

南宁职业技术学院人工智能学院

# 毕 业 设 计 报 告

课题名称： 数字生命系统

二级学院： 人工智能学院

专 业： 物联网应用技术

班 级： 21 物联网 3 班

学生姓名：

学 号：

指导教师： 李华洲

日 期： 2023-11-

## 摘 要

随着科技的不断发展，对于便利生活的科技需求不断增加，人机交互应运而生。根据社会需求，市场上出现许多语音助手。而我们也尝试做一个关于南宁职业技术学院的语音交互助手。它与市场上常见的语音助手不同的是，我们给予了这个系统一个南宁职业技术学院在读生的身份，并拥有基于电子显示屏的可视化形象。它结合了百度语音识别 API，百度语音合成 API，ChatGLM 生成式对话模型，Unity 引擎等技术的一款产品。通过百度语音识别 API 对同学的语音提问进行识别转换成文字，再由 ChatGLM 生成式对话模型对问题进行回答，通过百度语音合成进行语音输出。Unity 引擎对人物形象进行修改。

**关键词：**百度语音识别 API；百度语音合 API；ChatGLM 生成式对话模型；Unity 引擎

## Abstract

With the continuous development of technology, there is an increasing demand for convenient living and human-computer interaction. In response to this social need, many voice assistants have emerged in the market. We have also attempted to create a voice interactive assistant specifically for Nanning Vocational Technical College. What sets this system apart from common voice assistants on the market is that we have given it the identity of a student currently enrolled at Nanning Vocational Technical College, complete with a visual representation on an electronic display. This product combines technologies such as Baidu Speech Recognition API, Baidu Speech Synthesis API, ChatGLM generative dialogue model, and Unity engine. The Baidu Speech Recognition API is used to recognize and convert students' voice questions into text, which are then answered by the ChatGLM generative dialogue model. The Baidu Speech Synthesis API is used to output the answers as speech. The Unity engine is used to modify the character representation.

## 目录

Abstract .....	3
一. 概论 .....	5
1.1 项目背景 .....	5
1.2 项目发展现状及趋势 .....	7
1.3 项目设计目标及背景 .....	7
1.4 项目设计依据 .....	8
二.需求分析 .....	8
2.1 项目环境分析 .....	8
2.2 项目目标受众群分析 .....	8
2.3 需求分析总结 .....	8
三. 国内外先进案例对标 .....	9
3.1 国外案例分析 .....	9
3.2 国内案例分析 .....	10
3.3 国内外案例分析总结 .....	11
四. 项目解决方案 .....	12
4.1 项目设计思路 .....	12
4.2 项目总体框架 .....	12
4.3 项目功能分析 .....	13
五. 项目系统设计 .....	13
5.1 系统目标 .....	13
5.2 后台服务器 .....	14
5.2.1 ChatGLM .....	14
5.2.2 百度 TTS & STT .....	16
5.3 系统界面 .....	19
5.3.1 VRoid 建模 .....	19
5.3.2 Unity 引擎 .....	21
5.4 系统特色 .....	22
5.4.1 ChatGLM .....	22
5.4.2 百度语音合成（TTS）特色: .....	23

---

5.4.3 百度语音识别（STT）特色： .....	23
5.4.4 Unity 引擎 .....	24
总结 .....	25
致谢 .....	26
参考文献 .....	26

## 一. 概论

### 1.1 项目背景

数字生命系统的创造背景可以追溯到计算机科学和人工智能领

域的发展历程。在 20 世纪 40 年代至 50 年代，计算机科学快速发展，计算机硬件和软件技术的提升使得计算机能够处理更复杂的任务。随后，人工智能成为计算机科学的重要研究方向，20 世纪 50 年代末和 60 年代初，研究人员开始尝试模拟人类的智能行为和思维过程，涉及的技术包括机器学习、模式识别和推理等。

与此同时，计算机的处理能力和存储容量不断增强，为数字生命系统的创造提供了基础。这种增强的计算能力使得计算机能够处理更复杂和计算密集的任务。在计算能力的推动下，生物学和进化生物学的研究也取得了突破，人们对生物体的结构、行为和进化机制有了更深入的了解，这为数字生命系统的建模提供了理论基础。

同时，20 世纪 80 年代开始发展的虚拟现实和仿真技术为数字生命系统的创造提供了重要平台。虚拟现实和仿真技术使得人们能够创造和探索虚拟的环境和实体，包括数字生命体和生态系统。利用虚拟现实和仿真技术，研究人员能够以一种全新的方式观察和研究数字生命系统的行为和交互。

数字生命系统的创造是建立在这样的背景之上。它为科学研究提供了一个模拟和研究生命现象的平台，帮助科学家和研究人员更好地理解生态系统、进化过程以及生命的基本原理。同时，数字生命系统也在教育和学习领域发挥着重要作用，作为一个互动和沉浸式的学习工具，它帮助学生更好地理解生命现象，并提供实践经验和动手能力。

## 1.2 项目发展现状及趋势

目前，语音交互技术在中国的应用已经非常广泛，涵盖了智能音箱、智能助理、智能客服、智能驾驶、教育和医疗等众多领域。随着技术的不断进步，语音交互将进一步融入我们的生活，并为我们带来更多便捷和智能化的体验。

