

‘전문 지능 보조 서비스(IEA)’와 작업 지향적 규칙 기반 챗봇의 서비스 디자인 중심 비교분석 연구

김희석^{1*} · 설상훈²¹성균관대학교 서비스융합디자인협동과정 석사과정²성균관대학교 서비스융합디자인협동과정 교수

Service Design-Focused Comparative Analysis of Intelligent Expert Assistant (IEA) and Task-Oriented Rule-Based Chatbot

Huey Kim^{1*} · Sanghun Sul²¹Master's Course, Department of Service Design, Sungkyunkwan University, Suwon 16419, Korea²Professor, Department of Service Design, Sungkyunkwan University, Suwon 16419, Korea

[요약]

팬데믹 이후 챗봇의 사용이 증가하였지만, OpenAI의 ChatGPT를 포함 현대 챗봇 모델들의 이해력과 유연성 부족 문제가 드러나고 있다. 이에 대응해 본 연구는 초거대 멀티 모달 모델(LMM: Large Multimodal Models)과 벡터DB 기술을 통해 고도화된 전문 지능 보조 서비스인 IEA의 6가지 핵심 요소가 전문 지능을 제공하는 서비스 디자인에 필수적임을 확인한다. 연구 참여자들의 심층 인터뷰와 프로토타이핑을 통해 수집된 데이터를 분석한 결과, 응답 적응성, 유연성, 서비스 신뢰성, 사회적 실재감, 의인화 및 개인화 이상 6가지 IEA 요소가 전문 지능 보조 서비스 최적화에 중요한 역할을 한다는 것을 92%의 연관도로 확인했다. 결론적으로, 6가지 IEA 요소들이 사용자의 만족도와 재방문 의도를 향상시켜 전문 지능 보조 서비스를 최적화한다는 가설을 검증하였다.

[Abstract]

Amidst the rapid digital transformation after the pandemic, chatbots have become increasingly integral to global industries. However, current chatbot models (e.g., OpenAI's ChatGPT) reveal constraints in understanding and flexibility, coupled with hallucination phenomena, thereby leading to a pressing need for enhanced AI assistants. This study underscored the importance of Intelligent Expert Assistant (IEA) components integrated with large multimodal models and VectorDB for advanced service design. Rigorous testing sessions led by domain experts identified six key IEA elements (response adaptability, flexibility, service reliability, social presence, anthropomorphism, and personalization). These elements were crucial in optimizing service design, showing a 92% relevance in keyword analysis, and enhancing user satisfaction and service quality. The findings of this study validated the necessity of these six elements in optimizing IEA services.

색인어 : 전문 지능 보조 서비스, 서비스 디자인, 챗봇, 초거대 멀티 모달 모델, 벡터DB**Keyword** : IEA, Service Design, Chatbot, LMM, VectorDB<http://dx.doi.org/10.9728/dcs.2024.25.6.1443>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 06 May 2024; Revised 04 June 2024

Accepted 12 June 2024

*Corresponding Author; Huey Kim

Tel: 

E-mail: huey0605@gmail.com

I. 서론

2020년 팬데믹 이후, 디지털 변환은 더욱 가속화되었고 시대의 변화에 발맞춰 기업들은 인공지능 챗봇을 채택하려는 경향이 더 강해졌다[1]. 그러나, 작업 지향적 규칙 기반 챗봇은 사용자의 복잡한 요구에 유연하게 대응하는데 한계가 있다[2]. OpenAI의 ChatGPT 같은 생성형 인공지능 모델의 출시로 인해 챗봇 기술이 비약적으로 발전하였으나, 여전히 모델이 실제 데이터나 정보에 기반하지 않은 잘못된 또는 허구의 내용을 생성하는 환각 현상으로 인해 서비스에 대한 신뢰성이 낮은 문제의 해결이 필요하다[3].

그에 따라 본 연구에서는 LMM과 벡터DB 기술의 결합을 통해 전문 지능 보조 서비스인 IEA로 고도화하여 모델의 정확도와 신뢰성을 향상해 문제점을 개선하였다. IEA와 작업 지향적 규칙 기반 챗봇을 서비스 디자인 관점에서 비교 분석하여 전문 지식을 제공하는 인공지능 서비스 디자인의 방향성 결정에 필요한 요소들의 검증에 초점을 맞췄다. 작업 지향적 규칙 기반 챗봇의 인간적 특성 5가지에 IEA에 필요한 특성 3가지를 추가하여 제작한 24장의 카드덱을 이용하여 연구 참여자들을 대상으로 카드 소팅을 하여 심층 인터뷰를 진행하였고, 연구 참여자들이 실험에 참여하는 프로토타이핑 방법도 활용하였다. 전문가들이 심층 인터뷰와 실험 경험을 통해 IEA 모델의 고도화된 성능을 확인하여 궁극적으로 6가지 IEA 요소들이 전문 지능 보조 서비스를 최적화시킨다는 연구 가설을 검증하는 것이 목표이다.

II. 이론적 배경

2-1 챗봇의 개요

1) 챗봇의 정의

Chatterbot의 줄임말인 챗봇은 인간과 자연스럽게 대화하는 인공지능 소프트웨어로, 자연어 처리 기술을 사용하여 질문이나 요청을 이해하고 적절한 응답을 제공한다[4]. B2C와 B2B 환경에서 비용을 거의 들이지 않고, 시간과 위치에 제한 없이 고객 서비스를 제공하는 효율적인 도구로 활용된다. 인공지능 챗봇은 매일의 일상에서 다양한 업무를 자동화하여 기업의 고객 지원, 영업, 홍보 활동 등 모든 분야에 걸쳐 지원하고 있다.

글로벌 IT 기업의 경우, Apple의 Siri를 시작으로 챗봇 시장에 차례대로 뛰어들며, Google Now, Cortana, Alexa 등이 출시되어 음성인식과 인공지능 기술 기반으로 사용자의 요청에 대한 더 복잡하고 정교한 반응을 제공하게 되었다. ChatGPT 모델의 등장 이전에도 챗봇은 고객 서비스와 업무 효율성을 높이는 혁신적인 도구로 기대를 모았으나, 기술적 한계로 인한 성능 부족 때문에 많은 사용자가 실망감을 경험하였다.

