

생성형 AI 기반 '사용자 조사 인터뷰' 파생질문 생성 및 분석 연구

(Exploration and Analysis of Derivative Interview Questions Using Generative AI in User Experience Research)

박유진*, 이종섭**

(Yujin Park, Joong Sup Lee)

요약

본 연구는 사용자 경험(UX) 리서치 과정에서 생성형 AI를 활용한 인터뷰 파생 질문 생성의 효용성을 검증한다. 그동안 AI의 인터뷰 질문 생성 잠재력은 구체적인 평가 기준을 통해 체계적으로 검증되지 않았으며, 특히 파생 질문의 질적 특성과 탐색 심화 효과에 대한 분석은 미비했다. 이에 본 논문은 AI 에이전트 간의 시뮬레이션을 통해 구조화된 파일럿 인터뷰 응답을 도출하고, 이를 바탕으로 파생 질문을 자동 생성하는 모델을 구현하여 그 실효성을 평가했다. 분석 결과, 생성형 AI는 답변 내 핵심 키워드를 포착해 논리적 연속성(3.96점/5.0점)을 유지하는 데는 탁월했으나, 텍스트 이면의 잠재적 니즈와 통합적 맥락을 파악하는 통찰의 깊이(3.21점/5.0점)에서는 한계를 보였다. 결론적으로 생성형 AI는 인터뷰 질문 초안 작성의 효율성을 획기적으로 높일 수 있으나, 심층적 인사이트 도출을 위해서는 인간 연구자의 개입(Human-in-the-loop)이 필수적임을 확인하였다.

■ 중심어 : 생성형AI ; 인공지능 ; 인터뷰질문 ; 반구조화인터뷰 ; 사용자경험디자인

Abstract

This study evaluates the utility of Generative AI in creating derivative interview questions for user experience (UX) research. Previously, its potential in this area lacked systematic verification through specific evaluation criteria, particularly regarding the qualitative characteristics and exploratory depth of follow-up questions. Therefore, we simulated an AI-driven pilot interview to generate standardized responses and subsequently autonomously create derivative questions. Quantitative and qualitative analyses revealed that while AI excels at capturing key terms to maintain logical continuity (3.96/5.0), it shows clear limitations in grasping potential needs and holistic contexts behind the text (3.21/5.0). In conclusion, while Generative AI significantly improves the efficiency of drafting initial questions, human intervention (Human-in-the-loop) is essential for deriving in-depth insights.

■ keywords : Generative AI ; Artificial Intelligence ; Interview questions ; Semi-structured Interview ; UX Design

1. 서론

1.1. 연구 배경 및 목적

최근 생성형 인공지능(Artificial Intelligence, AI)의 발전으로 다양한 산업 분야에서 그 활용 가능성이 극대화되고 있다. 특히 디자인 분야에

서는 이미지 생성, 프로토타입 제작, 사용자 경험(UX) 설계 등 다양한 과정에서 중요한 역할을 하고 있다[1,2]. 그러나 대부분의 연구가 이미지 생성에 집중되어 있어 AI의 다른 잠재적 용도에 대한 탐구는 상대적으로 미흡한 실정이다[3,4].

* 준회원, 한동대학교 문화미디어디자인학과

** 정회원, 한동대학교 콘텐츠융합디자인학부

이 논문은 한동대학교 교내연구지원사업 제202500740001호에 의한 것임.

접수일자 : 2025년 12월 22일

수정일자 : 2026년 02월 04일

재제확정일 : 2026년 03월 08일

교신저자 : 이종섭 e-mail : joongsup@handong.edu

생성형 AI는 언어모델로서 UX 디자인의 초기 단계인 기획과 사용자 조사에서도 큰 잠재력을 보이고 있다[5]. 특히 질문을 생성하는 능력을 활용하여 사용자 조사에 적용될 수 있는 가능성이 있다. 그러나 사용자 조사에서 AI를 체계적으로 활용한 연구는 드물며, 주로 생성형 AI의 활용도를 아이디어 발상과 디자인 구체화에 초점을 맞추고 있어 UX 디자인의 초기 과정인 사용자 조사에서는 크게 활용되지 않고 있다[3,4]. 이는 디자인 프로세스에서 인간의 창조적 노력을 AI가 대체할 수 있을지에 대한 불확실성이 그 원인으로 보인다[6].

본 논문은 생성형 AI를 활용하여 인터뷰의 파생 질문을 자동 생성하는 방법을 탐구하고자 한다. 생성형 AI는 확률적으로 높은 단어의 조합을 택하여 답변을 생성하기 때문에 보편적이고 일반적인 답변을 제공하는 경향이 있다[7]. 이러한 특성을 활용하여, 인터뷰이의 페르소나가 부여된 AI에게 설계된 메인 질문에 응답하게 하여 보편적인 응답을 얻고, 해당 인터뷰 질문과 응답 내용을 또 다른 생성형 AI에게 입력하여 연관된 파생 질문을 생성하고자 한다.

본 연구를 통해 기존 사용자 조사 과정에서 파일럿 인터뷰(Pilot Interview)를 수행하여 파생 질문을 생성하고 교정했던 인터뷰 질문 설계 단계에서 생성형 AI를 활용하여 대체할 수 있는 가능성을 검증하고자 한다. 또한 생성형 AI가 인터뷰 질문 개발 및 최적화 과정에서 어떻게 효과적으로 활용될 수 있는지를 평가하고자 한다. 이를 통해 AI가 사용자 경험 조사, 특히 인터뷰 질문 개발 및 최적화 과정에서 어떻게 효과적으로 활용될 수 있는지를 제시하고, 디자이너와 연구자들이 AI를 보조 도구로서 활용하여 기존 업무 효율성을 개선할 수 있는 방안을 제안하고자 한다.

