JESK J Ergon Soc Korea 2026; 45(3): 213-237
https://dx.doi.org/10.5143/JESK.2026.45.3.213
https://jesk.or.kr eISSN:2093-8462

Decision Mechanism of User Switching Behavior in AI Service Failure: The Role of Cognitive Evaluation

GUO MENGLING, Jinho Yim

Kookmin University, Graduate School of Techno Design, Department of Smart Experience Design, Seoul, 02707

AI 서비스 실패 상황에서 사용자 전환 행동의 결정 메커니즘: 인지 평가의 역할

곽맹령, 임진호

국민대학교 스마트디자인학과

Corresponding Author
Jinho Yim
Kookmin University, Graduate School of Techno Design, Department of Smart Experience Design, Seoul, 02707
Email : hci.yim@kookmin.ac.kr

Received : April 01, 2026
Revised : June 01, 2026
Accepted : June 02, 2026

Copyright@2026 by The Human Factors & Ergonomics Society of Korea. All right reserved.

Objective: The purpose of this study is to investigate users' dual-path behavioral decision-making—continuing AI service use versus switching to human support—under AI service failure conditions, and to examine the role of cognitive evaluations in this process.

Background: Prior research has primarily focused on single behavioral intentions, such as continuance intention, while overlooking the trade-off and transition between alternative service channels under failure conditions. In problem-solving contexts, users often face a decision conflict between persisting with AI services and seeking human assistance.

Method: This study adopts a two-stage mixed-method approach. In the first stage, Latent Dirichlet Allocation (LDA) topic modeling was applied to analyze 1,500 real user complaints to identify six data-driven dimensions of AI service failure. In the second stage, attribution theory was used to screen the dimensions and select four constructs. Based on the Technology Acceptance Model (TAM), structural equation modeling (SEM) was employed to examine how these failure dimensions influence users' behavioral decisions through cognitive evaluation pathways.

Results: Perceived context discontinuity (PCD) exerts the strongest negative effect on perceived ease of use. Among cognitive evaluations, perceived ease of use has a stronger influence on switching intention than perceived usefulness. Multi-group analysis further reveals significant differences across groups with varying levels of AI usage frequency.

Conclusion: Users' responses to AI service failure reflect a cognitive cost-based decision-making process rather than simple withdrawal behavior. By adopting a data-driven approach based on real user complaints, this study enhances the contextual validity of failure dimensions and extends the applicability of TAM to failure contexts.

Application: Reducing interaction costs and maintaining conversational continuity are critical for preventing user switching and improving AI service retention. Enterprises should prioritize the development of context-aware dialogue systems and implement differentiated hybrid AI–human service recovery strategies based on users' experience levels.

Keywords: AI service failure, Technology acceptance model, Cognitive evaluation, Switching intention, LDA topic modeling

1. Introduction

인공지능(AI)과 자연어 처리(NLP) 기술의 발전에 따라, AI 챗봇은 24시간 서비스 제공, 즉각적인 응답, 그리고 비용 절감의 장점을 바탕으로 전자상거래, 금융 등 다양한 분야에서 널리 활용되고 있다(Huang and Rust, 2018). 특히 포스트 코로나 시대에서 AI 챗봇은 기업이 디지털 전환과 '비용 효율적 서비스 우수성(cost-effective service excellence)'을 달성하기 위한 핵심 수단으로 자리잡고 있다(Wirtz and Zeithaml, 2018). 그러나 다양한 소비 상황에서 AI 챗봇은 주관성이 높거나 맥락이 불명확한 사용자 요구를 정확히 이해하는 데 여전히 한계를 보인다. 시스템이 사용자의 문제에 적절히 대응하지 못하는 경우 서비스 실패(Service Failure)가 발생하며, 이는 사용자 경험에 부정적인 영향을 미친다(Huang and Dootson, 2022).

본 연구에서는 'AI 고객 서비스 실패'를 문제 해결 및 불만 처리(Problem Solving / Complaint Handling) 상황으로 한정하여 정의한다. 이러한 상황은 일반적인 정보 탐색에 비해 과업의 중요성과 실패 위험이 높기 때문에, 서비스 목표가 달성되지 않을 경우 사용자는 더 강한 불만을 느끼고 이후 행동을 변경할 가능성이 높다. 이와 같은 맥락에서 사용자 반응은 단순한 사용 지속 여부를 넘어, 'AI 서비스를 계속 사용할 것인지' 혹은 '인간 상담으로 전환할 것인지' 사이에서의 선택 문제로 나타난다. 따라서 AI 서비스 실패 이후의 사용자 반응은 본질적으로 행동 의사결정 과정으로 이해될 수 있으며, 그 결정 메커니즘을 규명하는 것은 중요한 이론적·실무적 의미를 지닌다.

기존 AI 챗봇 연구는 주로 자연어 처리 능력, 인터페이스 설계, 시스템 성능 등 기술적 구현에 초점을 두거나(Yang, 2022), 사용자 관점에서 사용 의도(Ma and Park, 2024), 서비스 품질(Jeon and Yoo, 2023), 의인화 설계(Hwang and Sung, 2023) 등을 다루어 왔다. 그러나 AI 고객 서비스의 확산과 함께 서비스 실패가 중요한 사용 맥락으로 부각되고 있음에도, 기존 연구는 주로 연구자 관점에서 실패를 분류해 왔으며(Cho and Jung, 2025) 실제 사용자 데이터에 기반한 체계적 분석은 부족하다. 또한 기존 실증 연구는 지속사용의도와 같은 단일 행동 결과에 초점을 맞추는 경우가 많아, '지속 사용'과 '인간 상담 전환' 간의 의사결정 과정을 통합적으로 설명하는 데 한계가 있다. 따라서 서비스 실패 상황에서 다중 행동 경로가 형성되는 메커니즘을 통합적으로 분석할 필요가 있다.

서비스 실패는 인지적 평가, 감정 반응, 귀인 판단이 복합적으로 작용하는 심리적 사건으로 이해할 수 있다(Smith and Bolton, 2002; Weiner, 1985). 그러나 감정 및 귀인 변수를 포함한 통합 모형을 구성하기에 앞서, 기술수용모형(TAM)과 같은 인지 평가 틀이 서비스 실패 상황에서도 유효하게 적용되는지를 검증할 필요가 있다. 이에 본 연구는 우선 감지된 유용성(perceived usefulness)과 감지된 용이성(perceived ease of use)에 초점을 맞추어, 서비스 실패 상황에서 이들이 사용자 행동 의사결정에 미치는 영향을 분석한다.

따라서 본 연구는 AI 고객 서비스 실패 상황을 배경으로, 실제 사용자 불만 데이터를 활용하여 LDA 토픽 모델링을 통해 서비스 실패 유형을 도출하고, 기술수용모형(TAM)을 기반으로 '서비스 실패–인지 평가–행동 선택'의 연구 모형을 구성한다. 본 연구는 사용자가 'AI 서비스 지속 사용'과 '인간 상담 전환' 사이에서 어떠한 의사결정을 내리는지를 분석함으로써, 서비스 실패 상황에서의 행동 형성 메커니즘을 설명하고, 기업의 서비스 회복 및 고객 유지 전략에 대한 이론적·실무적 시사점을 제시한다.

2. Literature Review

2.1 AI Chatbot service failure

2.1.1 Definition and characteristics

AI 챗봇의 서비스 실패는 사용자가 서비스 이용 과정에서 경험하는 실제 오류 또는 주관적으로 인식하는 문제를 의미한다(Liu et al., 2023). AI 고객 서비스는 자연어 처리(NLP) 기술에 기반한 인간-컴퓨터 상호작용을 통해 이루어지며, 자연어는 높은 맥락 의존성과 표현의 다양성을 가지기 때문에 현재 시스템은 맥락 이해 및 의도 인식 측면에서 일정한 한계를 보인다(Izadi and Forouzanfar, 2024). 시스템이 사용자의 의도를 정확히 이해하지 못하거나 부적절한 응답을 제공할 경우, 이는 의사소통 실패로 이어지며 사용자 불만을 유발할 수 있다(Ozuem et al., 2024). 또한 서비스 실패는 기술적 차원의 실제 오류(actual failure)뿐만 아니라 사용자의 주관적 인식에 기반한 실패(perceived failure)를 모두 포함한다. 즉, 시스템이 정상적으로 작동하더라도 서비스 결과가 사용자 기대에 미치지 못할 경

우 서비스 실패로 인식될 수 있다(Liu et al., 2023). 따라서 서비스 실패는 본질적으로 사용자 주관적 평가에 의해 규정되는 부정적 서비스 경험으로 이해할 수 있다.

2.1.2 Types of service failure

기존 연구에서는 AI 챗봇의 서비스 실패를 기능적 실패, 정보적 실패 등으로 구분하는 경향이 있다(Liu et al., 2023). 그러나 이러한 분류는 주로 연구자 관점에 기반하고 있어 실제 사용자 경험을 반영한 체계적 분석이 부족한 한계가 있다. 이에 본 연구는 LDA 토픽 모델링 기법을 활용하여 실제 사용자 불만 데이터를 기반으로 서비스 실패 유형을 도출하고, 사용자 실제 경험에 보다 부합하는 변수 체계를 구축하고자 한다. 또한 이를 바탕으로 서비스 실패가 사용자 인지 평가 및 행동 의사결정에 미치는 영향을 분석한다.

