AI辅助的在线作业平台软件系统测试计划和报告书

1. **引言**
   1. **编写目的**

通过编写测试计划报告书，可以有效地规划和管理测试活动，提高项目交付的质量和成功实施的可能性，还可以提供一个对AI辅助的在线作业平台软件的测试计划，包括对每项测试活动的内容、进度安排、测试用例的设计及评价准则。其中，主要目的包括：

（1）指导测试活动：提供测试团队在项目中的指导和规划，确保测试活动按照既定计划进行。

（2）确保质量：确保项目交付的软件具有高质量，符合用户需求和预期。

（3）风险管理：识别和评估测试过程中可能出现的风险，并制定相应的风险应对策略。

（4）资源规划：分配和管理测试资源，包括人力、时间、工具等，以确保测试活动的有效执行。

（5）沟通和协调：作为沟通工具，帮助项目团队成员和利益相关者了解测试计划和执行流程。

（6）决策支持：为项目管理人员提供决策支持，帮助他们评估测试进度和质量，及时调整项目方向。

（7）验收标准：确定测试完成的标准和验收条件，以便评估测试结果是否符合预期要求。

（8）文档化：将测试计划、方法和结果文档化，作为项目的重要文档，便于后续追踪和审查。

* 1. **背景**

**1.2.1 被测项目名称**

被测项目名称为：AI辅助的在线作业平台。

**1.2.2 开发该软件系统的目的**

首先，该项目需要提供在线编程作业的布置、完成、提交与评测的作业平台，老师可以通过平台布置作业并设置提交截至时间，学生可以通过该平台进行作业的提交，并通过系统得到正确有效的反馈。项目需要针对不同学生编码习惯不同，不利于老师批阅的问题，借助 AI 对学生在逻辑上错误的代码进行修复，自动完成代码的批阅。另外，该项目应该能够借助 AI 来对bugged code进行修复，更好地评测学生代码，并且生成代码质量和批改报告，向学生提供更全面的代码评价和批改信息。并且，AI需要根据代码质量给出一定权重的评分。

**1.2.3 用户特点**

学生：学生可以方便地查看题目、提交代码，进行评测，并查看AI对其代码的反馈和修复建议。

教师：教师可以方便地布置题目，浏览AI统计的题目结果，并进行平台上题目信息的维护。

**1.2.4 功能需求**

用户共同功能：登录平台。

（1）学生功能：

查看题目：学生能够方便地查看教师布置的题目。

提交代码：学生能够方便地提交自己的代码。

代码评测：系统应该能够自动评测学生提交的代码。

查看反馈和修复：学生能够查看AI对自己代码的反馈和修复建议。

（2）教师功能：

布置题目：教师可以方便地在平台上布置并查看题目信息。

查看统计结果：教师能够浏览AI统计的题目结果。

维护题目信息：教师能够方便地维护平台上的题目信息。

**1.2.5 非功能需求**

对于非功能需求，分为质量要求和约束性要求，质量要求又分为外部和内部两种。

从用户角度，外部质量要求包括软件运行的可靠性、性能、安全性、易用性、时效性和对计算机资源的有效利用等。对管理人员以及其他用户而言，软件的安全性涉及到知识产权、账号管理以及学生隐私的方方面面。

内部质量要求是对开发方而言，包括软件的可扩展性、可维护性、可互操作性等。平台经常会涉及功能的修改以及日常的维护，这对可扩展性和可维护性有较高的要求。可互操作性为人员间相互协作配合提供保障。

约束性要求包括开发进度要求、成本要求、技术选型、规模限制等各个方面。

**1.2.6 环境约束**

（1）软件环境：Ubuntu LTS 22.04。

（2）硬件环境：AI服务器需要48G显存，64G以上内存，10核以上CPU，X86架构。

* 1. **参考资料**

<https://ollama.com/>

<https://dl.acm.org/doi/pdf/10.1145/3551349.3560425>

<https://arxiv.org/pdf/2209.14876.pdf>

1. **测试计划**
2. **软件说明**

在线作业平台是为教师和学生提供方便的作业布置、提交、评测和反馈的平台。主要功能包括学生查看题目、提交代码、代码评测、查看反馈和修复建议；教师布置题目、查看统计结果、维护题目信息等。