1.2. 이론적 배경

가. 생성형 AI의 개요

생성형 인공지능(Generative AI)은 주어진 데

이터셋을 기반으로 새로운 콘텐츠를 생성하는 능력을 갖고 있다. 이 AI 모델들은 대규모 데이터셋을 학습하여 텍스트, 이미지, 음악 등 다양한 종류의 콘텐츠를 생성할 수 있다. 특히 GPT-4o와 같은 언어 모델은 자연어 처리(NLP) 작업에서 뛰어난 성능을 보이며, 단어의 출현 확률을 계산하여 문장을 생성한다. 이러한 기능은 창조적인 답변보다는 일반적이고 보편적인 답변을 제공함으로써, 파일럿 인터뷰와 같은 인터뷰 준비 단계에서 유용하게 활용될 수 있다[7]. 생성형 AI를 사용하면 일반적인 응답을 토대로 인터뷰 질문을 보다 효과적으로 다듬을 수 있다.

나. 인터뷰 유형과 질문의 유형

인터뷰는 준비된 질문의 정도에 따라 세 가지 유형으로 나눌 수 있다. 구조화된 인터뷰(Structured Interview)는 미리 준비된 질문으로 구성되며, 모든 인터뷰 대상자에게 동일한 질문을 한다. 비구조화된 인터뷰(Unstructured Interview)는 사전에 정해진 질문 없이 자유로운 대화를 통해 정보를 얻는다. 반구조화된 인터뷰(Semi-structured Interview)는 주요 질문(Main questions)과 사전에 설계된 파생 질문(Planned follow-up questions), 그리고 상황에 따라 즉흥적으로 묻는 질문(Spontaneous follow-up questions)을 포함하여 유연성을 가진다[8].

다. 선행 연구

사용자 경험(UX) 연구에서 인터뷰 질문 설계는 사용자의 잠재된 니즈를 파악하고 심층적인 인사이트를 도출하는 핵심적인 단계로 인식된다[9-11]. 그러나 전통적인 인터뷰 질문 구성은 연구자의 개인적 경험과 훈련 수준에 의존하는 경향이 짙으며, 수동으로 이루어지는 복잡한 과정이라는 한계가 있다[12]. 이러한 방식은 많은 시간과 비용을 소모할 뿐만 아니라, 연구자의 편향이 개입될 여지가 있어 새롭고 다양한 질문 풀(Pool)을 확보하는 데 제약이 된다[13-15]. 특히

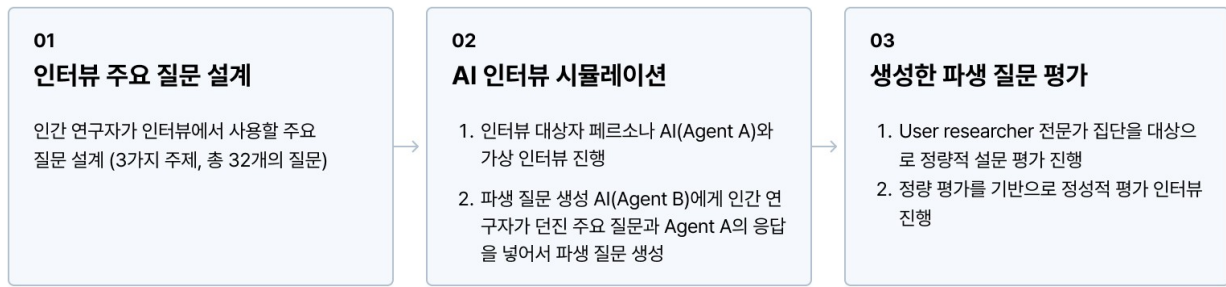


그림 1. 연구 프로세스

UX 디자인 분야에서는 연구자가 개방형 질문을 통해 탐색을 확장하면서도, 동시에 연구 목적에 맞는 질문의 방향성을 유지해야 한다는 점에서 그 난이도가 높다[16,17].

이러한 한계를 극복하기 위해 최근에는 인공지능 기술을 인터뷰 프로세스에 접목하여 효율성을 높이려는 연구가 활발히 진행되고 있다. 초기에는 이력서 기반의 맞춤형 질문 생성이나 기술 면접을 위한 질문 추천 시스템 등이 제안되었다[18]. 나아가 소규모 코퍼스를 이용하거나[19], 비동기식 인터뷰 상황에서 자동으로 후속 질문(Follow-up Question)을 생성하는 연구들이 등장하며[20] 실질적인 인터뷰 보조 도구로서의 가능성을 보여주었다.

특히 대규모 언어 모델(LLM)의 등장 이후, 다중 에이전트(Multi-agent) 상호작용을 기반으로 인터뷰 프로세스 전체를 설계하거나, ChatGPT를 활용해 텍스트 일관성과 상황 이해도가 높은 고품질의 질문 데이터셋을 생성하는 시도가 이루어지고 있다[21]. 선행 연구들은 AI 기반의 질문 생성이 인터뷰 준비의 효율성을 높이고, 보다 체계적인 프로세스를 구축하는 데 기여할 수 있음을 시사한다. 본 연구는 이러한 기술적 흐름에 발맞추어, 생성형 AI를 활용해 UX 인터뷰의 파생 질문을 생성하고 그 유효성을 검증함으로써 사용자 조사 방법론의 확장에 기여하고자 한다.

II. 연구 방법

본 논문은 생성형 AI를 활용하여 인터뷰 시뮬

레이션을 기반으로 한 파생질문 생성 과정을 탐구한다. 인터뷰의 주제는 '간호사의 문화적 배경 및 도메인 이해와 수간호사 근무표 제작 경험 파악'으로 선정되었다. 해당 주제는 암묵지와 전문용어 그리고 상충하는 복잡한 규칙이 혼재된 고난도 도메인으로서 AI가 단순 정보 생성을 넘어 복잡한 맥락을 이해하고 유의미한 질문을 도출할 수 있는지 검증하기 위한 스트레스 테스트의 성격을 갖는다. 세부 연구 과정은 [그림 1]과 같은 단계로 구성된다.

2.1. 인터뷰 주요 질문 설계

인터뷰 주제를 기반으로 사용자 경험(UX)의 3요소인 사용자, 맥락, 과업을 고려하여 배경 지식, 개인의 가치에 따른 근무 유형, 근무표 제작 고려 요소 등 총 6가지 영역으로 나누어 주요 질문을 설계하였다. 본 연구에서는 모든 영역을 다루기보다 인터뷰의 깊이와 답변의 다양성을 확보하기 위해 AI의 맥락 파악 능력을 검증하기 가장 적합한 핵심 영역을 선별적으로 채택하였다.

