붙임 2 **활용 사례**

제목	산지니AI의 신뢰도 제고를 위한 3단계 자가 검증 프롬프팅 방안 - 학습 · 분석 · 진로 준비에서의 활용 사례를 중심으로

<1. 개요>

1.1 활용 필요성 및 목적

ChatGPT, Claude 등 생성형 AI는 학습, 자료조사, 문서 작성, 취업 준비 등 대학 생활 전반으로 확산되었다. **부산대학교 역시 AI 윤리헌장과 AI 활용 가이드라인을 수립하였으며, 부산대 특화 AI 서비스인 산지니AI를 운영 중이다.** 이는 AI를 단순한 편의 도구가 아니라, 책임 있는 활용이 요구되는 보조 수단으로 자리매김하겠다는 방향성을 보여준다.

그러나 AI 답변에는 구조적 한계가 존재한다. 표면적으로는 유창하고 논리적이지만, 검증해 보면 사실과 다르거나 과장된 정보가 포함되어 있는 경우가 적지 않다. 특히 최신 정책, 기업 분석, 통계 기반 과제처럼 정확한 근거가 요구되는 영역에서는 이러한 문제가 두드러진다.

이 문제에 대한 접근은 AI를 더 정교하게 활용하는 방법론을 설계하는 것이었다. 단순히 사실 확인을 요청하는 수준을 넘어, 조사 설계 → 근거 기반 분석 → 자기 검증이라는 3단계 구조를 거치면 결과물의 신뢰도가 유의미하게 향상된다는 점을 반복 실험을 통해 확인하였다.

1.2 활용 분야

본 방법론은 다음과 같은 영역에 적용할 수 있다.

1)전공 수업 리포트 및 학술 자료 정리
2)자격증 학습(투자자산운용사, OPIc, ADSP 등)
3)기업 및 산업 분석 과제
4)취업 면접 준비 및 기관 조사

이는 부산대의 "AI를 사고력과 창의성을 확장하는 도구로 활용해야 한다"는 원칙과 부합한다. **핵심은 AI가 생성한 답변을 그대로 수용하는 것이 아니라, 조사와 검증 과정에서 AI를 보조 수단으로 활용하는 데 있다.**

<2. 프롬프트 설계 및 세부 내용>

2.1 설계 원칙

본 3단계 프롬프트의 설계 원칙은 다음 세 가지이다.

1)조사 설계 선행: 답변을 바로 요구하지 않고, 먼저 어떤 정보가 필요한지, 어떤 기준으로 판단해야 하는지를 AI가 구조화하도록 유도한다.

2)근거와 리스크의 병행 제시: 모든 주장에 출처를 달게 하고, 긍정적 근거와 부정적 근거(리스크)를 함께 제시하도록 강제한다.

3)자기 검증 유도: AI가 스스로 답변의 한계를 점검하고, 단정적 표현이나 출처 미확인 정보를 수정하도록 한다.

이러한 설계는 부산대 AI 가이드라인의 핵심 원칙, 즉 "AI를 활용하되 본인의 사고와 판단 과정을 생략해서는 안 된다"는 방침을 프롬프트 수준에서 구현한 것이다.

2.2 [1단계] 맥락 부여 및 목표 설정

[목적] AI에게 현재 상황과 목표를 명확히 전달하여, 본격적인 분석에 앞서 조사 프레임워크를 먼저 구조화하도록 유도한다.