2) ChatGPT의 단계별 학습

2022년 11월, OpenAI가 출시한 ChatGPT는 포춘(Fortune) 선정 500대 기업 중 80% 이상에서 채택될 정도로 큰 인기를 얻으며 새로운 챗봇 시대를 열었다. ChatGPT는 GPT(Generative Pre-trained Transformers)라는 사전 학습된 생성형 트랜스포머 아키텍처를 기반으로 하며, 대규모 텍스트 데이터 학습을 통해 발전하므로 양질의 데이터를 많이 학습할수록 성능이 향상되고 있다.

GPT 기반의 챗봇은 인간의 언어와 대화 맥락을 이해하고, 머신러닝과 자연어 처리 기술을 사용하여 사용자의 요구에 맞는 응답을 생성하는데 OpenAI의 경우, 아래 3단계와 같은 강화 학습 기반의 미세조정을 통해 GPT-3.5의 답변 정확도를 향상시켰다[5].

- 1단계: 학습 데이터 준비 및 지도 학습
대량의 질문과 답변 데이터를 활용해 초기 학습을 시행, 잠재적 사용자의 요구에 맞는 출력 생성에 초점을 맞춤. 이 단계는 선생님이 학생을 위한 문제집을 준비하는 과정에 비유됨.
- 2단계: 보상 모델 훈련
다양한 데이터 세트를 사용하여 보상 모델을 훈련 시키고, 여러 결괏값의 순위를 매김으로써 보상 모델을 개선함. 이 단계는 그룹 스터디에서 선생님이 학생들의 답변에 피드백을 제공하는 과정에 비유됨.
- 3단계: 강화 학습 알고리즘 적용
PPO(Proximal Policy Optimization) 알고리즘을 사용해 보상 모델의 정책을 최적화함. 이 단계는 논술 학원에서 글쓰기 훈련을 받고 피드백을 통해 작문 능력을 향상시키는 과정에 비유됨.

GPT-4 버전의 경우, 멀티 모달 기능을 통해 이미지를 인식하고 처리할 수 있으며, 이미지에 대한 텍스트 정보를 생성하는 능력이 추가되었다. 또한, 언어 이해와 처리 능력이 향상되어 영어에서 25,000개의 단어 처리와 장기 기억력에서는 64,000개의 단어 기억이 가능하다. 이러한 개선으로 GPT-4는 더 정교한 언어 이해와 맥락적 대응이 가능하며, 사실과 허구를 구분하는 능력에서도 개선이 이루어졌다. 최신 버전인 GPT-4O는 이미지와 동영상까지 분석할 수 있는 LMM 모델로, 인간의 지성에 가까운 다양한 언어 처리와 인식 기능을 제공한다[6].

물론, GPT 모델들에도 문제점이 존재한다. 대표적인 문제점은 인공지능, 특히 자연어 처리 모델에서 발생하는 문제로 모델이 실제 데이터나 정보에 기반하지 않은 잘못된 또는 허구의 내용을 생성하는 환각 현상이 존재하는데 이에 대한 완벽한 해결책이 필요하다.

3) 작업 지향적 규칙 기반 챗봇의 인간적 특성

챗봇은 크게 GPT 기술 도입 전과 후로 나누어 볼 수 있는데, GPT 기술 도입 전 모델인 작업 지향적 규칙 기반 챗봇의 인간

적인 특성부터 살펴보면 크게 아래 5가지로 구분할 수 있다.

첫째, 사회적 실재감(Social Presence)은 개인이 의사소통 과정에서 실제로 상대와 대면하는 것 같은 느낌이 드는 것을 의미하며, 이는 개인이 상대방과의 심리적 연결을 감지하는 척도로 이해된다[7]. 신체적 접촉 없이 다른 인간과 어떻게 연결될 수 있는지를 설명하며, 사용자가 매개적 인터페이스를 통해 타인의 존재를 인식하는 수준을 나타낸다[8],[9]. 인간과 비슷한 외형의 사회적 단서가 사회적 실재감을 증가시키며, 시각적 인지가 글로만 이루어진 인지보다 이를 강화한다고 입증되었다[10]. 인식 과정은 대부분 무의식적으로 이루어지며, 사용자들은 기계와 상호작용을 인간과 상호작용처럼 인식하고 반응하는 것이다[11].

둘째, 의인화(Anthromorphization)는 제품이나 시스템에 인간과 유사한 특성을 부여해 인간처럼 인식하고 반응하도록 만드는 개념이다[12]. 마케팅과 커뮤니케이션에서 사용되어 거부감을 줄이고 친숙함을 높이며, 제품에 대한 호의적인 태도와 구매 의도를 증가시킨다[13]. 이는 대상에 인간과 같은 성격, 의도, 감정을 부여함으로써 소비자 태도에 긍정적인 영향을 미친다[14].

셋째, 개인화(Personalization)는 정보나 서비스를 사용자의 특성이나 요구에 맞게 조정하여 제공하는 개념이다. 이는 고객에게 맞춤형 서비스를 제공하는 중요한 마케팅 전략으로, 개인화된 추천 서비스가 대표적인 예이다[15]. 개인화의 대표적인 방법으로 개인화 추천 서비스가 있는데, 예를 들어 챗봇의 성별 선택이나 고객 이름 사용 및 맞춤 답변 제공 등이 있으며, 이를 통해 고객에게 더 개인적이고 차별화된 경험을 제공한다[16].

넷째, 상호작용성(Interactivity)은 챗봇이 인간과 유사하게 반응하는 능력을 의미하며, 사용자 맞춤형 서비스 제공과 사용자 참여의 핵심적인 역할을 한다. 이는 챗봇이 민감한 주제에 대해 인간처럼 반응하는 정도를 나타내는 중요한 척도로 사용된다[17]. 챗봇의 대화 이해와 반응에 초점을 맞춘 이해성과 달리 상호작용성은 사용자의 통제력에 중점을 둔다. 이 두 요소는 사용자 경험을 결정하는 중요한 요인이다.

마지막으로 공감성(Empathy)은 기업이 개별 고객의 필요와 고민을 인식하고 이에 대응하는 능력이다. 챗봇이 사용자의 감정 상태와 필요를 적절히 인식하고 반응하는 능력으로, 이는 서비스 품질을 결정하는 중요한 요소로 작용한다[18].

2-2 인공지능의 개요

1) 인공지능의 정의

인공지능은 인간의 인지, 추론, 판단 능력을 모방해 기계가 인간처럼 작업할 수 있게 하는 기술이다. 많은 데이터를 효과적으로 처리할 수 있기에 다양한 학문을 바탕으로 발전한 인공지능은 데이터 및 자연어 처리 등을 수행하며, 자율주행 자동차, 의료 진단 등 여러 분야에서 활용된다.