표 1. 인터뷰 주요 문질 영역 구분

번호	질문 영역명	실험 채택 여부
1	배경 지식	0
2	개인의 가치에 따른 근무 유형	0
3	근무표 제작 고려 요소의 종류	0
4	고려 요소의 반영 원칙	X
5	근무표 제작 순서	X
6	근무표 수정 및 변동	X

(2) **홀리스틱성:** 질문이 단편적 사실 확인을 넘어, 사용자-맥락-과업을 입체적이고 포괄적인 관점에서 다루고 있는지 평가한다.

(3) **본질적 니즈 발견 가능성:** 질문이 사용자의 잠재적 니즈나 숨겨진 동기를 끌어낼 가능성이 높은지 평가한다.

정량적 평가 이후에는 심층 인터뷰를 진행하여, AI가 생성한 질문의 실무적 효용성과 한계에 대한 전문가 집단의 구체적인 의견을 수집함으로써 정량적 결과를 보완하였다. 이러한 평가 과정을 통해 생성형 AI가 사용자 인터뷰 질문 개발 및 최적화 과정에서 실제적인 보조 도구로 활용될 수 있는지 다각도로 검증하고자 하였다.

III. 실험 방법

제2장에서 논의된 연구 설계와 페르소나 설정 의도를 바탕으로, 본 장에서는 실제 AI 에이전트의 구현 내용과 데이터 수집 절차를 상세히 기술한다. [그림 2]는 본 실험의 전체적인 구조와 데이터 흐름을 3단계로 도식화한 것이다. 실험은 인간 연구자의 주요 질문 설계(1단계), Agent A를 통한 가상 인터뷰 진행(2단계), 그리고 Agent B를 통한 파생 질문 생성(3단계)의 순서로 진행되었으며, 각 단계별 구체적인 수행 방법과 프롬프트 구성은 다음과 같다.

3.1. 실험 환경 및 도구

본 실험은 2026년 1월, Google의 대규모 언어 모델(LLM)인 Gemini 3 Pro를 기반으로 수행되었다. 특히 실험의 정교함을 높이기 위해 Gemini Advanced 환경의 맞춤형 에이전트 제작 도구인 'Gems'를 활용하였다. 이를 통해 Agent A(인터뷰이)와 Agent B(인터뷰어) 각각에 고유한 시스템 프롬프트를 사전 이식함으로써, 대화가 길어지더라도 페르소나와 사고 과정이 일관되게 유지되도록 환경을 구축하였다.

3.2. 프롬프트 엔지니어링 및 에이전트 설정

제2장에서 설계한 도메인 복잡성과 추론 로직을 실제 모델에 적용하기 위해, 각 에이전트에는 다음과 같은 구체적인 지시어와 제약 조건이 입력되었다.

(1) Agent A 시스템 프롬프트 구성

Agent A에게는 대학병원 마취회복실(PACU)이라는 특수한 맥락을 구현하기 위해, 단순한 역할 부여를 넘어 [표 3]과 같은 강제적 제약 조건과 현장 은어 리스트를 입력하였다. 이를 통해 AI가 일반적인 간호 지식이 아닌 해당 조직의 특수성이 반영된 답변을 생성하도록 유도하였다.

표 3. Agent A 시스템 프롬프트 주요 설정

구분	프롬프트 입력 내용 (구체적 지시사항)
Role Definition	25년 차 대학병원 마취회복실(PACU) 수간호사 '박 수샘'
Hard Constraints	<ul style="list-style-type: none"> 프리셉터 매칭: 신규(프리셉터)와 사수(프리셉터) 동일 듀티 배치 온콜(On-call): 야간/주말 응급 대기 인원 2명 필수 근무 제한: 임신부 야간 열외, N-off-D 및 풍당풍당 근무 금지
Knowledge Base	<ul style="list-style-type: none"> 필수 사용 은어: 듀티, 워터드, 킵, 빵꾸, 액팅, 차지, 더블 등 상황 변수: 전공의 파업, 병동 폐쇄 등 돌발 상황 반영 허용
Output Guide	<ul style="list-style-type: none"> 딱딱한 문체 지양, 자연스러운 구어체 사용 단순 불평이 아닌 맥락과 동기가 포함된 200-400자 내외 스토리텔링

(2) Agent B 시스템 프롬프트 구성

Agent B에게는 입력된 답변을 분석하고 최적의 질문을 도출하는 전문가의 인지 과정을 모사하기 위해, CoT(Chain of Thought) 기법이 적용된 프롬프트를 입력하였다. [표 4]와 같이 사고의 단계를 명시적으로 구조화하여 무작위 생성을 차단하고 논리적 완결성을 갖춘 질문이 산출되도록 제어하였다.

표 4. Agent B 시스템 프롬프트 논리 구조

단계	프롬프트 입력 내용 (구체적 지시사항)
Role Definition	10년 차 Senior UX Researcher
Step 1 (Deep Analysis)	답변을 5가지 관점(Workaround, Pain Point, Ambiguity, Context, Gap)에서 스캔하여 핵심 단서 포착
Step 2 (Strategy Mapping)	<ul style="list-style-type: none"> Workaround 발견 시: Process Tracing(과정 추적) 전략 적용 Ambiguity 발견 시: Critical Incident(결정적 사건) 전략 적용 Context 발견 시: Perspective Taking(관점 전환) 전략 적용 등
Step 3 (Generation)	<ul style="list-style-type: none"> 선정된 전략에 기반하여 단 하나의 최적 파생 질문 생성 답변 내용을 인용(Relevance)하고 맥락을 확장(Holistic)할 것

3.3. 데이터 수집 절차

실험 데이터는 [그림 2]에 제시된 프로세스에 따라 다음과 같이 순차적으로 수집되었다.

