传统检索到深层搜索的发展方向、 RAG的作用及存在问题，以及知识获取和数据整理等任务。

[53:46](https://zhipu-ai.feishu.cn/minutes/obcnlu1q5du7jyoidvcyrsh4?t=3226000) 关于模型优化、行业支持及教育个性化的讨论

主要讨论了相关业务问题，涉及成交互优化、模型间对话与整合、企业数据服务、RAG需求、搜索结果判断、模型参数更新、教育领域应用如个性化教育及为学生配备助教等内容，还对一些相关实践及可行性进行了思考。

[01:05:36](https://zhipu-ai.feishu.cn/minutes/obcnlu1q5du7jyoidvcyrsh4?t=3936000) 关于教育与AI应用的思考及讨论

主要讨论了教育相关话题，涉及课程上线情况、助教作用、AI在教育中的应用如英文写作提升等，还谈及AI个性化问题，包括针对不同年龄段学生提供合适内容，以及提到AI知识库、虚拟教师等相关设想和教育中可能遇到的问题。

[01:17:38](https://zhipu-ai.feishu.cn/minutes/obcnlu1q5du7jyoidvcyrsh4?t=4658000) 关于AI陪伴、助手、记忆机制及教育应用的讨论

主要讨论了AI相关话题，提及核心问题包括陪伴、助手以及记忆机制，还谈到新技术首先在军事上突破，涉及教育付费、从基础模型到助手的布置等，此外还以物理、数学学习为例探讨了可视化功能的可能性及工程实现相关。

[01:25:41](https://zhipu-ai.feishu.cn/minutes/obcnlu1q5du7jyoidvcyrsh4?t=5141000) 关于项目知识可交互、模型泛化及数据分析的讨论

主要讨论了知识可交互项目，尝试以可交互方式展现理工科知识，如用编程教小孩。提到项目面临成本高、泛化难等问题，还探讨了大模型可解释性、数据质量等，以及燃烧器模型通用性、数据分析应用等相关困惑与思考。

## 3.3 录音

<https://zhipu-ai.feishu.cn/minutes/obcnlu1q5du7jyoidvcyrsh4>