2.2 Attribution theory: How users interpret service failure

귀인이론(Attribution Theory)은 개인이 사건 결과의 원인을 어떻게 판단하고 해석하는지를 설명하는 이론이다(Weiner, 1985). Weiner는 개인이 사건의 원인을 주로 인과 소재(Locus), 안정성(Stability), 통제 가능성(Controllability)의 세 가지 차원에서 귀인한다고 제시하였으며, 이러한 귀인 방식의 차이는 개인의 감정 반응과 후속 행동에 영향을 미칠 수 있다고 설명하였다.

서비스 실패 연구에서는 인과 소재(Locus)가 가장 널리 사용되는 귀인 차원 중 하나이다. 이 차원은 개인이 실패의 원인을 시스템 또는 서비스 제공 주체 내부의 문제로 해석하는 내부 귀인(internal attribution)과 외부 환경이나 상황적 요인으로 해석하는 외부 귀인(external attribution)으로 구분된다(Folkes, 1984). 사용자가 실패의 원인을 시스템 자체의 문제로 인식할 경우, 시스템의 능력과 서비스 품질에 대한 평가는 더욱 부정적으로 형성될 수 있으며, 반대로 실패 원인을 외부 요인으로 인식할 경우 부정적 평가의 정도는 상대적으로 완화될 수 있다(Bitner, 1990).

AI 서비스 상황에서 사용자는 단순히 실패 결과만을 평가하는 것이 아니라, 실패의 원인이 AI 시스템 자체에 있는지, 또는 그 배후에 있는 기업의 지원 체계나 외부 환경에 있는지를 함께 판단하게 된다(Leo and Huh, 2020). 이러한 귀인 방식의 차이는 사용자가 AI 시스템의 능력, 신뢰성 및 서비스 가치에 대해 평가하는 방식과 밀접하게 관련될 수 있다(Gu et al., 2024).

이에 본 연구에서는 귀인이론을 구조 모형의 직접 변수로 설정하기보다는, LDA 토픽 모델링을 통해 도출된 서비스 실패 유형을 이론적으로 분류하고 연구 변수로 구성하기 위한 해석 틀로 활용하였다. 구체적으로, 사용자가 실패 원인을 AI 시스템의 이해 능력, 상호작용 과정, 대화 연속성 및 인간 지원 접근성의 문제로 인식할 가능성이 높은 유형은 내부 귀인적 실패로 분류하였다. 반면, 플랫폼 규칙, 제도적 제한 또는 제3자 책임과 관련된 유형은 외부 귀인적 실패로 해석하였다.

2.3 Emotional reactions in service failure contexts

서비스 실패는 사용자의 인지적 평가뿐만 아니라 분노, 좌절감, 실망과 같은 부정적 감정 반응에도 영향을 미치며(Smith and Bolton, 2002; Chebat and Slusarczyk, 2005), 이러한 감정 반응은 서비스 실패 이후의 평가 및 행동 반응을 설명하는 중요한 심리적 요인으로 논의되어 왔다. 특히 AI 고객 서비스 실패 상황에서는 응답 오류, 반복적인 설명 요구, 문제 해결 지연 등으로 인해 사용자가 부정적 감정을 경험할 가능성이 높으며, 이는 AI 서비스의 지속 사용 여부 또는 인간 상담원 전환 의도와도 관련될 수 있다.

다만 본 연구는 감정 반응을 직접적인 실증 변수로 포함하기보다는, 기술수용모형(TAM)에 기반한 인지 평가 경로를 우선적으로 검증하는 데 초점을 둔다. 이는 AI 서비스 실패 상황에서 지각된 사용용이성과 지각된 유용성이 행동 의도와 어떠한 관련성을 가지는지를 확인하기 위한 이론적 선택이며, 감정 반응은 향후 연구에서 인지 평가 및 귀인 판단과 함께 통합적으로 다루어질 필요가 있는 확장 변수로 이해할 수 있다.

3. Research Method

본 연구는 두 단계 혼합 연구 방법(Two-stage Mixed Method)을 채택한다. 제1단계는 탐색적 연구로, 실제 사용자 불만 데이터를 기반으로 LDA 토픽 모델링을 활용하여 AI 서비스 실패의 구조를 식별하고 이를 사용자 인지 평가 및 행동 반응에 영향을 미치는 외부 상황 요인으로 구성한다. 제2단계는 검증적 연구로, 구조방정식모형(SEM)을 적용하여 서비스 실패 상황에서 사용자의 인지 평가(지각된 유용성 및 지각된 사용용이성)가 행동 의사결정(지속사용의도 및 인적 서비스 전환 의도)에 미치는 영향을 분석한다.

본 연구는 기술수용모형(TAM)이 제시하는 핵심 인지 경로에 초점을 두고, AI 서비스 실패 상황에서 해당 모형의 설명력과 안정성을 검증하고자 한다. 이를 통해 복합적인 정서 변수의 영향을 배제한 상태에서 인지 메커니즘을 보다 명확하게 규명하고, 향후 귀인 및 정서 변수를 통합한 확장 연구를 위한 이론적 기반을 제공하고자 한다.

3.1 Data collection

본 연구의 데이터는 중국 소비자 불만 접수 플랫폼인 '흑묘투소(黑猫投诉, Black Cat)'에서 수집되었다. 해당 플랫폼은 전자상거래, 금융, 물류 등 다양한 산업 분야의 소비자 불만 정보를 제공하며, 실제 서비스 이용 과정에서 발생한 사용자 경험과 문제 상황을 포함하고 있다. 특히 AI 고객 서비스 관련 불만 사례를 포함하고 있어, AI 서비스 실패 상황을 분석하는 데 적합한 데이터 환경을 제공하며 높은 생태 타당성과 연구적 가치를 지닌다.

데이터 수집 과정에서 본 연구는 "AI 고객 서비스(AI客服)"와 "지능형 고객 서비스(智能客服)"를 키워드로 활용하였으며, Python 기반 웹 크롤링을 통해 2023년부터 2025년까지 총 1,973건의 사용자 불만 텍스트 데이터를 수집하였다. 또한 다양한 서비스 환경에서의 사용자 경험을 반영하기 위해 특정 산업 분야를 제한하지 않았다.

데이터 전처리 이후 분석에 활용 가능한 유효 텍스트 코퍼스를 구축하였다. 이를 바탕으로, 본 연구는 단순 무작위 추출(simple random sampling) 방식을 적용하여 전체 유효 데이터 중 1,500건의 텍스트를 LDA 주제 분석 표본으로 선정하였다. 이 과정에서 각 텍스트는 동일한 확률로 선택되도록 하였으며, 전체 데이터의 분포 특성을 유지하면서 표본의 대표성을 확보하고자 하였다.

전체 원자료를 직접 사용하지 않은 이유는 다음과 같다. 첫째, 1,500건의 텍스트는 안정적인 주제 구조를 도출하기에 충분한 규모의 코퍼스를 형성한다. 둘째, LDA 모형의 계산 복잡도는 코퍼스 규모의 증가에 따라 상승하므로, 표본 규모를 적절히 통제하는 것이 모형 학습의 효율성과 수렴 안정성을 확보하는 데 유리하다. 셋째, LDA 결과는 키워드와 대표 텍스트를 기반으로 주제를 해석하고 명명하는 과정이 요구되므로, 코퍼스 규모를 적절히 제한하는 것이 주제 해석의 일관성과 분석의 실행 가능성을 높이는 데 기여한다.

3.2 Data preprocessing

주제 분석의 정확성을 제고하기 위하여, 본 연구는 LDA 주제 모델링에 앞서 텍스트 데이터에 대한 전처리 과정을 수행하였다. 먼저, 원시 데이터에서 중복된 불만 사례, 광고성 정보, 그리고 내용이 불완전하거나 해석이 어려운 텍스트를 제거하여 데이터의 품질을 확보하였다.

다음으로, 중국어 텍스트에 대해 분절 처리를 수행하였으며, Jieba 분절 도구를 활용하여 연속된 텍스트를 분석 가능한 단어 단위로 분리하였다. 분절 이후에는 중국어 일반 불용어 사전을 기반으로 하되, AI 고객 서비스 불만 데이터의 특성을 반영하여 불용어 목록을 보완하였다. 이를 통해 접속사, 조사 등 의미 기여도가 낮은 일반 단어를 제거함으로써, 주제 분포에 불필요한 영향을 미칠 수 있는 비관련 단어를 최소화하였다.

이와 같은 전처리 과정은 잡음 텍스트와 정보량이 낮은 단어가 LDA 모델 결과에 미치는 영향을 감소시키고, 주제 식별의 정확성과 의미적 일관성을 향상시키는 데 기여한다. 이러한 전처리 절차를 통해 본 연구는 LDA 주제 분석에 적합한 텍스트 코퍼스를 구축하였으며, 이후의 주제 모델링 및 주제 해석을 위한 기초 자료를 마련하였다.

3.3 LDA topic analysis

3.3.1 Topic identification

텍스트 데이터에 내재된 잠재적 토픽 구조를 식별하기 위해, 본 연구에서는 LDA (Latent Dirichlet Allocation) 토픽 모델을 활용하였다. LDA는 단어들의 공기(co-occurrence) 관계를 기반으로 비정형 텍스트로부터 의미적으로 일관된 잠재 토픽을 추출할 수 있는 대표적인 비지도 학습 방법이다(Blei et al., 2003).