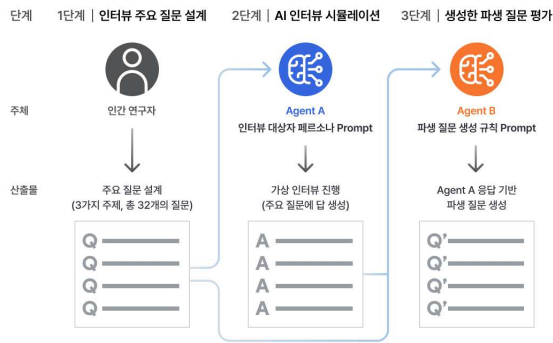


그림 2. 실험 구조도

가. 주요 질문 설계

연구자는 선행 연구 분석을 통해 도출한 '배경 지식', '개인의 가치관', '근무표 제작 고려 요소'의 3가지 영역에서 총 32개의 주요 질문(Q)을 설계하였다. 이는 실험의 독립 변인으로서 Agent A에게 입력되는 최초 자극 역할을 한다.

나. 가상 인터뷰 진행

설계된 32개의 주요 질문을 Agent A에게 순차적으로 입력하였다. Agent A는 입력된 질문에 대해 설정된 페르소나와 제약 조건을 바탕으로 구체적인 에피소드가 포함된 32개의 답변(A)을 생성하였다. 이 과정에서 연구자는 Agent A가 설정된 제약 조건(은어 사용, 근무 규칙 반영 등)을 준수하는지 모니터링하였다.

다. 파생 질문 생성

본 실험의 핵심 산출물인 파생 질문을 얻기 위해 주요 질문(Q)과 그에 대응하는 Agent A의 답변(A)을 하나의 텍스트 쌍으로 결합하여 Agent B에게 입력하였다. Agent B는 입력된 텍스트를 CoT 프로세스로 분석한 후, 해당 맥락에서 가장 심층적인 인사이트를 도출할 수 있는 단 1개의 파생 질문(Q')을 생성하였다.

기존 연구들이 다수의 질문을 생성한 후 선별

하는 방식을 취한 것과 달리 본 연구는 Agent B의 추론 능력을 활용하여 '가장 최적화된 하나의 질문'을 즉각적으로 도출하는 방식을 채택하였다. 이를 통해 최종적으로 32개의 주요 질문에 대응하는 32개의 파생 질문 데이터셋이 구축되었으며, 이는 후속 전문가 평가를 위한 기초 자료로 활용되었다.

IV. 평가 방법 및 결과

제3장을 통해 산출된 32개의 파생 질문 데이터셋에 대한 품질 검증을 수행하였다. 본 평가는 AI가 생성한 질문이 실제 UX 리서치 현장에서 전문가가 수행하는 심층 인터뷰 수준의 논리와 통찰을 갖추고 있는지 확인하는 데 주안점을 두었다. 이를 위해 전문가 집단을 통한 정량적 평가를 우선 실시하여 객관적 수치를 확보하고, 이후 동일한 전문가 집단을 대상으로 한 심층 인터뷰를 통한 정성적 평가를 병행하여 AI 에이전트의 실무적 효용성과 한계점을 다각도로 분석하였다.

4.1. 정량적 평가

가. 평가자 구성 및 선정 근거

본 연구의 평가에는 총 7명의 UX 전문가 집단이 참여하였다. 평가단은 UX/HCI 관련 전공 석사 학위 소지자 5명과 실무 경력 2년 이상의 현업 UX 리서처 2명으로 구성되었다([표 5] 참조).

본 연구에서 도메인 전문가(간호사)가 아닌 UX 전문가를 평가자로 선정 한 이유는, 본 실험

표 5. 전문가 집단 프로필

평가자	성별	구분	소속 및 직무	프로젝트 참여 횟수
평가자1	여	실무 전문가	UX 에이전시 소속 디자이너	14회
평가자2	여	실무 전문가	UX 에이전시 디자이너	10회
평가자3	여	연구원	UX 디자인 전공	14회
평가자4	남	연구원	UX 디자인 전공	13회
평가자5	여	연구원	UX 디자인 전공	14회
평가자6	여	연구원	UX 디자인 전공	12회
평가자7	남	연구원	UX 디자인 전공	10회

의 목적이 의료적 지식의 정오를 판별하는 것이 아니라 '인터뷰 질문의 구조적 적합성'과 '사용자 조사 도구로서의 효용성'을 검증하는 데 있기 때문이다. 따라서 인터뷰 방법론에 대한 체계적인 훈련을 받고, 질문의 논리적 흐름과 인사이트 도출 역량을 판별할 수 있는 UX 연구자를 선정하는 것이 타당하다고 판단하였다.

나. 평가 지표 및 절차

평가자에게는 [주요 질문] - [Agent A의 응답] - [Agent B가 생성한 파생 질문]으로 구성된 32개의 세트가 제공되었다. 평가자들은 각 세트를 읽고, 생성된 파생 질문이 인터뷰 맥락에서 얼마나 유효한지 다음 3가지 기준에 따라 리커트 5점 척도(5-point Likert Scale)로 평가하였다.

(1) 연속성

파생 질문이 인터뷰이의 직전 답변 내용과 논리적 맥락을 얼마나 유기적으로 유지하고 있는지 평가한다. 이는 AI가 답변을 잘못 해석하거나 맥락에서 이탈하지 않고, 대화의 흐름을 자연스럽게 이어가는지를 확인하는 지표이다.

- 1점(부적합): 직전 답변과 무관하거나 내용을 오인하여 질문함.

- 5점(최우수): 답변의 함의를 정확히 파악하여 새로운 관점으로 연결함.

(2) 홀리스틱성

질문이 단편적 사실 확인을 넘어, 사용자(User)-맥락(Context)-과업(Task)을 입체적이고 포괄적인 관점에서 다루고 있는지 평가한다. 이는 서비스 디자인의 핵심 가치인 '통합적 관점'이 질문 생성에 반영되었는지를 확인하는 기준이다.

- 1점(부적합): 예/아니오로 답하는 폐쇄형 질문이거나 단순 사실 확인에 그침.

- 5점(최우수): 개인의 문제를 넘어 조직, 인간 관계, 시스템 전체 생태계로 통찰을 확장함.

(3) 본질적 니즈 발견 가능성

질문이 사용자의 표면적 요구 뒤에 숨겨진 잠재적 니즈나 행동의 근본 원인을 끌어낼 가능성이 있는지 평가한다. 이는 실제 서비스 기획에 필요한 인사이트를 도출할 수 있는 실질적 가치를 측정하기 위해 본 연구에서 새롭게 도입한 지표이다.

