

基于对 Transformer、BERT 和 GPT-3 三篇文献的深度研读，以下为您整理的标准化的“文献综述矩阵”及“演进趋势小结”。

任务 1：文献综述矩阵

文献标题	发表年份	解决的核心痛点 (用一句话概括)	关键技术/创新点	局限性/未来方向
Attention Is All You Need Transformer	2017	解决了 RNN/CNN 在长距离依赖建模上的劣势，以及无法并行训练导致的效率瓶颈。	纯自注意力机制、多头注意力、编码器-解码器架构、位置编码。	计算复杂度随序列长度呈二次方增长 $SO(n^2)$ ，处理超长序列受限；未探索大规模无监督预训练范式。
BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding	2019	突破了单向语言模型（如 GPT）无法融合双向上下文的限制，提升了语言理解任务的性能。	深层 Transformer 编码器、掩码语言模型、下一句预测（NSP）、统一的“预训练+微调”架构。	掩码符号 [MASK] 仅在预训练阶段出现，导致与微调时的输入不匹配；生成能力弱于自回归模型；预训练成本高昂。

文献标题	发表年份	解决的核心痛点 (用一句话概括)	关键技术/创新点	局限性/未来方向
Language Models are Few-Shot Learners GPT - 3	2020	解决了传统微调范式需要针对每个任务收集大量标注数据，且模型难以适应新任务的泛化瓶颈。	1750亿参数超大规模模型、上下文学习、零样本/单样本/少样本学习、稀疏注意力机制。	训练与推理资源消耗极大（碳足迹问题）；存在明显的偏见、公平性及虚假信息生成风险；缺乏实体世界的多模态交互。

任务 2：演进趋势段落

2017年Transformer架构的提出，摆脱了RNN的序列计算限制，确立了自注意力机制的地位。2019年BERT通过双向掩码预训练，证明了深层双向表示对语言理解任务至关重要，将“预训练+微调”范式推向高潮。2020年GPT-3则进一步探索了超大规模模型的缩放定律，从依赖大规模标注数据的微调转向了仅需少量示例的上下文学习，展现了通用大模型的适应能力。这一演进表明NLP的核心驱动力正从算法架构创新，转向对模型规模与无监督数据利用的极致探索。

注:本文部分内容由AI生成,无法确保真实准确,仅供参考