<프롬프트 템플릿>

```
나는 현재 [나의 상황/역할]이다. [최종 목표]를 위해 [조사 주제]를 분석하려 한다.
이 주제를 완성도 높게 이해하려면:
(1) 어떤 핵심 정보 항목이 필요한가?
(2) 각 항목은 어디에서(공식 자료, 학술DB, 뉴스 등) 찾을 수 있는가?
(3) 정보의 신뢰성을 판단할 기준은 무엇인가?
(4) 분석 시 반드시 포함해야 할 관점(긍정/부정/중립)은?
위 4가지를 표 형태로 정리하라
```

<적용 예시>

```
"나는 금융공기업 취업을 준비 중인 경영학과 4학년이다. KDB 면접 대비를 위해 'AI 기술 도입이 KDB의 경쟁력에 미치는 영향'을 분석하려 한다. 이 주제를 완성도 높게 이해하려면 (1) 어떤 핵심 정보 항목이 필요한가? (2) 각 항목은 어디에서 찾을 수 있는가? (3) 정보의 신뢰성을 판단할 기준은? (4) 반드시 포함해야 할 분석 관점은? 위 4가지를 표로 정리하라."
```

<주의>

```
1) 역할과 목적이 구체적일수록 AI의 프레임 설정이 정교해진다.
2) "표 형태로 정리하라"를 추가하면 구조화된 출력을 유도할 수 있다.
3) 한 번에 답변을 요구하지 않고, 먼저 '조사 설계'만 요청하는 것이 핵심이다.
```

2.3 [2단계] 근거 기반 분석 및 리스크 검증

[목적] 1단계에서 정리된 항목별로 출처가 명시된 양면 분석을 수행하되, AI 스스로 확신도를 표기하게 하여 환각 위험을 가시화한다.

<프롬프트 템플릿 및 예시>

위에서 정리한 항목에 대해 분석하라. 반드시 아래 4가지 규칙을 준수할 것.

【규칙 1 - 출처 명시】 모든 수치와 주장에 "YYYY년 MM월 [기관명] [자료 유형]에 따르면" 형식으로 출처를 기재한다. 출처를 찾을 수 없는 정보는 "※ 출처 미확인"으로 표기한다.
【규칙 2 - 양면 분석】 각 항목마다 [긍정적 근거]와 [부정적 근거/리스크]를 모두 제시한다.
【규칙 3 - 표현 정확성】 "일반적으로", "대부분", "알려져 있다" 등 모호한 표현 대신 구체적 데이터나 사례로 뒷받침한다.
【규칙 4 - 확신도 표시】 각 주장 옆에 확신도를 표시한다.
[상] 공식 발표/학술자료 기반
[중] 신뢰 언론/기관 보고서 기반
[하] 참고 수준(블로그, 커뮤니티 등)
[미확인] AI 생성 추론(검증 필요)

<주의>

1) [미확인] 표시가 많으면 해당 부분은 반드시 직접 검색으로 확인해야 한다.
2) 웹검색 기능이 없는 AI에서는 출처를 생성(환각)할 수 있으므로 교차확인이 필수이다.

2.4 [3단계] 자기 검증 및 한계 고백

[목적] AI가 스스로 자신의 답변을 비판적으로 재검토하게 하여, 오류와 한계를 명시적으로 드러내도록 한다.

<프롬프트 템플릿>

이제 위 답변을 아래 체크리스트로 자가 검증하라.
☐ 사실 검증: 제시한 수치 중 공식 출처가 없는 것이 있는가? → "※ 검증 필요"로 수정
☐ 반대 검증: "[주제] 실패 사례" 또는 "[주제] 부작용" 관점에서 놓친 논점은 없는가? →추가
☐ 최신성 검증: 제시한 자료 중 1년 이상 경과한 것은 없는가? → 최신 업데이트 여부 명시
☐ 표현 검증: "확실히", "반드시" 같은 단정적 표현이 있는가? → 조건부 표현으로 수정
☐ 누락 검증: 원래 질문의 목적에 비추어 빠뜨린 핵심 논점은 없는가? → 보충

위 검증 결과를 [수정 전 → 수정 후] 형식으로 제시하고, 마지막에 "전체 신뢰도 자가평가: __/10"을 기재하라.

