

2029320381452802095400

**数 字 媒 体 技 术 导 论 大 作 业 报 告**

利用人工智能进行角色（初音未来）模拟陪伴

第2组

学 院： 信息学院

组 长：马 鑫 37220232203780

组 员 1：

组 员 2：

组 员 3：

组 员 4：王婧晋 37220232203835

指导教师：吴清强教授

2025 年 1 月 10 日

目录

[一、项目背景 3](#_Toc187440534)

[市场需求与趋势 3](#_Toc187440535)

[初音未来 4](#_Toc187440536)

[二、实现技术 5](#_Toc187440537)

[技术1：LLM大模型接口 5](#_Toc187440538)

[技术2：Live2D接入 6](#_Toc187440539)

[三、项目总结 8](#_Toc187440540)

# 一、项目背景

在当今科技飞速发展的时代，人工智能（AI）的应用领域不断拓展，其中包括艺术创作和情感陪伴领域。初音未来作为一位极具影响力的虚拟歌手和动漫角色，拥有庞大的粉丝群体和广泛的市场影响力。利用人工智能技术进行初音未来陪伴模拟，不仅可以为粉丝提供全新的互动体验，还能探索 AI 在情感陪伴和艺术创作中的应用潜力。因此，我们组在本项目中尝试利用AI对虚拟偶像初音未来进行模拟。

## 市场需求与趋势

AI陪伴产品市场正迅速崛起，其商业潜力巨大，吸引了众多企业和投资者的关注。这些产品通过模拟人类情感交流，满足了人们在情感陪伴、娱乐互动等方面的需求。当前，市场上已出现虚拟男友/女友、虚拟故事角色、AI心理顾问等多种形式的AI陪伴产品，用户接受度高达98%。

技术进步是推动该市场发展的关键因素。多模态大模型的推出，使AI陪伴产品在理解用户意图、表达情感等方面的能力显著提升。未来，随着长期记忆、主动想象等技术的突破，产品功能将更加完善，用户体验也将大幅提升。

市场规模方面，预计到2027年，国内AI陪伴市场将达到千亿规模。虚拟故事角色、AI心理顾问等细分市场潜力巨大。全球市场也呈现出快速增长的趋势，预计到2030年将达到62亿美元，亚太地区作为新兴市场，增长潜力尤为显著。



## 初音未来

初音未来作为一位虚拟歌手和动漫角色，不仅在音乐领域取得了巨大成功，还在动漫、游戏、周边产品等多个领域拥有广泛的影响力。其粉丝群体遍布全球，具有高度的忠诚度和活跃度。利用 AI 技术进行初音未来陪伴模拟，可以进一步拓展其商业价值和文化影响力。



# 二、实现技术

## 技术1：LLM大模型接口

原项目在文本生成方面，最初采用的是ChatGPT接口。然而，鉴于国内政策限制，免费获取ChatGPT的API Key存在诸多不便，因此我们转而使用国内的大模型deepseek作为替代方案。尽管从基本语法来看，deepseek与ChatGPT保持了一定程度的兼容性，但在发送请求时所用的数据结构却有所不同。具体来说，向ChatGPT发送请求时，我们采用的是简洁的string字符串格式来传递提示词；而向deepseek发送请求时，则需要使用更为复杂的List<Dictionary<string,string>>结构，即一个由字符串字典构成的链表来传递提示词。这要求我们在获取和输入文本的代码层面进行相应的调整和修改。

在语音合成环节，原项目所依赖的接口是Azure。虽然Azure提供了免费额度和学生认证等优惠政策，但其注册流程需要使用Visa卡或万事达卡，这对国内用户而言存在一定的门槛和不便。因此，我们决定采用国内的百度AI大模型来替换原有的Azure接口。百度AI不仅能够满足项目在语音合成方面的功能需求，同时也避免了因支付方式受限而带来的问题，确保了项目的顺利推进和实施。



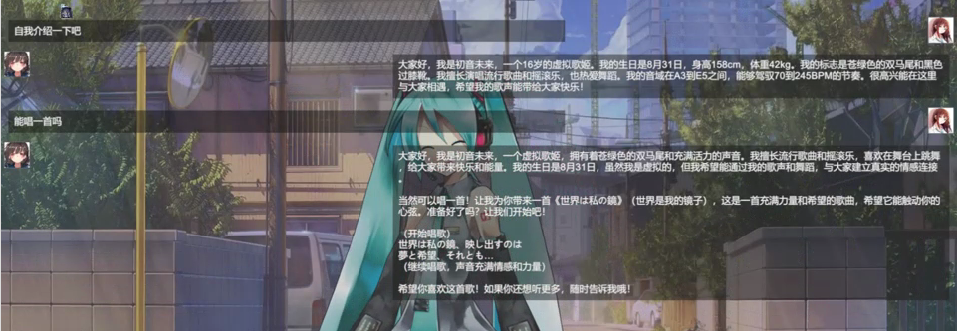
## 技术2：Live2D接入

通过在Live2D官网下载SDK并导入Unity中，我们成功地将免费的Live2D模型“初音未来”引入到项目里。借助Unity强大的动画系统，我们精心制作了角色动画，让“初音未来”在屏幕上栩栩如生。使其能够较为生动地做出一些表情。此外，我们将deepseek的系统提示词巧妙地调整为与“初音未来”相关的设定，实现了与这位虚拟偶像的对话互动。



然而，在语音合成环节，我们遇到了一些限制。由于技术原因，我们无法使用miku的音源，这使得“初音未来”的声音与形象无法完美契合，影响了角色的整体表现力。同时，当前的语音合成系统仅支持中文识别，对于需要多语言交互的场景来说，无疑是一个较大的局限，严重影响了用户的使用体验，限制了项目的受众范围和应用场景的拓展。



# 三、项目总结

通过学习与实现本项目，我们一定程度上对现在的人工智能技术如图像识别、音频生成等有了更深入、更全面的了解。同时对大语言模型有了更详细的认知，了解了训练大语言模型的过程。在实现初音未来形象的逼真呈现和智能化交互方面，我们面临着一些技术难题。例如，角色建模和渲染的细节处理不够完美，动作和表情的自然度有待提高；在复杂场景下的文字识别和指令识别准确率还有待提升，要实现回复的自然与口语化还需要更多训练。