그러나, 아직 인공지능은 여전히 더 많은 데이터를 학습해

야 하며, 한 번의 경험으로는 학습을 충분히 하기 어렵다. 연구자들은 인공지능 발전을 위해 인간 뇌의 작동 원리를 더 잘 이해할 필요가 있다고 주장한다. 현재 인공지능의 학습 방식이 완전히 이해되지 않았기 때문에, 인공지능을 ‘발명’했다 보다는 ‘발견’했다고 보는 것이 적절하다[19].

2) 주요 인공지능 LLM 모델과 산업

전 세계적으로 글로벌 IT 기업들과 연구기관들이 주도하여 인공지능 모델을 개발하고 있다. ChatGPT 출시 이후 다양한 회사들이 전략적 파트너십을 통해 시장 선점을 위한 움직임을 보이고 있다. 그중 대표적인 챗봇 서비스는 아래 표 1과 같이 크게 3가지 서비스로 볼 수 있다.

표 1. 주요 생성형 인공지능 서비스

Table 1. Major generative AI-based chatbot services

	Anthropic Claude 3	OpenAI GPT-4	Google Gemini
Multimodal processing	Text, image	Voice, text, image	Best in image
Understanding	Human level	Covers various industries	Less than Claude 3
Undergraduate level test	86.8%	86.4%	83.7%
Graduate level test	50.3%	35.7%	N/A
Elementary level test	95.0%	92.0%	94.4%
IQ	101	85	77.5

OpenAI의 경우, 마이크로소프트가 두 번의 대규모 투자를 집행하여 현재 최대 주주로 자리를 굳히면서, 마이크로소프트의 애저 클라우드에 OpenAI 서비스를 탑재하였고 역으로 OpenAI는 마이크로소프트의 애저 클라우드 인프라를 이용하고 있다.

2023년 5월, 구글과 SK텔레콤 등의 투자를 받은 엔트로픽은 OpenAI에서 퇴사한 개발자들이 설립한 스타트업으로, 2024년 3월 자사 인공지능 챗봇인 클로드3가 허깅페이스 리더보드에서 GPT-4를 이기고 1위를 차지했으며, 미국에서 최초로 IQ 100을 넘는 수치를 기록했다[20]. 엔트로픽은 다국어 지원 LLM을 개발하여 글로벌 시장에 진출할 계획이며, SK텔레콤과 협력해 국내에도 서비스를 제공할 예정이다[21].

구글 제미니는 구글이 개발 중인 인공지능 챗봇으로, 다국어 지원과 고도의 자연어 이해를 특징으로 한다. 사용자 맞춤형 대화 생성을 가능하게 하며, 구글의 다양한 서비스와 통합하여 사용자에게 개선된 경험을 제공한다. 제미니는 OpenAI의 ChatGPT에 대응하기 위해 고안되었으며, 구글의 강력한 데이터 분석 및 머신러닝 기술을 기반으로 발전된 인공지능 서비스를 목표로 하고 있다.

2-3 LMM과 벡터DB 기술 기반의 연구

그림 1과 같이 작업 지향적 규칙 기반 챗봇에서 고도화된 전문 지능 보조 서비스인 IEA를 이해하기 위해서는 인공지능의 최신 기술인 LMM과 벡터DB에 대한 이해가 필요하다. 먼저 벡터DB 기술은 벡터 임베딩을 활용하여 이미지, 텍스트, 센서 데이터 등의 구조화되지 않은 데이터를 고차원 벡터 형태로 저장하고 검색한다. 이 데이터베이스는 각 차원이 데이터의 특정 속성을 나타내며, 벡터 임베딩의 인덱싱, 거리 측정, 유사성 검색 기능을 통해 비정형 및 반정형 데이터 관리에 최적화되어 있다. 벡터DB는 전통적인 표 형식의 데이터 저장 방식과 달리, 유사성을 기반으로 쿼리 결과를 반환하여 더 빠르고 정확한 정보 처리가 가능하다[22]. 쉬운 예를 들어 설명하자면, 남자와 여자의 상관관계와 왕과 여왕의 상관관계를 인식하고 수치화하여 둘 사이의 유사성을 빠르게 인식하고 처리하는 방식이다.

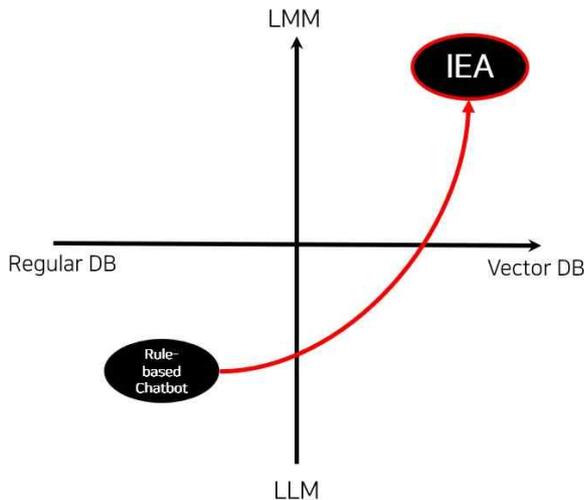


그림 1. IEA와 규칙 기반 챗봇의 차이
Fig. 1. Comparison between IEA and rule-based chatbot

다음 기술인 LMM은 대규모 언어 데이터를 학습하여 텍스트 기반 질의응답을 처리하는 초거대 언어 모델(LLM: Large Language Model)과 다르게 언어뿐만 아니라 이미지와 음성 등 다양한 데이터 형식을 이해하고 학습하여 정교한 결과물을 생성한다. 예를 들어, ChatGPT-4V 같은 LMM은 텍스트뿐만 아니라 이미지와 음성 데이터도 처리할 수 있어, 훨씬 다양하고 복잡한 입력에 대응한다. 이러한 기능은 OpenAI의 기술 발전에 따라 강화되었으며, 안전성과 기술적 문제 해결에 중점을 두고 있다[23].

