

测试日期: 2026年01月28日
报告版本: v2.0
系统版本: v1.0

总测试数: 46
通过数: 39 通过率: 84.8%
失败数: 7

自动化测试系统生成
保留所有权利

目录

1 执行摘要

1.1 概述

本次测试对医疗智能助手进行了全面的测试，包括单元测试、功能测试、性能测试和流程测试。

总体测试结果为：**良好**，通过率为84.8%。

1.2 关键发现

- 所有Skill均完成了测试覆盖，其中health-educator通过率达到100%
- 系统平均响应时间在毫秒级，性能表现良好
- 简单复杂度测试用例全部通过，复杂和边界情况存在改进空间

2 测试概述

2.1 测试统计

测试类型	总数	通过	通过率
单元测试	0	0	0.0%
功能测试	36	31	86.1%
性能测试	0	0	0.0%
流程测试	0	0	0.0%

表 1: 按测试类型统计

3 Skill测试结果分析

3.1 greeting-handler

该Skill测试结果为：**需改进**，通过率为 57.1%。

指标	数值
总测试数	7
通过数	4
失败数	3
通过率	57.1%

表 2: greeting-handler 测试统计

3.2 symptom-analyzer

该Skill测试结果为：优秀，通过率为 90.0%。

指标	数值
总测试数	10
通过数	9
失败数	1
通过率	90.0%

表 3: symptom-analyzer 测试统计

3.3 department-recommender

该Skill测试结果为：优秀，通过率为 100.0%。

指标	数值
总测试数	7
通过数	7
失败数	0
通过率	100.0%

表 4: department-recommender 测试统计

3.4 medication-advisor

该Skill测试结果为：优秀，通过率为 100.0%。

指标	数值
总测试数	7
通过数	7
失败数	0
通过率	100.0%

表 5: medication-advisor 测试统计

3.5 health-educator

该Skill测试结果为: 优秀, 通过率为 100.0%。

指标	数值
总测试数	6
通过数	6
失败数	0
通过率	100.0%

表 6: health-educator 测试统计

3.6 fallback-handler

该Skill测试结果为: 需改进, 通过率为 66.7%。

指标	数值
总测试数	9
通过数	6
失败数	3
通过率	66.7%

表 7: fallback-handler 测试统计

4 复杂度分析

4.1 各复杂度测试结果

复杂度	总数	通过	失败	通过率
simple	12	10	2	83.3%
medium	14	13	1	92.9%
complex	8	7	1	87.5%
edge_case	12	9	3	75.0%

表 8: 按复杂度统计

5 意图识别分析

5.1 置信度统计

统计量	值
样本数	46
平均值	0.5917
0.5333 标准差	0.5336 中位数
0.0000 最大值	2.9333 最小值 height

表 9: 置信度统计指标

6 性能分析

6.1 响应时间统计

性能指标	值
样本数	46
平均响应时间	0.69 ms 中位数响应时间
0.33 ms 最小响应时间	0.05 ms 最大响应时间
13.85 ms 标准差	2.00 ms P50响应时间
0.33 ms P95响应时间	1.42 ms P99响应时间
8.53 ms height	

表 10: 性能指标统计

6.2 各Skill性能

Skill	平均(ms)	最小(ms)	最大(ms)	请求数
greeting-handler	0.10	0.05	0.22	7
symptom-analyzer	1.92	0.27	13.85	10
department-recommender	0.60	0.21	2.03	7
medication-advisor	0.53	0.30	1.17	7
health-educator	0.42	0.23	0.97	6
fallback-handler	0.15	0.05	0.23	9

表 11: 各Skill性能统计

7 失败分析

7.1 失败用例详情

Skill	输入	预期	置信度
greeting-handler	你好...	greeting-handler	0.00
greeting-handler	hi...	greeting-handler	0.00
greeting-handler	嗨...	greeting-handler	0.00
fallback-handler	我想买一台电脑，你有什么推荐吗...	fallback-handler	0.30
fallback-handler	你好我想请问一下今天北京的天气怎么样，哦对了还有我头痛怎么办...	fallback-handler	0.95
symptom-analyzer	我headache好几天了...	symptom-analyzer	0.00
fallback-handler	我不头痛...	fallback-handler	0.53

8 测试用例详情

8.1 greeting-handler 测试用例

共7个测试用例，通过4个。

复杂度	输入	预期Skill	置信度	结果
simple	你好	greeting-handler	0.00	FAIL
simple	hi	greeting-handler	0.00	FAIL
medium	早上好，今天感觉不错	greeting-handler	0.95	PASS

复杂度	输入	预期Skill	置信度	结果
medium	谢谢你的帮助	greeting-handler	0.95	PASS
complex	你好啊, 请问你可以帮我吗? 我想咨询一些健康问题	greeting-handler	0.95	PASS
edge _{case}	嗨	greeting-handler	0.00	FAIL
edge _{case}	hello 你好	greeting-handler	0.95	PASS

