



第十七届全国大学生软件创新大赛

文档编号: SWC2024-哇哈哈



TA-交互式陪伴与抚慰平台

TA- Interactive companion and comfort platform

项目测试文档

Version: 1.0.0



哇哈哈

2024-3-21

All Rights Reserved

目录

1	测试计划	1
1.1	测试策略与目标	1
1.1.1	测试策略	1
1.1.2	测试目标	1
1.2	测试范围	2
1.3	测试环境	3
2	单元测试	5
2.1	再见一面模块	5
2.1.1	测试用例与结果分析	5
2.1.2	测试结果综合分析及建议	8
2.1.3	测试经验总结	9
3	功能测试	10
3.1	预生成功能	10
3.1.1	测试用例与结果分析	10
3.1.2	测试结果综合分析及建议	11
3.1.3	测试经验总结	11
3.2	语音对话功能	11
3.2.1	测试用例与结果分析	11
3.2.2	测试结果综合分析及建议	12
3.2.3	测试经验总结	12
3.3	视频对话功能	13
3.3.1	测试用例与结果分析	13
3.3.2	测试结果综合分析及建议	14
3.3.3	测试经验总结	14
4	系统测试	15
4.1	模型性能测试	15
4.1.1	测试用例与结果分析	15
4.1.2	测试结果综合分析及建议	15
4.1.3	测试经验总结	15

文档修订历史

序号	修订原因	版本号	作者	修订日期	备注
1	总结测试策略	V0.1.0	队员 1	2024.1.20	
2	更新单元测试	V0.2.0	队员 2	2024.3.10	
3	更新功能测试	V0.3.0	队员 2	2024.3.17	
4	格式修订与定稿	V1.0.0	队员 1	2024.3.23	

1 测试计划

1.1 测试策略与目标

1.1.1 测试策略

- 测试重点：
再见一面等功能的稳定性、响应特性，AI 小伴、心理咨询师等功能的回复稳定性，并测试是否所有功能都能正常运行
- 测试难点：
需要考虑到尽可能多的使用场景及情况，做到尽可能全面的测试
- 测试分层：
 1. 功能层面：测试正常情况下各功能能否稳定运行，各功能响应时间是否在可接受范围内。
 2. 边界层面：在用户输入超出正常范围时软件运行是否正常，是否有边界检测和保护机制。
 3. 性能层面：在用户数量增多的高并发场景下对软件性能的影响程度如何。

1.1.2 测试目标

在功能上，需要在测试中验证再见一面等功能的稳定性、响应特性，AI 小伴、心理咨询师等功能的回复稳定性，并测试是否所有功能都能正常运行。另外，还需要对各功能进行边界测试。

根据以往程序开发和测试经验，软件应用程序中往往存在预料不到的问题，许多隐藏错误只有在特定环境下才能暴露。因此，我们希望通过搭建内部测试环境，模拟本系统的实际运行环境，通过测试，尽可能多发现并排除本项目中潜藏的错误，证实所实现的产品功能与用户要求一致。通过集成测试、系统测试、性能测试之后，本系统将在真实环境中投入运行。

类别	目标
性能指标	事务响应时间：除再见一面功能外，其余事务响应时间应不超过 500ms
	成功率 SR：不低于 95%
	每秒事务数 TPS：100
执行目标	缺陷清除率：85%
	测试用例覆盖率：80%

	测试用例通过率：100%
--	--------------

1.2 测试范围

类别	测试范围	测试方法
功能测试	页面显示规则、各个功能模块的业务规则的验证，数据精确度、数据类型、业务功能等相关方面的正确性	采用黑盒测试、边界测试、等价类划分等测试方法，手工的方式进行检测
用户界面 UI 测试	APP 的窗口的对象和特征（logo、按钮、输入框、大小、位置、状态）都符合标准 符合操作方便、界面友好等一般性测试原则（布局、配色、字体、易用性等）界面与用户之间进行交互的互动界面测试（如所有业务可能的进入、退出方式，所有可能的按键操作，各类页面间所有可能的跳转、焦点切换、页面、菜单显示的正确性）	手工测试、目测
兼容性测试	使用不同硬件、操作系统、分辨率和各种运行软件等各种条件的组合测试	黑盒测试的方法，进行手工测试核实系统在不同软硬配置中的运行情况
安全性测试	用户、管理员的密码安全权限 用户的文件安全	采用相关的代码包、工具进行手工测试
流程测试	验证本系统的所有业务流程是否畅通，主要功能模块是否正常	手工测试
接口测试	验证本系统所用的各个接	采用 Postman、Jmeter

