基于昇思MindSpore的香橙派开发板聊天机器人

体验操作文档

## 昇思MindSpore介绍

昇思MindSpore是一个全场景深度学习框架，旨在实现易开发、高效执行、全场景统一部署三大目标。其中，易开发表现为API友好、调试难度低；高效执行包括计算效率、数据预处理效率和分布式训练效率；全场景则指框架同时支持云、边缘以及端侧场景。更多详情可以查看官方网址：<https://www.mindspore.cn/>

## MindSpore NLP套件介绍

MindSpore NLP是基于MindSpore的一个自然语言处理大模型套件，其中包含了许多自然语言处理的常用方法，提供了一个易用的NLP平台，旨在帮助研究人员和开发者更高效地构建和训练模型。感兴趣的同学可以点击访问我们的开源代码仓库和文档：

* 代码仓：<https://github.com/mindspore-lab/mindnlp>
* 文档：<https://mindnlp.cqu.ai/zh/>

## OrangePi AIpro香橙派开发板介绍

OrangePi AIpro 采用昇腾AI技术，搭载4核64位处理器和AI处理器，支持8-12及20TOPS AI算力，配备多种内存和eMMC扩展，支持双4K输出。官方系统镜像（openEuler和Ubuntu）目前已预置昇思MindSpore AI框架，支持昇思MindSpore官网教程所有模型，方便高效开发。体验环境已预配置，开发者可直接运行代码体验聊天机器人对话。

* 香橙派官网：<http://www.orangepi.cn/>

## 模型及任务简介

通义千问Qwen是阿里云自主研发的大语言模型，Qwen-1.5-0.5B 为其轻量版，拥有5亿参数，具备低成本和高效文本生成能力。适用于文字创作、文本处理、编程辅助、翻译服务和对话模拟等场景。

在本次任务中，我们将基于昇思MindSpore在香橙派开发板上运行通义千问Qwen-1.5-0.5B模型，体验和模型的对话互动，完成和聊天机器人对话，并star代码仓视为成功完成任务。具体操作如下：