- 1점(부적합): 이미 알고 있는 사실을 재확인하거나 답변의 깊이를 더하지 못함.

- 5점(최우수): 사용자가 무의식적으로 행동하는 이유나 숨겨진 동기를 말하게끔 유도함.

두 가지 평가 지표를 기반으로 정량적인 평가 방식으로 진행한 후 이를 기반으로 인터뷰를 통해 진행된 정성적 평가를 통해 통계적으로는 세밀하게 파악하기 어려웠던 생성형 AI의 인터뷰 질문 생성에 활용하는 것에 관한 실용적인 사용 경험 및 생각을 파악하였다.

다. 정량적 평가 결과 및 심층 분석

7명의 UX 전문가가 수행한 32개의 파생 질문 평가 데이터를 종합적으로 분석한 결과, 생성형 AI 에이전트의 수행 능력과 한계가 정량적 수치를 통해 명확히 드러났다. 본 절에서는 종합 성능, 지표 간 상관관계, 그리고 고득점 및 저득점 사례에 대한 심층 분석을 통해 AI의 실무적 효용성을 검증한다.

(1) 종합 성능 분석

3가지 평가 지표에 대한 전체 평균 점수를 비교 분석한 결과, 연속성이 5점 만점에 3.96점으로 가장 높게 나타났으며, 이어 홀리스틱성이 3.49점, 본질적 니즈 발견 가능성이 3.21점을 기록하였다.

이러한 결과는 본 연구에서 적용한 CoT(Chain of Thought) 기반의 프롬프트 엔지니어링이 AI의 논리적 맥락 유지 능력을 강화하는 데 주효했음을 시사한다. AI는 인터뷰이의 답변 내 핵심

키워드를 포착하고 대화의 흐름을 이어가는 구조적 과업에서는 준수한 성능을 보였으나, 표면적 진술 이면에 숨겨진 잠재적 동기나 욕구를 파악해야 하는 통찰 영역에서는 상대적으로 한계를 보였다. 이는 AI가 텍스트의 논리적 정합성은 담보할 수 있으나, 심층적인 질적 분석 역량은 여전히 인간 전문가의 보완이 필요함을 정량적으로 입증하는 결과라 할 수 있다.

(2) 평가 지표 간 상관관계 분석

평가 지표 간의 상관관계를 분석한 결과, 홀리스틱성과 본질적 니즈 발견 가능성 사이에 강한 양의 상관관계($r=0.86$)가 확인되었다.

이는 인터뷰 질문이 단편적인 사실 확인을 넘어 사용자, 맥락, 과업을 아우르는 거시적 관점을 견지할 때 비로소 실무적으로 유의미한 인사이트를 도출할 확률이 높아짐을 통계적으로 뒷받침한다. 즉, AI 에이전트가 단순한 대화의 유지를 넘어 가치 있는 정보의 발굴이라는 리서치 목적을 달성하기 위해서는 시스템적 사고를 유도하는 프롬프트 설계가 선행되어야 함을 시사한다.

(3) 사례 기반 심층 분석

평가 점수의 극단값(최고점 및 최저점)을 기록한 질문 사례를 텍스트 수준에서 정밀 분석함으로써 AI 질문 생성의 성공 요인과 실패 요인을 규명하였다.

먼저, 가장 높은 평가를 받은 질문들은 공통적으로 개인의 주관적 경험보다는 업무 프로세스나 시스템의 구조적 모순을 파고드는 경향을 보였다.

- **연속성 최고점(5.00점) 사례(Q2):** 근무표 작성 시 특정 간호사 조합을 배제하는 기준이 개별적인 관계에 의한 것인지, 혹은 연차 차이와 같은 일반화된 공통 기준에 의한 것인지를 구체적으로 묻는 질문이 선정되었다. 이 질문은 수간호사의 업무 고충에 공감하면서도, 자동화 서비

스의 로직 설계에 즉각적으로 활용 가능한 구체적 배치 규칙을 요구했다는 점에서 높은 평가를 받았다.

- **본질적 니즈 발견 가능성 최고점(4.43점)**

사례(Q26): 타 직군(의사)의 당직 일정 확정 시점과 간호사 근무표 작성 시점의 불일치로 인해 발생하는 역순의 업무 비효율을 지적하는 질문이 선정되었다. 이 질문은 간호 조직 내부의 문제를 넘어, 병원 전체 시스템의 프로세스 불일치라는 구조적 문제점을 예리하게 포착하였다는 점에서 탁월한 통찰력을 인정받았다.

반면, 최하위권 점수를 기록한 질문은 인터뷰의 본질적 목적인 과업 분석에서 벗어나는 경향을 보였다.

- **전 항목 최저점 사례(Q27):** 까다로운 의료진을 상대하는 업무 노하우를 묻는 과정에서 시스템적 해결책이 아닌 간호사 개인의 성격이나 처세술과 같은 개인적 속성을 묻는 질문이 선정되었다. 해당 질문은 연속성(2.29), 홀리스틱성(2.29), 본질적 니즈(2.00) 모두에서 가장 낮은 평가를 받았다. 이는 AI가 이전 답변의 비유적 표현에 매몰되어 서비스 기획에 반영하기 어려운 흥미 위주의 정보와 리서치 목적에 부합하는 핵심 정보를 구별하는 데 있어 여전히 한계가 있음을 보여준다.

(4) 평가자 간 일치도 및 신뢰성 분석

평가자 간의 점수 편차(표준편차)를 분석한 결과, 평가 기준이 구체적일수록 합의도가 높고 추상적일수록 불일치가 증가하는 경향이 확인되었다.

- **연속성(SD=1.05):** 가장 낮은 표준편차를 보여 전문가 간의 평가가 일치하는 경향을 보였다. 이는 질문의 논리적 연결성 여부가 비교적 객관적으로 판단 가능함을 의미한다.

- **본질적 니즈 발견 가능성(SD=1.34):** 가장 높은 표준편차를 기록하여 평가자 간 이견이 컸음을 나타냈다. 특정 질문이 도출할 수 있는 통

찰의 깊이는 평가자의 실무 경험과 관점에 따라 해석이 달라질 수 있는 주관적 영역이기 때문이다.