2023년 10월, 마이크로소프트는 GPT-4V를 활용한 다양한 실험을 진행하며 이를 공개했다. 이 실험들에는 이미지 속 수학 문제 해결, 도표 이해, 스토리 게임을 통한 추리, 인물 표정 분석 등이 포함됐다. 또한, GPT-4V는 엑스레이 이미지에서 골절 부위를 지목하고, 다양한 언어 기반의 이미지 텍스트

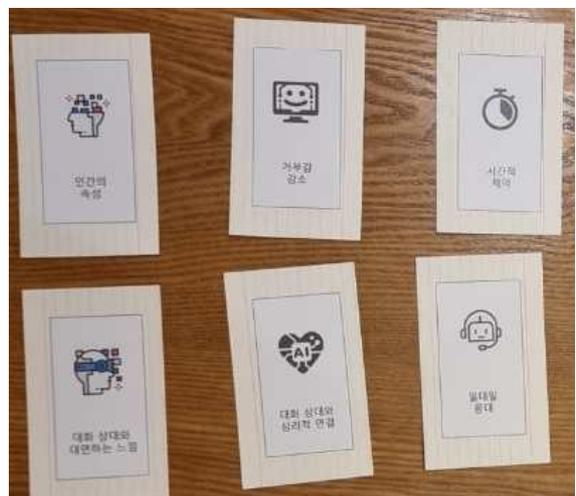
를 해석하여 설명하는 등의 작업을 수행했다. 이 외에도 김밥 만드는 순서를 정확히 나열하고, 손 글씨 이미지를 이해하는 등, 인공지능의 발전 속도와 그 범위의 확장을 실감할 수 있는 성과를 보여주었다[6].

III. IEA 연구 및 모델 정의

3-1 IEA의 연구 가설 정의

1) 사용자 심층 인터뷰를 통한 연구 가설 정의

본 연구에서는 서비스 디자인 측면에서 IEA의 6가지 핵심 요소들이 필요하다는 연구 가설을 검증하기 위해 규칙 기반 챗봇과 IEA를 비교 분석하였다. 본 논문의 2.1 챗봇의 개요의 3) 작업지향적 규칙 기반 챗봇의 인간적인 특성에서 언급된 5가지 요소에 3가지의 새로운 IEA 요소들인 응답 적응성(Response Adaptability), 유연성(Flexibility) 및 서비스 신뢰성(Service Reliability)을 포함해 총 8가지 요소를 기반으로 24개의 키워드로 구성된 카드덱을 제작하여 오픈 카드 소팅을 진행했다. 카드덱의 내용들은 빠른 연산, 신속한 처리, 민감한 반응도, 환각 현상 감소, 학습 데이터의 질, 학습 알고리즘, 시간적 제약, 다양한 데이터 학습, 자연스러운 대화, 인간의 속성, 친숙함 증가, 거부감 감소, 맞춤형 분석, 일대일 응대, 추천 서비스, 인공지능의 민감성, 사용자의 통제력, 높은 이해도, 고객의 고민 인식, 관심과 배려, 고객의 생각 이해, 대화 상대와 대면하는 느낌, 대화 상대와의 심리적 연결, 대화 상대의 시각적 인지 이상 24가지이다. 카드 소팅을 통해 취합한 80개의 키워드를 정량적으로 분석하였고, 그 후 인터뷰 대상자들이 프로토타이핑에 참여하여 IEA의 6가지 핵심 요소의 필요성에 대한 가설을 검증했다.



* Actual card deck used for card sorting method was prepared in Korean language only.

그림 2. 카드 소팅 용 카드 예시
Fig. 2. Example cards for card sorting

2) 연구 모형 및 절차

본 연구의 연구 모형은 그림 3과 같이 작업 지향적 규칙 기반 챗봇과 전문 지능 보조 서비스인 IEA의 비교 분석을 기반으로 서비스 디자인 측면에서 인공지능 서비스의 핵심 요소들을 검증했다. 선행 연구를 통해 작업 지향적 규칙 기반 챗봇의 5가지 요소를 정의하고, LMM과 벡터DB 기술 도입을 통해 고도화된 IEA의 6가지 요소로 재정의했다. 전문가들의 심층 인터뷰를 기반으로 IEA 관련 키워드를 분석하고, 카드 소팅과 프로토타이핑을 통해 모델의 서비스 사용 의도를 강화하는 요소들을 검증했다. 연구는 신뢰성 있는 정량적 평가를 도출하기 위해 전문가 의견을 종합적으로 반영하는 방식을 채택했다.

연구 절차는 그림 4와 같이 선행 연구를 시행하여 연구의 방향성으로 결정하고, 카드 소팅과 1차 델파이 조사인 심층 인터뷰를 통해 인공지능 서비스에 대한 사용자들의 요구 사항을 확인하였다. 다음으로 인터뷰 대상자들이 직접 실험에 참여하여 프로토타이핑을 진행 후 2차 델파이 조사인 심층 인터뷰를 진행하였다. 1, 2차 델파이 조사를 거쳐 취합된 키워드들 총 80개와 6가지 IEA 요소의 서비스 디자인 측면 관련도를 정량적으로 분석하여 가설 검증을 진행하였다.