	口是否能够正常运行，返回预期的结果	模拟调用接口进行测试
性能测试		手工测试、自动化测试

1.3 测试环境

➤ 服务器环境：

资源类型	操作系统	系统配置	软件配置	备注
Web 应用	CentOS 7	CPU:E5-268x 2 核 内存：2G 硬盘：RAID10 50G	JDK 1.8 Docker 25.0.4 Redis 3.2.12 RabbitMQ 3.11	
数据库	CentOS 7	CPU:E5-268x 2 核 内存：2G 硬盘：RAID10 50G	MySQL 5.7	
AIGC 服务端				

➤ 客户端环境：

资源类型	操作系统	系统配置	软件配置	备注
APP	ColorOS 14 (Android 14)	CPU：高通骁龙 888 内存：16G 存储空间：256GB	Web view 90.0.4430.61	

➤ 网络环境：

资源类型	IP 地址	端口号	代理服务器	备注
Web 应用	Inkslab.icu 202.146.220.165	443	无	
数据库	Inkslab.icu 202.146.220.165	3306	无	

AIGC 服务端	内网 IP	8080	202.146.220.165	AIGC 服务端的 IP 地址为内网 IP，因此采用内网穿透的方式提供服务
----------	-------	------	-----------------	---------------------------------------

➤ 测试工具：

1. 对于后端应用程序，采用 **Junit** 进行单元测试，测试函数以及接口
2. 使用 **Postman** 模拟前端，来向后端应用程序接口发送请求，来测试接口是否可以正常工作。
3. 借助 **WireShark** 对网络请求进行抓包、分析，排查网络方面的问题
4. 使用 **JMeter** 对系统进行压力测试，测试系统并发性、可用性、效率等性能。

2 单元测试

2.1 再见一面模块

2.1.1 测试用例与结果分析

单元测试用例：

用例编号	001			
测试单元描述	预生成中能上传文件操作			
用例目的	测试用户能否正常上传文件，能否预防用户上传非法文件			
前提条件	用户已经登录，且具有预生成次数，进入了预生成界面			
特殊的规程说明	无			
用例间的依赖关系	执行预生成用例的必须前置			
具体步骤	输入	期望输出	实际输出	备注
1	合法的业务类型 biz、合法的文件	文件在云存储中的 Key	文件在云存储中的 Key	
2	非法的业务类型 biz、合法的文件	异常信息：业务类型不合法	异常信息：业务类型不合法	
3	合法的业务类型 biz、非法的文件	异常信息：文件不合法	异常信息：文件不合法	具体的异常信息由文件决定，可能有：格式错误、超出大小限制等
4	非法的业务类型 biz、非法的文件	异常信息：业务类型不合法	异常信息：业务类型不合法	

测试结果分析：

测试结果均正常，用户可以正常将文件上传至云存储，且文件与具体的业务类型适配。文件在云存储中的 Key 也合法

单元测试用例：

用例编号	002			
测试单元描述	向消息队列中发送预生成消息			
用例目的	测试后端向消息队列发送消息是否正常			
前提条件	用户已经上传了相关文件，并提交了上传的请求			
特殊的规程说明	需开启消息队列，并创建对应的消息队列以及交换机			
用例间的依赖关系	上传文件之后的后置用例			

