多模态大模型原理与实践提纲

一、 培训需要解决的问题

移动云盘紧跟前沿技术用 AI 全面重塑云盘“采传存处用”场景，探索对文本、图片、笔记、视频、音频等数字内容的智能化赋能。通过本次培训，拓展对多模态 AI 应用领域的视野，帮助团队聚焦 AI 赋能创新突破，提升对资产管理 AI 服务的技术认知与理解，更好地助力业务智能化业务建设。

二、 培训要求

已开展两期的大语言模型培训，在此基础上开展一期多模块方面结合大语言模型的通用生成类技能培训。

培训课程重点内容：① CLIP、SD；②结合中移的用户资产对“视频生成，音频生成和音频总结”部分可以进行前瞻性的技术分析和研讨；③希望结合公司业务来讲解。

基础知识部分可不讲或略讲，文生视频技术目前不太成熟，只略讲概念部分

三、 培训时长

1 天

四、 培训提纲

第 1 部分：多模态学习概述

多模态学习的定义

多模态学习的意义

多模态数据类型：文本、图像、视频、音频等

多模态学习的应用领域（自然语言处理、计算机视觉、推荐系统等）

第 2 部分：ViT、Beit 与 CLIP/BLIP

ViT 模型架构概述

Patch Embedding 与 Positional Encoding

Beit 与 ViT 的比较

Beit 在自监督学习中的应用

Beit 在多模态任务中的优势

实践演示：利用 ViT 和 Beit 进行图文转化的效果

CLIP 模型介绍：从图像到文本的跨模态嵌入

BLIP 模型架构：结合 CLIP 的多模态模型

CLIP/BLIP 在多模态任务中的应用：图像-文本匹配、图像标注等

实践演示：使用 CLIP 进行图像-文本匹配任务

第 3 部分：Stable Diffusion 及 SD XL

Stable Diffusion 模型概述：生成模型在图像生成中的应用

SD 的原理推导

SD 模型的架构

Stable Diffusion XL：扩展的 Stable Diffusion 模型

微调扩散模型：DreamBooth

微调扩散模型：Textual-Inversion

微调扩散模型：LoRA

微调扩散模型：Hypernetworks

Stable Diffusion 在艺术创作和设计中的应用

实践演示：使用 Stable Diffusion 生成图像

第 4 部分：微调与 RLHF 方法

微调的基本概念

SFT：监督微调方法

PEFT 的概念

P-tuning v2 / LoRA / Freeze 等

微调方法在多模态学习中的应用

实践演示：对多模态大模型进行微调

第 5 部分：与人类偏好对齐

强化学习基础概述

DPO：直接偏好优化

PPO：近端策略优化

llama-factory 简介

实践演示：利用 llama-factory 对大模型进行 RLHF

第 6 部分：多模态大模型

qwen_vl_chat
Yi_vl_chat
LLaVa
open-sora
chatTTS

实践演示：使用 qwen_vl 和 Yi_vl_chat 进行视觉问答任务

第 7 部分：结合中移业务的开放讨论

用户资产管理所需的多模式模型

各种 AI 技术在用户资产管理中的应用