

# 通用大模型原理及训练实践

## 实验课①：环境搭建及模型准备



# 实验内容

- 5人一组，自由组队（请尽快完成，组队成功后发放账号）
- 每个小组可使用1张H800或A100-40G GPU
- 目标：基于已有**基座模型**，根据应用需求自行**构建指令**，并对模型进行**指令微调**，最终实现：
  - 模型具备通用问答能力
  - 模型对某个特定NLP任务较为擅长
  - 模型具备明确的自我认知
    - 问模型“你是谁”，可正确回答名字

# 实验平台

## ■ 信息高铁智算算力网AI平台

- 每个小组一个账号
- 每个小组可使用1张H800或A100-40G GPU
- 用于代码调试、模型训练
- <https://aicloud.oneainexus.cn:30013/aicloud/#/login>
- 文档：登录后右上角“帮助文档”

AI制品仓库

帮助文档

zhouyan



# 信息高铁智算算力网AI平台：账号登录

- 用户名与密码将发放给每个小组组长

短信登录 | 账号登录

请输入用户名/手机号/邮箱

请输入用户名/手机号/邮箱

密码

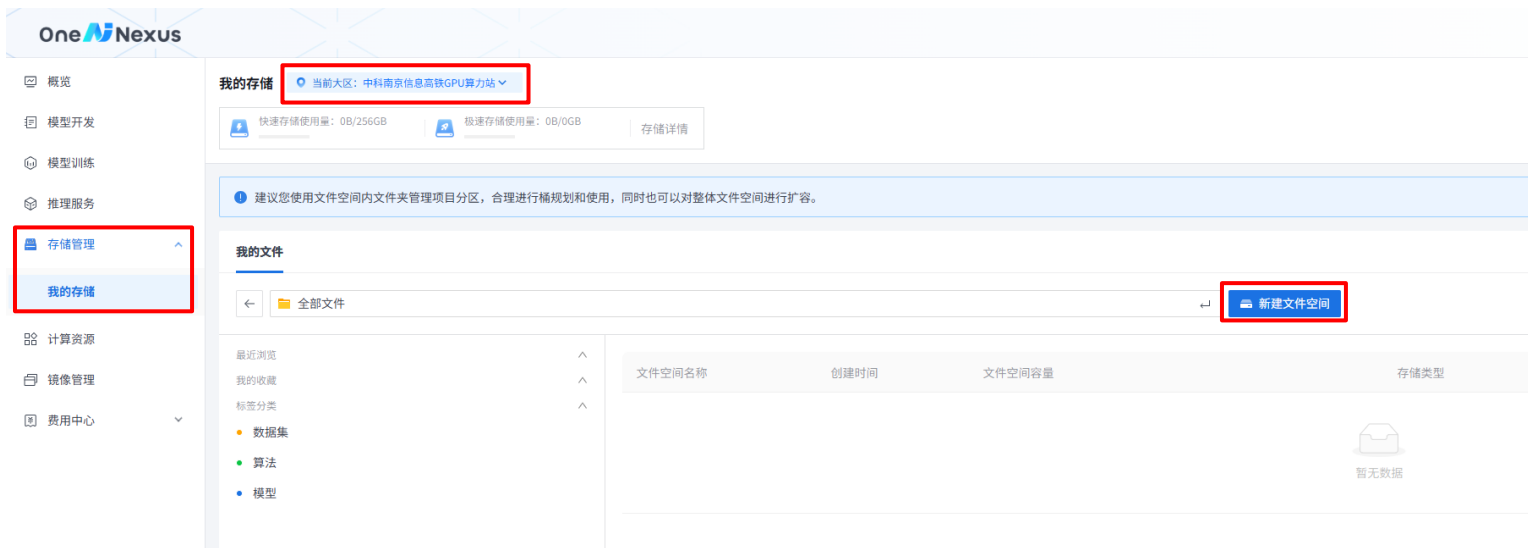
请输入密码

登 录

[忘记密码](#) | [注册](#)

# 信息高铁智算算力网AI平台：新建文件空间

- 点击【存储管理】→【我的存储】，大区选择：中科南京信息高铁GPU算力站
- 点击【新建文件空间】



# 信息高铁智算算力网AI平台：新建文件空间

- 选择快速存储，填写名称与容量（每个账号最多256GB存储）

新增文件空间

\* 存储类型:

