

定制化沟通记录表

AI 草稿待人工复核：当前内容不得作为正式结论或执行依据。

户号	1-录音测试	面积	待补充	风格	待补充
时间	2026年7月28日	地点	未记录	次数	第1次
设计	未记录				
管家	未记录				
营销	未记录				
总述	会议主题与重点：本次会议为录音测试沟通，主要探讨下一代模型参数量级（150B-250B）的对标对象，以及推理层编译语言从CUDA生态向Triton等高级语言迁移对效率的影响。				
颜色说明	● 已明确/已确认	● 方向性要求/建议	● 待确认/责任待定	● 风险或重要限制	● 报价、增项或费用相关 ● 已取消事项
会议主题	1. 本次会议为录音测试沟通，主要探讨下一代模型参数量级（150B-250B）的对标对象，以及推理层编译语言从CUDA生态向Triton等高级语言迁移对效率的影响。 发言：多人对话				
其他核心观点 (角色待确认)	1. 采用Triton等高级编译语言替代PTX等底层语言，不会导致推理效率降低或产生短期负面影响，反而能大幅提高效率。这被视为摆脱CUDA生态依赖的技术机遇，且代码编写更简洁快速。 发言：多人对话 2. 即使目前Triton代码主要由人工编写，其开发效率已显著高于原有的C/CUDA代码。结合AI辅助编程，底层硬件执行效率不受负面影响，整体属于技术发展的重大机会而非需弥补的缺点。 发言：多人对话				
行动项	1. 本次会议暂无明确行动项。				
待确认事项	1. 本次会议暂无待确认事项。				

正式转写证据 2 段 · 发言角色 多人对话 · 生成时间 2026-07-28 07:43:41 · 版本 V1