具体步骤	输入	期望输出	实际输出	备注
1	格式正确的消息	任务处理模块收到消息的通知	任务处理模块收到消息的通知	
2	格式正确，但是内容不存在的消息	异常：此消息未查询到数据。并拒绝消息	异常：此消息未查询到数据。并拒绝消息	
3	格式错误的消息	异常：消息非法。并拒绝此消息	异常：消息非法。并拒绝此消息	
4	空消息	异常：消息非法。并拒绝此消息	异常：消息非法。并拒绝此消息	

测试结果分析：

测试结果均正常，消息格式正确时，消息能成功发送到对应的消息队列。
当消息格式、内容错误的时候，此消息会被拒绝，无法添加到消息队列

单元测试用例：

用例编号	003			
测试单元描述	后端向 AIGC 模块发送预生成请求			
用例目的	测试后端能否正常向 AIGC 服务端发送预生成请求，以及请求超时等异常情况下能否合适处理			
前提条件	任务处理模块成功从消息队列中取到消息			
特殊的规程说明	AIGC 服务端需开放对应的接口服务			
用例间的依赖关系	向消息队列发送消息后的后置用例			
具体步骤	输入	期望输出	实际输出	备注
1	正确的 id、正确的音频 URL、正确的照片 URL	响应码为：200，预生成文件在存储桶的 Key	响应码为：200，预生成的文件在存储桶的 Key	AIGC 服务端响应为 HTTP Response
2	正确的 id、空的音频 URL、空的照片 URL	响应码为 400	响应码为 400	400 响应码表示请求参数不合法，缺少必要的参数
3	正确的 id、超出文件大小的照片、音频 URL	响应码为 400	响应码为 400	
4	错误的 id、合法的照片、音频 URL	响应码为 400	响应码为 400	

测试结果分析：

测试结果均正常，只有当参数全部合法的情况下，AIGC 服务端才会执行生

成操作，否则会返回响应码 400

单元测试用例：

用例编号	004			
测试单元描述	从数据库中查询预生成任务的状态			
用例目的	测试能否正确的从数据库中查询出来相关的预生成任务的生成状态以及生成信息			
前提条件	用户已经提交了预生成的请求			
特殊的规程说明	无			
用例间的依赖关系	预生成用例为此用例的前置用例			
具体步骤	输入	期望输出	实际输出	备注
1	正确的对话 Id	任务生成状态 status, 生成信息 execMessage	任务生成状态 status, 生成信息 execMessage	ExecMessage 的内容要符合任务生成状态 status
2	格式错误的对话 Id	异常: MessageId 不合法	异常: 未查询到 Messgae!	
3	不存在的对话 Id	异常: 未查询到 Messgae!	异常: 未查询到 Messgae!	

测试结果分析：

经过测试，发现当对话 Message 的 Id 格式错误的时候存在错误，不会输出 MessageId 不合法的异常，而是未查询到 Messgae! 的异常。

原因是因为未对 MessageId 进行格式校验，而是仅对凭此 MessageId 查询到的结果进行校验，进行了不必要的操作。

单元测试用例：

用例编号	005			
测试单元描述	通过文字生成 AI 的语音回复			
用例目的	测试能否正确的根据用户的文字输入来生成 AI 的语音回复			
前提条件	预生成任务状态成功，且用户通过语音输入了相关的内容			
特殊的规程说明	AIGC 服务需要提供相应的端口服务、文本大模型需要正常运行			
用例间的依赖关系	预生成用例为此用例的前置用例			
具体步骤	输入	期望输出	实际输出	备注
1	正确的对话 Id、合法的用户输入	文字生成结果、语音在云存储中的 Key	文字生成结果、语音在云存储中的 Key	需要检查根据此 Key 能否得到文件
2	格式错误的对话 Id、合法的用户输入	异常: MessgaeId 非法	异常: 未查询到对话 Message	
3	不存在的对话 Id、合法的	异常: 未查询到对话 Message	异常: 未查询到对话 Message	

