# **Meetup-03**

# **12.09 TOOD**

1. 收集benchmark数据网页形式（多样），给出note，描述网页的特点. （done）
2. HTML element 可见性、属性是否需要比较（基于Todo4 进行细化）
3. 局部模块（CSS＋位置）测评 (done)
4. DOM Tree构建+比较 (done)
5. GPT4打分的测评prompt构建，和人类打分，计算皮尔逊一致性和相关系数 (done)
6. 数据集搜集

脚本１：通过网址批量爬取代码文件的脚本 (done)

脚本２：通过构建复杂网页生成的prompt，引导GPT４产生多元化的一批网页数据 (done)

# **12.14 TODO**

1. 整理代码，给出完整的测评脚本和得分
2. 确定多模态大模型评测baseline（GPT4、qwenVL等）
3. Benchmark数据集继续扩充
4. 这周给出简单版测评结果，根据测评结果分析如何设计一些CoT提升能力