快速存储 HDD+NVMe

\* 文件空间名称:

data

\* 文件空间容量(GB):

100

添加标签:

请选择标签

取消

创建文件空间

# 信息高铁智算算力网AI平台：创建实验环境

- 点击【模型开发】→【+新建开发环境】



# 信息高铁智算算力网AI平台：创建实验环境

## ■ 区域选择：中科南京信息高铁GPU算力站

1 基本信息

2 数据及算法代码配置

3 算力资源配置

---

\* 任务名称:

llmcourse

描述:

请输入描述

0 / 150

\* 区域:

中科南京信息高铁GPU算力站



# 信息高铁智算算力网AI平台：创建实验环境

## ■ 存储配置：默认挂载，全选先前创建的文件空间

1 基本信息      2 数据及算法代码配置      3 算力资源配置

\* 存储配置：

☒ 默认挂载

☐ 自定义挂载

文件空间名: data 0B/100GB

运行时路径: \${WORK\_PATH}data

文件空间名: shared 13.2GB/100GB

运行时路径: \${WORK\_PATH}shared

☒ 全选

# 信息高铁智算算力网AI平台：创建实验环境

- 资源池：专属资源池
- 资源规格：唯一一项
- 环境类型：vscode
- 任务镜像：codeserver
- SSH远程开发：开启

基本信息 — 数据及算法代码配置 — 3 算力资源配置

\* 资源池: 专属资源池

\* 资源规格: ml.g1hxl.3xlarge

\* 环境类型: vscode

\* 任务镜像: codeserver

SSH远程开发: ☒ 开启 ☐ 关闭

# 信息高铁智算算力网AI平台：进入实验环境

## ■ 点击“连接”进入在线vscode

模型开发

新手入门 使用指南

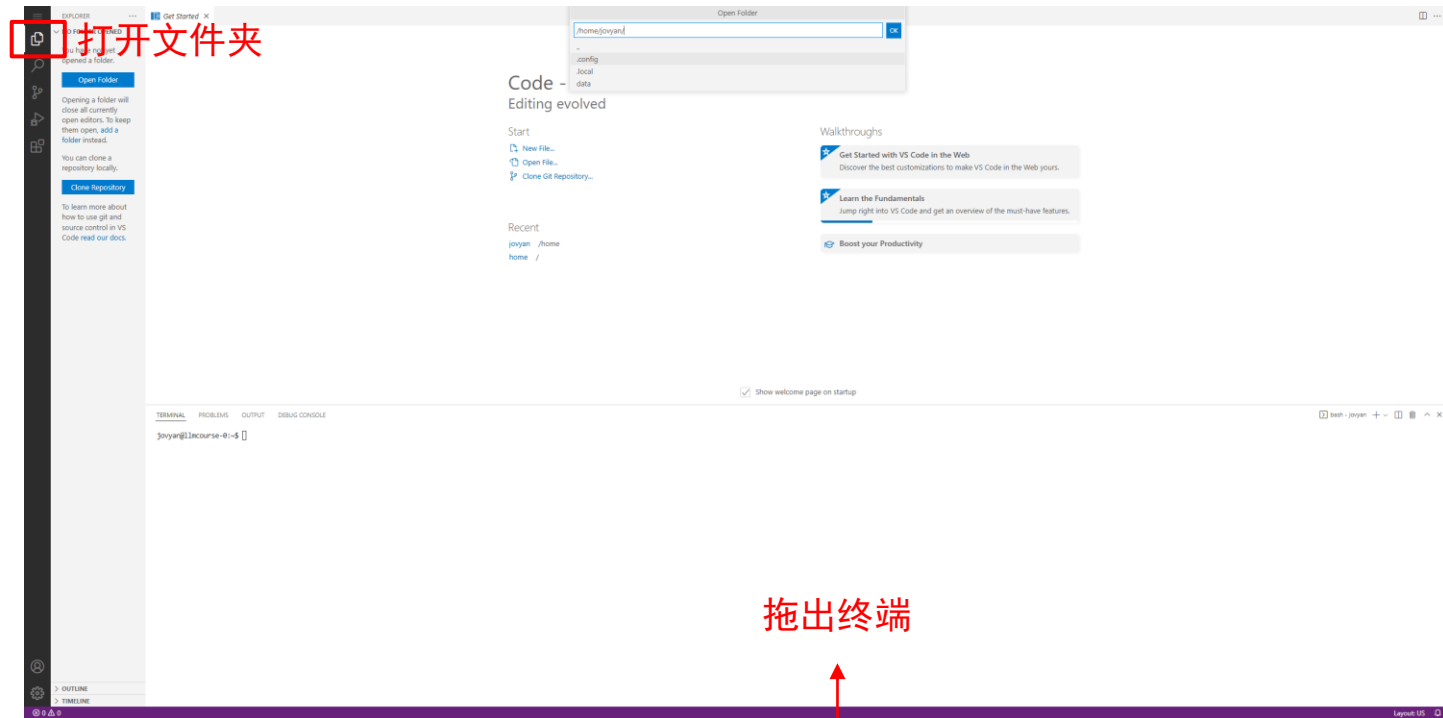
+ 新建开发环境 全部 1 创建中 0 错误 0 运行中 1 等待中 0 已停止 0 提交失败 0