	用户输入			
4	正确的对话 Id、空的用户输入	异常：用户输入不可以为空	异常：用户输入不可以为空	

测试结果分析：

经过测试，当参数均合法的时候，此用例能够得到符合期望的、正确的输出，但是我们也发现当对话 Message 的 Id 格式错误的时候存在错误，不会输出 MessageId 非法的异常，而是未查询到 Messgae 的异常。

原因与编号 004 的情况相同，是因为未对 MessageId 进行格式校验，而是仅对凭此 MessageId 查询到的结果进行校验，进行了不必要的操作

单元测试用例：

用例编号	006			
测试单元描述	通过文字生成 AI 的视频回复			
用例目的	测试能否正确的根据用户的文字输入来生成 AI 的视频回复			
前提条件	预生成任务状态成功，且用户通过语音输入了相关的内容			
特殊的规程说明	AIGC 服务需要提供相应的端口服务、文本大模型需要正常运行			
用例间的依赖关系	预生成用例为此用例的前置用例			
具体步骤	输入	期望输出	实际输出	备注
1	正确的对话 Id、合法的用户输入	文字生成结果、视频在云存储中的 Key	文字生成结果、视频在云存储中的 Key	需要检查根据此 Key 能否得到文件
2	格式错误的对话 Id、合法的用户输入	异常：MessgaeId 非法	异常：未查询到对话 Message	
3	不存在的对话 Id、合法的用户输入	异常：未查询到对话 Message	异常：未查询到对话 Message	
4	正确的对话 Id、空的用户输入	异常：用户输入不可以为空	异常：用户输入不可以为空	

测试结果分析：

经过测试，当参数均合法的时候，此用例能够得到符合期望的、正确的输出，此用例功能正常。

2.1.2 测试结果综合分析及建议

经过测试，系统整体表现良好，错误较少。然而，在测试过程中发现两处错

误是由于没有对 `MessageId` 进行校验而导致的。这表明在系统设计和开发过程中，对于输入数据的验证和处理还存在一些不足之处。

针对此情况，以下是一些建议：

- **强化输入数据的验证：**在系统设计和开发阶段，应该对所有输入数据进行严格的验证，包括 `MessageId` 等重要参数。确保输入的数据符合预期的格式、范围和规则，以避免类似错误的再次发生。
- **实施更严格的单元测试：**在编写代码时，加强对于关键功能和边界条件的单元测试。通过单元测试，可以及早发现代码中的潜在问题，并在上线前进行修复，提高系统的稳定性和可靠性。

2.1.3 测试经验总结

- **充分理解需求：**在进行测试之前，需要对系统的需求和功能进行充分理解。只有理解了需求，才能够有效地设计测试用例和评估系统的表现。
- **制定详细的测试计划：**在开始测试之前，应该制定详细的测试计划，明确测试的范围、目标、方法。
- **设计全面的测试用例：**在设计测试用例时，应该考虑到各种可能的情况，包括正常情况、边界情况和异常情况。测试用例应该覆盖系统的所有功能和模块，以确保系统的完整性和稳定性。

3 功能测试

3.1 预生成功能

3.1.1 测试用例与结果分析

测试用例：

用例编号	001			
功能描述	用户输入相关信息，即可进行预生成，生成模型以及视频			
用例目的	测试用户能否正常进行预生成，以及在异常条件下系统的适应性			
前提条件	用户需进行登录操作并具有预生成次数			
特殊的规程说明	云存储、AIGC 服务端模型需正常运行			
用例间的依赖关系	语音、视频聊天的必须前置			
具体步骤	输入	期望结果	实际结果	备注
1	合法的照片、音频文件、合法的称谓以及简介	预生成成功的响应，以及正确的预生成视频	预生成成功的响应，以及正确的预生成视频	前端会提示：生成状态为完成，用户可进入到聊天界面并显示预生成视频
2	空照片、音频文件、合法称谓以及简介	异常：请上传完整文件	异常：请上传完整文件	
3	不含人脸的照片、合法的音频、称谓以及简介	预生成状态失败，请重试	预生成状态失败，请重试	
4	含有人脸的照片、不含人声的音频、合法的称谓以及简介	预生成状态失败，请重试	预生成状态失败，请重试	