中国政府制定了《新一代人工智能发展规划》，明确了人工智能的发展目标和政策导向。其中包括加快智能语音技术和应用的研究，推动语音助手的创新和普及。中国政府对于语音助手的发展给予了政策支持，并制定了一系列政策措施，以促进其技术研发、规范管理和市场应用。政策支持有助于推动语音助手在中国市场的普及和发展。总体来看，语音助手在政策支持下的需求及门类会越来越多，实现对便利生活、互联互通、更优化的技术集成和更深入的智慧服务趋势发展。

## 1.3 项目设计目标及背景

本项目主要设计了一款基于语音识别的问答助手，是一款专门为帮助同学便利校园生活而设计的，可以向她提出关于学习、学校或者个人困惑的问题，能够解决同学们向她提出的问题。它结合了百度语音识别 API；百度语音合 API；ChatGLM 生成式对话模型；Unity 引擎等技术的一款产品，通过百度语音识别 API 对同学的语音提问进行识别转换成文字，再由 ChatGLM 生成式对话模型对问题进行回答，通过百度语音合成进行语音输出。Unity 引擎对人物形象进行修改。

## 1.4 项目设计依据

Designing Voice User Interfaces: Principles of Conversational Experiences (作者: Cathy Pearl)

# 二.需求分析

## 2.1 项目环境分析

语音交互助手市场将继续保持强劲的增长，并在智能音箱、智能手机等领域逐渐成为主流。随着技术的进一步发展和用户需求的不断增加，语音助手将在更多垂直应用领域发挥其巨大潜力，并为用户提供更好的交互体验和服务。

## 2.2 项目目标受众群分析

语音交互助手的目标群体包括使用智能手机和智能音箱的个人用户、商业场景中的企业和服务提供商、智能家居用户和驾驶者。随着技术的进步和用户需求的增加，语音交互助手的应用和目标群体还将继续扩大。

## 2.3 需求分析总结

语音交互助手是一种创新的技术和产品，市场潜力巨大。市场规模正在快速增长，预计在未来几年将继续保持强劲的增长。消费者的



采用率也在增加，越来越多的人开始使用智能手机、智能音箱等设备上的语音助手。

## 三. 国内外先进案例对标

### 3.1 国外案例分析

Amazon Alexa: Alexa 是亚马逊推出的语音助手，被嵌入了多种设备，如 Echo 智能音箱和 Fire TV。用户可以通过语音与 Alexa 进行互动，查询天气、播放音乐、控制智能家居设备等。Alexa 还提供了 Skills 功能，允许第三方开发者为其添加各种应用和服务。

Google Assistant: Google Assistant 是谷歌开发的语音助手，搭载在谷歌的智能手机、智能音箱等设备上。用户可以通过语音与 Assistant 进行交流，获取信息、设置提醒、查看日历等。Google Assistant 还具有人工智能的特点，能够根据用户的需求和上下文提供个性化的回应。

Apple Siri: Siri 是苹果公司的语音助手，在苹果设备上广泛使用。用户可以使用语音命令与 Siri 进行交流，如发送短信、提醒、查找信息等。Siri 还与其他苹果应用程序和服务整合，为用户提供更方便的操作体验。

Microsoft Cortana: Cortana 是微软开发的语音助手，搭载在 Windows 和 Windows 手机等设备上。用户可以通过语音与 Cortana 进

行对话，查询信息、设置提醒、发送电子邮件等。Cortana 还能够学习用户的习惯，并提供个性化的建议和推荐。

**Samsung Bixby:** Bixby 是三星公司的语音助手，可以通过语音和图像进行交互。用户可以使用语音指令来控制三星设备，如发送消息、拍照、调整设置等。Bixby 还提供了可自定义的快捷命令功能，让用户更好地个性化设置。

### 3.2 国内案例分析

**小爱同学 (XiaoAI):** 小爱同学是小米公司推出的语音助手，嵌入在小米智能设备中，如小米手机、小米电视和小米智能音箱等。用户可以通过语音与小爱同学进行交互，从天气查询、音乐播放到智能家居控制等多个领域都得以支持。小爱同学通过深度学习和大数据分析为用户提供个性化的服务与建议。

**百度度秘 (Baidu DuerOS):** 百度度秘是百度公司开发的语音助手，广泛应用于百度智能音箱、手机、汽车等产品中。度秘通过语音交互提供多样化的服务，如语音搜索、音乐播放、电影票订购等。而且百度还开放了 DuerOS 开放平台，允许第三方开发者开发自己的技能，丰富度秘的功能和应用。

**阿里云天猫精灵 (Tmall Genie):** 阿里云天猫精灵是阿里巴巴公司推出的智能语音助手，主要嵌入在天猫精灵智能音箱中。用户可以通过语音与天猫精灵进行互动，实现智能家居控制、购物、音乐播放等多个功能。天猫精灵还拥有语音识别、语义理解和对话管理等强大

的技术能力，提供高质量的语音交互体验。

**腾讯小微（Tencent Xiaowei）：**腾讯小微是腾讯公司推出的智能语音助手，主要应用于腾讯智能音箱和腾讯手机等产品中。用户可以通过语音与小微进行交互，包括查询天气、播放音乐、控制智能家居设备等多个功能。小微还具备智能聊天和情感识别的能力，能够更好地理解和适应用户的需求。