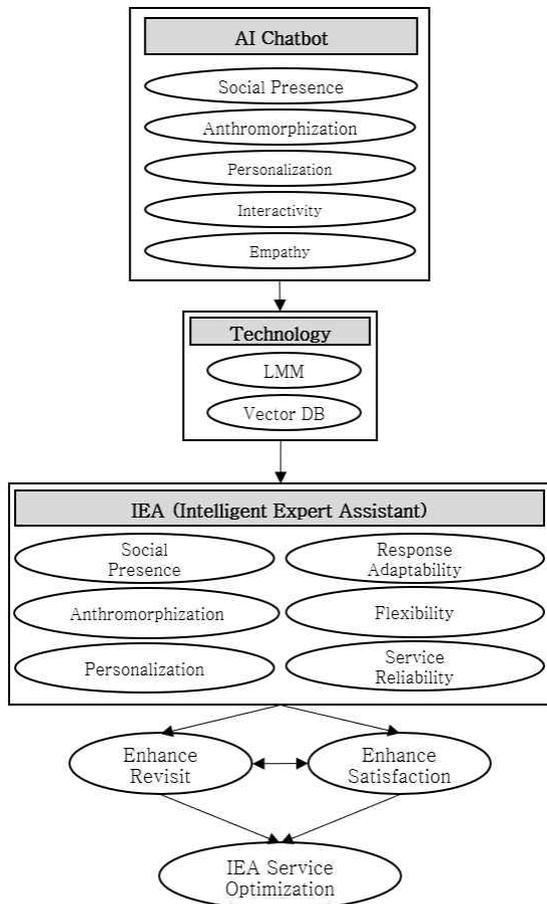


그림 3. 연구 모형
Fig. 3. Research model

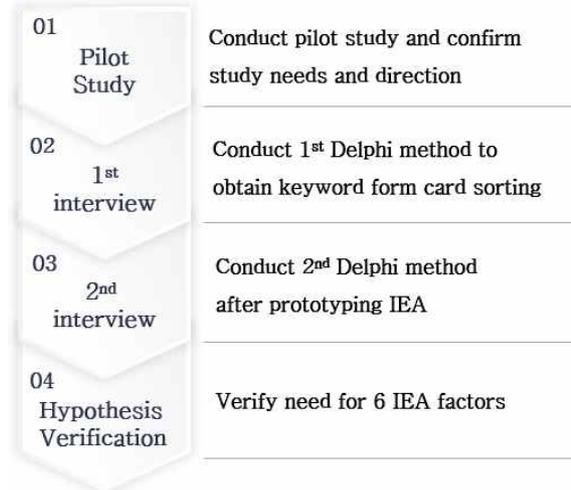


그림 4. 연구 절차
Fig. 4. Research procedure

3) 연구 참여 대상

본 연구는 IT 기술직과 서비스 기획 분야의 전문가 10명을 대상으로 IEA의 이해도와 활용 가능성을 평가하기 위한 심층 인터뷰를 진행하였다. 연구 참여자들은 모두 현재 인공지능 서비스의 사용이 활발한 업종에서 활동하며, 30대 6명과 40대 4명이며, 남성 7명과 여성 3명이고, 학력은 학사 5명, 석사 4명 및 박사 1명으로 구성되어 있다. 경력은 10년 이하 6명, 10-14년 2명에 15-20년 2명으로 분포되어 있으며, 정량적 비교 분석을 위해 IT 기술자 5명과 기획 전문가 5명을 대상으로 진행하였다.

인터뷰를 총 6시간 42분에 걸쳐 진행하여 35,411단어를 확보하였고, 결과물을 음성, 동영상 및 사진으로 기록하여 수집하였다. 심층 인터뷰는 1차와 2차 델파이 조사로 나누어 진행되었는데, 1차는 카드 소팅 완료 후 인공지능 챗봇에 대한 요구 사항을 질문하였다. 2차는 인터뷰 참여자들이 직접 IEA 서비스를 경쟁 서비스인 ChatGPT와 CHATPDF와 비교 사용해 보는 프로토타이핑을 진행한 후 IEA 모델의 경험에 대해 질문하였다.

3-2 IEA 모델의 정의

1) IEA의 특징

IEA는 아래 그림 5와 같이 LMM과 벡터DB 기술 도입을 통해 고도화되어 작업 지향적 규칙 기반 인공지능 챗봇의 5가지 요소인 개인화, 의인화, 상호작용, 공감성 및 사회적 실재감 중에서 상호작용과 공감성이 부분적으로 융합하여 응답 적응성으로 개선되었고, 유연성 및 서비스 신뢰성이 추가되어 총 6가지 요소로 재정의하였다.

응답 적응성은 챗봇이나 IEA가 사용자와의 상호작용 중에 상황에 맞게 적절하게 반응하고, 변화하는 사용자 요구나 맥락에 따라 응답을 조정할 수 있는 능력을 의미한다. 이는 시

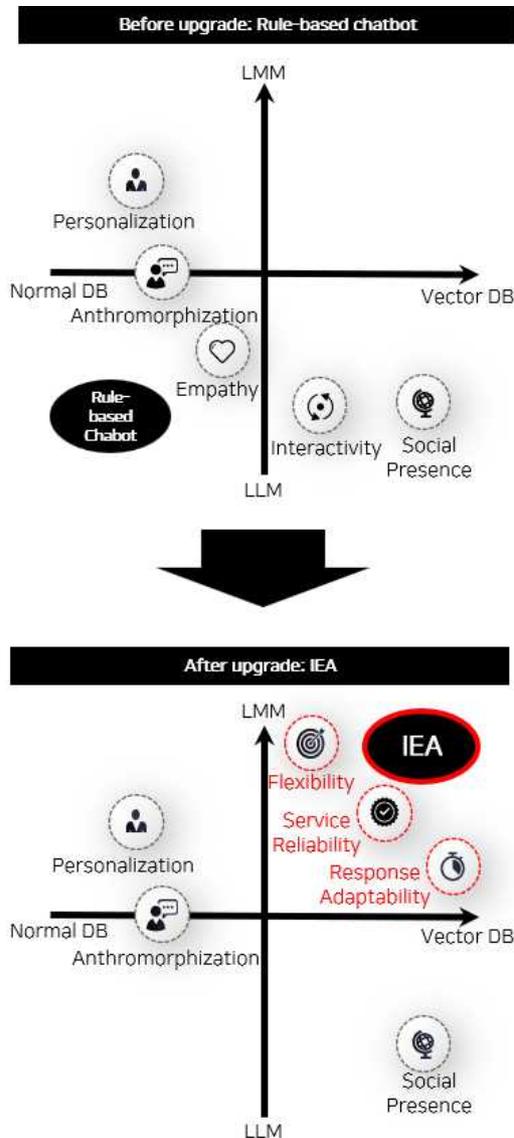


그림 5. 고도화 전 인공지능 챗봇과 고도화 후 IEA의 비교
 Fig. 5. Comparison between rule-based chatbot and IEA

시스템이 고정된 응답을 제공하는 대신, 사용자 입력에 따라 동적으로 반응하고 맞춤형 응답을 생성하는 능력을 포함한다. 응답 적응성의 측정 방법으로는 정확성 평가, 반응 시간 측정, 문맥 인식 능력 평가 및 사용자 만족도 조사가 대표적이며 본 연구에서는 위 측정 방법들을 종합적으로 활용하여 IEA의 응답 적응성을 측정하였다.

유연성은 챗봇이나 IEA 서비스가 다양한 상황과 사용자의 요구에 적응하고, 변화하는 환경에 따라 동적으로 대응할 수 있는 능력을 의미한다. 이는 시스템이 사전에 정의된 규칙이나 시나리오에만 의존하지 않고, 새로운 상황에서도 적절한 반응을 생성할 수 있는 능력을 포함한다. 유연성의 측정 방법으로는 다양한 입력 처리 능력 평가, 예측 불가능한 상황 대응 능력 평가, 지속적 학습과 개선 평가 및 다기능성 평가 등이

있다. 본 연구에서는 인터뷰 참여자들이 경험한 IEA의 성능에 대한 측정을 위해 입력 처리 능력 평가를 주로 활용하였다.