**本手册最后附录含昇思MindSpore相关资料，昇思MindSpore香橙派开发板申请表，及昇思MindSpore开源社区实习申请表，欢迎感兴趣的开发者了解参与。**

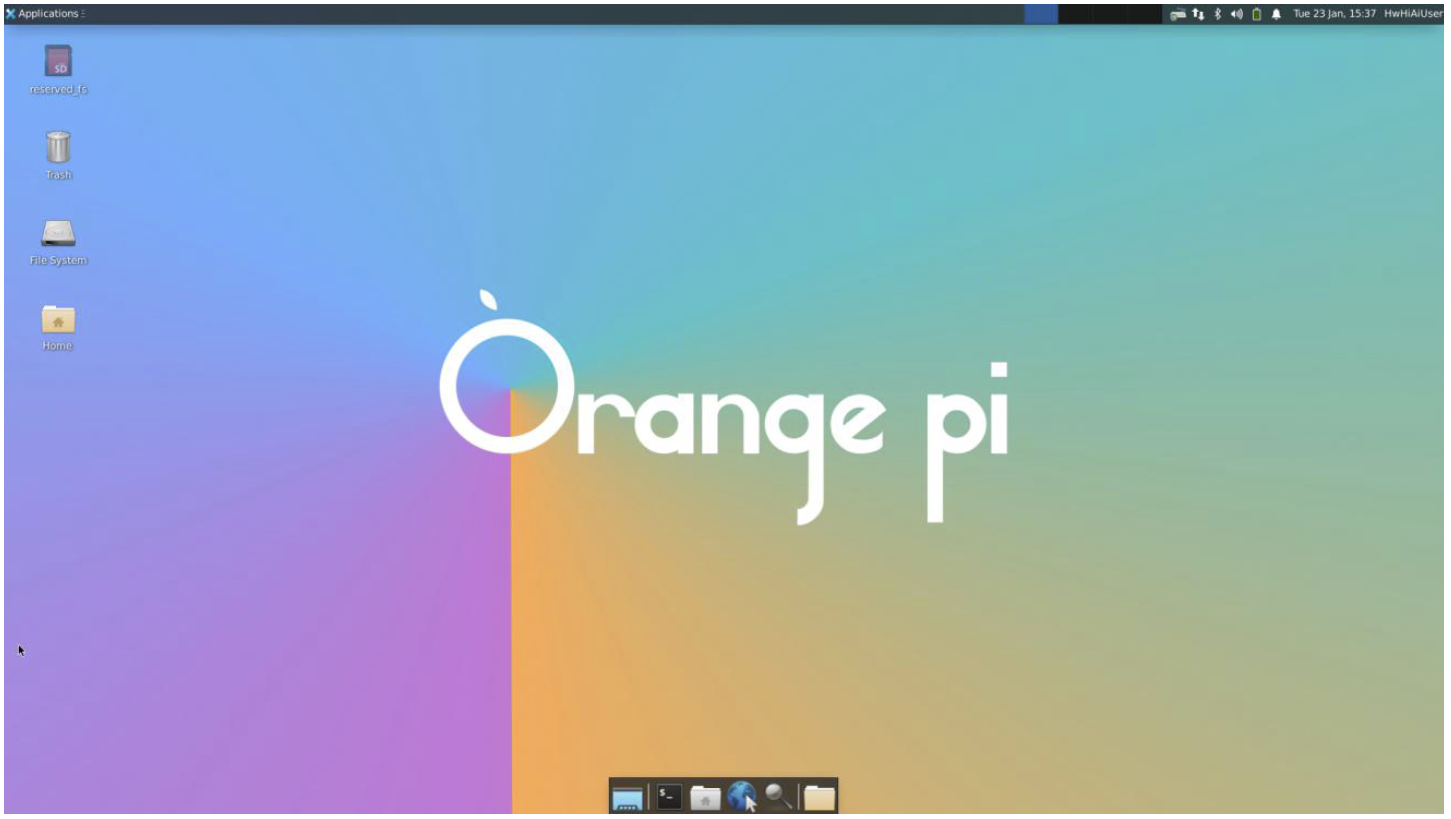
## 操作指引

### 5.1启动开发板

开发板连接电源后，需要等待一段时间，HDMI显示器才会显示Linux系统的登录界面，登录界面如下图所示：



默认登录用户为HwHiAiUser，登录密码为Mind@123。成功登录后显示的系统桌面如下图所示，如果在体验时屏幕显示已进入系统桌面，则跳过此步。



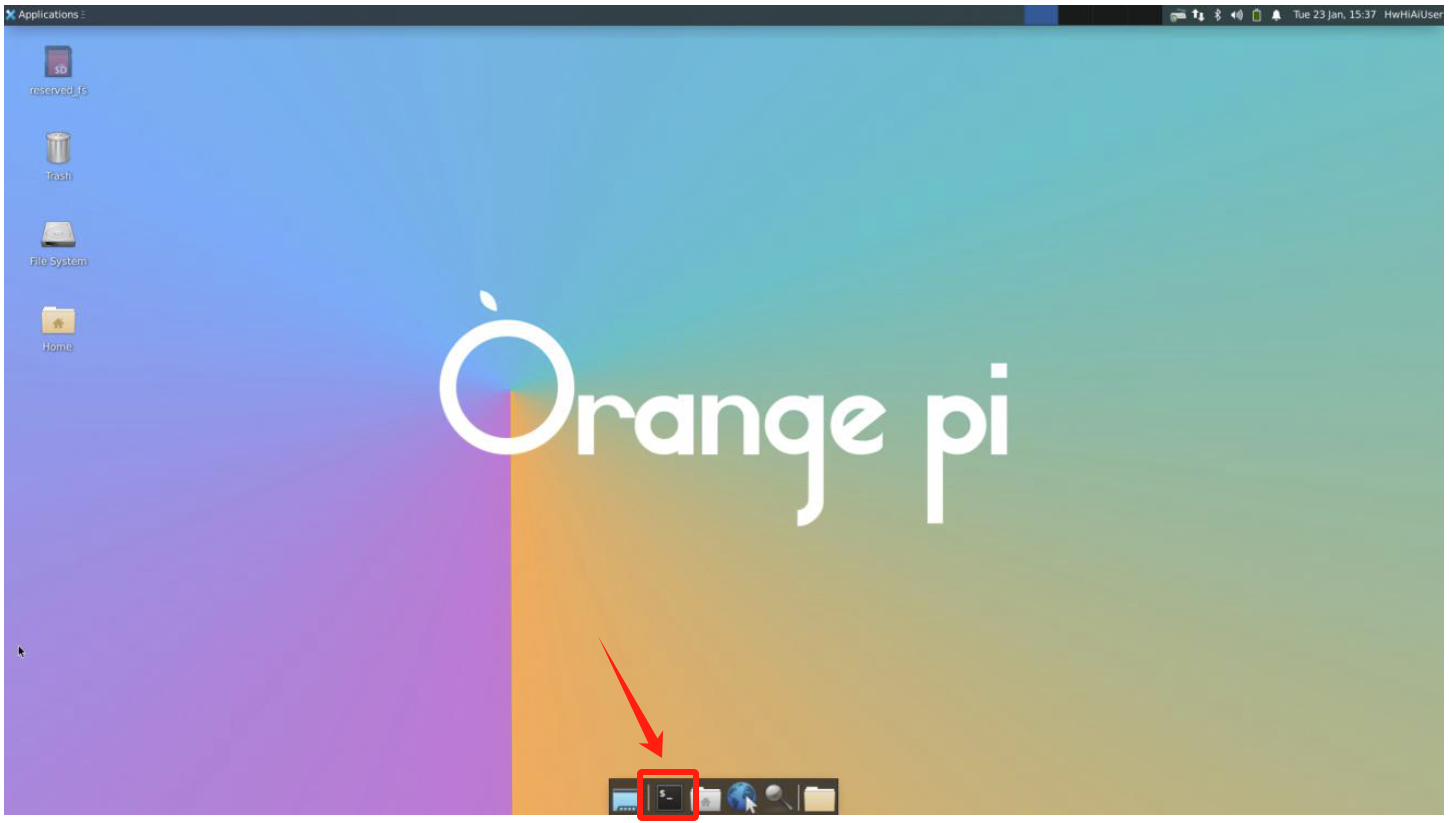
**注意：请勿随意移动香橙派开发板，拔插连接线等，以免开发板出现意外关机的情况。如出现黑屏，风扇停止转动等问题，请联系现场布道师处理。**