### 3.3 国内外案例分析总结

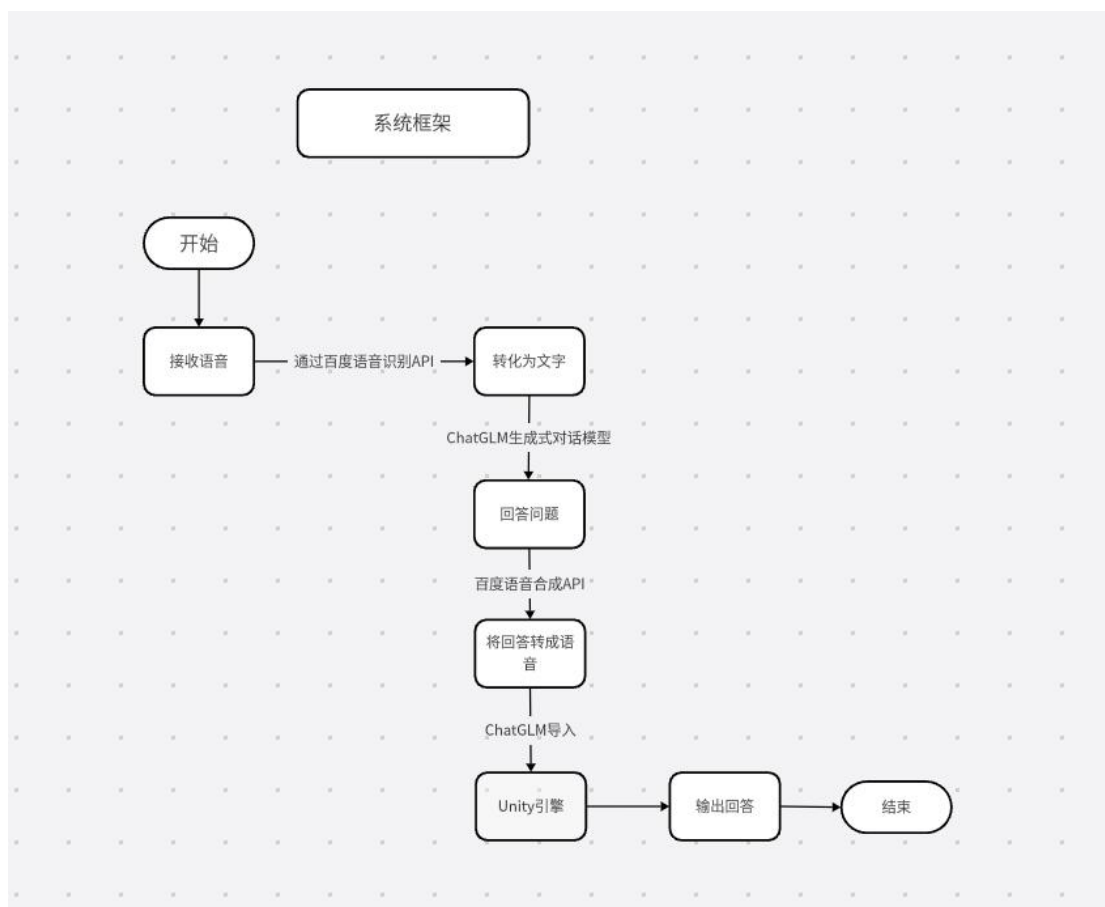
语音交互系统已经渐入人们的视野，被人们所接受所推崇，更便捷更经济的语音交互成为人们迫不及待的需求。语音交互系统的智能化直接影响到用户的使用体验。那么语音交互就要借助于各种各样的信息进行训练并测试，提高对于问题的灵敏性，回答时的精确性。对问题回答的反应时间控制在一个拟人的一个合理反应时间内，为用户提高使用体验感。不仅比传统的工作方式节省时间，还让人们能够像和人类交流一样自然地与设备进行交互，无需使用键盘、鼠标或触摸屏等传统输入设备。这种自然的交互方式让用户更容易接受和使用，特别是对于那些不熟悉或无法使用传统输入设备的人群来说更加便利，也将会成为人们生活必需品。

## 四. 项目解决方案

### 4.1 项目设计思路

通过我们在校学习的物联网知识根据需求分析和先进案例分析来设计基于 ChatGLM 的人工智能对话系统，它结合了百度语音识别 API 将提出的问题语音转化为文字，ChatGLM 生成式对话模型对问题进行回答，百度语音合成 API 将回答转换为语音，Unity 引擎制作人物模型等技术的一款产品。用户可通过显示屏查看历史问题及回答，是一款方便并用法简便的对话系统。

### 4.2 项目总体框架



### 4.2.1 系统结构框架图

## 4.3 项目功能分析

1. 百度语音识别 API： 识别用户语音，转换为文字
2. ChatGLM 生成式对话模型： 对用户提出的问题进行回答
3. 百度语音合成 API： 将回答的文字转为语音输出
4. Unity 引擎： 提供一个便于交流的界面