본 연구는 Python 환경에서 Gensim 라이브러리를 활용하여 LDA 모형을 구현하였다. 모형 구축 과정에서 전처리된 텍스트 자료를 문서–단어 행렬(document–term matrix)로 변환하여 모형의 입력 자료로 사용하였다. 문서–주제 분포의 디리클레 사전 분포 파라미터(α)와 주제–단어 분포의 디리클레 사전 분포 파라미터(β)는 모두 "auto"로 설정하여, 모형이 학습 과정에서 데이터에 기반하여 비대칭 사전 분포를 자동으로 추정하도록 하였다. 또한 모형의 반복 횟수는 200회로 설정하였으며, 결과의 안정성과 재현 가능성을 확보하기 위해 난수 시드(random_state = 42)를 고정하였다.

최적의 주제 수(K 값)를 결정하기 위해, 본 연구는 K=2부터 K=10까지의 다양한 후보 모형을 비교·분석하였으며, 모형 평가 지표로는 퍼플렉시티(Perplexity)와 일관성 점수(Coherence Score)를 활용하였다. 퍼플렉시티는 모형의 데이터 적합도를 측정하는 지표로, 값이 낮을수록 모형의 일반화 성능이 우수함을 의미한다. 반면, 일관성 점수는 각 주제에 포함된 상위 단어들 간의 의미적 일관성을 평가하는 지표로, 값이 높을수록 주제의 해석 가능성이 높음을 의미한다(Röder et al., 2015). 특히, 주제 모형은 명확한 정답 레이블이 존재하지 않기 때문에, 일관성 점수는 주제의 의미적 품질을 평가하는 중요한 내부 지표로 간주된다(Chang et al., 2009).

각 K 값에 따른 모형 평가 결과는 Figure 1에 제시하였다. 분석 결과, K 값이 증가함에 따라 퍼플렉시티는 전반적으로 감소하는 경향을 보였으며, K=6에서 최저값을 나타낸 이후 K>6 구간에서는 다시 증가하는 양상을 보였다. 동시에, 일관성 점수는 K=6에서 국지적 최대값(약 0.415)을 기록하였고, K=7에서 소폭 감소한 이후 다시 완만하게 증가하는 경향을 보였으나, K=6부터 K=10까지의 전체 변동 폭은 0.01 이하로 제한되어 추가적인 개선 효과는 크지 않은 것으로 나타났다.

또한, 주제 수가 과도하게 증가할 경우 주제 간 내용이 과도하게 세분화되어 전체적인 의미 해석의 명확성이 저하될 수 있다는 점을 고려하여, 본 연구는 모형 적합도, 주제 일관성 및 의미적 해석 가능성을 종합적으로 고려한 결과, 최종적으로 K=6을 최적의 주제 수로 선정하였다.

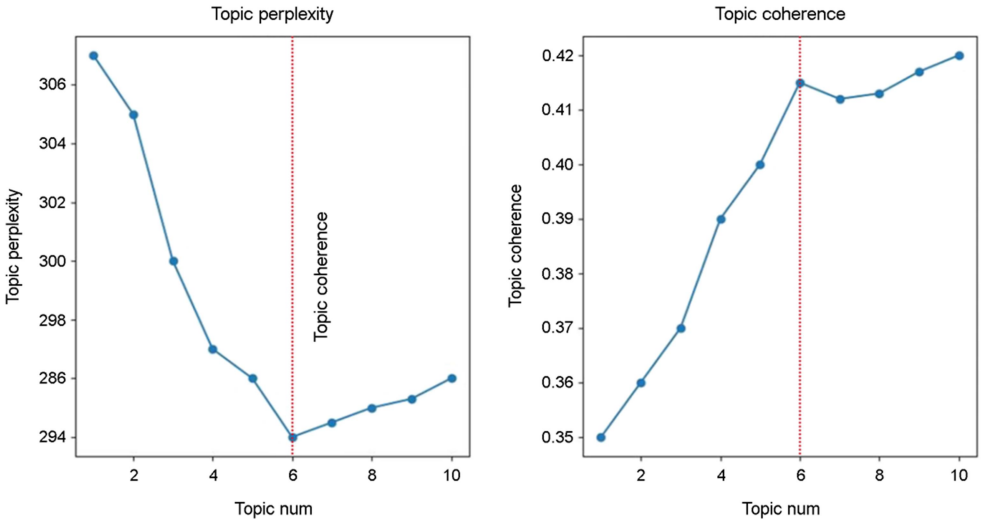

Figure 1. Perplexity and Coherence curves

3.3.2 LDA topic visualization and interpretation

주제 모형 결과에 대한 해석을 보완하기 위해, 본 연구는 PyLDAvis를 활용하여 LDA 결과를 시각화하였다. 해당 방법은 주제 간 거리를 2차원 공간에 투영하고 각 주제의 주요 키워드를 함께 제시함으로써, 서로 다른 주제 간의 관계를 직관적으로 이해하는 데 도움을 준다. Figure 2에 제시된 바와 같이, 여섯 개 주제는 2차원 공간 상에서 비교적 명확하게 분포되어 있으며, 주제 간 중첩 정도가 낮게 나타났다. 이는 각 주제가 일정 수준의 구분 가능성을 지니고 있음을 시사한다. 다만, PyLDAvis 시각화 결과는 주제 해석을 위한 보조적 도구로 활용되었으며, 주제 모형의 타당성을 검증하는 주요 근거로 사용되지는 않았다.

본 연구는 주로 일관성 점수(Coherence Score)와 주제 키워드의 의미적 해석 가능성을 중심으로 주제 결과의 적절성을 종합적으로 평가하였다. 일관성 점수 측면에서, K=6일 때 Coherence Score는 0.415로 나타났으며, 이는 K=2-5 구간에 비해 뚜렷한 향상을 보였다.

이후 K 값이 증가함에 따라 일관성 점수는 소폭 감소 후 다시 증가하는 경향을 보였으나, K=6부터 K=10까지의 전체 변동 폭은 0.01 이내로 제한되어 추가적인 개선 효과는 크지 않은 것으로 나타났다. 이러한 결과는 K=6에서 의미적 일관성과 주제 간 구분성 간의 균형이 비교적 잘 이루어졌음을 시사한다.

의미적 해석 가능성 측면에서, 본 연구는 각 주제의 핵심 키워드와 이에 대응하는 대표 텍스트를 추가적으로 검토하였다. 그 결과, 각 주제 내부의 키워드들은 높은 수준의 의미적 일관성을 보였으며, 실제 AI 고객 서비스 상호작용 과정에서 사용자가 경험하는 다양한 문제 유형과 대응되는 것으로 나타났다. 이에 따라, 키워드와 텍스트 맥락을 종합적으로 고려하여 각 주제를 명명하고 해석하였으며, 구체적인 결과는 제3.3.3절에 제시하였다.

종합하면, 본 연구는 Coherence Score와 의미적 해석 가능성에 기반한 이중 검증을 통해 LDA 모형 결과의 타당성을 평가하였으며, 해당 주제 구조가 AI 고객 서비스 실패 상황에서 나타나는 주요 문제 유형을 비교적 잘 반영하며, 양호한 의미적 해석력과 현실 상황과의 대응성을 가지는 것으로 판단된다.

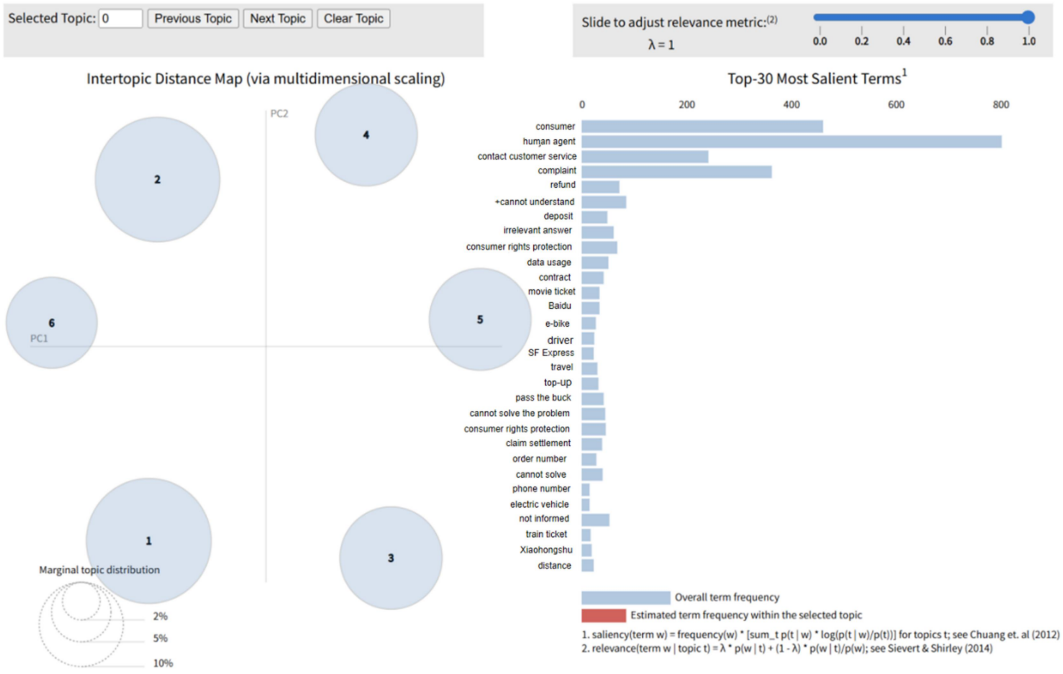

Figure 2. Topic visualization results

3.3.3 Topic naming and attribution classification

주제 식별 결과를 바탕으로, 본 연구는 각 주제의 고빈도 키워드와 대표 텍스트 내용을 분석하여 LDA를 통해 도출된 여섯 개 주제에 대한 의미적 명명을 수행하였다. 구체적으로, 키워드의 공출현 특성과 전형적인 텍스트 맥락을 종합적으로 검토함으로써 각 주제가 반영하는 사용자 문제 상황을 도출하고 주제명을 확정하였다. 또한 주제 해석의 신뢰성을 확보하기 위해 연구자 간 검토를 실시하였으며, 핵심 의미에 대해서는 높은 수준의 일치가 확인되었고 일부 표현상의 차이는 논의를 통해 조정하였다.