고도화된 IEA는 다양한 형태의 비정형과 정형 데이터를 다차원 벡터로 변환하고 저장할 수 있게 되었다. 이를 통해 사용자는 신속하고 정확한 답변을 받고, 개인화된 서비스를 경험할 수 있다. 이러한 통합은 고객 만족도와 서비스 재사용률을 높이며, 인공지능 시스템의 성능 극대화에 이바지하여 서비스에 대한 신뢰성을 향상시킨다.

2) 평가 척도

본 연구는 IT 기술직과 서비스 기획 분야 전문가들을 대상으로 심층 인터뷰를 진행했다. 인터뷰는 인공지능 서비스인 IEA의 핵심 요소들의 중요성을 평가하기 위해 이루어졌으며, 이해력, 결과물의 정확성과 적절성, 학습 능력, 사용자 맞춤화, 다기능성, 의인화, 그리고 보안 및 프라이버시를 평가 척도로 설정했다. 다양한 언어 이해를 포함해 사용자 질문의 의도와 맥락을 정확히 파악하여 복잡한 작업을 처리하고, 이를 바탕으로 사용자에게 개인화된 응답을 인간처럼 친숙하게 제공할 수 있어야 한다. 결과물의 정확성과 적절성은 지속적인 데이터 학습과 알고리즘 최적화를 통해 더욱 향상되어, 사용자 경험을 크게 개선하고 서비스 효율성을 높일 것이다. 동시에 그와 같은 서비스 제공이 사용자의 데이터를 안전하게 처리하는 개인 정보 보호 속에서 이루어져야 한다.

IV. IEA 모델의 실증

4-1 실험 제품 설계 및 검증

본 연구에서는 실험용 전문 지능 보조 서비스인 챗파일(Chatfile)을 개발하여 IEA를 검증하였다. 이 서비스는 LMM과 벡터DB 기술을 중심으로, OpenAI의 ChatGPT-3.5와 4.0V, Pinecone 및 네이버 광학 글자 인식(OCR: Optical Character Recognition) 기술 등을 통합하여 설계되었다. 기술 스택 중 GPT-4V는 LMM을, Pinecone은 벡터DB를 대표하는 서비스들이다. 네이버 광학 글자 인식 기술이 이미지 내 한국어 글자를 정확하게 인식하여 인식률을 향상시켰다.

프로토타이핑을 진행하기 위한 비교 대상 서비스로는 2023년 9월 기준 전 세계 생성형 인공지능 서비스 월 사용자 순위 1위인 OpenAI의 ChatGPT-4와 28위인 CHATPDF 서비스들이었다. 실제 사용된 실험 자료는 수원지방법원 성남지원의 판결정본 PDF 파일로 사건 번호는 2022가단239673이다. 판결정본을 사용한 이유는 법원 판결문이 실제 사례를 기반으로 작성되어 매우 구체적이고 현실적인 데이터를 포함하기 때문에 인공지능 챗봇이 현실 세계에서 어떤 성능을 발휘하는지 평가하기 적합하며 법적 용어와 복잡한 논리 구조를 포함하고 있어 자연어 처리 능력을 검증하는 데 이상적이다.



* Chatfile answer was generated in order to show the OCR (Optical Character Recognition) capability of the Chatfile service in Korean language.

그림 6. 챗파일의 답변
Fig. 6. Chatfile answer

또한, 판결정본을 선택한 기준은 판결문은 법적, 논리적, 서술적 요소를 모두 포함하고 있어 다양한 언어적 도전을 제공하기 때문이다. 실험 진행 결과, ChatGPT-4는 PDF 파일의 내용을 아예 인식하지 못하였고, CHATPDF는 관련 없는 거짓 답변을 제공하는 환각 현상을 보였다. 반면, 챗파일은 그림 6과 같이 문서를 정확히 인식하고 적절한 요약 답변을 제공하였다. 챗파일은 원고의 계약 해제 시도와 관련된 법원의 판단을 정확하게 요약하여, 사용자의 질문에 맞는 정보를 제공하며 기술적 우수성을 입증하였다.

4-2 실험 제품 사용 경험 분석 및 검증

심층 인터뷰를 통해 취합한 키워드의 정량적 분석 결과, 총 80개(100%) 중 74개(92%)가 IEA의 6가지 요소와 직접적으로 관련이 있었으며, 그림 7과 같이 서비스 디자인 측면에서는 서비스 시스템(Service System)이 40%(응답 적응성 30% + 유연성 10%)로 가장 높았다. 총 80개의 키워드들 중 가장 많이 언급된 순서대로 살펴보면 전문 지식 노동자 페르소나에 해당하는 학습 데이터와 의인화가 각각 9회와 7회, 이해 관계자 중심 서비스에 해당하는 연산 속도와 사용자 맞춤화가 각각 7회와 6회를 기록하며 서비스 디자인 요소들과 직접적인 관련도를 보였다.

또한, 인터뷰 대상에 따라 다른 결과가 나왔다. IT 기술직 5명은 대체로 시스템 기반의 기술에 대해 강조하고 있지만, 서비스 측면에서는 전체적인 그림을 잘 파악하지 못하였다. 반면, 기획 전문가들 5명은 전반적으로 기술적인 지식 정도는 낮지만, 서비스의 사용자 입장에서 기능을 고려하고 전체적인 서비스 측면을 강조하며 의견을 제시하였다.

그 예로 IEA 프로토타이핑에 대해 IT 기술직 중 한 명은 “IEA는 벡터DB를 사용하여 전체 데이터를 잘 인덱싱하며 빠

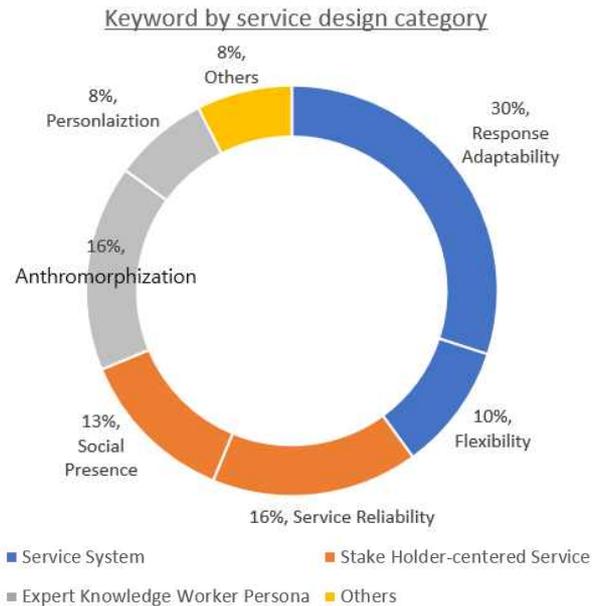


그림 7. 키워드 분석 내용
Fig. 7. Keyword analysis

르게 분석할 준비가 되어있다.”라고 언급했으며, 또 다른 IT 기술자는 “질 높은 답변을 제공하므로 인공지능 서비스들이 지향하는 방향이 맞지만, 비용 문제를 해결해야 범용으로 자리 잡을 수 있다.”라고 답변했다.