# 五. 项目系统设计

## 5.1 系统目标

在当今数字时代，智能对话系统正在迅速崛起，成为信息获取和日常沟通的关键工具。本项目的愿景是通过高效的语言模型微调技术，打造一个富有知识深度和人物设定的虚拟角色——“知微学姐”。这个虚拟角色的设立旨在提供用户更加亲切、专业的在线咨询体验，将智能对话系统推向一个新的高度。

在实现与 AI 进行语音和文字对话的同时，“知微学姐”还能通过语音回答用户的问题，为对话增加更为生动的交互性。更进一步，她能够做出相应的动作，使得对话过程更加富有趣味性和真实感。这种全方位的互动体验将用户带入一个仿佛在与真实人物交流的感觉，为在

线咨询提供了一种全新的、更加人性化的形式。

通过这一项目，我们将数字智能与人性化交互有机结合，为用户创造一个既充满知识深度又富有趣味性的对话伙伴，提升在线咨询体验的质量和深度。

## 5.2 后台服务器

### 5.2.1 ChatGLM

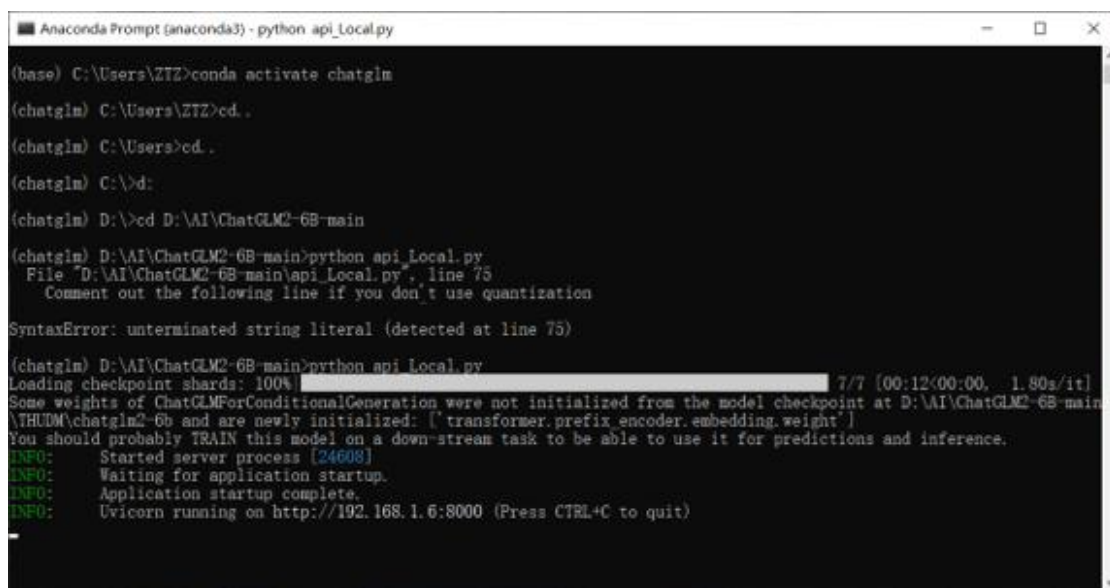
1. ChatGLM 是由清华大学 KEG 实验室和智谱 AI 公司于 2023 年共同训练的语言模型 GLM-130B 开发而成，它采用了深度学习技术，能够对自然语言文本进行建模，并且具备语言生成和对话的能力。ChatGLM-6B 是一个开源的、支持中英双语的对话语言模型，基于 General Language Model (GLM) 架构，具有 62 亿参数。ChatGLM-6B 使用了和 ChatGPT 相似的技术，针对中文问答和对话进行了优化。经过约 1T 标识符的中英双语训练，辅以监督微调、反馈自助、人类反馈强化学习等技术的加持，62 亿参数的 ChatGLM-6B 已经能生成相当符合人类偏好的回答。为了方便下游开发者针对自己的应用场景定制模型，GLM 同时实现了基于 P-Tuning v2 的高效参数微调方法，INT4 量化级别下最低只需 7GB 显存即可启动微调。

2. 使用 Windows10 平台的计算机进行 ChatGLM2-6B 大语言模型的部署，可以使用带有 CUDA 核心的英伟达 GPU 部署与 CPU 部署两种方式，演示电脑基于英伟达 GeForce RTX 4070 12GB 部署，CUDA 版本

为 12.2，安装有 Anaconda3，并创建名为 chatglm 的环境。下载项目模型导入项目文件夹内。

3. 对 ChatGLM 进行微调。选择 ChatGLM2-6B 作为基础模型的原因在于其强大的语言理解和生成能力。通过 P-Tuning v2 的微调，我们不仅能够精细化调整模型以契合“知微学姐”的角色设定，还能大幅降低运行模型所需的硬件资源，使得项目具有更广泛的应用潜力。