软件环境需要使用Ubuntu LTS 22.04，硬件环境需要48G显存，64G以上内存，10核以上CPU，X86架构。

测试范围包括了功能测试、性能测试、界面测试、压力测试以及健壮性测试。

测试人员为：顾田、邢益鸣、赵晚甜、孙龙灵、戴威扬。

测试时间为：6月1日。

1. **测试内容**

功能测试

|  |  |  |  |
| --- | --- | --- | --- |
| 功能A描述 | 登录平台 | | |
| 用例目的 | 确保用户能够登录平台。 | | |
| 前提条件 | 用户进入登录界面。 | | |
| 输入/动作 | | 期望的输出/相应 | 实际情况 |
| 输入用户名和正确密码，点击登录。 | | 系统提示成功登录，显示执行下一步操作的界面。 |  |
| 输入正确用户名和错误密码，点击登录。 | | 系统提示“密码错误”，提示重新登录。 |  |
| 输入错误用户名和密码，点击登录。 | | 系统提示“用户不存在”的错误信息，提示重新登录。 |  |
| 输入正确用户名，不输入密码，点击登录。 | | 系统提示“密码未输入”的错误信息，提示重新登录。 |  |
| 不输入用户名，只输入密码，点击登录。 | | 系统提示“用户名未输入在”的错误信息，提示重新登录。 |  |
| 功能A描述 | 学生查看题目：学生能够查看教师布置的题目。 | | |
| 用例目的 | 确保学生能够正确地查看题目信息。 | | |
| 前提条件 | 学生已登录到在线作业平台。 | | |
| 输入/动作 | | 期望的输出/相应 | 实际情况 |
| 学生登录到在线作业平台后，点击“作业”选项卡。 | | 显示学生所需完成的所有作业列表，包括题目名称和截止日期。 |  |
| 学生选择特定作业，点击查看详情。 | | 展示该作业的详细信息，包括题目描述、要求、附件等。 |  |
| 学生查看题目时尝试搜索查看不存在的题目。 | | 系统提示“题目不存在”的错误信息，要求学生重新选择有效的题目。 |  |
| 学生查看题目时尝试查看空白题目。 | | 系统提示“题目内容为空”的错误信息，要求学生联系管理员修复题目内容。 |  |
| 学生查看题目时尝试查看包含特殊字符的题目。 | | 系统能够正确显示包含特殊字符的题目内容，确保内容展示的准确性。 |  |
| 功能B描述 | 学生提交代码：学生能够提交自己的代码。 | | |
| 用例目的 | 验证学生能够成功提交代码。 | | |
| 前提条件 | 学生已登录到在线作业平台并选择了要提交的题目。 | | |
| 输入/动作 | | 期望的输出/相应 | 实际情况 |
| 学生登录到在线作业平台后，选择特定作业并点击“开始作业”。 | | 弹出作答界面，学生可以开始编写代码或回答问题。 |  |
| 学生在作答界面中编写完整的代码后，点击“提交代码”按钮。 | | 提示代码提交成功，并显示提交时间和状态为“已提交”。 |  |
| 学生尝试提交超出系统限制大小的代码文件。 | | 系统提示“代码文件大小超出限制”的错误信息，要求学生重新提交符合大小要求的代码文件。 |  |
| 学生尝试提交空白的代码文件或空文件。 | | 系统提示“代码文件不能为空”的错误信息，要求学生提交有效的代码文件。 |  |
| 功能B描述 | 代码测评：系统自动评测学生提交的代码。 | | |
| 用例目的 | 确保系统能够正确评测学生的代码。 | | |
| 前提条件 | 学生已提交代码并系统已完成评测。 | | |