이후 데이터 기반 결과에 대한 이론적 해석을 강화하기 위해, 본 연구는 귀인이론(Attribution Theory)을 적용하여 식별된 주제를 내부 귀인과 외부 귀인으로 분류하였다. 귀인이론에 따르면, 사용자는 서비스 실패 상황에서 문제의 원인을 특정 대상에 귀속시키며, 이러한 귀인 방식은 이후의 인지적 평가와 행동 반응에 영향을 미친다(Weiner, 1985). 특히 서비스 실패 연구에서는 인과 소재(locus of causality)가 핵심 차원으로 간주되며, 사용자는 실패 원인을 서비스 제공자의 내부 요인 또는 외부 환경 요인으로 구분하여 인식하는 경향이 있다(Folkes, 1984; Bitner, 1990).

분석 결과, 의미 이해 실패, 과정 비효율, 맥락 연속성 단절과 같은 주제는 AI 고객 서비스 상호작용 과정에서 나타나는 시스템의 성능 및 기능적 한계를 반영하므로 내부 귀인으로 분류하였다. 또한 Topic 1(인간 지원 접근 불가능성)은 표면적으로 기업의 운영 방식과 관련되어 있으나, 사용자가 이를 AI 서비스 수행 능력의 부족으로 인식할 가능성이 있다는 점에서 내부 귀인에 포함하였다(Folkes, 1984; Leo and Huh, 2020).

반면, 시스템 규칙 제한 및 제3자 책임과 관련된 주제는 플랫폼 정책, 제도적 설계 또는 외부 서비스 주체와 관련된 문제로 해석되어 외부 귀인으로 분류하였다. 구체적으로, Topic 4는 시스템 제한, 규칙 제약 및 권한 문제와 관련되며, Topic 6은 책임 전가, 제3자 개입 및 서비스 운영 태도와 관련된 특성을 포함한다. 귀인이론의 인과 소재 기준에 따르면(Folkes, 1984; Bitner, 1990), 이러한 외부 귀인 요인은 AI 시스템 자체에 대한 사용성 및 유용성 평가와 직접적으로 연결되기 어렵다. 따라서 본 연구는 해당 주제들을 후속 구조 모형 분석에서 제외하였다. 각 주제의 주요 키워드, 귀인 분류 및 대응 변수에 대한 상세 내용은 Figure 3에 제시하였다.

Topic	Representative Keywords	Topic Label	Attribution Type	Corresponding Variable	Included in Model
Topic 1	human agent, unable to reach, transfer to human agent, transfer failure, unable to transfer, human agent unavailable, no response, queueing, re-queueing, hidden entry point, unable to find the button	Perceived Human Support Unavailability	Internal Attribution	PHSU	Included
Topic 2	irrelevant answers, answering the wrong question, does not understand, unable to understand, recognition error, repeated explanation, explained many times, chatbot reply, template response, wasted effort, human-AI service, cannot connect, incapable, unable to handle	Perceived Understanding Failure	Internal Attribution	PUF	Included
Topic 3	repetition, looping, repeated cycles, same responses, automated reply, going in circles, no progress, useless, ineffective, no outcome, no change, no advancement, futile effort, wasted processing, too many steps, starting over, cumbersome operation, re-operation, repeated input, disconnection	Perceived Process Inefficiency	Internal Attribution	PPI	Included
Topic 4	processing failure, system limitation, task incompletion, submission failure, cancellation failure, appeal failure, query failure, rule-based restriction, lack of support, operational restriction, missing functionality, permission restriction, system overload, software system constraint, account abnormality, system error	Perceived System Rule Restriction	External Attribution	—	Excluded
Topic 5	repeated description, re-input, say it again, no record, context loss, conversation history cleared, resubmission, history missing, unable to view records, incoherent, disconnected	Perceived Context Discontinuity	Internal Attribution	PCD	Included
Topic 6	shirking responsibility, inaction, lack of accountability, passing the buck, poor attitude, cold handling, third party, neglect, misleading, vague response	Perceived Third-party Responsibility Evasion	External Attribution	—	Excluded

Figure 3. LDA topic and attribution-based classification

추가적으로, 본 연구에서 도출된 네 가지 내부 귀인 주제는 단일한 실패 유형이 아니라 과정적 실패(process failure)와 결과적 실패(outcome failure)로 구분될 수 있다. 서비스 실패 연구에 따르면, 과정적 실패는 서비스 상호작용 과정에서 발생하는 문제와 관련되며, 결과적 실패는 서비스 목표 달성 여부와 관련된다(Smith et al., 1999). 이에 따라, 의미 이해 실패, 과정 비효율, 맥락 연속성 단절은 AI 고객 서비스 상호작용 과정에서 발생하는 문제로서 과정적 실패에 해당한다.

반면, 인간 지원 접근 불가능성은 사용자의 문제 해결 및 서비스 회복 가능성과 직접적으로 관련되므로 결과적 실패로 분류하였다. 종합하면, 본 연구는 네 가지 내부 귀인 주제를 핵심 외생 변수로 구성하였다. 구체적으로, 의미 이해 실패(Perceived Understanding Failure, PUF), 과정 비효율(Perceived Process Inefficiency, PPI), 맥락 연속성 단절(Perceived Context Discontinuity, PCD), 그리고 인간 지원 접근 불가능성(Perceived Human Support Unavailability, PHSU)을 포함하였다. 각 변수의 정의, TAM과의 이론적 대응 관계 및 측정 방식은 제3.4절에 제시하였다.

3.4 Operationalization and measurement of research constructs

제3.3절의 LDA 주제 분석 결과와 귀인 분류를 기반으로, 본 연구는 선별된 네 가지 외생 변수를 조작적으로 정의하고 측정하였다. 구체적으로, 의미 이해 실패(Perceived Understanding Failure, PUF), 과정 비효율(Perceived Process Inefficiency, PPI), 맥락 연속성 단절(Perceived Context Discontinuity, PCD), 그리고 인간 지원 접근 불가능성(Perceived Human Support Unavailability, PHSU)은 모두 LDA를 통해 식별된 내부 귀인 주제에서 도출된 변수로, AI 고객 서비스 실패 상황에서 사용자가 인식하는 시스템의 성능 및 서비스 수행 한계를 반영한다.

의미 이해 실패(PUF)는 사용자가 AI 고객 서비스와 상호작용하는 과정에서 시스템이 자신의 질문이나 의도를 정확하게 이해하지 못한다고 인식하는 정도를 의미하며, 질문과 관련 없는 응답이나 반복적인 설명 요구와 같은 상황을 포함한다. 과정 비효율(PPI)은 문제 해결 과정에서 사용자가 인지하는 절차적 복잡성과 비효율성을 의미하며, 반복적인 조작이나 불필요한 절차 등의 상황을 포함한다. 맥락 연속성 단절(PCD)은 상호작용 과정에서 정보 단절이나 맥락 상실이 발생한다고 인식하는 정도를 의미하며, 동일한 내용을 반복 설명해야 하거나 대화의 연속성이 유지되지 않는 상황 등을 포함한다. 인간 지원 접근 불가능성(PHSU)은 AI 고객 서비스가 문제를 해결하지 못하는 상황에서 사용자가 인간 상담원의 지원에 접근하기 어렵다고 인식하는 정도를 의미하며, 상담원 연결의 제한이나 응답 지연 등의 상황을 포함한다.

측정 항목은 기존의 검증된 척도를 기반으로 AI 고객 서비스 실패 상황에 맞게 수정·보완하였다. 의미 이해 실패(PUF)와 과정 비효율(PPI)의 문항은 Chen et al. (2022)의 AI 고객 서비스 품질 연구를 참고하였으며, 본 연구의 LDA 분석 결과에서 도출된 주제 특성을 반영하여 수정하였다. 맥락 연속성 단절(PCD)의 문항은 Chatbot Usability Scale (BUS-11)을 기반으로 하여 AI 서비스 실패 상황에 맞게 재구성하였다(Borsci et al., 2022). 인간 지원 접근 불가능성(PHSU)의 문항은 서비스 품질 연구의 접근성(access) 차원을 참고하여 AI 고객 서비스 환경에 맞게 수정하였다(Parasuraman et al., 1985). 또한 감지된 사용용이성(PEOU)과 감지된 유용성(PU)은 TAM 척도를 기반으로 측정하였으며(Davis, 1989), 지속사용의도(CI)는 Bhattacherjee (2001), 전환 의도(SI)는 Bansal et al. (2005)의 연구를 참고하였다. 모든 측정 항목은 기존 척도의 핵심 표현을 유지하면서 AI 고객 서비스 맥락에 맞게 조정하였다.