P-Tuning v2 作为一种高效的参数微调技术，通过引入可训练的 prompt 参数，极大地减少了传统微调中需要调整的模型参数量。结合模型量化技术，我们进一步减少了运行模型所需的内存。通过 Gradient Checkpoint 等先进的内存优化技术，最终实现了仅需 7GB 显存的轻量化运行环境。项目组精心编写了一系列训练数据集，涵盖了与“知微学姐”设定相关的知识点和对话场景。数据集的编写考虑了角色的背景、语言风格和专业知识，确保“知微学姐”能够以真实且合情合理的方式进行交流。未来，我们计划引入更多的个性化和上下文适应性训练，以及利用用户反馈进行持续的模型优化，进一步增强“知微学姐”的对话能力和用户体验。



```
Anaconda Prompt (anaconda3) - python api_Local.py
(base) C:\Users\ZTZ>conda activate chatglm
(chatglm) C:\Users\ZTZ>cd.
(chatglm) C:\Users\ZTZ>cd.
(chatglm) C:\Users\ZTZ>cd.
(chatglm) D:\>cd D:\AI\ChatGLM2-6B-main
(chatglm) D:\AI\ChatGLM2-6B-main>python api_Local.py
File "D:\AI\ChatGLM2-6B-main\api_Local.py", line 75
    Comment out the following line if you don't use quantization
SyntaxError: unterminated string literal (detected at line 75)
(chatglm) D:\AI\ChatGLM2-6B-main>python api_Local.py
Loading checkpoint shards: 100% 7/7 [00:12<00:00, 1.80s/it]
Some weights of ChatGLMForConditionalGeneration were not initialized from the model checkpoint at D:\AI\ChatGLM2-6B-main\THUDM\chatglm2-6b and are newly initialized: ['transformer.prefix_encoder.embedding.weight']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
INFO: Started server process [24608]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: Uvicorn running on http://192.168.1.6:8000 (Press CTRL+C to quit)
```

### 5.2.1.1 ChatGLM 启动后的界面

## 5.2.2 百度 TTS & STT

### 关于百度语音转文字（STT）

百度语音转文字（STT）技术的卓越性体现在多个方面，将其打造成一款业界领先的语音识别工具。首先，其采用了国际领先的流式端到端语音语言一体化建模方法，结合了百度自然语言处理技术，使得在近场中文普通话识别中达到了惊人的 98% 准确率。这不仅代表了该技术的高度成熟，更意味着用户能够享受到极其准确和流畅的语音转文字体验。

百度 STT 提供了自助训练专属模型的独特功能，用户可以在语音自训练平台上进行模型训练，只需上传词汇文本即可完成零代码训练。这种灵活性使得用户能够针对具体业务领域进行模型优化，实现



5-20%的词汇识别率提升，并且模型还可以专属使用，更好地适应项目的个性化需求。

在接入方面，百度 STT 支持 API 及多种 SDK 接入，用户可以基于 Demo 快速简便地完成接入。采用最新的识别解码技术，该工具在保持高准确率的同时，大幅度提升了识别速度。这种简单快速的接入方式，为开发者提供了更加高效便捷的开发体验。

对于企业级用户而言，百度 STT 提供了专有服务集群，确保企业级的稳定服务。其弹性灵活的高并发承载及高可靠性保障，为用户提供了无忧的使用体验。这种高效稳定的特性使得百度 STT 成为企业在语音转文字领域的理想选择，无论是在产品开发、客户服务还是其他应用场景中都能够发挥其强大的作用。

### 关于百度语音合成（TTS）

百度语音合成（TTS）技术的卓越性在于其多方面的强大功能，使其成为一款广泛应用于多语言场景的语音合成工具。首先，它支持多语言多音色，包括中文、英文以及中英文混读合成。百度 TTS 提供了基础音库和精品音库，总计 14 种音库供用户选择，为产品提供了更加个性化的声音选择。这种多样性的音库选项为用户提供了更大的自由度，使其能够根据产品的定位和用户喜好选择最合适的声音。

在语音参数方面，百度 TTS 支持多种参数配置，用户可以根据不同场景的需求对发音人的语速、音调、音量进行灵活的设置。这使得用户能够实现更加精细化的个性化需求，确保语音合成的效果更符合

项目的整体氛围。这种灵活的参数配置为用户提供了更加定制化的语音合成体验。

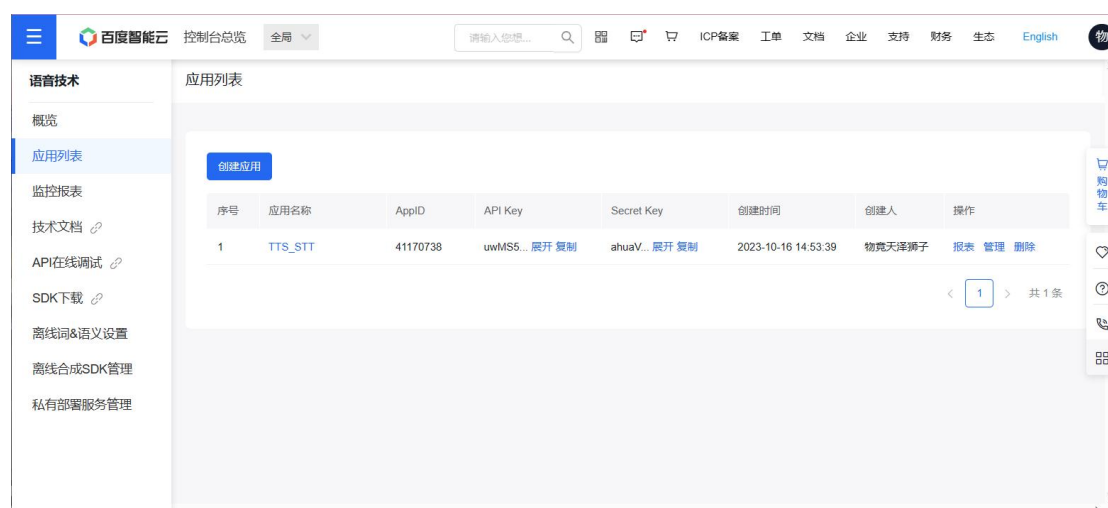
此外，百度 TTS 还支持多音字标注，用户可以通过标注拼音和音调的方式自行定义多音字的发音。这项功能为处理中文中常见的多音字问题提供了解决方案，确保语音合成更加准确、自然。

