

测试日期：2026年01月28日

版本：v1.0

总测试数：	46
通过数：	34
失败数：	12
通过率：	73.9%

自动化测试系统生成

目录

1 测试概述

1.1 总体统计

指标	数值
总测试数	46
通过数	34
失败数	12
通过率	73.9%

表 1: 总体测试统计

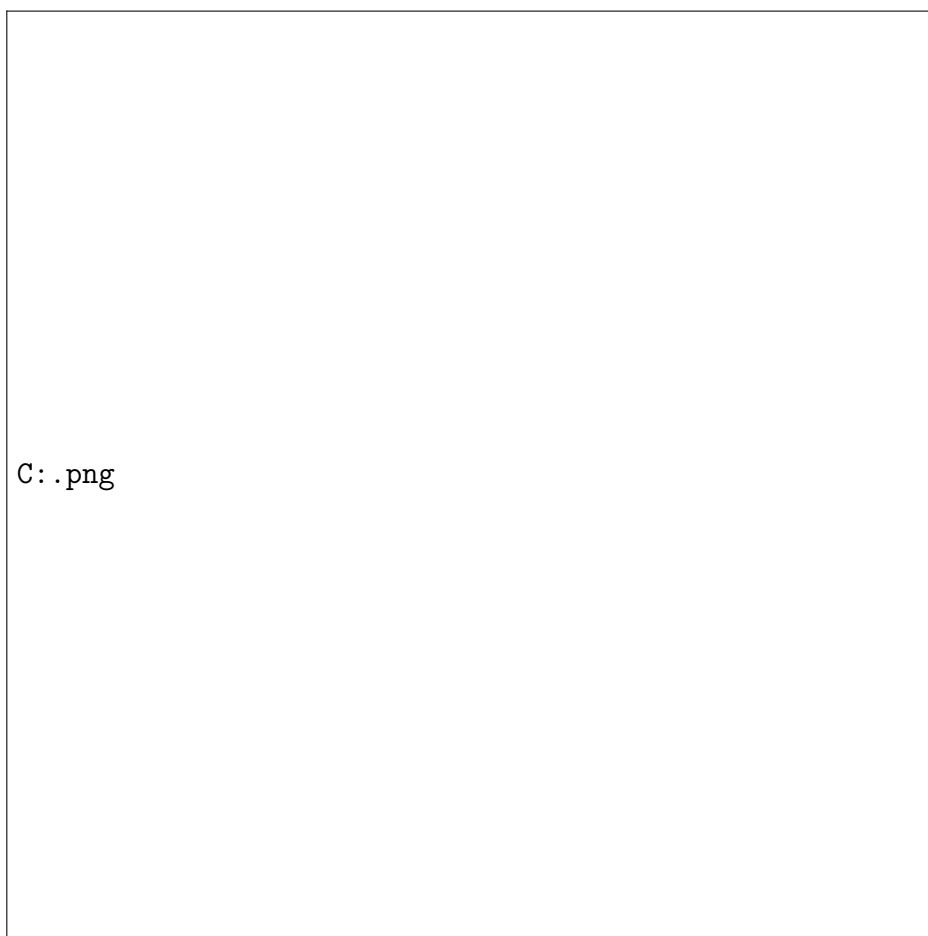


图 1: 测试概览

2 Skill测试结果



图 2: Skill性能分析

2.1 greeting-handler

指标	数值	通过率
总测试数	7	-
通过数	6	85.7%
失败数	1	-

表 2: greeting-handler 测试统计

2.2 symptom-analyzer

指标	数值	通过率
总测试数	11	-
通过数	6	54.5%
失败数	5	-

表 3: symptom-analyzer 测试统计

2.3 department-recommender

指标	数值	通过率
总测试数	7	-
通过数	5	71.4%
失败数	2	-

表 4: department-recommender 测试统计

2.4 medication-advisor

指标	数值	通过率
总测试数	7	-
通过数	6	85.7%
失败数	1	-

表 5: medication-advisor 测试统计

2.5 health-educator

指标	数值	通过率
总测试数	6	-
通过数	6	100.0%
失败数	0	-

表 6: health-educator 测试统计

2.6 fallback-handler

指标	数值	通过率
总测试数	8	-
通过数	5	62.5%
失败数	3	-

表 7: fallback-handler 测试统计

3 复杂度分析



图 3: 复杂度分析

复杂度	总测试	通过	通过率
simple	12	12	100.0%
medium	14	11	78.6%
complex	8	4	50.0%
edge _{case}	12	7	58.3%