측정 항목의 내용 타당성을 확보하기 위해, 본 연구는 기존 척도와 LDA 분석 결과 간의 정합성을 중심으로 문항을 검토하였다. 이후 AI 고객 서비스 이용 경험이 있는 사용자(n=13)를 대상으로 사전조사를 실시하여 문항의 이해 가능성과 적용 가능성을 점검하였으며, 그 결과 일부 표현을 수정·보완하였다. 이러한 절차를 통해 기존 척도의 이론적 기반을 유지하면서도 AI 서비스 실패 맥락에 적합한 측정 도구를 구성하고자 하였다.

모든 문항은 Likert 7점 척도를 사용하여 측정하였으며, 1은 "전혀 동의하지 않음", 7은 "매우 동의함"을 의미한다. 각 변수의 측정 항목과 출처는 Table 1에 제시하였다.

Table 1. Questionnaire items

Variable	Item	Reference
Perceived Understanding Failure (PUF)	1. The AI customer service does not accurately understand my query / what I am trying to say. 2. The responses from the AI customer service do not match my questions. 3. The AI customer service fails to provide accurate answers to my questions.	Chen et al. (2022)

Table 1. Questionnaire items (Continued)

Variable	Item	Reference
Perceived Process Inefficiency (PPI)	1. The AI customer service process feels slow and inefficient. 2. It takes too much time to resolve my problem through the AI customer service. 3. The AI customer service requires too many unnecessary steps and does not simplify the process.	Chen et al. (2022)
Perceived Context Discontinuity (PCD)	1. The chatbot failed to remember information from earlier in the conversation. 2. The responses were not consistent with the previous conversation context. 3. After the conversation is interrupted, I usually need to explain my problem to the AI customer service again.	Borsci et al. (2022)
Perceived Human Support Unavailability (PHSU)	1. I found it difficult to locate the option for connecting to a human customer service agent. 2. Even when a channel to a human customer service agent was provided, I still found it difficult to actually reach one. 3. When the chatbot could not solve my problem, I was unable to obtain further assistance from a human customer service agent.	Parasuraman et al. (1985)
Perceived Ease of Use (PEOU)	1. The AI customer service is easy to use. 2. Interacting with the AI customer service is easy. 3. I can use the AI customer service without much difficulty	Davis (1989)
Perceived Usefulness (PU)	1. The AI customer service is helpful in solving my problems. 2. Using AI customer service improves the efficiency of problem solving. 3. Overall, the AI customer service is useful to me.	Davis (1989); Bhattacherjee (2001)
Switching Intention (SI)	1. I want to switch to a human agent to solve my problem. 2. I would prefer a human agent rather than continuing to use AI customer service. 3. In a similar situation, I would choose a human agent first.	Bansal et al. (2005)
Continuance Intention (CI)	1. I would still consider using AI customer service in the future. 2. If I need customer service in the future, I would still be willing to try AI customer service. 3. In similar situations, I would still consider using AI customer service again.	Bhattacherjee (2001)

3.5 Research model and hypotheses

3.5.1 Theoretical basis for model specification

본 연구는 서비스 실패가 행동 의도에 직접적으로 영향을 미치기보다는, 감지된 사용용이성(PEOU)과 감지된 유용성(PU)과 같은 인지적 평가를 통해 간접적으로 영향을 미친다고 본다. 이에 따라 본 연구는 '서비스 실패 → 행동 의도'의 직접 경로를 설정하지 않고, 기술수용모형(TAM)의 이론적 논리에 기반하여 인지적 평가 변수의 매개 역할에 초점을 두었다.

TAM에 따르면, 외부 변수는 행동 의도에 직접 영향을 미치기보다 인지적 평가를 통해 간접적으로 작용한다(Davis, 1989). AI 고객 서비스 실패 상황에서도 사용자는 실패 경험 이후 시스템의 사용 가능성과 기능적 가치에 대한 재평가를 수행하며, 이를 바탕으로 후속 행동 의사결정을 형성할 수 있다(Chebat and Slusarczyk, 2005). 따라서 본 연구는 서비스 실패와 행동 의도 간 관계를 설명하는 주요 경로로 PEOU와 PU를 설정하였다.

또한 본 연구는 제3.3절에서 제시한 과정적 실패(process failure)와 결과적 실패(outcome failure)의 구분을 바탕으로, 서비스 실패 유형

에 따라 인지적 평가에 미치는 영향 경로가 다를 수 있다고 보았다. 과정적 실패는 사용자와 시스템 간 상호작용 과정에서 발생하는 조작 경험 및 인지적 부담과 관련되므로 PEOU에 보다 밀접하게 영향을 미칠 가능성이 있다(Chebat and Slusarczyk, 2005). 반면, 결과적 실패는 문제 해결 성과 및 서비스 결과 평가와 관련되므로 PU에 보다 직접적으로 작용할 가능성이 있다(Bhattacherjee, 2001).

이에 따라 이해 실패, 절차 비효율, 대화 맥락 단절과 같은 상호작용 과정상의 문제는 사용자의 인지적 부담을 증가시켜 PEOU를 저하시킬 수 있으며(Chebat and Slusarczyk, 2005), 문제 해결 실패나 적절한 서비스 보완의 부재는 PU 평가를 저하시킬 가능성이 있다(Bhattacherjee, 2001; Huang and Rust, 2018). 다만 본 연구는 특정 서비스 실패 유형이 단일 인지 변수에만 영향을 미친다고 가정하지 않으며, 이론적 명확성을 위해 주요 영향 경로를 중심으로 연구 모형을 구성하였다.

마지막으로, 본 연구는 다수의 외생 변수를 포함하고, 이중 인지 경로를 통해 지속사용의도(Continuance Intention, CI)와 전환 의도(Switching Intention, SI)에 영향을 미치는 구조를 설정하였다. 이러한 구조는 다수의 변수와 경로를 포함하여 설명력을 확장하는 TAM 확장 연구의 흐름과 일치한다(Venkatesh et al., 2003).

본 연구의 전체 연구 모형은 Figure 4에 제시하였다.

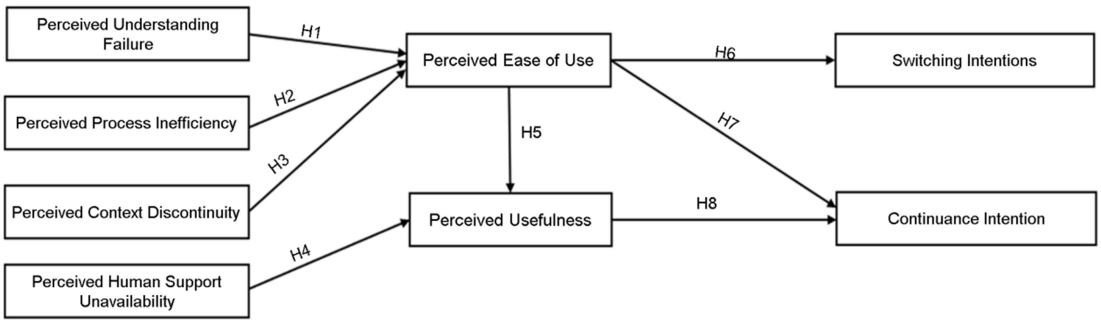

Figure 4. Structural model

3.5.2 Hypotheses development

Service failures and perceived ease of use

AI 챗봇과의 상호작용 과정에서 발생하는 이해 실패, 프로세스 저효율, 맥락 단절 등의 문제는 사용자의 인지적 부하와 조작 부담을 증가시키며, 이는 시스템 용이성에 대한 평가 저하로 이어진다. 기술수용모델(TAM)에 따르면, 지각된 용이성은 시스템 사용 과정에서의 조작 난이도와 인지적 부담에 대한 사용자의 주관적 평가를 반영한다(Davis, 1989; Venkatesh and Davis, 2000). 따라서 상호작용 과정에서 정보의 반복 입력이 요구되거나 과업 수행이 어려워질 경우, 인지적 부담이 증가하게 되고, 사용자가 체감하는 용이성은 하락하게 된다. 이에 본 연구는 다음과 같은 가설을 설정한다.

H1: 지각된 이해 실패(PUF)는 지각된 용이성(PEOU)에 부(-)의 영향을 미칠 것이다.
H2: 지각된 프로세스 저효율(PPI)은 지각된 용이성(PEOU)에 부(-)의 영향을 미칠 것이다.
H3: 지각된 맥락 단절(PCD)은 지각된 용이성(PEOU)에 부(-)의 영향을 미칠 것이다.

Service failure and perceived usefulness

지각된 유용성은 특정 시스템이 사용자의 문제 해결 및 과업 수행에 실질적인 도움을 제공할 것이라는 기대에 기반한 평가를 의미

한다(Davis, 1989). AI 고객 서비스에서 시스템이 사용자의 문제를 해결하지 못하고, 인간 상담원으로의 전환 또한 원활하지 않을 경우, 사용자는 해당 시스템이 문제 해결에 있어 충분한 가치를 제공하지 못한다고 인식하게 된다. 이러한 결과적 실패는 시스템의 기능적 가치에 대한 평가를 직접적으로 저하시킬 수 있으며, 이는 지각된 유용성의 하락으로 이어진다(Bhattacherjee, 2001). 이에 본 연구는 다음과 같은 가설을 설정한다.

H4: 지각된 인간 상담원 지원 불가성(PHSU)은 지각된 유용성(PU)에 부(-)의 영향을 미칠 것이다.

