모델 테스트 계획 및 결과 보고서

# 1. 개요

본 문서는 사용자에게 특허 관련 서비스를 제공하기 위한 AI 시스템의 구성요소(Qwen3 메인 모델, RAG 시스템, LangChain Agent)의 통합 구조에 대한 테스트 계획 및 그 결과를 정리한 보고서입니다. 각 기능은 특허 초안 작성, 아이디어 보완, 법률 기반 챗봇 상담 등으로 구성되어 있으며, 각 기능에 대해 시스템 정확성, 응답 일관성, 처리 속도 등을 평가하였습니다.

# 2. 시스템 구성

|  |  |
| --- | --- |
| 구성 요소 | 역할 |
| Qwen3 모델 | 특허 초안 작성, 응답 생성의 주 모델 |
| RAG (Retriever-Augmented Generation) | 심사기준, 법률, KIPRIS 데이터 기반 정보 검색 |
| LangChain Agent | 사용자 질의 분석, 도구 선택 및 응답 최종 조율 |
| Tool 구성 | `law\_retriever`, `similar\_patent\_search`, `feedback\_evaluator` 등 총 5개 |

# 3. 테스트 항목 및 시나리오

## 3.1 테스트 목적

- Agent 기반 도구 선택 및 조합 정확도 평가  
- 사용자 요구에 따른 응답 품질 평가  
- 시스템 전체 응답 시간 및 안정성 측정

## 3.2 테스트 시나리오

|  |  |  |  |
| --- | --- | --- | --- |
| 시나리오 번호 | 기능 | 테스트 내용 | 입력 예시 |
| TC-01 | 특허 초안 작성 | Qwen3가 형식에 맞는 초안을 작성하는지 | 음성 감정을 분석하는 기술에 대해 초안을 작성해줘 |
| TC-02 | 초안 피드백 | Agent가 심사기준 기반 피드백을 줄 수 있는지 | 초안 텍스트 입력 |
| TC-03 | 유사 특허 추천 | RAG + Qwen3 + Agent 흐름으로 유사 특허 제공 및 보완 제안 | 화재 위험 감지를 위한 IoT 기술이 있는데 특허 가능성 있어? |
| TC-04 | 특허 상담 | 법률/가이드 기반 챗봇이 정확히 응답하는지 | 컴퓨터 프로그램은 특허가 가능한가요? |
| TC-05 | 비정형 입력 대응 | 비형식적 입력에 대한 Agent 대응 능력 | 이 기술이 등록 가능할까요? |

# 4. 테스트 결과 요약

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 항목 | 평가 지표 | 기대 기준 | 결과 | 비고 |
| 초안 생성 정확도 | 항목 누락 여부, 표현 적절성 | 90% 이상 정합 | 92% | 형식 맞춤 성능 양호 |
| 피드백 정밀도 | 항목별 평가 정확성 | 기준 일치율 85% | 88% | Agent가 RAG 활용 잘함 |
| 유사 특허 검색 품질 | 검색된 특허의 관련도 | 3개 중 2개 이상 관련 | 만족 | 검색 + 생성 결합 양호 |
| 응답 일관성 | 반복 질문에 대한 동일 응답 | ±1 이내 편차 | 유지 | Qwen3 + Agent 안정 |
| 처리 속도 | 응답 지연 시간 | 5초 이내 | 평균 3.7초 | API 캐싱 도움 |

# 5. 주요 개선점 및 리스크

|  |  |  |
| --- | --- | --- |
| 항목 | 내용 | 대응 방안 |
| 질의 분류 오차 | Agent가 의도를 간혹 잘못 분류함 | Prompt 강화 및 임베딩 기반 질의 클러스터링 추가 |
| 유사 특허 설명 다양성 부족 | Qwen3가 비슷한 형식 반복 | Post-Processing 템플릿 다양화 |
| 초안 일부 표현 과다 일반화 | "기술적 특징"이 추상적일 수 있음 | 체크리스트 기반 피드백 추가 예정 |

# 6. 결론

Agent 중심의 Qwen3 + RAG 통합 시스템은 대부분의 사용자 요청에 대해 안정적이고 정확한 응답을 제공하는 것으로 확인되었습니다. 특히 특허 초안 작성 및 피드백 기능에서 실무자 관점에서의 활용성이 높았으며, Agent를 통한 도구 연계 및 응답 조율의 성능도 만족스러운 수준이었습니다. 향후 사용자 피드백과 실전 적용을 기반으로 지속적인 개선이 필요합니다.

# 7. 부록

- 테스트에 사용된 예시 질의 리스트 (20개)  
- 초안 작성 결과 샘플 3종  
- RAG 검색 결과 로그 예시