在集成方式方面，百度 TTS 提供了 REST API 接口、离线 SDK、在线 SDK 等多种集成方式，充分满足不同场景下的语音合成需求。无论是在线应用还是离线场景，百度 TTS 都能提供流畅自然的合成体验。这种灵活而丰富的集成方式使得百度 TTS 成为各种应用场景下语音合成的首选工具。

申请百度 TTS & STT 的 API 用于接入项目实现对话功能百度的 TTS API 为项目提供了强大的文本转语音功能，通过这一技术，我们能够为用户提供更自然、流畅的语音输出，使得对话更具真实感和亲和力。这对于改进项目的用户体验和提升交互性至关重要。通过百度 TTS API，我们可以借助先进的语音合成技术，为用户提供高质量、多样化的语音输出。

同时，百度 STT API 的引入将使得项目具备先进的语音识别能力，将用户的语音输入迅速准确地转化为文本。这项技术对于实现更自然、方便的对话界面至关重要。用户能够通过语音进行交互，而项目则能够迅速理解并作出相应的响应，从而提升整体的用户友好性和便捷性。通过整合百度 TTS 和 STT API，项目将获得更强大的语音交互能力，为用户提供更加智能、个性化的体验。这不仅提升了项目的功能性，

还在很大程度上满足了用户对于更自然交流方式的需求。借助这些先进的语音技术，我们将为项目赋予更高层次的人机交互能力，进一步推动其在市场上的竞争力。



### 5.2.2.1 百度 API

## 5.3 系统界面

### 5.3.1 VRoid 建模

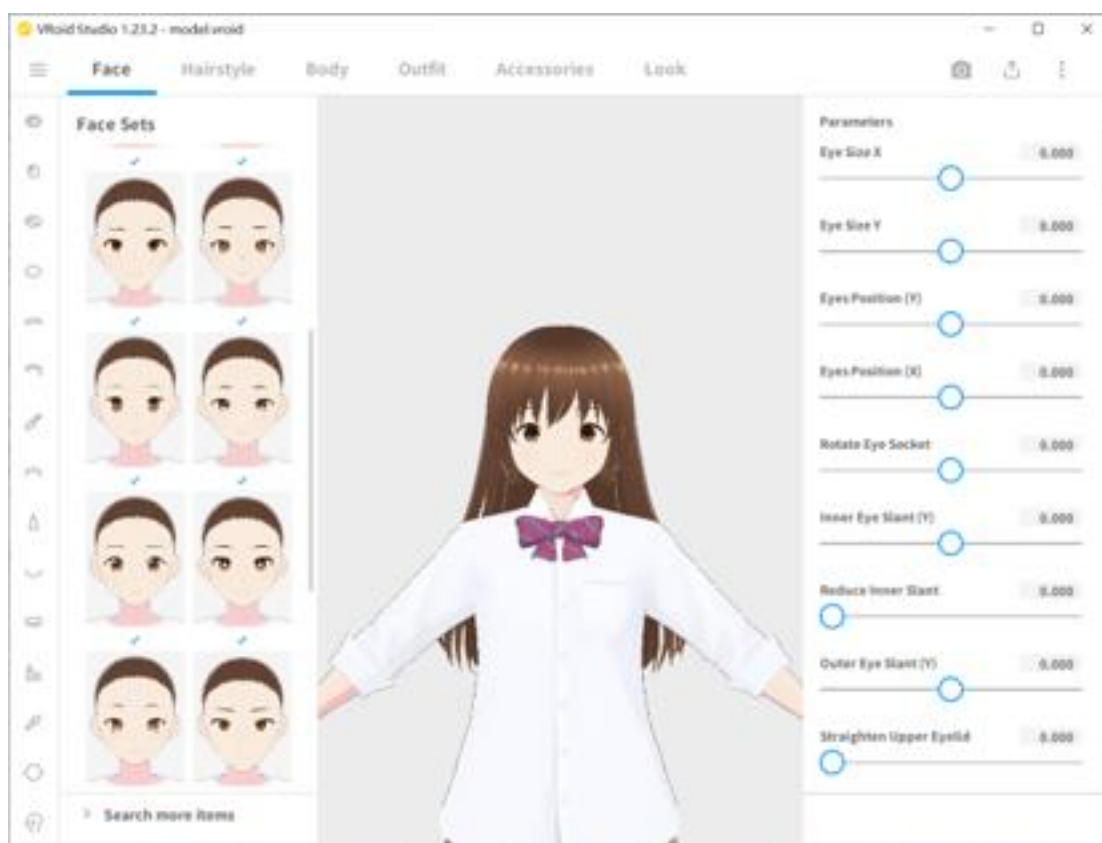
VRoid Studio, 是 pixiv 于 2018 年 3 月 29 日推出的一款免费软件, 旨在为画师提供更为便捷的 3D 角色建模工具。通过这一工具, 画师能够更轻松而直观地塑造独特的 3D 角色形象。