반면, 기획 전문가 중 한 명은 “생산성을 많이 향상시켜준다. 다만 팩트 체크가 필요하고 프롬프트 엔지니어링도 자동화되는 등 UI가 개선됐으면 좋겠다.”라고 언급했고, 또 다른 기획 전문가는 “IEA는 문서를 빠르게 요약하여 업무 효율성이 극대화되고 편리하다. 많은 문서를 업로드할 수 있도록 UI를 개선했으면 좋겠다.”라고 답변했다.

4-3 IEA 모델 요소의 정의

카드 소팅, 심층 인터뷰 및 프로토타이핑 실험 참여를 통해 취합한 데이터를 분석하여 다음과 같이 IEA 요소들이 서비스 디자인에 미치는 영향을 정의하였다. 총 6가지 중 서비스 시스템에 해당하는 응답 적응성은 서비스가 사용자의 상황과 요구에 신속하게 적응하여 유용한 반응을 제공하는 능력이고, 유연성은 실시간 정보를 기반으로 상황과 맥락을 빠르게 이해하여 사용자에게 맞춤 서비스를 제공한다.

이해 관계자 중심 서비스에 해당하는 서비스 신뢰성은 서비스가 일관되고 정확한 정보를 제공하여 사용자의 신뢰를 구축하는 능력이고, 사회적 실재감은 사용자가 서비스 이용 시 실제 사회적 상호작용을 하는 것처럼 느끼게 하는 요소이다.

마지막으로 전문 지식 노동자 페르소나에 해당하는 의인화는 서비스나 제품에 인간적인 요소를 더해 사용자와의 상호작용을 자연스럽게 만드는 능력이고, 개인화는 사용자의 과거 데이터와 프로필에 중점을 두고 장기간에 걸쳐 학습하여 경험을 제공하는 능력이다.

4-4 IEA의 최종 모델

IEA의 최종 모델은 그림 8과 같이 서비스 시스템인 응답 적응성과 유연성이 40%, 이해 관계자 중심 서비스인 서비스 신뢰성과 사회적 실재감이 29%, 그리고 전문 지식 노동자 페르소나인 의인화와 개인화가 24%를 기록하며 사용자 만족도와 재사용 의도 강화에 미치는 영향 중 핵심 요소로 분석되며, 기술과 서비스 중심의 전문 지능 보조 서비스로 발전 가능성을 확인했다.

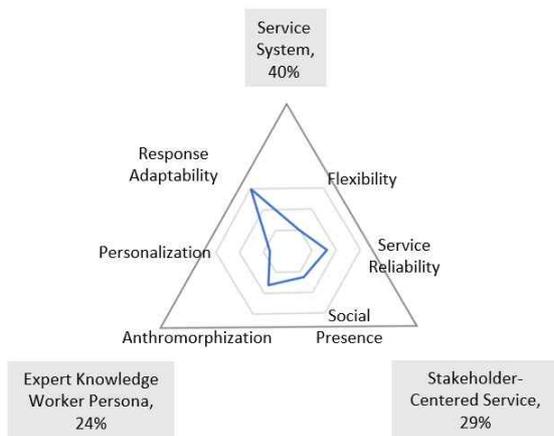


그림 8. IEA 모델 검증 및 정의
 Fig. 8. Validation and definition of IEA model

V. 결론 및 제언

디지털 전환 가속화로 인공지능 챗봇 수요가 증가했지만,

ChatGPT 모델을 포함한 현대 챗봇들의 한계로 인해 더 발전된 인공지능 서비스 개발의 필요성이 대두되고 있다. 그런 가운데 본 연구는 카드 소팅, 델파이 심층 인터뷰 조사를 통해 LMM과 벡터DB를 활용한 전문 지능 보조 서비스인 IEA의 필요성을 확인했고, 프로토타이핑을 통해 사용자의 요구를 정확히 파악하고 만족시키는 IEA의 핵심 요소들을 도출했다. 심층 인터뷰의 분석 결과, 총 80개의 키워드들과 6가지 IEA 요소들 사이 92%의 관련성을 보였다. 결론적으로, 6가지 IEA 요소들은 사용자의 전문 지식 제공 서비스를 최적화하므로 전문 지능 보조 서비스 디자인에 반드시 필요하다는 것을 검증하였다.

이러한 결과는 LMM과 벡터DB 기술 기반의 전문 지능 보조 서비스가 사용자들의 요구를 충족시키며, 인공지능서비스가 앞으로 나아가야 할 방향임을 보여주었다. 프로토타이핑과 같이 IEA 서비스를 법률 분야에 적용하는 예로 볼 경우, 법률 컨설팅을 받을 때 변호사 상담 관련 금전적인 부담을 덜어주며, 24시간 서비스 제공이 가능하므로 사용자에게 편리성을 제공할 수 있다.

그러나, 방대한 양의 여러 포맷의 정형 및 비정형 데이터 기반의 자료를 분석할 경우, 높은 비용 문제가 있다. 또한, 인터뷰 대상자가 10명이므로 연구 결과의 일반화 가능성과 신뢰성에 영향을 미칠 수 있는 한계점이 존재한다. 이러한 한계점을 보완하기 위해 더 큰 표본 크기와 다양한 배경을 가진 인터뷰 대상자를 포함하는 것이 필요하므로 향후 추가 연구를 진행할 수 있기를 바란다.

감사의 글

본 논문은 제 1 저자의 석사학위논문 ‘인공지능전문가(Expert Assistant(IEA))와 작업 지향적 규칙 기반 챗봇의 서비스 디자인 중심 비교분석 연구’를 발췌하여 수정 및 보완한 것이다.