### 5.2 CANN和MindSpore环境搭建

#### 5.2.1 CANN升级

##### 5.2.1.1 Toolkit升级

使用CTRL+ALT+T快捷键或点击界面下方带“$\_”的图标打开终端控制台。



切换root用户，root用户密码：Mind@123。

# 打开开发板的一个终端，运行如下命令

(base) HwHiAiUser@orangepiaipro:~$ su – root

Password:

(base) root@orangepiaipro: ~#

删除镜像中已安装CANN包释放磁盘空间，防止安装新的CANN包时报错磁盘空间不足。

cd /usr/local/Ascend/ascend-toolkit

rm -rf \*

打开[昇腾CANN官网访问社区版资源](https://www.hiascend.com/developer/download/community/result?module=cann)，下载所需版本的toolkit包，该处以8.0.RC3.alpha002版本为例，如下图：



执行如下命令，根据环境实际输出信息选择aarch64或是x86\_64对应的软件包。

uname -a

进入Toolkit包下载目录。

cd /home/HwHiAiUser/Downloads

# Orange Pi AI Pro浏览器文件默认下载目录：/home/HwHiAiUser/Downloads，用户在更换保存路径时请同步修改上述命令中的路径

给CANN包添加执行权限。

chmod +x ./Ascend-cann-toolkit\_8.0.RC2.alpha003\_linux-aarch64.run

执行以下命令升级软件。

./Ascend-cann-toolkit\_8.0.RC2.alpha003\_linux-aarch64.run --install --install-for-all

进入安装流程，根据回显页面输入y或Y接受协议，输入其他任意字符为拒绝协议，确定接受协议后开始安装。安装完成后，若显示如下信息，则说明软件安装成功。

[INFO] Ascend-cann-toolkit install success

安装升级后的路径（以root用户默认安装升级路径为例）：/usr/local/Ascend/ascend-toolkit/

配置环境变量。

source ~/.bashrc

##### 5.2.1.2 Kernels升级

二进制算子包Kernels依赖CANN软件包Toolkit，执行升级时，当前环境需已安装配套版本的Toolkit，并使用同一用户安装。

打开终端，并切换root用户。root用户密码：Mind@123。

# 打开开发板的一个终端，运行如下命令

(base) HwHiAiUser@orangepiaipro:~$ su – root

Password:

(base) root@orangepiaipro: ~#

执行如下命令，获取开发板NPU型号。

npu-smi info

打开[昇腾CANN官网访问社区版资源](https://www.hiascend.com/developer/download/community/result?module=cann)，下载与CANN包版本一致，并且匹配NPU型号的kernel包，如下图：