이러한 분석 결과는 AI가 생성한 질문 초안에 대해 인간 연구자가 개입하여 선별하고 정제하는 Human-in-the-loop 프로세스가 필수적으로 동반되어야 함을 강력히 시사한다. AI는 논리적 구조를 생성하는 데 탁월하지만 그 질문의 가치를 판단하고 결정하는 것은 여전히 인간 전문가의 몫임을 알 수 있다.

4.2. 정성적 평가

가. 평가자 구성 및 선정 근거

본 정성적 평가는 앞선 정량적 평가에 참여했던 총 7명의 UX 전문가 집단을 동일하게 대상으로 하였다. 이들은 32개의 질문 세트를 직접 평가하는 과정을 통해 AI 에이전트의 수행 능력과 패턴을 이미 상세히 인지하고 있는 상태이다. 따라서 새로운 피실험자를 모집하는 것보다 해당 도구를 직접 경험한 전문가들에게 심층적인 피드백을 청취하는 것이 평가의 연속성을 확보하고 수치 이면의 구체적인 맥락을 파악하는 데 효과적이라고 판단하였다. 특히 실무 경력자와 학술 연구원이 혼합된 평가단의 구성은 이론적 정합성과 현업 활용성을 균형 있게 진단할 수 있는 최적의 표본이라 할 수 있다.

나. 평가 설계 및 인터뷰 절차

정량적 평가 결과가 보여주는 수치적 경향성의 원인을 규명하고, 실무 적용 가능성을 타진하기 위해 1:1 심층 인터뷰(In-depth Interview)를 설계하였다. 인터뷰는 정량 평가 직후 진행되었으며, 답변의 흐름에 따라 유연한 탐색이 가능한 반구조화 방식을 채택하여 1인당 평균 30분간 수행되었다. 인터뷰 프로토콜은 평가자가 부여한 점수의 근거를 추적하고, AI의 효용성을 다각도로 검증하기 위해 다음의 3가지 핵심 차원으로 구성되었다. 첫째, 평가 근거의 규명이다. 평가 점

수의 고득점 및 저득점 원인을 파악하기 위해, 평가자의 판단에 결정적인 영향을 미친 특정 단어나 맥락이 무엇인지 질의하였다. 둘째, 논리적 완결성 및 품질 검증이다. AI가 생성한 질문이 실제 전문가의 화법과 비교하여 이질감이 없는지, 그리고 CoT(Chain of Thought) 프롬프트가 논리적 비약을 얼마나 효과적으로 제어했는지 확인하였다. 셋째, 실무 활용성의 타진이다. 본 연구의 궁극적 목적인 현업 도입 가능성을 확인하기 위해 실제 프로젝트에서의 활용 의향과 구체적인 협업 시나리오를 탐색하였다.

다. 정성적 평가 결과 및 심층 분석

전문가 심층 인터뷰 데이터를 종합 분석한 결과, 생성형 AI 에이전트의 특성은 구조적 완결성 및 심층 탐색 능력, 맥락적 해석의 한계, 보조 도구로서의 실무적 가치로 요약될 수 있다.

(1) 논리적 구조와 구체적인 파고들기의 탁월성:

평가자들은 본 연구의 AI 에이전트가 보여준 질문 생성 능력 중, 답변 내 특정 키워드를 포착하여 심층적인 후속 질문으로 연결하는 역량을 높게 평가했다. 대다수의 전문가는 AI가 인터뷰이의 답변에 포함된 전문 용어나 특정 행동을 정확히 인지하고, 이를 기반으로 구체적인 꼬리 질문을 생성하는 과정이 매우 정교했다고 언급했다. 특히 인터뷰이가 언급한 비공식적인 업무 처리 방식을 놓치지 않고 프로세스의 허점을 파고드는 질문은 숙련된 리서처 수준의 예리함을 보였다. 또한, CoT 프로세스의 적용으로 인해 질문 간의 논리적 연결성이 강화되었으며, 이는 초심 연구자가 범하기 쉬운 논리적 비약이나 맥락 이탈 없이 안정적인 인터뷰 구조를 설계하는 데 기여한 것으로 분석된다.

(2) 맥락적 유연성 부족과 감정 해석의 한계:

반면, 텍스트의 이면에 숨겨진 비언어적 맥락이나 감정선을 해석하는 데에는 기술적 한계가

뚜렷하게 나타났다. 평가자들은 AI가 인터뷰의 비유적 표현을 문자 그대로 해석하여 맥락에 부합하지 않는 질문을 생성하거나, 정서적 공감 이 필요한 시점에서 기능적인 사실 확인에 치중하는 경향을 지적했다. 이러한 현상은 심층 인터뷰의 핵심인 라포(Rapport) 형성을 저해하고 몰입도를 떨어뜨리는 요인으로 작용할 수 있음이 확인되었다. 아울러 생성된 질문의 문장이 지나치게 길거나 번역투의 어조를 띠어 구어체 대화 환경에서 이질감을 준다는 점 또한 개선이 필요한 부분으로 언급되었다. 이는 정량 평가에서 홀리스틱성과 본질적 니즈 발견 가능성 지표가 상대적으로 낮게 측정된 원인과 일맥상통한다.

(3) 인간 참여형(Human-in-the-loop) 협업 모델로서의 실무적 활용:

전문가들은 AI를 인간 연구자의 대체재가 아닌, 인지적 부하를 경감하고 연구의 질을 높이는 보조 도구로서 정의하였다. 구체적인 활용 방안으로는 크게 세 가지가 제시되었다. 첫째, 시각지대 보완이다. 연구자가 자신의 가설에 매몰되어 간과할 수 있는 관점을 AI가 제3의 시각에서 제시함으로써, 보다 객관적이고 다각적인 리서치 설계를 가능하게 한다는 것이다. 둘째, 초안 작성의 효율화이다. 인터뷰 준비 단계에서 AI가 생성한 방대한 질문 리스트를 인간 연구자가 선별하고 수정하는 방식을 통해, 리서치 설계에 소요되는 시간과 노력을 획기적으로 절감할 수 있다는 의견이다. 셋째, 교육 및 훈련 도구로서의 활용이다. AI의 탁월한 논리 구조 생성 능력은 경험이 부족한 주니어 연구자들이 인터뷰 흐름을 학습하고 논리적인 질문 구성 방식을 익히는 데 유용한 트레이닝 도구가 될 수 있다는 가능성이 제기되었다. 결론적으로, 본 연구의 AI 에이전트는 단독 수행보다는 인간 전문가의 검수와 개입이 동반되는 'Human-in-the-loop' 프로세스 하에서 최적의 성능을 발휘하며 리서치 업무의 효율성과 품질을 동시에 제고하는 강력한 협업 파트

너가 될 수 있음을 시사한다.