참고문헌

[1] C.-G. Yang, "A Study on the Satisfaction and Dissatisfaction in AI Chatbot," *Asia-Pacific Journal of Business Venturing and Entrepreneurship*, Vol. 17, No. 2, pp. 167-177, April 2022.

[2] C. Jeong, "A Study on the Service Integration of Traditional Chatbot and ChatGPT," *Journal of Information Technology Applications and Management*, Vol. 30, No. 4, pp. 11-28, August 2023. <https://doi.org/10.21219/jitam.2023.30.4.011>

[3] S. Yin, C. Fu, S. Zhao, T. Xu, H. Wang, D. Sui, ... and E. Chen, "Woodpecker: Hallucination Correction for Multimodal Large Language Models," arXiv:2310.16045, October 2023. <https://doi.org/10.48550/arXiv.2310.16045>

- [4] E. Elsholz, J. Chamberlain, and U. Kruschwitz, “Exploring Language Style in Chatbots to Increase Perceived Product Value and User Engagement,” in *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval (CHIIR '19)*, Glasgow, UK, pp. 301-305, March 2019. <https://doi.org/10.1145/3295750.3298956>
- [5] OpenAI. Introducing ChatGPT [Internet]. Available: <https://openai.com/index/chatgpt>.
- [6] Z. Yang, L. Li, K. Lin, J. Wang, C.-C. Lin, Z. Liu, and L. Wang, The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision), Microsoft Corporation, Technical Report, 2023. <https://doi.org/10.48550/arXiv.2309.17421>
- [7] R. E. Rice, “Media Appropriateness: Using Social Presence Theory to Compare Traditional and New Organizational Media,” *Human Communication Research*, Vol. 19, No. 4, pp. 451-484, June 1993. <https://doi.org/10.1111/j.1468-2958.1993.tb00309.x>
- [8] D. Gefen and D. W. Staub, “Consumer Trust in B2C e-Commerce and the Importance of Social Presence: Experiments in e-Products and e-Services,” *Omega*, Vol. 32, No. 6, pp. 407-424, December 2004. <https://doi.org/10.1016/j.omega.2004.01.006>
- [9] C. N. Gunawardena and F. J. Zittle, “Social Presence as a Predictor of Satisfaction within a Computer-Mediated Conferencing Environment,” *American Journal of Distance Education*, Vol. 11, No. 3, pp. 8-26, 1997. <https://doi.org/10.1080/08923649709526970>
- [10] L. Qiu and I. Benbasat, “Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems,” *Journal of Management Information Systems*, Vol. 25, No. 4, pp. 145-182, 2009. <https://doi.org/10.2753/MIS0742-1222250405>
- [11] C. Nass, J. Steuer, and E. R. Tauber, “Computers are Social Actors,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*, Boston: MA, pp. 72-78, April 1994. <https://doi.org/10.1145/191666.191703>
- [12] J. Short, E. Williams, and B. Christie, *The Social Psychology of Telecommunications*, London, UK: John Wiley & Sons, 1976.
- [13] L. Jiang, J. Hoegg, and D. W. Dahl, Seeing Smiles: Consumers’ Adoption of Anthropomorphized New Product, in *NA - Advances in Consumer Research (Vol. 38)*, Duluth, MN: Association for Consumer Research, pp. 157-158, 2011.
- [14] S. Byun and C. Cho, “The Effect of the Anthropomorphism Level and Personalization Level on AI Financial Chatbot Recommendation Messages on Customer Response,” *The Korean Journal of Advertising and Public Relations*, Vol. 22, No. 2, pp. 466-502, April 2020. <https://doi.org/10.16914/kadpr.2020.22.2.466>
- [15] A. F. Smeaton and J. Callan, “Personalisation and Recommender Systems in Digital Libraries,” *International Journal on Digital Libraries*, Vol. 5, No. 4, pp. 299-308, August 2005. <https://doi.org/10.1007/s00799-004-0100-1>
- [16] SSRN. Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents [Internet]. Available: <https://ssrn.com/abstract=2569637>.
- [17] E. Go and S. S. Sundar, “Humanizing Chatbots: The Effects of Visual, Identity and Conversational Cues on Humanness Perceptions,” *Computers in Human Behavior*, Vol. 97, pp. 304-316, August 2019. <https://doi.org/10.1016/j.chb.2019.01.020>
- [18] W. H. DeLone and E. R. McLean, “The DeLone and McLean Model of Information Systems Success: A Ten-Year Update,” *Journal of Management Information Systems*, Vol. 19, No. 4, pp. 9-30, 2003. <https://doi.org/10.1080/07421222.2003.11045748>
- [19] M.-H. Cho, “A Study on the History, Classification and Development Direction of Artificial Intelligence,” *Journal of the KIECS*, Vol. 16, No. 2, pp. 307-312, April 2021.
- [20] Maximum Truth. AIs Ranked by IQ; AI Passes 100 IQ for First Time, with Release of Claude-3 [Internet]. Available: <https://www.maximumtruth.org/p/ais-ranked-by-iq-ai-passes-100-iq>.
- [21] ChosunBiz. ChatGPT and Gemini are Nervous about the Release of Claude 3, which has “Human-Level Understanding”... Anthropic Shakes Up the Generative AI Landscape [Internet]. Available: <https://biz.chosun.com/it-science/ict/2024/03/12/XDRFR5TOYVFWRCGJKGIBCCYECQ/>.
- [22] Elastic. What is Vector Search? [Internet]. Available: <https://www.elastic.co/what-is/vector-search/>.
- [23] OpenAI. GPT-4V(ision) System Card [Internet]. Available: https://cdn.openai.com/papers/GPTV_System_Card.pdf.



김희석(Huey Kim)

2024년 : 성균관대학교 서비스융합디자인협동과정(협동과정 석사)

※ 관심분야 : 서비스융합디자인, AI챗봇, 빅데이터분석



설상훈(Sanghun Sul)

2012년 : 성균관대학교 기계공학과(공학석사)

2016년 : 성균관대학교 기계공학과(공학박사)

2018년~2019년: LG전자 자문교수

2019년~2019년: 문화체육관광부 전문위원

2019년~2020년: 국토교통부 전문위원

2019년~2020년: 아모레퍼시픽 자문교수

2018년~현 재: 성균관대학교 기계공학과 교수

※ 관심분야 : 서비스융합디자인, 데이터디자인, 브랜드전략