请输入训练任务名称

任务名称	状态	环境类型	镜像	资源规格	创建时间	停止时间	运行总时长	SSH	描述	操作
llmcourse	运行中	vscode	...43/notebook/codeserver-rootv1.6.1	ml.g1hx.3xlarge cpu*12个 memory*200GB NVIDIA-H800...	2025-06-27 15:32:06	-	12分24秒	开启	-	<a href="#">连接</a> <a href="#">停止</a> <a href="#">更多</a>

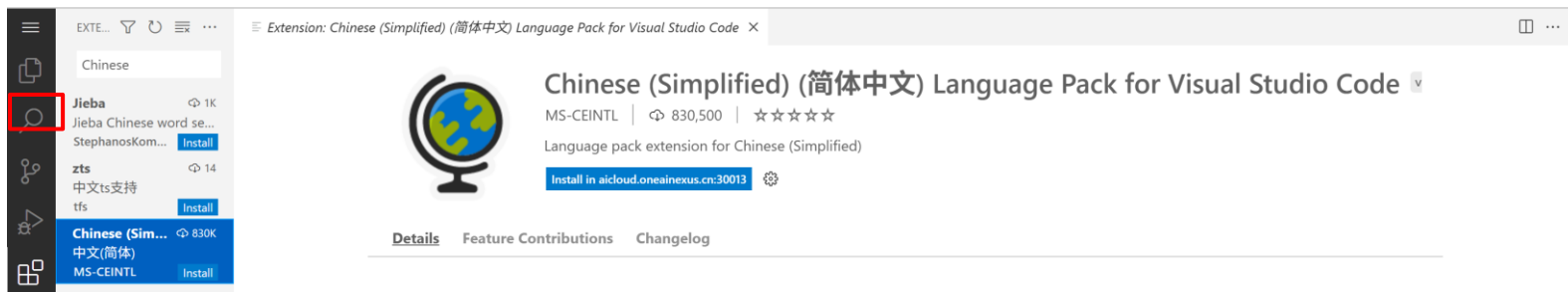
# 信息高铁智算算力网AI平台：进入实验环境

- 点击左侧最上图标可打开文件夹，底部可拖出终端



# 添加中文支持

- 在 vscode 中，点击左侧 Extensions，搜索 Chinese，安装简体中文包
- 安装后右下角出现弹窗，按提示重启 vscode
- 终端中输入汉字，能成功显示即可



```
jovyan@example-0:~$ echo 中文
中文
```