V. 결론 및 제언

5.1. 연구 요약 및 결론

본 연구는 사용자 경험 리서치의 초기 단계인 심층 인터뷰(In-depth Interview) 과정에서 생성형 AI의 활용 가능성을 탐구하기 위해, 최신 LLM인 Gemini 3를 기반으로 과생 질문 자동 생성 모델을 설계하고 그 유효성을 검증하였다. 특히 도메인 지식의 복잡성이 높은 '간호사 근무표 제작'을 실험 주제로 선정하여, AI가 전문적인 맥락과 암묵지를 얼마나 효과적으로 파악하고 심층 질문으로 연결할 수 있는지에 대한 스트레스 테스트를 수행하였다.

연구 결과, CoT(Chain of Thought) 기법이 적용된 AI 에이전트는 인터뷰의 논리적 구조를 수립하고, 답변 내의 특정 행동이나 전문 용어를 포착하여 구체적인 사실 관계를 파고드는 능력에서 탁월한 성과를 보였다. 정성적 평가 결과, 전문가들은 AI가 연구자의 편향을 배제하고 놓치기 쉬운 시각지대를 보완하는 데 유용하다는 점에 동의하였다.

그러나 동시에 텍스트 이면에 숨겨진 비언어적 맥락, 감정선, 그리고 조직 내 미묘한 정치적 역학 관계를 통합적으로 해석하는 홀리스틱 역량에서는 여전히 인간 전문가에 미치지 못하는 한계가 드러났다. 이는 생성형 AI가 인터뷰의 구조와 논리를 담당할 수는 있으나, 공감과 직관이 요구되는 영역에서는 인간 연구자의 개입이 필수적임을 시사한다.

5.2. 연구의 시사점

본 연구가 UX 디자인 및 리서치 분야에 제공하는 학술적·실무적 시사점은 다음과 같다.

첫째, 질문 생성을 위한 방법론적 프레임워크를 제시하였다. 기존 연구들이 AI를 페르소나(답변자)로 활용하는 데 그쳤다면, 본 연구는 AI에

게 UX 전문가의 사고 과정(분석-전략 매핑-질문 생성)을 학습시킴으로써 능동적인 인터뷰어로서의 가능성을 입증하였다. 이는 향후 AI 기반 리서치 도구 개발에 있어 구체적인 알고리즘 설계의 기초 자료가 될 것이다.

둘째, 'Human-in-the-loop(인간 참여형)' 기반의 협업 프로세스를 정립하였다. 연구 결과는 AI를 인간의 대체재가 아닌, 증강 도구로 정의한다. AI는 인터뷰 준비 단계에서 방대한 질문 초안을 생성하여 발산을 돕고, 인간 연구자는 이를 문맥에 맞게 선별하고 다듬어 수렴시키는 협업 모델이 실무 효율성을 극대화하는 현실적인 방안임을 확인하였다.

셋째, 주니어 연구자를 위한 교육 도구로서의 가치를 발굴하였다. 경험이 부족한 신진 연구자들은 인터뷰 설계 시 논리적 흐름을 놓치거나 구체적인 과정을 파고들지 못하는 경향이 있다. 이때 AI가 보여주는 체계적인 질문 구조와 과정 추적 전략은, 이들이 올바른 질문법을 익히고 논리적 오류를 방지하도록 돕는 효과적인 교육용 시뮬레이터 역할을 수행할 수 있다.

5.3. 연구의 한계 및 향후 연구 제언

본 연구는 생성형 AI의 인터뷰 진행 능력을 다각도로 검증하였으나, 다음과 같은 한계점을 가지며 이를 바탕으로 후속 연구를 제언한다.

첫째, 도메인의 확장성 검증이 필요하다.본 연구는 복잡도가 높은 의료/간호 도메인에 한정하여 실험을 진행하였다. 향후 연구에서는 금융, 커머스, 엔터테인먼트 등 상이한 특성을 가진 도메인으로 실험 범위를 확장하여 AI 질문 생성 능력의 범용성을 검증할 필요가 있다.

둘째, 소규모 표본으로 인한 정량적 지표의 통계적 일반화 한계이다. 본 연구의 정량적 평가에 참여한 UX 전문가 표본의 크기(N=7)가 상대적으로 작아, 평가 지표 간 도출된 높은 상관관계수($r=0.86$)를 모집단 전체의 특성으로 확정적으로 일반화하기에는 무리가 따른다. 이러한 제한점

을 고려하여, 본 연구에서의 정량적 수치는 AI의 질문 생성 패턴과 경향성을 파악하기 위한 보조적 지표로 제한하여 해석하였다. 대신, 핵심적인 학술적 통찰은 동일한 전문가 집단을 대상으로 한 정성적 심층 인터뷰를 통해 상호 보완적으로 도출하였다. 향후 연구에서는 평가 표본의 크기를 확대한 대규모 정량 조사를 통해 통계적 유의성을 보다 엄밀하게 검증할 필요가 있다.

셋째, 실제 인간 대상의 검증이 요구된다.본 실험은 변수 통제를 위해 AI 페르소나를 대상으로 시뮬레이션을 진행하였다. 후속 연구에서는 실제 사용자를 대상으로 AI가 생성한 질문을 적용해 보고, 실제 인터뷰 상황에서의 라포(Rapport) 형성 여부와 답변의 질적 변화를 측정하는 실증 연구가 이어져야 할 것이다.