Cognitive evaluation and behavioral intentions

기술수용모델(TAM)에 따르면, 지각된 용이성은 행동 의도에 직접적인 영향을 미칠 뿐만 아니라, 지각된 유용성을 매개로 하여 행동 의도에 간접적인 영향을 미친다(Davis, 1989). 서비스 실패 상황에서 시스템에 대한 사용자의 인지적 평가가 낮아질 경우, 사용자는 문제 해결을 위해 인간 상담원으로 전환하는 등의 회피 행동을 취할 가능성이 높아진다. 반면, 시스템의 용이성과 유용성이 높게 지각될 경우, 사용자는 해당 시스템을 지속적으로 이용하려는 의도를 유지할 가능성이 높다. 이에 본 연구는 다음과 같은 가설을 설정한다.

H5: 지각된 용이성(PEOU)은 지각된 유용성(PU)에 정(+)의 영향을 미칠 것이다.
H6: 지각된 용이성(PEOU)은 인간 상담원 전환 의도(SI)에 부(-)의 영향을 미칠 것이다.
H7: 지각된 용이성(PEOU)은 지속 이용 의도(CI)에 정(+)의 영향을 미칠 것이다.
H8: 지각된 유용성(PU)은 지속 이용 의도(CI)에 정(+)의 영향을 미칠 것이다.

3.6 Participants and data collection

본 연구는 최근 6개월 이내 AI 고객 서비스 이용 경험이 있으며, AI 챗봇의 서비스 실패를 경험한 중국 사용자를 대상으로 하였다. 데이터는 온라인 설문조사를 통해 수집되었으며, 응답자들은 실제 서비스 실패 경험을 바탕으로 설문에 참여하였다. 불성실 응답을 제외한 총 375부의 유효 표본이 최종 분석에 활용되었다. 표본의 지역적 다양성을 확보하기 위해 베이징, 광둥, 허난, 허베이, 산둥 등 다양한 경제 발전 수준의 성(省) 및 직할시를 포함하였으며, 응답자의 구체적인 인구통계학적 특성은 <Table 2>에 제시된 바와 같다.

Table 2. Demographic characteristics of respondents

Variable	Category	Frequency	Percentage (%)
Gender	Male	191	50.9
	Female	184	49.1
Age	20~30	91	24.3
	31~40	86	22.9
	41~50	74	19.7
	51~60	68	18.1
	Over 60	56	14.9
AI service usage frequency	Rarely (Less than 2 times per month)	41	10.9
	Occasionally (2~4 times per month)	208	55.5
	Frequently (More than once a week)	126	33.6
Total		375	100

3.7 Data analysis procedure

본 연구는 제안된 연구 모형 및 가설을 실증적으로 검증하기 위해 SmartPLS 3.0을 활용한 부분최소제곱 구조방정식 모델(Partial Least Squares Structural Equation Modeling, PLS-SEM) 분석을 실시하였다. PLS-SEM은 복잡한 경로 관계 모델 분석에 적합하며, 특히 예측 지향적 연구 및 탐색적 연구에서 높은 적용성을 지니고 있어 정보시스템 및 서비스 경영 관련 연구에서 널리 활용되고 있다.

4. Results

4.1 Measurement model results

본 연구는 반사적 측정 모델(Reflective Measurement Model)의 신뢰성과 타당성을 검증하였다(Hair et al., 2019). 분석 결과는 <Table 3> 및 <Table 4>와 같다. 먼저, 모든 측정 항목의 외부 로딩(Outer Loadings)은 0.889~0.926으로 기준치인 0.70을 상회하여 지표 신뢰성을 확보하였다. 내부 일관성 신뢰성을 나타내는 Cronbach's alpha는 0.883~0.908, 합성신뢰도(CR)는 0.928~0.942로 모두 0.70 이상으로 나타났다. 또한 평균분산추출값(AVE)은 0.810~0.845로 기준치인 0.50을 상회하여 집중 타당성이 입증되었다. 마지막으로 판별 타당성 확인을 위해 실시한 HTMT (Heterotrait-Monotrait Ratio) 분석 결과, 모든 구성개념 간 수치가 0.305~0.535로 기준치인 0.90 미만으로 나타나 판별 타당성이 확보되었다(Henseler et al., 2015). 이상의 결과를 통해 본 연구의 측정 모델은 구조 모델 분석을 수행하기에 적합한 수준의 신뢰성과 타당성을 갖춘 것으로 판단된다.

Table 3. Measurement model assessment results

Variable	Indicator	Factor loading	α	CR	AVE
Continuance intention	CI1	0.900	0.883	0.928	0.810
	CI2	0.905			
	CI3	0.896			
Perceived context discontinuity	PCD1	0.924	0.908	0.942	0.845
	PCD2	0.916			
	PCD3	0.917			
Perceived ease of use	PEOU1	0.921	0.897	0.936	0.830
	PEOU2	0.907			
	PEOU3	0.904			
Perceived human support unavailability	PHSU1	0.904	0.898	0.936	0.831
	PHSU2	0.911			
	PHSU3	0.918			
Perceived process inefficiency	PPI1	0.891	0.893	0.933	0.823
	PPI2	0.918			
	PPI3	0.913			
Perceived understanding failure	PUF1	0.904	0.896	0.935	0.828
	PUF2	0.913			
	PUF3	0.913			

Table 3. Measurement model assessment results (Continued)

Variable	Indicator	Factor loading	α	CR	AVE
Perceived usefulness	PU1	0.926	0.908	0.942	0.845
	PU2	0.910			
	PU3	0.922			
Switching intention	SI1	0.914	0.895	0.934	0.826
	SI2	0.923			
	SI3	0.889			

Table 4. HTMT criterion

	CI	PCD	PEOU	PHSU	PPI	PUF	PU	SI
CI								
PCD	0.430							
PEOU	0.424	0.535						
PHSU	0.426	0.440	0.415					
PPI	0.405	0.378	0.403	0.305				
PUF	0.441	0.436	0.451	0.378	0.365			
PU	0.411	0.439	0.456	0.399	0.409	0.379		
SI	0.485	0.468	0.497	0.411	0.373	0.496	0.398	

추가적으로, 본 연구는 모든 측정 항목의 외부 로딩 값이 0.889~0.926 범위에 집중되고, 제안된 8개 가설이 모두 유의하게 지지된 결과에 대해 구성개념 잉여성(construct redundancy) 및 개념적 중복(conceptual overlap) 가능성을 추가로 검토하였다. 이에 대해 본 연구는 다음과 같은 세 가지 측면에서 검증을 수행하였다.

첫째, 이론적·개념적 측면에서 본 연구의 네 가지 외생 변수(PUF, PPI, PCD, PHSU)는 LDA 토픽 모델링을 통해 실제 사용자 불만 데이터로부터 도출된 것으로, 각각 의미 이해 실패, 과정 비효율, 맥락 연속성 단절, 인간 지원 접근 불가능성이라는 서로 다른 실패 차원을 반영한다(제3.3절 참조). 또한 측정 항목은 각각 Chen et al. (2022), Borsci et al. (2022), Parasuraman et al. (1985), Davis (1989), Bhattacherjee (2001), Bansal et al. (2005) 등 서로 다른 이론적 기반을 가진 검증된 척도를 토대로 구성되었으며, 본 연구의 맥락에 맞게 수정·보완(adaptation)되었다(제3.4절 참조). 더불어 사전조사(n=13)와 동료 연구자 검토를 통해 문항 표현의 명확성과 구성개념 간 개념적 경계를 점검하였다. 이러한 절차를 통해 구성개념 간 이론적 독립성이 확보되었다.

둘째, 판별 타당성 측면에서 모든 구성개념 간 HTMT 값은 0.305~0.535 범위로, 보수적 기준치인 0.85 (Henseler et al., 2015)를 크게 하회하였다. 이는 구성개념 간 충분한 개념적 구분이 존재함을 시사한다. 특히 HTMT 값이 전반적으로 낮은 수준으로 나타난 것은 구성개념 간 개념적 중복 가능성이 제한적임을 보여준다.

셋째, 다중공선성 측면에서 구조 모형 내 모든 VIF 값은 1.000~1.258 범위로, 보수적 기준치인 3.3 (Kock, 2015)을 크게 하회하였다. 이는 구성개념 간 통계적 중복이 존재하지 않음을 의미한다.

외부 로딩 값이 상대적으로 좁은 범위(0.889~0.926)에 집중된 것은 다음과 같은 요인에 기인한 것으로 해석할 수 있다. 첫째, 본 연구

의 측정 항목은 각 구성개념당 3개의 정제된 문항으로 구성되어 있으며, 모두 기존의 잘 확립된 척도를 기반으로 하여 의미적 명확성과 응답 일관성이 높다. 둘째, 본 연구 표본은 최근 6개월 이내 AI 고객 서비스 실패를 경험한 중국 사용자로 한정되어 응답자의 경험적 동질성이 비교적 높다. 셋째, 7점 리커트 척도의 사용 또한 일관된 응답 패턴 형성에 기여할 수 있다. 이러한 요인들이 종합적으로 작용하여 외부 로딩 값이 상대적으로 좁은 범위에 분포한 것으로 판단된다.

종합하면, 본 연구의 구성개념들은 이론적·통계적 측면에서 충분한 독립성을 가지며, 모든 가설이 유의하게 지지된 결과는 구성개념 잉여성에 기인한 것이 아니라 서비스 실패 상황에서 TAM 기반 인지 평가 메커니즘의 이론적 일관성과 설명력을 반영하는 것으로 해석된다.