在 VRoid Studio 中进行建模, 不仅仅是技术的展示, 更是一场关于创意和想象力的冒险。画师们可以通过简单的操作, 灵活地改变角

色的外貌，塑造出符合需求的理想形象。这一软件不仅提供了广泛的定制选项，还为用户提供了丰富的素材库，使得创作者能够更加轻松地表达他们的创意和独特视觉。

VRoid Studio 的推出，不仅仅是为了简化 3D 建模的技术流程，更是为了激发创作者们的无限创意。画师们不再受限于复杂的技术操作，而是能够专注于展现他们的独特美学和创意构想。这一工具的免费提供，更是为绘画社群带来了更广阔的创作空间。

总体而言，VRoid Studio 的问世为画师们提供了一个富有创意和自由度的平台，让他们能够更加便捷地打破传统，创造出属于自己的独特世界。这一软件的涌现，无疑推动了数字艺术领域的发展，让创作者们更加轻松地将他们的想象力转化为视觉的奇迹。



### 5.3.1.1VRoid 界面

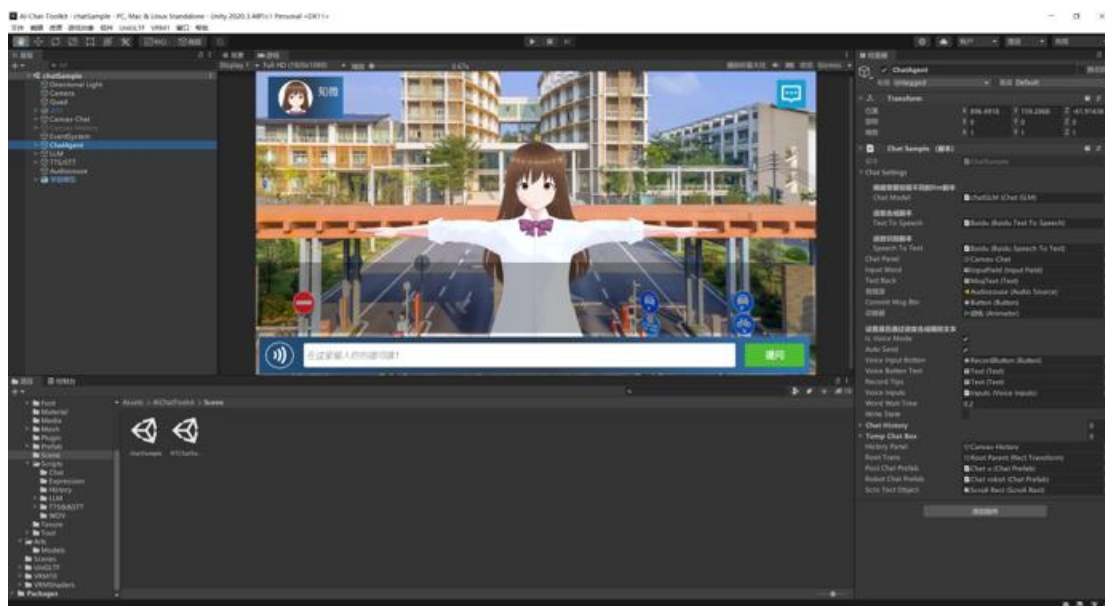
### 5.3.2 Unity 引擎

Unity 是一款强大的实时 3D 互动内容创作和运营平台，为各领域的创作者提供了一个统一的创作工具，涵盖游戏开发、美术设计、建筑规划、汽车设计、影视制作等多个领域。Unity 的使命在于将创意转化为现实，为创作者们提供一个全面的解决方案。

这一平台为创作者们提供了丰富而完整的软件工具，可用于创作、运营和变现各类实时互动的 2D 和 3D 内容。无论是在手机、平板电脑、PC、游戏主机，还是在增强现实和虚拟现实设备上，Unity 都提供了全方位的支持。这使得创作者们能够在各种平台上呈现出统一而高质量的创意作品。

在项目中，通过利用 Unity 引擎制作界面，对项目的界面进行美化和调整，可以提升用户体验，使其更为简单易用。这种界面的优化不仅仅是外观的提升，更是为了确保用户能够更顺畅地使用工具，从而更有效地表达他们的创意。通过精心设计的界面，Unity 引擎为创作者们提供了一个直观而高效的工作平台，使其能够更专注地将创意付诸实践。

总体而言，Unity 作为一款综合性的创作平台，不仅连接了各个创作者的创意和实现，还通过界面的美化和调整，进一步提升了创作过程的便捷性和愉悦感，为实时互动内容的创作提供了强有力的支持。