进入Kernels包下载目录。

cd /home/HwHiAiUser/Downloads

# Orange Pi AI Pro浏览器文件默认下载目录：/home/HwHiAiUser/Downloads

给kernels包添加执行权限。

chmod +x ./Ascend-cann-kernels-310b\_8.0.RC2.alpha003\_linux.run

执行以下命令升级软件。

./Ascend-cann-kernels-310b\_8.0.RC2.alpha003\_linux.run --install

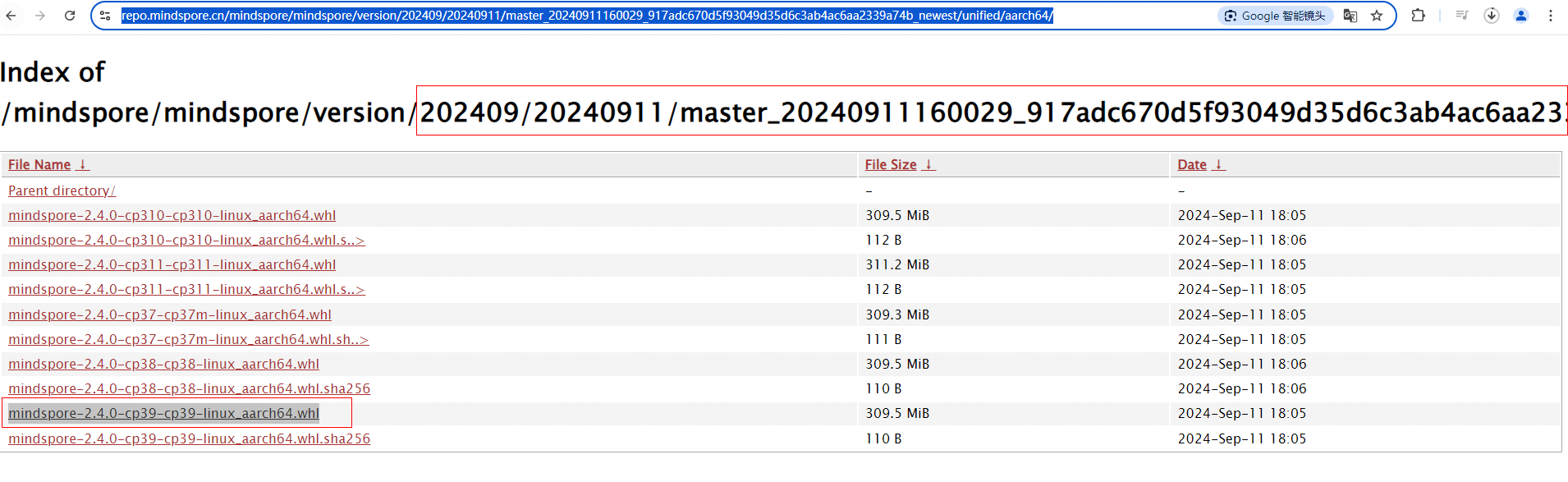
#### 5.2.2 MindSpore升级

安装MindSpore daily包（以9月11日daily包为例）

香橙派开发板支持自定义安装MindSpore daily包，可从[[此链接]](https://repo.mindspore.cn/mindspore/mindspore/version/)获取到对应日期的软件包。

* 目标 daily whl包具体查找过程如下：

1. 进入以master为前缀的目录。若是出现多个目录前缀是master时，推荐进入日期更靠后的目录。
2. 进入unified目录。
3. 根据实际操作系统信息，进入对应目录。由于香橙派开发板默认操作系统为linux-aarch64，所以进入aarch64目录。
4. 根据实际python版本信息，找到对应daily whl包。由于香橙派开发板默认为python3.9，所以目标daily包为mindspore-2.4.0-cp39-cp39-linux\_aarch64.whl。



下载whl包进行安装，终端运行如下命令。

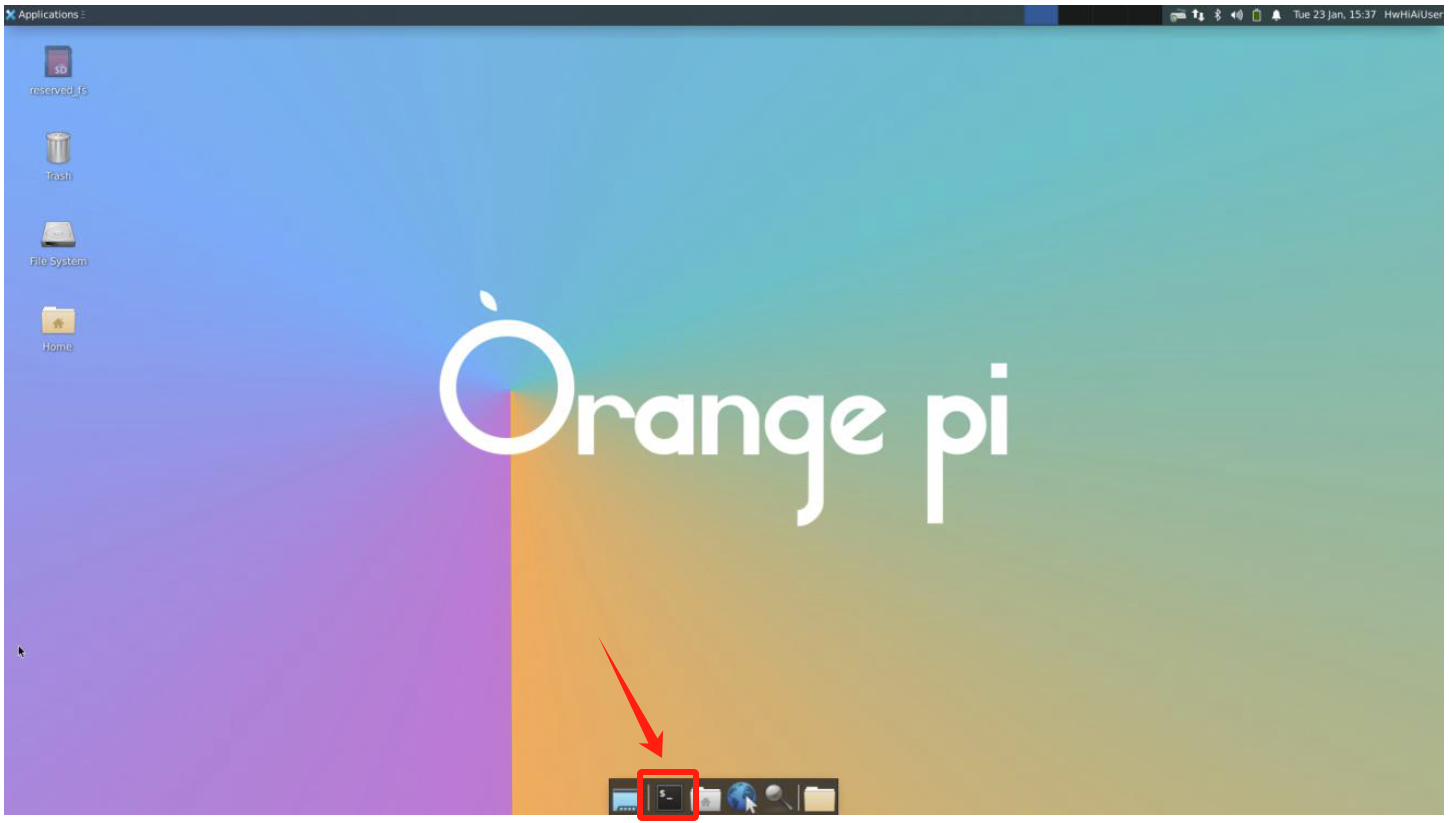
# wget下载whl包

wget https://repo.mindspore.cn/mindspore/mindspore/version/202409/20240911/master\_20240911160029\_917adc670d5f93049d35d6c3ab4ac6aa2339a74b\_newest/unified/aarch64/mindspore-2.4.0-cp39-cp39-linux\_aarch64.whl