8.2 symptom-analyzer 测试用例

共10个测试用例，通过9个。

复杂度	输入	预期Skill	置信度	结果
simple	我头痛	symptom-analyzer	0.53	PASS
simple	最近一直咳嗽	symptom-analyzer	0.20	PASS
medium	我头痛好几天了, 还有点发热	symptom-analyzer	0.73	PASS
medium	剧烈头痛, 伴有恶心呕吐	symptom-analyzer	0.60	PASS
complex	我这三天一直头痛, 特别是左边太阳穴位, 非常疼, 还有点眩晕的感觉, 严重影响了睡眠	symptom-analyzer	0.53	PASS
edge _{case}	感觉不舒服, 不知道哪里不对劲	symptom-analyzer	0.20	PASS

8.3 department-recommender 测试用例

共7个测试用例，通过7个。

复杂度	输入	预期Skill	置信度	结果
simple	头痛挂什么科	department-recommender	0.53	PASS
simple	肚子疼去哪个科	department-recommender	0.87	PASS
medium	我最近总是咳嗽，应该去看哪个科室	department-recommender	1.10	PASS
medium	皮肤过敏是挂皮肤科吗	department-recommender	0.47	PASS

复杂度	输入	预期Skill	置信度	结果
complex	我有关节痛还有皮肤红肿，应该是风湿科还是皮肤科	department-recommender	0.47	PASS
edge_case	脚趾甲长进肉里了要挂什么科	department-recommender	0.53	PASS
medium	头痛是不是应该挂神经内科？	department-recommender	0.73	PASS

8.4 medication-advisor 测试用例

共7个测试用例，通过7个。

复杂度	输入	预期Skill	置信度	结果
simple	阿莫西林怎么吃	medication-advisor	0.30	PASS
simple	布洛芬有什么副作用	medication-advisor	0.97	PASS
medium	成人阿莫西林一次吃多少，一天几次	medication-advisor	0.30	PASS
medium	阿莫西林和布洛芬能一起吃吗	medication-advisor	0.93	PASS
complex	我有胃溃疡，可以吃布洛芬止痛吗？有没有替代的药物？需要注意什么？	medication-advisor	1.10	PASS
edge_case	那个消炎药一天吃几次	medication-advisor	0.80	PASS
medium	我吃了3天药了	medication-advisor	0.80	PASS

8.5 health-educator 测试用例

共6个测试用例，通过6个。

复杂度	输入	预期Skill	置信度	结果
simple	怎么预防高血压	health-educator	0.93	PASS
simple	高血压不能吃什么	health-educator	0.57	PASS
medium	糖尿病患者有什么运动建议	health-educator	1.40	PASS
medium	高血压患者日常生活中要注意什么	health-educator	0.27	PASS
complex	我有高血压和糖尿病，饮食和运动方面有什么需要注意的吗？能不能推荐一些适合的运动项目？	health-educator	1.40	PASS
edge_case	怎么样才能保持健康	health-educator	0.30	PASS

8.6 fallback-handler 测试用例

共9个测试用例，通过6个。

复杂度	输入	预期Skill	置信度	结果
simple	今天天气怎么样	fallback-handler	0.00	PASS
simple	asdfgh	fallback-handler	0.00	PASS
medium	那个东西怎么用	fallback-handler	0.00	PASS

复杂度	输入	预期Skill	置信度	结果
medium	我想买一台电脑，你有什么推荐吗	fallback-handler	0.30	FAIL
complex	你好我想请问一下今天北京的天气怎么样，哦对了还有我头痛怎么办	fallback-handler	0.95	FAIL
edge _{case}		fallback-handler	0.00	PASS
edge _{case}	@#\$%^& *()	fallback-handler	0.00	PASS
edge _{case}	痛痛痛痛痛	fallback-handler	0.00	PASS
edge _{case}	我不头痛	fallback-handler	0.53	FAIL

9 测试结论

9.1 总体评估

本次测试结果为：**良好**。大部分功能模块运行正常，存在少量需要优化的问题。

9.2 各测试类型结论

- 功能测试: 31/36 通过 (86.1%)

9.3 改进建议

- 优化复杂和边界情况下的意图识别准确率
- 增加对长文本和特殊字符输入的处理能力

3. 重点优化以下Skill的准确性: greeting-handler, fallback-handler
4. 继续增加测试用例覆盖范围
5. 完善错误处理和兜底机制

10 附录

10.1 测试环境

项目	信息
系统	医疗智能助手 v1.0
测试框架	Python Asyncio + Pytest
MCP协议	自定义MCP协议
测试日期	2026年01月28日