<예시>

"이제 위 답변을 아래 체크리스트로 자가 검증하라.
☐ 사실 검증: 제시한 수치 중 공식 출처가 없는 것이 있는가? → '※ 검증 필요'로 수정
☐ 반대 검증: 'KDB의 AI 도입 실패 사례' 또는 'AI 도입 부작용' 관점에서 놓친 논점은 없는가? → 추가
☐ 최신성 검증: 제시한 자료 중 1년 이상 경과한 것은 없는가? → 최신 업데이트 여부 명시
☐ 표현 검증: '확실히', '반드시' 같은 단정적 표현이 있는가? → 조건부 표현으로 수정
☐ 누락 검증: 'KDB 면접 대비'라는 원래 목적에 비추어 빠뜨린 핵심 논점(예: 면접관이 중시할 공공성·수익성 균형 관점)은 없는가? → 보충

위 검증 결과를 [수정 전 → 수정 후] 형식으로 제시하고, 마지막에 '전체 신뢰도 자가평가: __/10'을 기재하라."

<주의>

1) 자가평가 점수가 7 이하이면 해당 주제는 직접 자료조사가 필수이다.
2) [수정 전 → 수정 후] 형식을 요구하면 어떤 부분이 약한지 한눈에 파악할 수 있다.
3) 3단계까지 완료한 후에도 핵심 수치와 출처는 반드시 원문에서 직접 확인해야 한다.

<3. 단순 요청과 3단계 프롬프트의 결과 비교>

동일한 주제에 대해 두 가지 접근 방식을 비교하면 다음과 같다.

비교 항목	*단순 요청*	*3단계 프롬프트*
분석 구조	없음.	조사 항목 정리 후 항목별 분석
출처 제시	거의 없음.	시점/기관명/확신도 표시
리스크 포함	긍정 측면 위주	긍정/부정 양면 분석 필수
자기 검증	없음.	5개 체크리스트 기반 검증
소요 시간	짧음.	길지만 신뢰도 향상
환각 위험	높음(검증 장치 부재)	감소(완전 제거는 아님)

[비교 예시] "마이데이터가 금융업에 미치는 영향을 설명해 달라"고 **단순 요청하면**, AI는 "편의성 증가, 맞춤형 서비스 확대"와 같은 일반론만 제시한다.

그러나 3단계 프롬프트를 적용하면, 먼저 제도 현황/이용 규모/사용자 편익/사업성 및 리스크라는 4가지 분석 틀이 정리된다. 이후 "2024년 2월 기준 69개 사업자, 1억 1,787만 명(금융위원회 공식 발표, [상])"이라는 구체적 수치가 나오며, 동시에 "사업자 수익모델 부재로 KB핀테크, NHN페이코 등이 사업자 자격 반납"이라는 리스크도 함께 제시된다.

<4. 활용 사례>

4.1 자격증 학습: 마이데이터 산업의 금융권 영향

[1단계] "투자자산운용사, ADSP 자격증을 준비 중이며, 마이데이터가 금융권에 미치는 영향을 공부하고 있다"고 맥락을 제시하자, AI는 제도 현황, 이용 규모, 사용자 편익, 사업성 및 리스크의 4가지 분석 틀을 먼저 제시하였다.

[2단계] 금융위원회 공식 보도자료(2024년 4월)를 근거로 69개 사업자, 1억 1,787만 명이라는 구체적 수치가 제시되었으며, 동시에 자산 상세조회 미흡, 오프라인 가입 제한, 중복 동의 절차 등의 개선 과제도 병행 도출되었다.

[3단계] AI가 "사용자 편익은 크지만, 수익모델과 서비스 차별화가 부족하면 산업 전체 성과는 제한될 수 있다"로 결론을 조건부 표현으로 수정하였다.

4.2 기업 분석: AI 기술 도입이 금융공기업 가치에 미치는 영향

[1단계] 도입 목적, 비용 절감 가능성, 규제 환경, 리스크 관리 개선이라는 판단 기준이 정리되었다.

[2단계] 금융위원회의 2024년 12월 "금융권 생성형 AI 활용 지원 방안" 발표를 긍정적 근거로 제시하되, 규제 불확실성, 도입 비용, AI 모델 리스크를 동시에 언급하였다.