4.2 Structural model results

측정 모형의 신뢰성과 타당성이 검증된 이후, 본 연구는 PLS-SEM 분석 절차에 따라 구조 모형을 평가하였다. 구조 모형 평가는 경로계수의 크기와 유의성, 내생 변수의 설명력(R^2), 그리고 다중공선성(VIF)을 중심으로 이루어졌다(Hair et al., 2019). 구조 모형의 전체 분석 결과는 Figure 5에 제시하였다. 각 경로의 표준화 경로계수와 주요 내생 변수의 설명력(R^2)을 제시하고 있으며, 이를 통해 AI 서비스 실패 상황에서 사용자의 인지 평가가 지속사용의도(CI)와 전환 의도(SI)에 어떠한 영향을 미치는지를 직관적으로 확인할 수 있다. 경로계수의 유의성은 Bootstrapping 기반의 비모수 검증 방법을 통해 검증하였다.

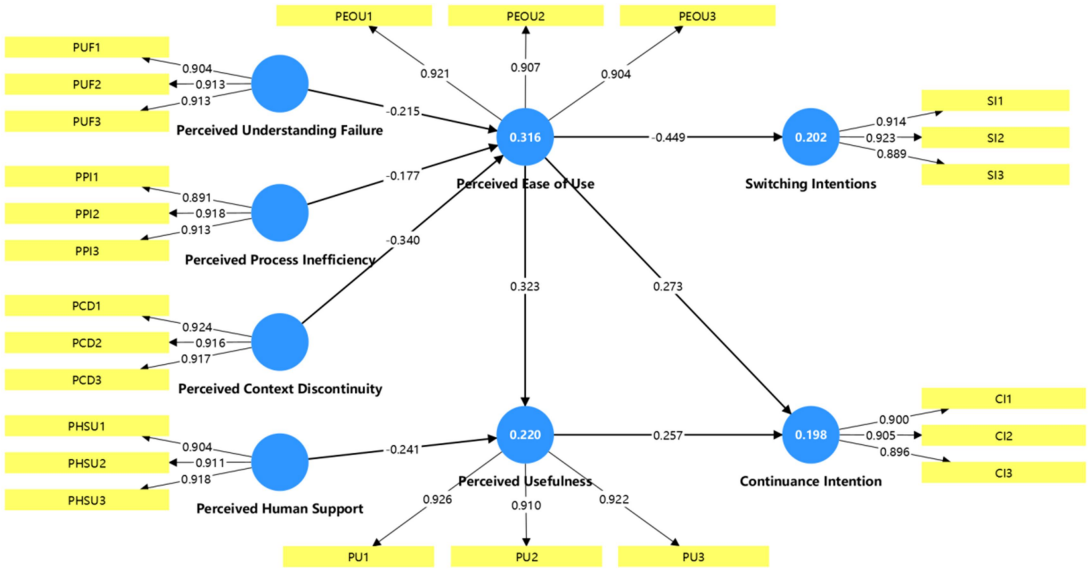

Figure 5. Structural model results

가설 검증 결과는 <Table 5>에 제시하였다. 분석 결과, 본 연구에서 제안한 8개의 가설은 모두 지지되었다. 서비스 실패 요인 중 지각된 이해 실패(β = -0.215, p < 0.001), 지각된 절차 비효율(β = -0.177, p = 0.001), 지각된 맥락 단절(β = -0.340, p < 0.001)은 지각된 사용용이성에 유의한 부(-)의 영향을 미치는 것으로 나타나 H1, H2, H3은 지지되었다. 또한 지각된 인간 지원 불가는 지각된 유용성에 유의한 부(-)의 영향을 미치는 것으로 나타나 H4는 지지되었다(β = -0.241, p < 0.001).

한편, 지각된 사용용이성은 지각된 유용성에 유의한 정(+)의 영향을 미치는 것으로 나타나 H5는 지지되었다(β = 0.323, p < 0.001). 또한 지각된 사용용이성은 전환 의도에 부(-)의 영향을, 지속사용의도에는 정(+)의 영향을 미치는 것으로 나타나 H6과 H7은 지지되었다(β = -0.449, p < 0.001; β = 0.273, p < 0.001). 마지막으로 지각된 유용성은 지속사용의도에 유의한 정(+)의 영향을 미치는 것으로

나타나 H8 역시 지지되었다(β = 0.257, p < 0.001).

Table 5. Hypothesis testing results

Hypothesis	Path	β	*t*-value	*p*-value	Decision
H1	PUF → PEOU	-0.215	4.200	***	Supported
H2	PPI → PEOU	-0.177	3.298	0.001	Supported
H3	PCD → PEOU	-0.340	6.359	***	Supported
H4	PHSU → PU	-0.241	3.935	***	Supported
H5	PEOU → PU	0.323	5.096	***	Supported
H6	PEOU → SI	-0.449	8.208	***	Supported
H7	PEOU → CI	0.273	4.387	***	Supported
H8	PU → CI	0.257	4.1971	***	Supported

Note: *p < 0.05, **p < 0.01, ***p < 0.001

인지 평가의 매개효과를 보다 명확히 검증하기 위하여, 본 연구는 Bootstrapping 기반의 specific indirect effect 분석을 수행하였다. 분석 결과는 <Table 6>에 제시하였다. 분석 결과, 모든 간접효과 경로의 p-value가 0.05 이하로 나타났으며, 95% bias-corrected confidence interval에 0이 포함되지 않는 것으로 확인되었다. 이는 본 연구에서 제안한 인지 평가 기반의 매개 메커니즘이 통계적으로 유의함을 의미한다.

특히, 지각된 맥락 단절(PCD)은 지각된 사용용이성(PEOU)을 매개로 전환 의도(SI)에 가장 강한 간접효과를 미치는 것으로 나타났다(β = 0.152, p < .001). 이는 AI 고객 서비스에서 대화 맥락의 연속성이 훼손될 경우 사용자의 사용용이성 인식이 저하되고, 결과적으로 인간 상담으로 전환하려는 경향이 강화될 수 있음을 시사한다.

또한 지각된 이해 실패(PUF)와 지각된 프로세스 비효율(PPI) 역시 PEOU를 매개로 전환 의도(SI)에 유의한 간접효과를 미치는 것으로 나타났다. 이러한 결과는 AI 서비스 실패 경험이 사용자의 인지적 평가 과정을 통해 행동 의도 형성에 영향을 미친다는 점을 보여준다.

한편, PEOU가 PU를 통해 CI에 영향을 미치는 간접효과 역시 유의한 것으로 나타났다. 예를 들어, PEOU → PU → CI 경로는 유의한 정(+)의 간접효과를 보였으며(β = 0.083, p = .001), 이는 사용용이성이 지각된 유용성을 향상시키고, 궁극적으로 지속사용의도를 강화할 수 있음을 의미한다.

종합적으로, 이러한 결과는 본 연구가 제안한 "서비스 실패 → 인지 평가 → 행동 의도"의 구조적 메커니즘을 추가적으로 지지하며, 인지 평가 변수가 AI 서비스 실패 상황에서 사용자 행동결정 과정에 중요한 매개 역할을 수행함을 보여준다(Table 6 참고).

Table 6. Results of specific indirect effect analysis

Category	Indirect path	β	STDEV	*t*-value	*p*-value	95% Bias-Corrected CI [LL, UL]
Indirect Effects on Continuance Usage Intention (CI)	PPI → PEOU → CI	-0.048	0.019	2.480	0.013	[-0.094, -0.017]

Table 6. Results of specific indirect effect analysis (Continued)

Category	Indirect path	β	STDEV	*t*-value	*p*-value	95% Bias-Corrected CI [LL, UL]
	PUF → PEOU → CI	-0.059	0.021	2.739	0.006	[-0.109, -0.025]
	PUF → PEOU → PU → CI	-0.018	0.007	2.501	0.012	[-0.036, -0.008]
	PCD → PEOU → PU → CI	-0.028	0.010	2.762	0.006	[-0.055, -0.013]
	PPI → PEOU → PU → CI	-0.015	0.007	2.114	0.035	[-0.033, -0.005]
	PEOU → PU → CI	0.083	0.026	3.219	0.001	[0.040, 0.142]
	PHSU → PU → CI	-0.062	0.025	2.440	0.015	[-0.122, -0.022]
	PCD → PEOU → CI	-0.093	0.028	3.334	0.001	[-0.151, -0.044]
Indirect Effects on Perceived Usefulness (PU)	PCD → PEOU → PU	-0.110	0.031	3.509	< 0.001	[-0.181, -0.058]
	PPI → PEOU → PU	-0.057	0.023	2.495	0.013	[-0.113, -0.022]
	PUF → PEOU → PU	-0.069	0.023	3.047	0.002	[-0.122, -0.033]
Indirect Effects on Switching Intentions (SI)	PCD → PEOU → SI	0.152	0.034	4.433	< 0.001	[0.092, 0.223]
	PPI → PEOU → SI	0.079	0.027	2.956	0.003	[0.031, 0.137]
	PUF → PEOU → SI	0.097	0.028	3.427	0.001	[0.046, 0.155]

Note: Indirect effects were tested using bootstrapping procedures with bias-corrected 95% confidence intervals. Mediation effects are considered significant when the confidence interval does not include zero

내생 변수의 R^2 분석 결과는 <Table 7>과 같다. R^2 값 0.25, 0.50, 0.75는 각각 약한, 중간, 강한 설명력을 의미하며(Hair et al., 2019), PLS-SEM 맥락에서 $R^2 \geq 0.10$이 수용 가능한 최소 기준으로 제시된 바 있다(Falk and Miller, 1992). 본 연구의 분석 결과, 모든 내생 변수의 R^2 값이 이 기준을 상회하였다.