测试结果分析：

经过测试，此功能正常。

在用户正确的输入的条件下，均能得到正常的输出。当用户输入非法时，系统会对错误进行相应的提示。但是此提示的内容不够详细，对于照片不含人脸以及音频不含人声的情况，系统的错误提示是相同的，缺乏针对性，可以加以改进。

3.1.2 测试结果综合分析及建议

本次功能测试，目的是验证用户在输入相关信息后，系统是否能够正常执行预生成，生成模型以及视频的功能，并测试系统在面对异常输入时的适应性与错误处理能力。测试覆盖了正常和异常的输入情况，包括合法的照片、音频文件、称谓以及简介的组合，以及缺失或不合规范的输入情况。

测试结果表明，在合法输入的情况下，系统能够正常响应并生成预期的结果，满足用例目的。在遇到异常输入，如空白文件、不含人脸的照片、或不含人声的音频文件时，系统能够给出错误提示，阻止了错误的操作继续进行。表明系统具有一定程度的错误处理能力。

然而，测试中也暴露出系统错误提示的不足之处。对于不同类型的错误输入，系统给出的错误提示相同且不够详细，缺乏针对性。这可能会使用户难以理解错误的具体原因，从而无法有效地纠正输入错误。

下一步希望能够进一步优化错误提示信息或者引入错误提示分类机制，建立一个更细致的错误分类系统，对不同类型的错误进行分类。同时前端应当加强验证，减轻服务器端的压力，并提升用户体验

3.1.3 测试经验总结

在测试前制定详细的测试用例，包括正常和异常情况的输入，可以帮助我们系统地覆盖各种场景，确保不遗漏任何重要的测试点。此次测试中，我们采用边界值分析法以及等价类划分法的方式设计用例，帮助我们有效地识别了正常情况下的系统表现以及异常输入下的错误处理机制。

3.2 语音对话功能

3.2.1 测试用例与结果分析

测试用例：

用例编号	002			
功能描述	用户通过语音输入相关的内容，AI 会对内容以语音的形式进行回复			
用例目的	测试用户能否正常与 AI 进行语音交流，以及在异常条件下系统的适应性			
前提条件	用户已经进行了预处理操作，且预生成成功			
特殊的规程说明	AIGC 服务端需正常运行			
用例间的依赖关系	预生成用例为前置			
具体步骤	输入	期望结果	实际结果	备注
1	合法的内容	AI 正常语音回复	AI 正常语音回复	

2	空	异常：输入不能为空	异常：输入不能为空	
3	不合法的输入（如脏话、广告等敏感内容）	异常：请文明语言	AI 正常语音回复	

测试结果分析：

经过测试，此功能基本正常。

在用户正确的输入的条件下，均能得到正常的输出。当用户输入非法时，系统会对错误进行相应的提示。但是当用户的输入不合法的时候（脏话、广告等敏感内容），内容识别有时候会不够精确，有的不合法语句不会正常识别出来，不合法内容会被直接发送给文本生成模型，造成生成内容偏差较大。因此内容识别功能需要进一步完善。

3.2.2 测试结果综合分析及建议

根据测试结果，语音交流功能在处理合法内容方面表现良好，能够达到预期的回应效果。然而，在处理非法输入（如脏话、广告等敏感内容）时，系统的表现并不完美。尽管在一定情况下能够提示用户输入不合适的内容，但不是所有的不合法输入都能被有效识别和阻止。表明了内容识别机制在精确度和覆盖面上存在不足，可能会导致不恰当内容的传递和生成，从而影响用户体验和平台形象。