| 输入/动作 | | 期望的输出/相应 | 实际情况 |
| 学生提交代码后，系统自动进行代码评测和批改。 | | 系统根据预设的评分标准对代码进行评分，并生成评分报告。 |  |
| 学生查看作业详情页面，找到已提交的作业并点击查看批改结果。 | | 显示批改结果，包括评分、错误提示、修改建议等信息。 |  |
| 学生阅读批改结果后，可以选择继续改进代码并重新提交。 | | 提示代码重新提交成功，并系统重新进行评测和批改。 |  |
| 学生在作答界面中编写代码时，出现保存草稿并退出的情况。 | | 提示草稿保存成功，学生可以随时回到该作业并继续编辑。 |  |
| 学生在评分报告中看到了自己的代码逻辑错误的具体位置。 | | 提供详细的错误提示和建议，帮助学生理解错误原因并改进代码。 |  |
| 学生提交作业代码，包含异常值的代码示例（如错误语法、逻辑错误等）。 | | 系统能够识别异常值的代码示例，给出相应的错误提示和建议修复方案。 |  |
| 学生提交作业代码，包含耗时较长的代码示例。 | | 系统能够在合理的时间内完成对耗时较长代码的评测，不影响用户体验。若时间过长，提示“代码运行超时”，要求重新提交代码。 |  |
| 学生提交作业代码，尝试提交包含恶意代码的文件。 | | 系统检测到恶意代码并拒绝评测，警示同学，同时提示老师及管理员学生违规行为。 |  |
| 功能C描述 | 查看反馈和修复：学生能够查看AI对自己代码的反馈和修复建议。 | | |
| 用例目的 | 验证学生能够获取到代码的反馈和修复建议。 | | |
| 前提条件 | 学生已查看代码评测结果。 | | |
| 输入/动作 | | 期望的输出/相应 | 实际情况 |
| 学生查看作业详情页面，找到已提交的作业并点击查看批改结果。 | | 显示批改结果，包括评分、错误提示、修改建议等信息。 |  |
| 学生阅读批改结果后，发现代码中存在逻辑错误或语法错误。 | | 提供具体的错误提示和建议，帮助学生理解错误原因并进行修复。 |  |
| 学生根据批改结果进行代码修复，并重新提交作业。 | | 提示代码重新提交成功，并系统重新进行评测和批改。 |  |
| 学生在查看批改结果时，发现评分不符合预期或有疑问。 | | 提供反馈机制，学生可以提交申诉或问题反馈，系统会及时处理并回复。 |  |
| 学生在修复代码后，重新查看批改结果并确认修复是否成功。 | | 显示修复后的代码评分和反馈信息，确保学生能够及时了解自己的改进情况。 |  |
| 学生查看作业评分报告，尝试查看未完成的作业评分。 | | 系统提示“作业尚未完成，暂无评分报告”的信息，提醒学生先完成作业。 |  |
| 学生查看作业评分报告，输入不存在的作业ID。 | | 系统提示“作业ID不存在”的错误信息，不显示任何评分报告。 |  |
| 学生查看作业评分报告，输入异常格式的作业ID。 | | 系统提示“作业ID格式错误”的信息，要求重新输入符合格式要求的作业ID。 |  |
| 学生查看作业评分报告，输入空白的作业ID。 | | 系统提示“作业ID不能为空”的错误信息，要求输入有效的作业ID。 |  |
| 学生查看作业评分报告，尝试查看已删除的作业评分。 | | 系统提示“作业已删除，无法查看评分报告”的信息，提醒学生该作业已不可用。 |  |
| 功能D描述 | 教师布置题目：教师可以在平台上布置题目。 | | |
| 用例目的 | 确保教师能够成功创建和发布题目。 | | |