# 信息高铁智算算力网AI平台：配置环境

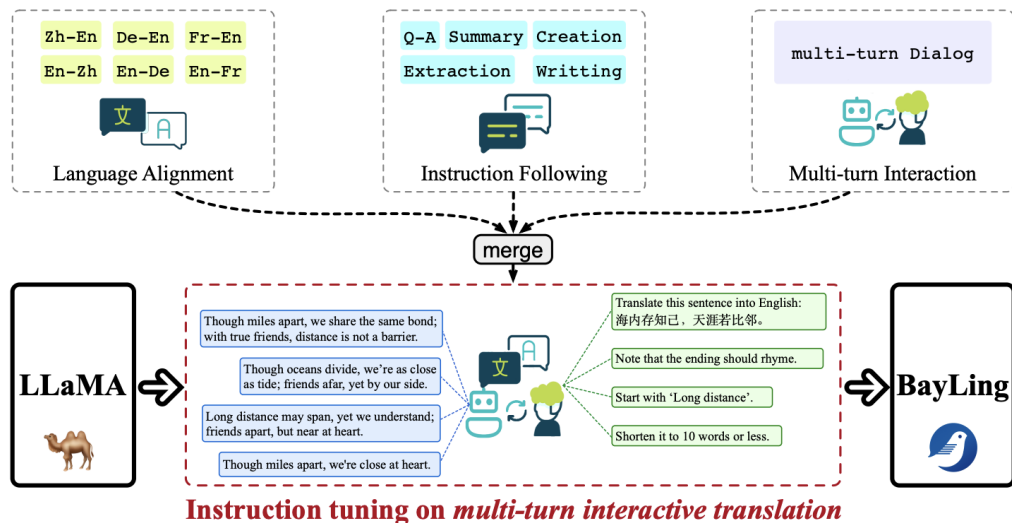
- 在vscode中新建终端
- （可选）在终端中键入命令：su - 然后输入密码123456，即可进入root用户；安装多终端管理程序tmux：apt-get update然后apt-get install tmux即可。tmux使用参考：<https://zhuanlan.zhihu.com/p/335818087>
- 在终端中键入命令：./Anaconda3-2022.05-Linux-x86\_64.sh（存储在shared目录下，通过cp命令拷贝到小组自己的目录下再运行），安装包管理工具conda
- shared目录为公共资源目录，请不要进行修改

# 信息高铁智算算力网AI平台：配置环境

- 在终端环境中激活conda，使用pip或conda安装Python库
- 推荐python=3.10，如conda create -n xxx python=3.10
- Conda国内镜像：<https://zhuanlan.zhihu.com/p/628870519>，可用清华源
- 如果安装速度较慢，可将pip源设置为清华源
  - python -m pip install -i https://pypi.tuna.tsinghua.edu.cn/simple --upgrade pip
  - pip config set global.index-url <https://pypi.tuna.tsinghua.edu.cn/simple>
- 或者临时使用
  - pip install -i <https://mirrors.tuna.tsinghua.edu.cn/pypi/web/simple> xxx

# 基座模型：百聆2-7B（BayLing2-7B）

- 百聆2-7B（<https://github.com/ictnlp/BayLing>，中科院计算所）
  - 基于LLaMA-2-7B，以多轮交互翻译任务为核心，进行指令微调
  - 以经济友好、内存节约的方式实现了多语言人机交互能力





# 安装相关依赖库、准备模型文件

- `conda create -n xxx python=3.10`
- `conda activate xxx`
- `git clone https://github.com/ictnlp/BayLing`
- `cd BayLing/`
- `pip install -r requirements.txt`
- `pip install -U torch torchvision torchaudio transformers accelerate google`
- `pip install protobuf==3.19.0`
- BayLing-2-7B模型文件存储在 `shared/bayling-2-7b` 目录下，拷贝到小组的用户目录下：`cp -r shared/bayling-2-7b/ ~/`

- ```
(base) jovyan@llmcourse-0:~$ cd BayLing/
(base) jovyan@llmcourse-0:~/BayLing$ python chat.py --model-path ~/bayling-2-7b/ --style rich --load-8bit
Welcome to interact with BayLing!
- Enter your question and press 'Enter' to interact with BayLing.
- Enter '!!' to clear the interaction history.
Loading checkpoint shards: 100%|██| 7/7 [00:00<00:00, 21.85it/s]
USER:
你好

BayLing:
你好！有什么我可以帮助你的吗？

USER:
请帮我把这句话翻译成英文：我今天吃了披萨

BayLing:
I ate pizza today.

USER:
以today开头

BayLing:
Today I ate pizza.
```

- ## 冯洋（中科院计算所）

# GPU使用情况

- `nvidia-smi`
- `gpustat: pip install gpustat`
- `watch -n 1 gpustat`

TERMINAL

PROBLEMS

OUTPUT

DEBUG CONSOLE

Every 1.0s: gpustat

```
course-0      Fri Jun 27 09:48:18 2025  535.129.03  
[0] NVIDIA H800 | 33  C,   0 % |      2 / 81559 MB |
```

# 课后任务

- 尽快完成分组，5人一组，确立小组长，在在线文档中登记
- 搭建实验环境，掌握信息高铁智算算力网AI平台的使用方法
- 安装相关依赖库
- 能够在命令行里与BayLing-2-7B模型交互

谢谢！