pip install mindspore-2.4.0-cp39-cp39-linux\_aarch64.whl

### 5.3 环境检查

使用CTRL+ALT+T快捷键或点击界面下方带“$\_”的图标打开终端控制台。



依次输入如下内容进行检查：

1. 检查mindspore是否为2.4.0版本

pip show mindspore

# 如版本不匹配，请联系在场布道师进行处理

1. 检查mindnlp是否为0.4.0版本：

pip show mindnlp

# 如版本不匹配，可执行如下命令进行安装，或联系在场布道师进行处理

pip install git+https://github.com/mindspore-lab/mindnlp.git

1. 检查gradio是否为4.44.0版本：

pip show gradio

# 如版本不匹配，可执行如下命令进行安装，或联系在场布道师进行处理

pip uninstall gradio --y

pip install gradio==4.44.0 -i https://pypi.tuna.tsinghua.edu.cn/simple

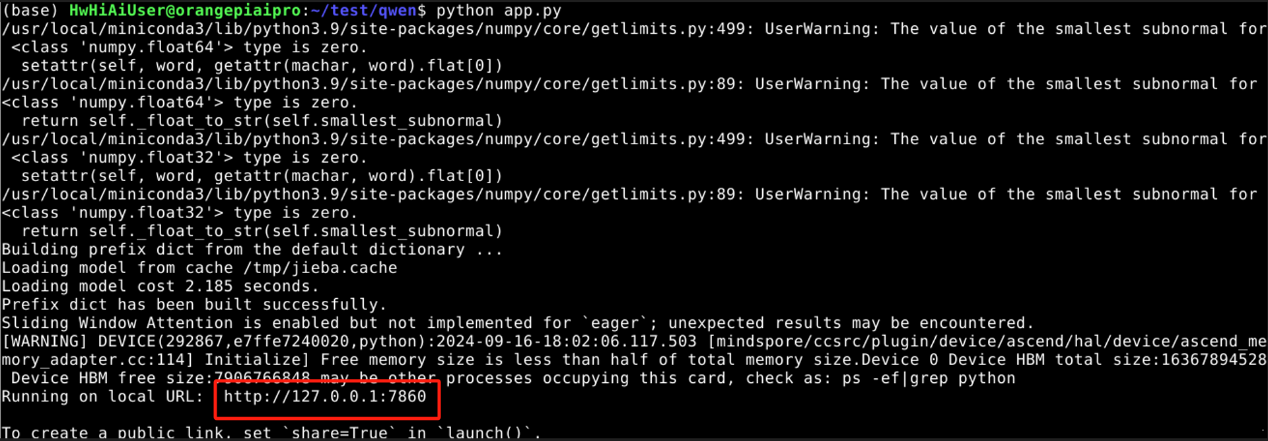
### 5.4 运行案例

1. 到Qwen案例目录下，直接运行脚本文件app.py

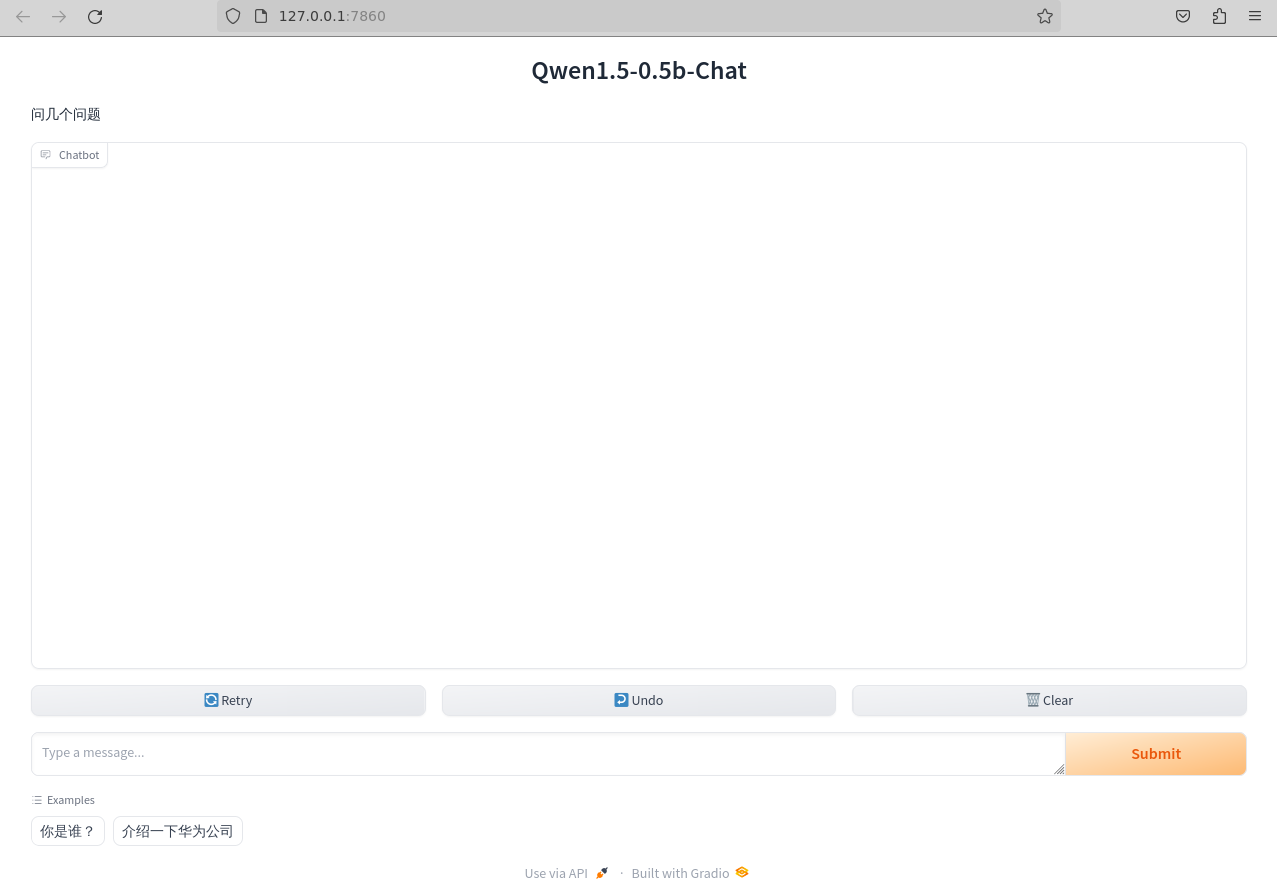