表 8: 按复杂度统计

4 意图识别分析

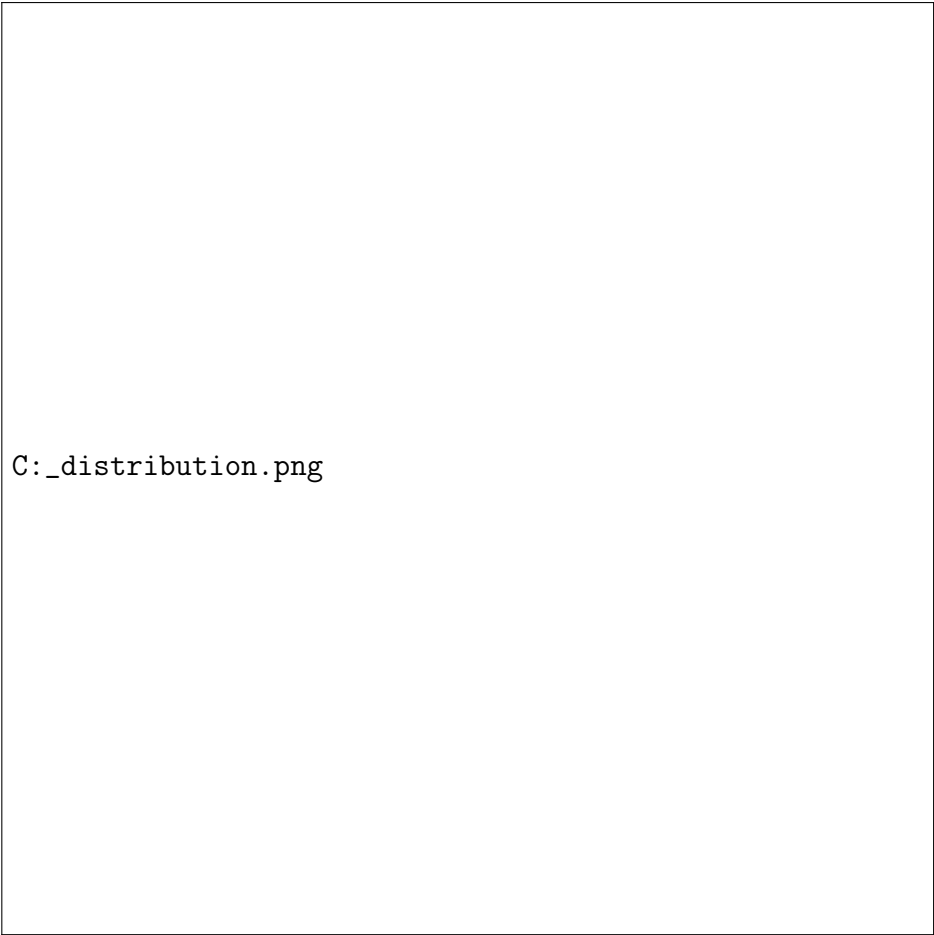


图 4: 置信度分布

统计量	值
平均值	0.626 中位数
0.600 标准差	0.545 最小值
0.000 最大值	3.100 height

表 9: 置信度统计

5 性能分析



图 5: 响应时间分析

性能指标	值
平均响应时间	0.74 ms
中位数响应时间	0.30 ms
最小响应时间	0.03 ms
最大响应时间	18.32 ms
标准差	2.64 ms
P95响应时间	1.04 ms
P99响应时间	10.83 ms

表 10: 性能指标统计

5.1 MCP工具性能

Skill	平均(ms)	最小(ms)	最大(ms)	请求数
greeting-handler	0.07	0.04	0.10	7
symptom-analyzer	2.20	0.14	18.32	11
department-recommender	0.29	0.15	0.45	7
medication-advisor	0.50	0.32	0.94	7
health-educator	0.36	0.13	0.70	6
fallback-handler	0.19	0.03	0.78	8

表 11: MCP工具性能统计

6 测试用例详情

6.1 greeting-handler 测试用例

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

复杂度	输入	预期Skill	置信度	结果
edge _{case}	hello 你好	greeting-handler	0.95	FAIL

6.2 symptom-analyzer 测试用例

复杂度	输入	预期Skill	置信度	结果
simple	我头痛	symptom-analyzer	0.70	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	0.90	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置，非常疼，还有点眩晕的感觉，严重影响了睡眠	symptom-analyzer	0.70	PASS
edge _{case}	感觉不舒服，不知道哪里不对劲	symptom-analyzer	0.20	PASS
edge _{case}	头痛	symptom-analyzer	0.20	FAIL
edge _{case}	我headache好几天了	symptom-analyzer	0.00	FAIL

复杂度	输入	预期Skill	置信度	结果
edge _{case}	痛痛痛痛痛	symptom-analyzer	0.00	FAIL
complex	我头痛，发热，咳嗽， 恶心， 呕吐， 腹泻， 失眠， 乏力， 疼痛， 痒， 不适， 难受， 不 舒服， 症状	symptom-analyzer	3.10	FAIL
complex	如果头痛发热应该怎 么办	symptom-analyzer	0.40	FAIL

6.3 department-recommender 测试用例

复杂度	输入	预期Skill	置信度	结果
simple	头痛挂什么科	department-recommender	0.70	PASS
simple	肚子疼去哪个科	department-recommender	1.20	PASS
medium	我最近总是咳嗽，应 该去看哪个科室	department-recommender	1.15	PASS
medium	皮肤过敏是挂皮肤科 吗	department-recommender	0.20	PASS
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	0.50	FAIL
edge _{case}	脚趾甲长进肉里了要 挂什么科	department-recommender	0.70	PASS
medium	头痛是不是应该挂神 经内科？	department-recommender	0.20	FAIL

6.4 medication-advisor 测试用例

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	0.60	PASS
edge _{case}	那个消炎药一天吃几次	medication-advisor	0.30	PASS
medium	我吃了3天药了	medication-advisor	0.30	FAIL

6.5 health-educator 测试用例

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

复杂度	输入	预期Skill	置信度	结果
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

6.6 fallback-handler 测试用例

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
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

复杂度	输入	预期Skill	置信度	结果
edge _{case}	我不头痛	fallback-handler	0.70	FAIL

7 测试结论

7.1 总体评估

本次测试结果为：**需要改进**。部分功能模块存在问题，需要进行修复和优化。

7.2 详细结论

- 功能测试: 33/36 通过 (91.7%)