可以增强内容过滤机制，利用机器学习模型，根据不断更新的数据集训练，以更好地识别和过滤不合法的输入。

实时更新敏感词库，定期更新敏感词和不当表达的数据库，以覆盖新出现的或先前未纳入的表达方式。

增加用户反馈机制，建立一套用户反馈机制，允许用户报告不当回复。这不仅可以帮助及时纠正特定实例，还可以为系统学习提供数据，不断优化内容过滤算法。

3.2.3 测试经验总结

测试用例需要不断更新和完善，以确保测试覆盖所有重要方面。

同时也要重视用户的反馈，用户的实际体验和反馈是评估功能表现的重要指标。在测试过程中应充分利用这一点，以指导产品的持续改进

同时也要注意安全性，在测试和开发过程中，始终要将用户数据的安全性和合规性放在首位，确保所有功能和改进都符合相关法律法规的要求

3.3 视频对话功能

3.3.1 测试用例与结果分析

测试用例：

用例编号	003			
功能描述	用户通过语音输入相关的内容，AI 会对内容以视频的形式进行回复			
用例目的	测试用户能否通过语音正常与 AI 进行视频交流，以及在异常条件下系统的适应性			
前提条件	用户已经进行了预处理操作，且预生成成功			
特殊的规程说明	AIGC 服务端需正常运行			
用例间的依赖关系	预生成用例为前置			
具体步骤	输入	期望结果	实际结果	备注
1	合法的内容	AI 正常视频回复	AI 正常视频回复	
2	空	异常：输入不能为空	异常：输入不能为空	
3	不合法的输入（如脏话、广告等敏感内容）	异常：请文明语言	AI 正常视频回复	
4	长内容（诱导 AI 生成长文字的语句）	AI 正常视频回复	异常：生成超时，请重试	当文字过长时，AI 视频相应的时长也会增加，导致响应超时

测试结果分析：

与视频对话聊天功能相同，经过测试，此功能基本正常。

在用户正确的输入的条件下，均能得到正常的输出。当用户输入非法时，系统会对错误进行相应的提示。但是当用户的输入不合法的时候（脏话、广告等敏感内容），内容识别有时候会不够精确，有的不合法语句不会正常识别出来，不合法内容会被直接发送给文本生成模型，造成生成内容偏差较大。因此内容识别功能需要进一步完善。

另外当用户输入长段文字或者诱导 AI 生成长段文字的语句时，AI 生成的时间会变长，从而导致响应超时，得不到生成结果。据此，可以限制 AI 生成的字数长度

3.3.2 测试结果综合分析及建议

根据测试结果，语音交流功能在处理合法内容方面表现良好，能够达到预期的回应效果。然而，在处理非法输入（如脏话、广告等敏感内容）时，系统的表现并不完美。尽管在一定情况下能够提示用户输入不合适的内容，但不是所有的不合法输入都能被有效识别和阻止。表明了内容识别机制在精确度和覆盖面上存在不足，可能会导致不恰当内容的传递和生成，从而影响用户体验和平台形象。

可以增强内容过滤机制，利用机器学习模型，根据不断更新的数据集训练，以更好地识别和过滤不合法的输入。

实时更新敏感词库，定期更新敏感词和不当表达的数据库，以覆盖新出现的或先前未纳入的表达方式。

增加用户反馈机制，建立一套用户反馈机制，允许用户报告不当回复。这不仅可以帮助及时纠正特定实例，还可以为系统学习提供数据，不断优化内容过滤算法。

3.3.3 测试经验总结

测试用例需要不断更新和完善，以确保测试覆盖所有重要方面。

同时也要重视用户的反馈，用户的实际体验和反馈是评估功能表现的重要指标。在测试过程中应充分利用这一点，以指导产品的持续改进

同时也要注意安全性，在测试和开发过程中，始终要将用户数据的安全性和合规性放在首位，确保所有功能和改进都符合相关法律法规的要求

4 系统测试

4.1 模型性能测试

4.1.1 测试用例与结果分析

预期性能指标测试用例：

用例编号	001			
性能描述				
用例目的				
前提条件				
特殊的规程说明				
用例间的依赖关系				
具体步骤	输入/动作	期望的性能 (平均值)	实际的性能 (平均值)	备注
1				
2				
...

测试结果分析：

4.1.2 测试结果综合分析及建议

4.1.3 测试经验总结