

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
<b>Attention Is All You Need</b> 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 $\$O(n^2 \$)$ ，处 理超长 序列受 限；未 探索大 规模无 监督预 训练范 式。
<b>BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding</b>	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
<b>Language Models are Few-Shot Learners</b> 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