| 前提条件 | 教师已登录到在线作业平台并进入题目管理界面。 | | |
| 输入/动作 | | 期望的输出/相应 | 实际情况 |
| 老师登录系统后，进入作业管理页面，点击“创建新作业”按钮。 | | 弹出作业创建界面，老师可以填写作业标题、描述、截止日期等信息。 |  |
| 老师在作业创建界面中选择添加编程题目，并填写题目要求和测试用例。 | | 系统保存题目信息，老师可以预览题目内容并进行修改。 |  |
| 老师在创建作业时设置了自动批改功能，并选择AI辅助批改。 | | 系统提示自动批改功能已开启，老师可以设置评分标准和批改规则。 |  |
| 老师创建完作业后，点击发布按钮将作业布置给学生 | | 系统提示作业发布成功，学生可以在系统中看到该作业并开始完成。 |  |
| 老师在作业管理页面查看已发布的作业列表，找到某个作业并点击编辑。 | | 弹出作业编辑界面，老师可以修改作业内容、截止日期等信息。 |  |
| 老师添加题目到作业中，尝试添加重复的题目。 | | 系统提示“题目已存在于作业中”的信息，不允许重复添加同一题目。 |  |
| 老师编辑作业要求，输入超长的作业要求描述。 | | 系统能够正确截取或提示超出字数限制，确保作业要求不会过长。 |  |
| 老师设置作业最大提交次数，输入负数。 | | 系统提示“提交次数格式错误”的信息，要求重新输入有效的提交次数。 |  |
| 老师设置作业最大提交次数，输入非数字。 | | 系统提示“提交次数格式错误”的信息，要求重新输入有效的提交次数。 |  |
| 老师设置作业最大提交次数，输入超大数值。 | | 系统提示“提交次数过大”的信息，要求重新输入有效的提交次数。 |  |
| 老师创建新的编程作业，输入空白的作业名称。 | | 系统提示“作业名称不能为空”的错误信息，要求输入有效的作业名称。 |  |
| 老师创建新的编程作业，输入含有不可识别字符的作业名称。 | | 系统提示“作业名称不规范”的错误信息，要求输入有效的作业名称。 |  |
| 老师设置作业截止日期，输入边界值的日期时间（正常时间）。 | | 成功设置截止时间。 |  |
| 老师设置作业截止日期，输入边界值的日期时间（0月几号）。 | | 系统提示“截止日期设置错误”的信息，要求重新输入有效的截止日期。 |  |
| 老师设置作业截止日期，输入边界值的日期时间（13月几号）。 | | 系统提示“截止日期设置错误”的信息，要求重新输入有效的截止日期。 |  |
| 老师设置作业截止日期，输入边界值的日期时间（闰年2月29号）。 | | 系统提示“截止日期设置错误”的信息，要求重新输入有效的截止日期。 |  |
| 老师设置作业截止日期，输入边界值的日期时间（2、4、6、9、11月的31号）。 | | 系统提示“截止日期设置错误”的信息，要求重新输入有效的截止日期。 |  |
| 老师设置作业截止日期，输入边界值的日期时间（正常月份32号）。 | | 系统提示“截止日期设置错误”的信息，要求重新输入有效的截止日期。 |  |
| 老师设置作业截止日期，输入边界值的非日期格式的日期时间。 | | 系统提示“截止日期格式设置错误”的信息，要求重新输入有效的截止日期。 |  |
| 老师设置作业截止日期，输入不可识别的字符。 | | 系统提示“截止日期格式设置错误”的信息，要求重新输入有效的截止日期。 |  |
| 功能E描述 | 教师查看统计结果：教师能够浏览AI统计的题目结果。 | | |