cd /home/HwHiAiUser/HC\_demo/qwen/

python app.py

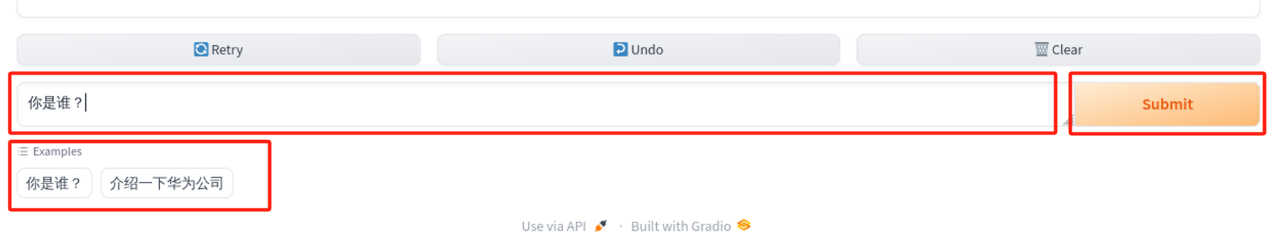
1. 在浏览器中打开链接http://127.0.0.1:7860



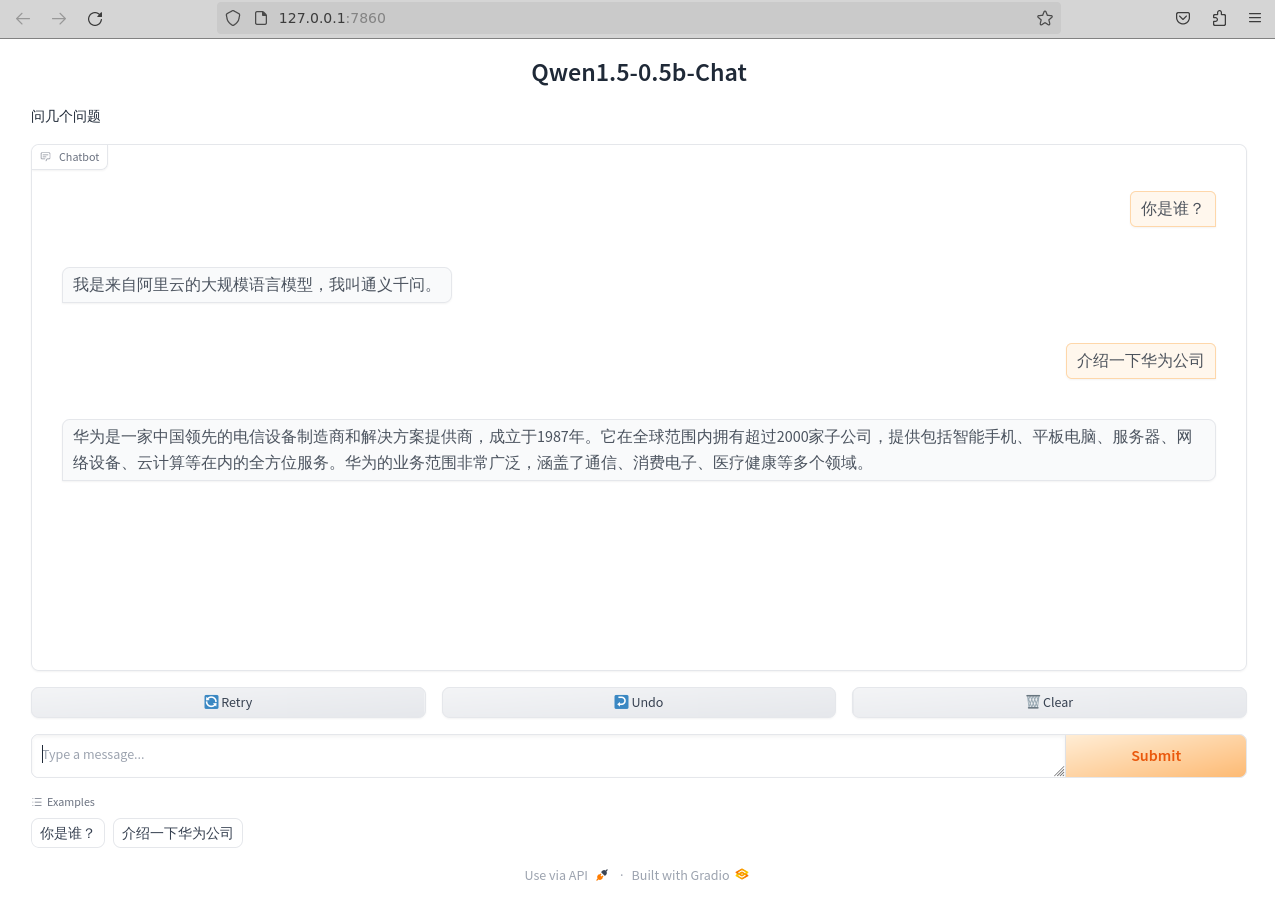
进入网页交互界面：



1. 开始聊天，可在页面下方消息输入框Type a message…中输入任何问题，或者点击下方Examples中设置好的问题，然后点击右侧的Submit按钮，Qwen模型将对此进行回答。



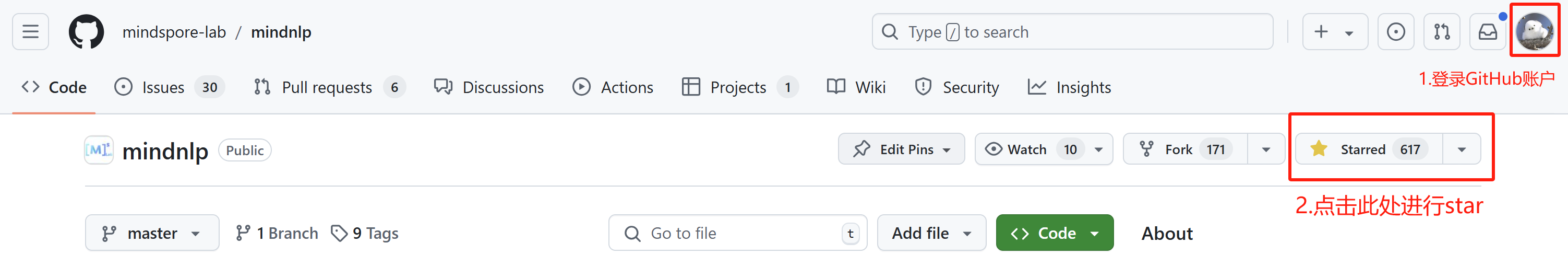
1. 第一次回答需要较长时间加载，请耐心等待。回答将显示在上方聊天框中：



1. 如果出现Error，可以点击retry按钮重新发送上一条消息，并让模型重新回答；点击undo按钮可撤回上一条消息；点击clear按钮将清空聊天框中的对话。

### 5.5 star代码仓

1. 扫描**【6.1 资源一览】**的MindSpore NLP代码仓二维码，使用自己的GitHub账号进行登录（无GitHub账号需进行注册）。
2. 点击代码仓右上角的星星图标，完成操作。最终向布道师出示对话及完成star的页面，进行活动打卡。



## 相关资讯

### 资源一览

**Star代码仓后完成打卡**

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **昇思MindSpore官网** | **香橙派官网** | **MindSpore NLP代码仓** |
|  |  |  |
| **昇思MindSpore香橙派案例** | **昇思MindSpore开源实习** | **昇思MindSpore大模型平台** |

### 昇思MindSpore香橙派AIpro申请活动

有想法基于昇思MindSpore构建开发板应用，但苦于手头没有开发板资源的同学，可以通过填写下方的申请表，说明申请人基本信息、项目信息和使用时长，[发送至wuyiqun3@huawei.com](mailto:发送至wuyiqun3@huawei.com)，即可有机会在申请的使用时间内，免费获取香橙派开发板进行项目开发。香橙派申请表详见附件，可自行获取。

### 昇思MindSpore开源实习

昇思MindSpore开源实习是MindSpore社区携手openEuler社区等多个社区共同发起的线上实习项目，旨在鼓励在校学生积极参与开源社区，在实际的开源环境中提升实践能力。由昇思MindSpore社区提供实习任务，并提供导师辅导，学生通过实习申请后，可在社区领取任务，每完成一个任务可获得相应积分，积分累计达规定量后，可获得实习证明和实习工资。

