一、论文理论分析分工 （一周）

模块	成员A任务（理论侧重）	成员B任务（实践侧重）	协作节点
1. 背景与动机	- 撰写NLP预训练技术发展脉络（Word2Vec→ELMo→GPT） - 整理GPT在2018年的技术突破点	- 统计论文引用的关键文献（图表化引用网络） - 对比GPT与同期模型（BERT等）的差异	共同确认技术定位准确性
2. 模型架构	- 推导Transformer数学公式（式2.1-2.4） - 绘制模型结构图（Figure 1复现）	- 验证位置编码有效性（正弦vs学习式） - 分析多头注意力参数占比（12头 vs 其他配置）	交叉检查公式推导正确性
3. 训练策略	- 解释无监督预训练目标函数（式2.1） - 证明最大似然估计的收敛性	- 复现学习率预热曲线（Figure 3） - 统计BooksCorpus数据分布（词频/长度）	联合设计消融实验方案
4. 实验结果	- 分析各任务AUC提升原因（表1） - 解释zero-shot性能（图2右）	- 复现主要实验结果（表1关键指标） - 可视化注意力模式（如Figure 4）	共同撰写实验分析段落
5. 局限与展望	- 总结模型计算效率缺陷 - 提出理论改进方向（稀疏注意力等）	- 测试模型在长文本的泛化性 - 设计知识蒸馏实验方案（加分项）	联合制定未来研究路线图

二、代码复现分工 （两周）

阶段	成员A任务（框架搭建）	成员B任务（实验优化）	协作产出
1. 数据处理	- 实现BPE分词器（原版编码） - 构建文本滑动窗口（context=512）	- 预处理BooksCorpus子集（10%数据） - 生成TFRecord格式训练文件	可复用的数据管道
2. 模型核心	- 编写Transformer Block（含LayerNorm） - 实现位置编码（正弦函数）	- 添加混合精度训练支持（AMP） - 实现梯度累积（应对显存限制）	模块化模型代码库
3. 训练循环	- 搭建语言模型损失函数 - 配置Adam优化器（β1=0.9, β2=0.999）	- 实现学习率调度器（带预热） - 添加wandb训练监控	可运行的训练脚本
4. 微调实验	- 适配文本分类任务（如CoLA） - 修改输入格式（[start]+text+[extract]）	- 对比不同微调策略（全参数/顶层微调） - 测试batch size对准确率的影响	微调性能对比报告
5. 优化扩展	- 实现稀疏注意力（Blockwise） - 集成Flash Attention加速	- 知识蒸馏：用12层模型指导6层训练 - 量化模型（FP16→INT8）	改进方案代码与benchmark

三、文档撰写分工 （两周）

章节	成员A负责内容	成员B负责内容	整合方式
1. 引言	- NLP预训练发展背景 - GPT的核心贡献陈述	- 同期模型对比表格 - 论文技术影响力分析（引用数/应用场景）	A起草文本，B补充数据
2. 方法	- 模型架构数学推导 - 训练目标函数解释	- 代码结构图（类图+数据流） - 工程实现细节（如并行化策略）	交叉验证公式与代码对应性
3. 实验	- 主实验结果分析（表1） - 消融实验讨论（层数/头数影响）	- 复现结果对比图表 - 训练资源消耗统计（显存/时间）	共同确保数据一致性
4. 讨论	- 理论局限分析（单向注意力缺陷） - 未来研究方向建议	- 实际部署挑战（模型压缩需求） - 改进方案实验效果	合并观点后逻辑串联
5. 附录	- 完整公式推导过程 - 超参数配置表	- 代码仓库使用说明 - 复现环境依赖文件	分别提交后合并