| 用例目的 | 验证教师能够查看题目的统计信息。 | | |
| 前提条件 | 教师已登录到在线作业平台并选择查看统计结果。 | | |
| 输入/动作 | | 期望的输出/相应 | 实际情况 |
| 老师登录系统后，进入作业管理页面，找到已批改的作业列表。 | | 显示已批改作业的列表，包括每个学生的提交情况和评分。 |  |
| 老师点击某个已批改作业，进入作业详情页面查看学生的代码和批改结果。 | | 显示学生提交的代码、批改结果、评分以及修改建议等信息。 |  |
| 老师在作业详情页面中点击“查看统计结果”按钮。 | | 弹出统计结果页面，显示该作业的总体评分分布、平均分、及格率等统计信息。 |  |
| 老师在统计结果页面中选择查看具体学生成绩分布情况。 | | 显示学生成绩的分布图表，包括优秀、良好、及格、不及格等分数段的人数统计。 |  |
| 老师在统计结果页面中发现某个学生的评分异常，点击学生姓名查看详细信息。 | | 显示该学生的作业详情，包括提交的代码、批改结果、评分和修改建议，帮助老师了解评分情况。 |  |
| 老师查看某次作业的得分统计结果，输入不存在的作业ID。 | | 系统提示“作业ID不存在”的错误信息，不显示任何统计结果。 |  |
| 老师查看某次作业的得分统计结果，输入异常格式的作业ID。 | | 系统提示“作业ID格式错误”的信息，要求重新输入符合格式要求的作业ID。 |  |
| 老师查看某次作业的统计结果，输入空白的作业ID。 | | 系统提示“作业ID不能为空”的错误信息，要求输入有效的作业ID。 |  |
| 功能F描述 | 教师维护题目信息：教师能够维护平台上的题目信息。 | | |
| 用例目的 | 确保教师能够成功修改和管理题目信息。 | | |
| 前提条件 | 教师已登录到在线作业平台并进入题目管理界面。 | | |
| 输入/动作 | | 期望的输出/相应 | 实际情况 |
| 老师登录系统后，进入题库管理页面，找到需要维护的题目。 | | 显示题库中的题目列表，包括题目标题、描述和难度等信息。 |  |
| 师点击某个题目，进入题目详情页面查看题目要求和测试用例。 | | 显示题目的详细信息，包括题目描述、输入输出要求、示例测试用例等内容。 |  |
| 老师在题目详情页面中发现题目描述有误，点击编辑按钮进行修改。 | | 弹出编辑界面，老师可以修改题目描述、输入输出要求等信息，并保存修改。 |  |
| 老师在题目详情页面中添加新的测试用例，以丰富题目的覆盖范围。 | | 系统提示测试用例添加成功，老师可以查看新的测试用例并进行必要的调整。 |  |
| 老师在题库管理页面中选择删除某个题目。 | | 系统提示确认删除该题目，老师确认后题目从题库中移除。 |  |
| 老师在题库管理页面中搜索特定关键词，查找相关题目。 | | 显示符合搜索关键词的题目列表，帮助老师快速找到需要维护的题目。 |  |
| 老师删除题目，输入不存在的题目ID。 | | 系统提示“题目不存在”的错误信息，不执行删除操作。 |  |
| 老师编辑题目要求，输入超长的描述内容。 | | 系统能够正确截取或提示超出字数限制，确保题目要求不会过长。 |  |
| 老师搜索题目关键词，输入包含特殊字符的搜索词。 | | 系统能够正确处理特殊字符，显示符合搜索词的题目列表。 |  |
| 老师添加新的测试用例，输入异常格式的测试数据。 | | 系统提示“测试数据格式错误”的信息，要求重新输入符合格式要求的测试数据。 |  |