### 5.3.2.1 项目的 Unity 界面

## 5.4 系统特色

### 5.4.1 ChatGLM

**自然语言处理：**ChatGLM 具有先进的自然语言处理能力，能够理解和处理人类语言表达的含义。它可以识别和解析句子的结构和语义，从而更准确地回答用户的问题。

**上下文理解：**ChatGLM 能够记住对话的上下文，以更好地理解用户的意图。它能够根据之前的对话内容提供连贯而连贯的回答，并确保回答与上下文相关。

**多领域覆盖：**ChatGLM 在多个领域上都有广泛的知识覆盖，包括科学、技术、历史、艺术等。无论是什么主题的问题，它都可以提供

相关的信息和答案。

**创造性回答：**ChatGLM 还具有一定的创造性，它可以提供独特和有趣的回答，而不仅仅是简单的事实陈述。这使得与 ChatGLM 的对话更为有趣和有趣。

#### 5.4.2 百度语音合成（TTS）特色：

**多语言多音色支持：**百度 TTS 支持中文、英文以及中英文混读的语音合成，提供基础音库和精品音库，共计 14 种音库，为用户提供更多个性化的声音选择。

**语速音调可调节：**提供多种参数配置，用户可以根据场景需求调整发音人的语速、音调和音量，以满足个性化的需求。

**多音字标注：**中文多音字可以通过标注拼音和音调的方式进行自定义发音，提高对多音字的准确识别。

**集成方式灵活：**提供 REST API 接口、离线 SDK 和在线 SDK 等多种集成方式，以满足不同场景下的语音合成需求，确保提供流畅自然的合成体验。

#### 5.4.3 百度语音识别（STT）特色：

**技术领先：**使用领先国际的流式端到端语音语言一体化建模方法，融合了百度自然语言处理技术，实现近场中文普通话识别准确率达到 98%。

**自助训练专属模型：**支持在语音自训练平台上进行模型训练，上

传词汇文本即可零代码完成训练，以提高业务领域词汇识别率 5-20%。

**简单快速：**支持 API 及多种 SDK 接入，可基于 Demo 快速简单接入，采用最新识别解码技术，提高识别速度。

**高效稳定：**提供专有服务集群，具备企业级的稳定服务，支持弹性灵活的高并发承载，以及高可靠性保障。

#### 5.4.4 Unity 引擎

**跨平台支持：**Unity 引擎支持多个平台，包括 PC、移动设备、主机游戏机和虚拟现实（VR）和增强现实（AR）设备等。开发者可以使用一套代码和资源创建适用于不同平台的游戏和应用程序，极大地提高了开发效率和可移植性。

**强大的编辑器和工具：**Unity 提供了一个直观且功能丰富的可视化编辑器，使开发者能够轻松地创建、编辑和组织场景、游戏对象和资源。它还提供了大量的工具和件，用于优化性能、调试和实时预览。

**灵活的脚本语言和 API：**Unity 支持多种脚本语言，包括 C#、JavaScript 和 Boo。开发者可以使用其所熟悉的语言编写游戏逻辑和交互行为。此外，Unity 还提供了丰富的 API 库，用于处理图形、物理、音频、网络和其他方面的功能。

**内置的物理引擎：**Unity 集成了一款强大的物理引擎，可以模拟真实的物理行为，如碰撞、重力、刚体动力学等。开发者可以使用它来创建逼真的物理效果，从而增强游戏的真实感和交互性。



## 总结

1. 对话处理：使用 ChatGLM 作为对话系统的核心，接收用户输入的文本并生成相应的回复。ChatGLM 可以通过自然语言处理技术理解和解析用户的意图，并提供准确的回答。
2. 语音合成（TTS）：将 ChatGLM 生成的回复文本转换为语音输出，可以使用百度 TTS 服务将文本转化为自然语音。百度 TTS 提供了多种合成音色和语言选择，可以根据需求选择合适的音色和语言。
3. 语音识别（STT）：将用户输入的语音转换成文本，以便传递给 ChatGLM 进行处理。可以使用百度 STT 服务将语音转化为文本。百度 STT 利用语音识别技术，可以将用户输入的语音准确地转换为文本。
4. Unity 引擎集成：将 ChatGLM、百度 TTS 和 STT 集成到 Unity 引擎中。用户可以通过 Unity 应用程序与对话系统进行交互，并实现语音和文本输入输出的功能。
5. 用户界面设计：在 Unity 应用程序中设计用户界面，提供可视化的输入和输出界面，使用户可以通过文本输入或语音输入与对话系统进行交互，并获取文本或语音回复。可以使用 Unity 的 UI 系统和音频系统来实现用户界面的设计和显示。

通过上述功能的组合，这个基于 ChatGLM、百度 TTS 和 STT 及 Unity 引擎的对话系统可以实现用户输入的文本或语音的解析和回复，并通过语音合成和语音识别技术实现语音的输入和输出。用户可以通过 Unity 应用程序与对话系统进行交互，并获取回复的文本或语音结果。

## 致谢

## 参考文献

- [1] ChatGLM 项目: <https://github.com/thudm/chatglm2-6b>
- [2] Hugging Face 大语言模型: <https://huggingface.co/THUDM/chatglm2-6b>
- [3] Unity 项目: <https://gitee.com/DammonSpace/unity-ai-chat-toolkit>