테스트 계획 및 결과 보고서

# 테스트 계획 (Test Plan)

■ 테스트 목적   
\* Gemma3 기반 파인튜닝된 LLM이 다음 조건에 맞는지 검증:  
 - 규정에 맞는 항공 관련 정보 제공  
 - 정보가 없는 국가에 대한 정보 요청에 대해 일관된 거절 응답 처리  
  
■ 테스트 범위   
\* 정답 응답 케이스: 항공 화물, 위험물, 반입 가능 품목, 여행지 정보 등   
\* 거절 응답 케이스: 자료가 존재하지 않는 국가별 문화·예절·금지 품목 질문  
  
■ 테스트 구성   
\* 정답 케이스: 14건   
\* 거절 케이스: 10건   
- 총합: 24건  
  
■ 테스트 기준   
\* 정답 응답: 기대 응답 내용과 동일하거나 유사한 항공 관련 정보 포함   
\* 거절 응답: "해당 국가에 대한 정보는 제공하지 않습니다." 포함   
\* 실패 기준: 기대와 다른 내용, 오답, 엉뚱한 정보, 무응답 등  
  
■ 테스트 방식   
\* RetrievalQA 체인을 통해 질문 입력 → 모델 응답 → 기대값과 비교   
\* 수동 평가로 통과/실패 판별  
  
**테스트 결과 보고서 (Test Report)**  
■ 테스트 요약   
\* 총 케이스: 24 중 통과 9/24 통과율 37.5%  
  
■ 통과/실패 분포  
\* 통과된 케이스 (9건)   
 - QA001 ~ QA007: 항공 정보 관련 응답 정확   
 - QA011: 일본 관광지 추천   
 - QA012: 대만 음식 추천   
  
\* 실패된 케이스 (15건)   
 - QA008~QA014: 일부 엉뚱한 정보 또는 미흡한 설명   
 - NG001~NG010: 거절 응답 실패 (실제 국가 정보 제공)  
  
■ 개선 포인트 요약   
\* 거절 실패: "정보 제공하지 않음" 응답 대신 실제 정보 포함   
\* 정답 불명확 or 혼란: 규정 기준 애매함, 수치 오류   
\* 문맥 일탈: 질문과 무관한 국가 정보 출력   
\* 일부 응답 포맷은 적절 (QA001~QA007)  
  
■ 개선 권장 사항   
\* 거절 응답 정책 강화   
\* 정답 응답 정제 및 기준 통일 (Wh 기준 등)