性能测试

|  |  |  |  |
| --- | --- | --- | --- |
| 性能1描述 | 负载测试：负载测试用于评估系统在不同负载条件下的表现，包括正常、峰值和超负载情况下的性能表现。 | | |
| 用例目的 | 确保系统在高负载情况下仍能够保持稳定运行，不会因为并发请求过多而导致系统崩溃或性能下降。 | | |
| 前提条件 | 系统已部署并可同时接受多个用户请求。 | | |
| 输入数据 | | 期望的性能（平均值） | 实际性能（平均值） |
| 正常负载：100个同时在线用户提交作业。 | | 正常负载下，系统响应时间应保持在2秒以内。 |  |
| 峰值负载：500个同时在线用户提交作业。 | | 峰值负载下，系统响应时间应保持在3秒以内。 |  |
| 超负载：1000个以上同时在线用户提交作业。 | | 超负载情况下，系统能够优雅地处理请求，并在合理时间内响应用户。 |  |
| 性能B描述 | 响应时间测试：响应时间测试用于测量系统对用户请求的响应时间，包括页面加载时间、代码评测时间等。 | | |
| 用例目的 | 确保系统在用户操作时能够迅速响应，提升用户体验。 | | |
| 前提条件 | 系统已部署并可访问，测试环境已准备就绪。 | | |
| 输入数据 | | 期望的性能（平均值） | 实际性能（平均值） |
| 提交作业页面加载时间 | | 提交作业页面加载时间应保持在1秒以内。 |  |
| 代码评测时间 | | 代码评测时间应根据代码复杂度在3秒到3分钟之间。 |  |
| 性能2描述 | 吞吐量测试：吞吐量测试用于评估系统在单位时间内能够处理的请求量。 | | |
| 用例目的 | 确定系统在高负载情况下的处理能力，以及系统的性能瓶颈。 | | |
| 前提条件 | 系统已部署并可访问，测试环境已准备就绪。 | | |
| 输入数据 | | 期望的性能（平均值） | 实际性能（平均值） |
| 并发用户数：100 | | 请求频率：每秒10个请求 |  |
| 并发用户数：500 | | 请求频率：每秒50个请求 |  |
| 并发用户数：1000 | | 请求频率：每秒100个请求 |  |
| 性能3描述 | 稳定性测试：稳定性测试旨在评估系统在持续负载条件下的表现，检测系统在连续运行期间是否会出现内存泄漏、资源耗尽或其他稳定性问题。 | | |
| 用例目的 | 确保系统在长时间运行时能够保持稳定，不会出现内存泄漏、资源泄露等问题，保障系统的可靠性和稳定性。 | | |
| 前提条件 | 系统已部署并可访问，测试环境已准备就绪。 | | |
| 输入数据 | | 期望的性能（平均值） | 实际性能（平均值） |
| 持续运行时间：1小时、2小时、4小时、8小时、12小时、24小时 | | 在持续运行期间，系统能够稳定运行，不出现系统崩溃或异常退出的情况。 |  |
| 并发用户数：100、200、500 | | 系统资源占用保持稳定，不会出现内存泄漏或资源泄露现象。 |  |
| 提交作业频率：每分钟10次、每分钟50次、每分钟100次 | | 在不同的并发用户数和提交作业频率下，系统能够保持稳定的响应时间，不会因为负载增加而导致系统性能下降。 |  |
| 性能4描述 | 资源利用率测试：资源利用率测试用于监测系统在不同负载条件下的资源利用情况，包括CPU、内存、网络带宽等，以确定系统在高负载情况下的资源消耗情况。 | | |
| 用例目的 | 确保系统在高负载情况下能够有效地利用计算机资源，避免资源瓶颈导致系统性能下降或崩溃。 | | |
| 前提条件 | 系统已部署并可访问，测试环境已准备就绪。 | | |
| 输入数据 | | 期望的性能（平均值） | 实际性能（平均值） |
| 并发用户数：50、100、200 | | CPU利用率：在不同并发用户数和作业数量下，系统的CPU利用率应该保持在合理范围内，不会出现过高的CPU利用率。 |  |
| 提交作业频率：每分钟10次、每分钟50次、每分钟100次 | | 网络带宽利用率：系统在高负载情况下，网络带宽利用率应该适度增加，但不会超出系统承受范围，不会出现网络拥堵或阻塞现象。 |  |
| 作业数量：100、500、1000 | | 内存利用率：系统在不同负载条件下，内存利用率应该保持在合理范围内，不会出现内存泄漏或内存耗尽现象。 |  |

