

课题组细分方向介绍与基础了解-2026

针对我们现在主要研究的两个方向：**System for Embodied AI** 和 **Future Generative OS**，我们整理了一个细分方向介绍与基础了解。主要包含两个方面：

1. 相关论文阅读。

首先，我们在两个方向里面各选择了几篇相关的论文，你可以任选其中一个方向的其中一篇论文阅读，以此来了解我们的研究方向。论文阅读完成后，请你用你喜欢的方式告诉我：

- ◇ 研究背景是什么？
- ◇ 解决的问题是什么？
- ◇ 如何解决的问题？
- ◇ 针对这篇论文你认为未来还可以进一步有哪些工作可以做？

2. 动手尝试

我们针对不同方向需要用到的能力，整理了一个动手尝试，“Keep your hands dirty”对于我们计算机的同学来说是很重要的。你可以借此体会一下相关的技术栈是否熟悉，当然最重要的是兴趣，技术栈不会的可以未来我们再和大家一起学习。不同方向的动手尝试的要求不同，详情见下面标准。同样，你也可以用你喜欢的方式告诉我你在动手尝试中做了什么、有什么心得和体会☺

根据自己兴趣可选择如下方向之一进行测试。

■ System for Embodied AI

1. 相关阅读列表（选其一阅读即可）

- [1] AAI'26 - DigimonGPT: An Evolvable Agent with Hierarchical Human-like Memory for Video Question Answering.
- [2] ACL'25 - MobiLoRA: Accelerating LoRA-based LLM Inference on Mobile Devices via Context-aware KV Cache Optimization.

2. 动手尝试

- 借助 Pytorch 实现一个 Transformer。
- 阅读 llama.cpp 的文档，尝试在本地(PC,手机均可)部署基于 LoRA 的 llama2-7b 模型。
- 在上述基础上，尝试实现一个 AI 模型的量化、剪枝或你知道的任何一种加速方法

■ Future Generative OS

1. 相关论文阅读

- [1] HotNets '25 - Man-Made Heuristics Are Dead. Long Live Code Generators!

[2] ASPLOS'26 - PF-LLM: Large Language Model Hinted Hardware Prefetching

2. 动手尝试

(可选 1)

- 使用现有的 OpenClaw、LangChain 等智能体编程平台或 Skill、MCP 等智能体技术，实现一个 AI 智能体应用（内容不限）

(可选 2)

- 按照你对于操作系统的理解，提出或实现一个 AI for OS 的小想法