[3단계] 결론이 "AI 도입 → 가치 상승"이라는 단정에서, "AI 도입 + 비용 효율화 + 규제 대응 성공이라는 조건 하에서 가치 상승이 가능하다"로 조정되었다.

4.3 취업 준비: 예금보험공사의 역할과 최근 과제

[1단계] 설립 목적, 핵심 업무, 최근 제도 변화, 향후 과제의 4가지 관점이 정리되었다.

[2단계] KDIC 공식 자료([상])를 통해 "예금보호한도 1인당 1억 원, 2025년 9월 1일 시행"을 명확히 제시하였다.

[3단계] "공식 자료는 제도 설명에 적합하지만, 비판적 쟁점이나 외부 평가가 부족하다"는 한계를 AI가 스스로 지적하고, 면접 준비 시 신문 기사나 정책 토론 자료로 보완할 것을 권고하였다.

<5. 기대효과>

1) 답변 신뢰도 향상: 출처가 명시되고, 다각도 분석이 이루어지며, AI가 스스로 약점을 인정함으로써 과제나 면접 준비에 바로 활용할 수 있는 수준의 결과물이 도출된다.

2) 비판적 AI 활용 습관 형성: AI의 출력을 무비판적으로 수용하는 대신, 조사 → 검증 → 수정이라는 사고 과정을 체득하게 된다. 이는 부산대가 제시한 "사고력을 확장하는 도구로서의 AI"라는 비전에 부합한다.

3) 실무 역량과의 연결: 정보 수집과 판단을 동시에 수행해야 하는 실무 환경과 유사한 과정을 반복함으로써, 특히 정책 변화가 잦고 공식 자료 확인이 중요한 금융 분야에서의 업무 적응력을 높일 수 있다.

<6. 한계 및 유의사항>

본 방법론이 갖는 한계를 명확히 인식해야 한다.

1) 환각의 감소이지, 제거가 아니다.
3단계를 거치더라도 AI가 출처를 생성(날조)하거나 수치를 왜곡할 가능성은 남아 있다. 최종 단계에서 반드시 원문 자료를 직접 확인해야 한다.

2) 웹검색 기능 부재 시 2단계의 실효성이 저하된다.
웹검색을 지원하지 않는 AI 환경에서는 출처를 지어내는 것이 가장 위험한 환각 유형이다. 이 경우 AI 답변의 출처를 별도로 검색하여 교차확인하는 작업이 필수적이다.

3) 간단한 질문에는 과도한 방법론이 될 수 있다.
단순 개념 확인이나 짧은 정보 요청에 3단계를 모두 적용하면 비효율적이다. 근거가 중요한 과제, 보고서, 면접 준비 등에 선별적으로 적용하는 것이 바람직하다.

4) 확신도 등급은 AI가 아닌 사용자가 최종 판단해야 한다.
AI에게 출처 신뢰도를 매기라고 하면 자의적으로 높은 등급을 부여할 수 있다. AI가 표기한 확신도는 초안 수준의 참고 지표이며, 사용자가 검토하여 최종 조정해야 한다.

<7. 결론>

AI가 점점 더 정교해지는 만큼, 사용자 역시 AI를 더 체계적으로 활용하는 역량이 요구된다. **좋은 질문이란 단순히 긴 질문이 아니라, "무엇을 먼저 조사할 것인가", "어떤 근거로 판단할 것인가", "내가 놓친 것은 무엇인가"를 체계적으로 구성하는 능력에서 비롯된다.**

본 3단계 프롬프팅 방법론은 산지니AI, ChatGPT, Claude 등 어떤 생성형 AI에서든 적용할 수 있다.

부산대가 제시한 "AI를 책임감 있게 활용해야 한다"는 원칙은, AI가 생성한 답변을 그대로 수용하는 것이 아니라 사용자가 주도적으로 조사와 판단 과정을 이끌어 가는 것을 의미한다. 이 방법론이 그러한 실천의 구체적인 출발점이 될 수 있기를 기대한다.