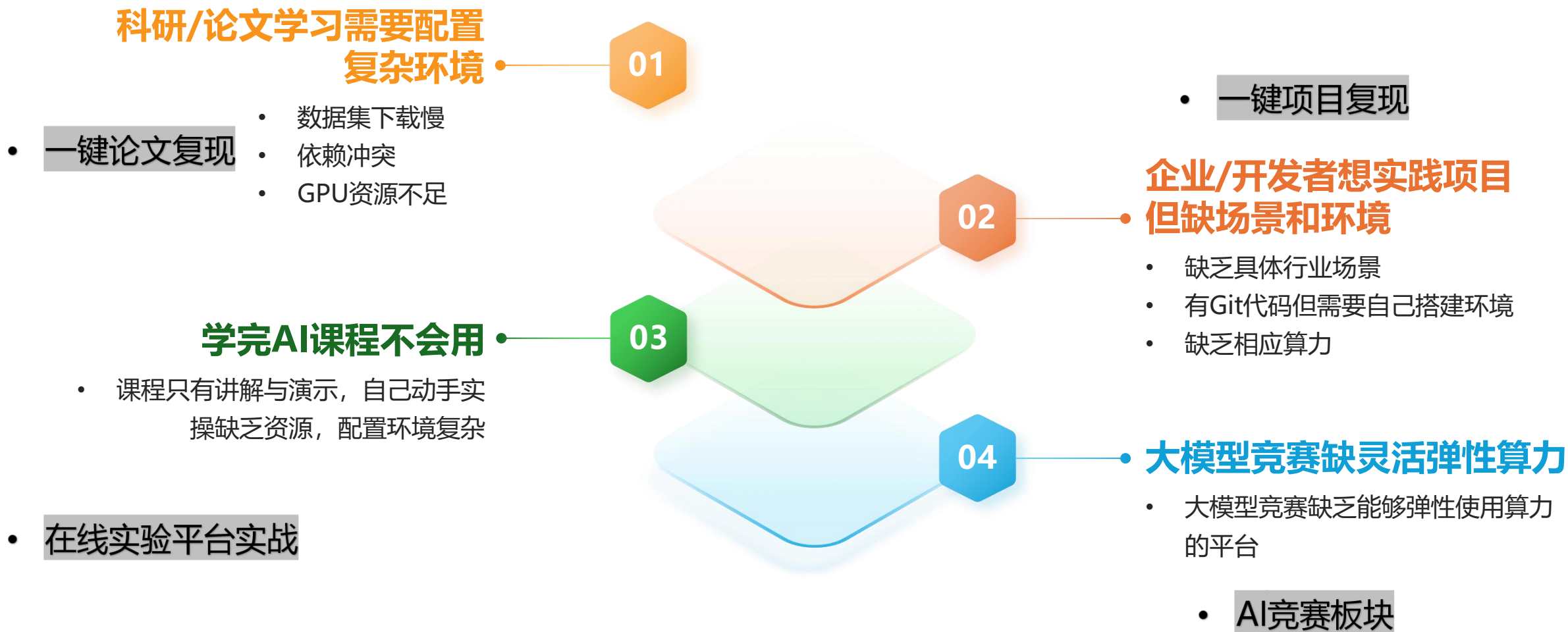


九章云极DataCanvas

人工智能基础设施提供商
领先的智算云厂商

www.datacanvas.com

大模型实验室——AI大模型学习和实践的痛点与解决方案



复现论文来源

论文清单

- <https://lab4ai-hub.github.io/ReproHub/>

自选论文

- 适用于H卡的论文;
- 能够根据论文或Github项目复现预训练或微调或推理的步骤;
- 如您有想复现的论文, 可以发给我们进行审核筛选, 包括: 论文名称、论文链接、年份、github地址等;
- 顶级会议 / 期刊论文优先;
- 将筛选出的待复现论文相关信息填入表格: 2-选题阶段-论文筛选表.xlsx 中

复现要求

推荐的复现步骤

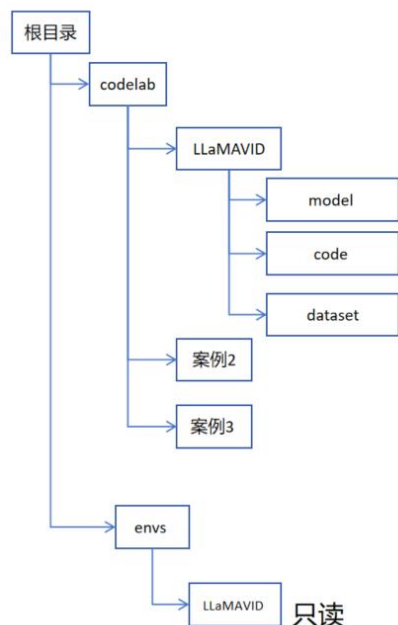
- ✓ 微调+评估+推理
- ✓ 微调+推理
- ✓ 预训练：从0开始的预训练模型，从中抽取部分数据做预训练

文件格式要求

账号和环境

- 使用 <https://www.lab4ai.cn/home> 注册并登录自己账号使用Jupyter进行项目复现

文件格式要求



1. 虚拟环境：各项目创建单独的虚拟环境，并存放到 /env/xxxx 文件夹下；
Conda create -n ttt python=3.12 -y
Conda env list
Conda activate ttt
2. 各案例单独建立文件夹，内部3个文件夹：model、code、dataset分别存放模型、代码和数据。由于我们在后续平台发布时，model和dataset文件夹对用户来只有只读权限，code文件夹有读写权限。所以，执行代码保存的模型或者文件可以保存到 code文件夹下；
3. 如何在ipynb中使用自己的虚拟环境：
创建conda环境后，在新conda环境安装pip install ipykernel
然后执行下面注册内核的操作
示例conda环境：aaa
python -m ipykernel install --prefix=/usr/local --name aaa --display-name "python(aaa)"
刷新下jupyter notebook界面，就有了

需要提交的输出物

1. 提供各项目文件路径及其他详细信息，后续我们根据路径进行项目迁移，详细内容见表格：[3-成果提交-项目上架表.xlsx](#)
2. 最后需要提交一份ipynb文件，该文件保存在code文件夹中。文件中说明其他人如何用你的代码、模型、数据复现整个流程。比如：可以标清楚你的模型在什么文件夹、数据文件夹在什么位置、启动什么文件就能执行微调或推理；提供资源配置，比如什么阶段选了什么卡；如他人按照您提交的文件中的步骤复现的话，每个步骤大约花费多长时间？可参考这个文件：[4-代码参考-Notebook示例.ipynb](#)



4-代码参考-Notebook示例.ipynb

3. 提供安装依赖包步骤，供复现组成员【迁移项目】和【发布】使用

THANKS



九章云极DataCanvas
人工智能基础设施提供商



400-805-7188

北京市·海淀区和盈中心C座6层&8层
上海市·静安区市北壹中心1号楼10层
济南市·奥盛大厦1号楼503室