界面测试

|  |  |
| --- | --- |
| 检查项 | 测试人员的类别及其评价 |
| 窗口切换、移动、改变大小时正常吗？ |  |
| 各种界面元素的文字正确吗？（如标题、提示等） |  |
| 各种界面元素的状态正确吗？（如有效、无效、选中等状态） |  |
| 各种界面元素支持键盘操作吗？ |  |
| 各种界面元素支持鼠标操作吗？ |  |
| 对话框中的缺省焦点正确吗？ |  |
| 数据项能正确回显吗？ |  |
| 对于常用的功能，用户能否不必阅读手册就能使用？ |  |
| 执行有风险的操作时，有“确认”、“放弃”等提示吗？ |  |
| 操作顺序合理吗？ |  |
| 有联机帮助吗？ |  |
| 各种界面元素的布局合理吗？美观吗？ |  |
| 各种界面元素的颜色协调吗？ |  |
| 各种界面元素的形状美观吗？ |  |
| 字体美观吗？ |  |
| 图标直观吗？ |  |
| 标题、按钮文字、提示信息等是否准确、清晰易懂？ |  |
| 文字拼写、语法和语义是否正确？ |  |
| 输入框、按钮等交互元素，其可用性和反馈状态是否符合预期？ |  |
| 系统在不同浏览器（如Chrome、Firefox、Safari等）显示效果和交互性能是否一致？ |  |

压力测试

|  |  |  |  |
| --- | --- | --- | --- |
| 极限名称A | 最大并发用户数量 | | |
| 前提条件 | 系统已部署并可访问，测试环境已准备就绪。 | | |
| 输入/动作 | | 输出/响应 | 是否能正常运行 |
| 100个用户并发操作 | | 系统响应时间应保持在2秒以内。 |  |
| 500个用户并发操作 | | 系统响应时间应保持在3秒以内。 |  |
| 1000个用户并发操作 | | 系统响应时间应保持在10秒以内。 |  |
| 2000个用户并发操作 | | 系统响应时间应保持在1分钟以内。 |  |
| 5000个用户并发操作 | | 系统响应时间应保持在5分钟以内。 |  |
| 极限名称B | 持续负载测试 | | |
| 前提条件 | 系统已部署并可访问，测试环境已准备就绪。 | | |
| 输入/动作 | | 输出/响应 | 是否能正常运行 |
| 持续运行时间：8小时 | | 在持续运行期间，系统能够运行，不出现系统崩溃或异常退出的情况。 |  |
| 持续运行时间：12小时 | | 在持续运行期间，系统能够运行，不出现系统崩溃或异常退出的情况。 |  |
| 持续运行时间：24小时 | | 在持续运行期间，系统能够运行，不出现系统崩溃或异常退出的情况。 |  |
| 持续运行时间：48小时 | | 在持续运行期间，系统能够运行，不出现系统崩溃或异常退出的情况。 |  |
| 持续运行时间：72小时 | | 在持续运行期间，系统能够运行，不出现系统崩溃或异常退出的情况。 |  |

1. **评价标准**

该项目的评价标准可以包括以下几个方面，用来判断测试工作是否能通过的评价尺度：

（1） 容许偏离范围：对于一些功能或性能指标，可以容许一定的偏离范围。例如，页面加载时间可以在1-2秒内波动，负载较小时响应时间可以在2-5秒内波动，负载较大时响应时间可以在1-3分钟内波动，但不能超出预期的合理范围。

（2）输出结果的类型合理性：检查输出结果的类型和格式是否符合预期，例如，评测系统的输出结果应该包括通过、未通过、错误信息等类型，以便正确反映学生代码的评测结果。

（3）系统稳定性和可靠性：测试工作还需评估系统在不同负载和异常情况下的稳定性和可靠性，包括系统是否崩溃、错误率是否增加、响应时间是否合理等方面。

（4）安全性和兼容性：对于在线作业平台这类涉及用户隐私和安全的系统，还需要评估系统的安全性和兼容性，包括数据传输的加密、不同浏览器和设备的兼容性等。

（5）用户体验和易用性：用户体验和系统的易用性也是重要评价标准之一。评估系统的界面设计是否清晰、操作是否直观、反馈信息是否友好等方面。

1. **测试结果及发现**

以表格的形式列出每一项测试结果，陈述其中的各项发现，给出是否达到预定目标，通过测试的结论。