넷째, 멀티모달(Multi-modal) 데이터의 활용이다. 현재의 텍스트 기반 분석을 넘어, 음성의 톤이나 표정 등 비언어적 데이터를 함께 분석할 수 있는 멀티모달 AI 모델을 도입한다면, 본 연구에서 한계로 지적된 감정 맥락 파악 능력을 획기적으로 개선할 수 있을 것으로 기대한다.

REFERENCES

- [1] 김동윤, 연명흠, “디자인 프로세스 내 생성형 AI 도구 및 생성형 AI 기반 디자인 툴의 활용가능성 탐색”, *한국디자인학회 학술발표대회 논문집*, pp. 212-213, 한국, 2023년 10월.
- [2] 윤하린, 이해인, 백병수, 정은빈, 진수진, “생성형 AI 도구를 사용하는 UX 실무자들의 사용자 부담 측정: 연차별 실무자들의 사용 경험 비교”, *한국 HCI학회 학술대회*, pp. 1140-1145, 한국, 2024년 1월.
- [3] 정은희, 최정민, “AI 디자인 툴의 디자인 업무 지원 방향성 제안”, *Archives of Design Research*, 제35권, 제4호, pp. 269-282, 2022년 11월.
- [4] 최정인, 이태일, “생성형 AI를 활용한 시각화 서비스 UI/UX 디자인에 있어서 사용자 경험 특성 분석”, *한국HCI학회 학술대회*, pp. 284-290, 한국, 2024년 1월.
- [5] 김유근, 최혜윤, 오아름, 전하진, 김경홍, “UX 실무에서 생성형 AI의 활용 가능성 탐구 : 대형 실내 시설용 로봇 컨셉 시나리오 발굴 사례를 중심으로”, *Design Works*, 제7권, 제1호, pp. 29-44, 2024년 3월.
- [6] 이현정, 홍세영, 박찬솔, 김영철, “인공지능 할루시네이션에 대응하는 질의개선”, *ICT플랫폼학회 하계학술발표대회논문집*, pp. 31-35, 한국, 2023년 12월.
- [7] AI가 창조할 수 있는 비밀: 생성형 AI 이해하기 (2023).
<https://blog-ko.superb-ai.com/secrets-ai-can-generate-understanding-generative-ai/> (accessed 12월 1일, 2025년).
- [8] H. Taherdoost, “How to Conduct an Effective Interview; A Guide to Interview Design in Research Study,” *International Journal of Academic Research in Management (IJARM)*, vol. 11, no. 1, pp. 39-51, May 2022.
- [9] J. Park, S. Han, H. Kim, and Y. Cho, “Developing elements of user experience for mobile phones and services: survey, interview, and observation approaches,” *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 23, no. 4, pp. 271-293, July 2013.
- [10] H. Pitariu, D. Andrei, and A. Guran, “Social research methods used in moving the traditional usability approach towards a user-centered design approach,” *Int. J. Inf. Technol. Web Eng.*, vol. 4, no. 4, pp. 16-28, Oct. 2009.
- [11] Y. Jung and A. Anttila, “How to look beyond what users say that they want,” *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, pp. 2099 - 2104, San Jose, USA, Apr. 2007.
- [12] R. Roberts, “Qualitative interview questions: guidance for novice researchers,” *The Qualitative Report*, vol. 25, no. 9, pp. 3185-3203, Sep. 2020.
- [13] R. Chenail, “Interviewing the investigator: strategies for addressing instrumentation and researcher bias concerns in qualitative research,” *The Qualitative Report*, vol. 16, no. 1, pp. 255-262, Jan. 2011.
- [14] H. Cairns-Lee and J. Lawley, “Enhancing researcher reflexivity about the influence of leading questions in interviews,” *The Journal of Applied Behavioral Science*, vol. 58, no. 1, pp. 164-188, Aug. 2021.
- [15] E. Ambele, “Developing interview guide in qualitative research: problems and solutions from a needs analysis doctoral study,” *rEFLECTIONS*, vol. 32, no. 1, pp. 196-216, Jan. 2025.
- [16] T. Mattelmäki, E. Brandt, and K. Vaajakallio, “On designing open-ended interpretations for collaborative design exploration,” *CoDesign*, vol. 7, no. 2, pp. 79-93, June 2011.
- [17] M. Grasso, M. Daquino, and G. Renda, “From ontology design to user experience: a methodology to design interfaces for information seeking purposes,” *Umanistica Digitale*, vol. 8, no. 18, pp. 53-85, Dec 2024.
- [18] S. Datta, P. Mallick, S. Patil, I. Bhattacharya, and G. K. Palshikar, “Generating an optimal interview question plan using a knowledge graph and integer linear programming,” *Proc. of the 2021 Conf. of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pp. 1995 - 2005, Online, June 2021.
- [19] M. Su, C. Wu, K. Huang, Q. Hong, and H. Huang, “Follow-up question generation using pattern-based seq2seq with a small corpus for interview coaching,” *Proc. Interspeech 2018*, pp. 1006-1010, Hyderabad, India, Sep. 2018.
- [20] SB, P., Agnihotri, M., & Jayagopi, D. (2020). Automatic follow-up question generation for asynchronous interviews. <https://aclanthology.org/2020.intellang-1.2.pdf> (Dec 1, 2025)
- [21] P. SB, M. Agnihotri, and D. Jayagopi, “Automatic follow-up question generation for asynchronous interviews,” *Proc. of the 1st Workshop on Intelligent Interviewing and Recruiting (InteLLang)*, pp. 10-20, Santiago de Compostela, Spain. May 2020.

저 자 소 개

**박유진(준회원)**

2024년 한동대학교 콘텐츠융합디자인
학부 학사 졸업.

2026년 한동대학교 문화미디어디자인
학과 석사 졸업.

<주관심분야 : AI디자인, 디자인 전략,
서비스디자인, UX디자인>

**이중섭(정회원)**

2001년 한동대학교 콘텐츠융합디자인
학부 학사 졸업..

2006년 한동대학교 문화미디어디자인
학과 석사 졸업.

2014년 아이오와 주립대학교 그래픽
디자인 MFA

<주관심분야 : 서비스디자인, 그래픽
디자인, UX디자인>