* 开源实习详情：<https://www.mindspore.cn/internship>， 或可扫描【6.1 资源一览】中的对应二维码了解详细信息

**【昇思MindSpore】香橙派AIpro申请表**

**（学生版）**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **申请人基本信息** | **姓名** |  | | | | | **电话** | |  | | | |
| **学校** |  | | | | | **学院** | |  | | | |
| **专业** |  | | | | | **学历** | |  | | | |
| **邮箱** |  | | | | | **指导老师** | |  | | | |
| **项目基本信息** | **项目名称** | | |  | | | | | | | | |
| **应用是否基于MindSpore框架训练推理？** | | | | |  | **项目的MindSpore版代码是否全部开源到github/gitee平台** | | | | |  |
| **项目主要研究内容和技术方案介绍** | | |  | | | | | | | | |
| **项目计划和当前进展** | | |  | | | | | | | | |
| **规格** | **香橙派规格** | | | |  | | | | | | | |
| **使用开始时间** | | | |  | | **使用结束时间** | | | |  | |
| **个人承诺书** | **我承诺申请使用的华为云代金券仅用于上述应用案例项目基于昇思MindSpore框架开展实验，代码将在github/gitee昇思社区仓库进行开源。** | | | | | | | | | | | |
| **承诺人** | |  | | | | | **日期** | |  | | |

**注：**

**（1）本申请可以由学生主要填写申请资源，指导老师负责指导学生基于MindSpore完成代码实现并开源。**

**openEuler开源实习活动声明**

**在您报名参加openEuler开源实习活动（以下称“本活动”）前，请确保您已年满18周岁，并请您务必阅读本声明，如您同意接受本声明内容，请在本声明底部签名并随申请材料一起发送至本活动邮箱intern@openeuler.io，否则请您立即停止报名本活动。**

1. **活动内容**

本活动是openEuler社区联合openGauss社区、openLooKeng社区、MindSpore社区与合作单位（“中国科学院软件研究所”）共同发起的线上实习项目，旨在鼓励在校学生积极参与开源社区，在实际的开源环境中提升实践能力。您向本活动邮箱发送申请资料后，openEuler社区将以邮件反馈审核结果，审核通过后，实习合作单位（“中国科学院软件研究所”）会与您签订6个月实习期的劳务协议。由openEuler社区、openGauss社区、openLooKeng社区、MindSpore社区提供实习任务，并提供导师辅导，您可在社区领取任务，每完成一个任务可获得相应积分，在实习期内积分累计达规定量后，可获得由实习合作单位（“中国科学院软件研究所”）发放的盖章实习证明和实习工资。

（注：上述实习证明仅用于证明您参与了本次开源实习，并达到了本次线上实习合格标准（满60积分），不代表对您技能水平的评价，也不保证能被您所在学校或其他组织、机构认可。）

1. **隐私保护**

**2.1 数据收集类型**

您了解并同意，为提供服务之必需，您同意授权我们收集和使用您的个人信息或个人敏感信息，包括：

（1）姓名、学校、专业、年级、联系邮箱、个人项目经验、个人技能，用于实习资质考核与后续实习相关事务联络以及本活动的统计分析；

为了保证沟通的及时性和便利性，另收集QQ号、联系电话用于进行实习事务联络。

（报名时提供给intern@openeuler.sh 邮箱）

（2）身份证扫描件、学生证或学籍证明扫描件，用于确认身份以及在中国科学院软件研究所报备实习生流程；个人银行卡信息以及银行卡扫描件，用于发放实习工资。

（以上扫描件均加上“仅用于开源实习”字样的水印，在签署实习劳务协议后[提供给interninfo@isrc.iscas.ac.cn](mailto:提供给interninfo@isrc.iscas.ac.cn)邮箱）

您理解上述信息均是为您提供服务之必需。如您选择不提供或不同意我们收集、使用以上信息，将导致您无法参加本活动。

* 1. **数据存储**

您的个人信息将被存储于中国大陆境内的服务器。我们将会在本声明所述目的所需的期限内保留您的个人信息，除非按照法律要求需要延长保留期。保留期可能会基于不同的处理目的而有所差异。

在您的个人信息超出保留期限后，且没有法律要求我们继续处理您的特定个人信息的情况下，我们将会根据可适用法律的要求删除您的个人信息，或进行匿名化处理。

**2.3 数据保护**

我们重视您的个人信息安全。我们采用适当的物理、管理和技术保障措施来保护您的个人信息。例如，我们会使用加密技术确保信息的机密性；我们会使用保护机制防止信息遭到恶意攻击；我们会部署访问控制机制，确保只有授权人员才可访问对于您的上述信息。

根据适用的法律法规，您可能有访问、更改、删除、拒绝、限制或移植您上述信息的权利。

**3.声明与保证**

**3.1 双方声明**

您了解并同意，您报名参加本活动符合法律法规的要求。我们仅按照您的指示，为您提供标准活动内容，并不对您参加本活动的合法合规性负责。

**3.2你的承诺**

您了解并同意，您在本活动中提供的信息应真实、准确、合法、有效。如您提交的信息违反前述承诺或违背本活动声明内容，您应自行承担由此带来的全部后果和责任。

**4.通知与更正**

您了解并同意，为适应政策和业务的需要，我们将在必要时对本活动声明内容进行更新并在本网站发布，请您及时参阅本服务声明更新，开展业务。

本人已阅读并同意上述声明

签名：

日期：

（注：请将本文件打印下来手写签名，回传扫描版签名文件）