구체적으로, 지각된 사용용이성(PEOU)의 R^2는 0.316으로 가장 높게 나타났으며, 이는 의미 이해 실패(PUF), 과정 비효율(PPI), 맥락 연속성 단절(PCD)이라는 세 가지 서비스 실패 요인이 사용용이성 평가 변량의 약 31.6%를 설명함을 의미한다. 이는 사용용이성 형성에 다양한 외생 요인이 작용함을 고려할 때, 사회과학 기반 사용자 행동 연구 맥락에서 수용 가능한 수준의 설명력을 가지는 것으로 판단된다.

지각된 유용성(PU)의 R^2는 0.220으로, 인간 지원 접근 불가능성(PHSU)과 PEOU가 유용성 평가에 대해 의미 있는 설명력을 가지는 것으로 나타났다. 전환 의도(SI)와 지속사용의도(CI)의 R^2는 각각 0.202와 0.198로, Falk and Miller (1992)가 제시한 최소 기준($R^2 \geq 0.10$)을 상회하였다.

본 연구의 R^2 수준은 기존 TAM 기반 연구들과 비교할 때 다소 보수적인 수준이다. 예를 들어, 정보시스템 지속사용 연구에서 지속사용의도의 R^2는 약 0.41로 보고된 바 있으며(Bhattacherjee, 2001), TAM2 연구에서는 지각된 유용성의 R^2가 0.40~0.60 수준으로 나타났다(Venkatesh and Davis, 2000). 이러한 차이는 본 연구가 일반적인 기술 수용 상황이 아닌 '서비스 실패'라는 부정적 맥락에 초점을 두고 있다는 점에서 기인할 가능성이 있다. 서비스 실패 상황에서는 사용자 행동 의사결정이 인지적 평가 외에도 감정 반응, 귀인 판단, 신뢰 훼손 등 다양한 심리적 요인의 영향을 받기 때문에(Smith and Bolton, 2002), 인지 평가 변수만으로는 행동 의도의 변량을 완전히 설명하기 어렵다.

따라서 본 연구의 R^2 수준은 모형의 한계라기보다는, 인지 평가 메커니즘의 기초적 작동 방식을 우선적으로 검증하고자 한 연구 설계의 결과로 이해될 수 있으며, 동시에 향후 연구에서 감정 및 귀인 변수를 통합한 확장 모형의 필요성을 시사한다.

Table 7. R^2 and adjusted R^2 values

Endogenous construct	R^2	Adjusted R^2
Continuance intention	0.198	0.194
Perceived ease of use	0.316	0.311
Perceived usefulness	0.220	0.216
Switching intention	0.202	0.200

경로의 통계적 유의성과 더불어, 각 외생 변수가 내생 변수의 설명력에 기여하는 상대적 영향력을 평가하기 위해 효과크기(f^2)를 분석하였다. Cohen (1988)에 따르면, f^2 값 0.02, 0.15, 0.35는 각각 작은(small), 중간(medium), 큰(large) 효과크기를 의미한다. 분석 결과는 <Table 8>에 제시하였다.

분석 결과, PEOU → SI 경로의 f^2 값이 0.253으로 가장 크게 나타났으며, 이는 중간 수준의 효과크기로서 모든 경로 중 가장 강력한 영향력을 보였다. 이어서 PCD → PEOU (f^2 = 0.134)와 PEOU → PU (f^2 = 0.115)가 상대적으로 높은 효과크기를 나타냈으며, 작은 효과크기와 중간 효과크기의 경계 수준에 위치하였다. 나머지 경로들(PUF → PEOU: 0.054, PPI → PEOU: 0.038, PHSU → PU: 0.064, PEOU → CI: 0.077, PU → CI: 0.068)은 모두 작은 효과크기 범위에 속하였으나, Cohen (1988)이 제시한 최소 기준($f^2 \geq 0.02$)을 상회하여 실질적인 영향력을 가지는 것으로 확인되었다.

특히 주목할 만한 발견은 다음과 같다. 첫째, PCD → PEOU 경로가 PEOU에 영향을 미치는 세 가지 서비스 실패 요인(PUF, PPI, PCD) 중 가장 큰 효과크기를 보였다는 점이다. 이는 AI 대화형 서비스의 특성상 대화 맥락의 연속성이 사용자 경험의 핵심 요소로 작용하기 때문으로 해석된다. 맥락 연속성 단절은 사용자가 동일한 정보를 반복 입력하거나 문제 상황을 다시 설명하도록 만들며, 이러한 반복적 상호작용 과정은 사용자의 인지적 부담을 직접적으로 증가시킨다. 결과적으로 이러한 누적적 인지 부담이 사용용이성 평가를 크게 저하시킬 가능성이 있는 것으로 판단된다.

둘째, PEOU → SI (f^2 = 0.253)의 효과크기는 PEOU → CI (f^2 = 0.077)보다 약 3배 이상 크게 나타났다. 이는 서비스 실패 상황에서 사용자의 시스템 사용용이성에 대한 부정적 인식이 지속사용의도(CI)보다 전환 의도(SI)에 보다 직접적인 영향을 미칠 가능성을 시사한다. 특히 실패 상황에서 사용자는 현재 상호작용 과정에서 경험하는 인지적 부담과 사용 편의성에 더욱 민감하게 반응할 수 있으며, 이러한 요인들이 전환 의도 형성에 보다 직접적으로 영향을 미칠 가능성이 있는 것으로 해석된다.

종합적으로, 효과크기 분석은 단순한 통계적 유의성을 넘어 서비스 실패 유형별 및 인지 평가 경로별 상대적 영향력의 차이를 보여준다. 또한 AI 서비스 실패 상황에서 우선적으로 관리 및 개선이 요구되는 핵심 사용자 경험 요인, 특히 대화 맥락 유지와 사용성 개선의 중요성을 시사한다.

Table 8. Effect size test f^2

	f^2
PCD → PEOU	0.134
PEOU → CI	0.077
PEOU → PU	0.115

Table 8. Effect size test f^2 (Continued)

	f^2
PEOU → SI	0.253
PHSU → PU	0.064
PPI → PEOU	0.038
PUF → PEOU	0.054
PU → CI	0.068

마지막으로 다중공선성을 검증하기 위해 VIF 값을 확인한 결과, 모든 값이 1.000~1.258 범위로 나타나 기준치인 5 이하로 확인되었다(Hair et al., 2019). 따라서 다중공선성 문제는 없는 것으로 판단된다.

4.3 Multi-group analysis results

집단 간 구조 경로의 차이를 검증하기 위해 본 연구는 다집단 분석(Multi-Group Analysis, MGA)을 실시하였다. 집단 비교에 앞서 MICOM (Measurement Invariance of Composite Models) 절차를 통해 측정불변성을 검증하였다(Henseler et al., 2016; Hair et al., 2019). 집단 구분 변수로는 성별, 연령, 그리고 AI 챗봇 사용 빈도를 사용하였다. 이 중 사용 빈도의 경우 저빈도 집단의 표본 수가 상대적으로 적어 저빈도와 중빈도 집단을 통합하여 저빈도 집단과 고빈도 집단 간 비교 분석을 실시하였다(Table 9 참고).

Table 9. Group distribution

Grouping variable	Category	N
Gender	Male	191
	Famale	184
	20~40	177
	41+	198
Usage frequency	Low frequency (Rarely + Occasionally)	249
	High frequency (frequency)	126
Total		375

본 연구는 MICOM (Measurement Invariance of Composite Models) 절차에 따라 설정 동일성(Configural Invariance)과 합성 동일성(Compositional Invariance)의 2단계 검증을 통해 측정 동일성을 확보하였다(Henseler et al., 2016). 향후 진행될 다집단 분석(MGA) 결과 해석의 근거가 되는 AI 챗봇 이용 빈도별 집단(저빈도 이용자 vs 고빈도 이용자)에 대한 MICOM 검증 결과는 <Table 10>에 제시하였다.

Table 10. MICOM results for compositional invariance across length-of-stay groups

Construct	Original correlation	5.0% quantile	Permutation *p*-value
CI	1.000	0.996	0.671

Table 10. MICOM results for compositional invariance across length-of-stay groups (Continued)

Construct	Original correlation	5.0% quantile	Permutation *p*-value
PCD	1.000	0.998	0.926
PEOU	1.000	0.999	0.426
PHSU	0.999	0.996	0.305
PPI	0.999	0.995	0.482
PUF	0.999	0.997	0.213
PU	1.000	0.999	0.732
SI	0.998	0.998	0.057

다집단 분석(Multi-Group Analysis, MGA) 결과(Table 11 참조)는 AI 서비스 실패 상황에서 사용 빈도에 따라 사용자 집단 간 차이가 존재함을 보여준다. 본 연구에서는 가설 H4(지각된 인간 지원 불가성 PHSU → 지각된 유용성 PU)에 주목하였다. 분석 결과, 해당 경로는 집단 간 유의한 차이가 있는 것으로 나타났으며, 고빈도 집단과 저빈도 집단 간 경로계수 차이는 p = 0.014로 나타나 통계적으로 유의한 차이가 확인되었다.

집단별 경로 분석 결과를 살펴보면, 고빈도 사용자 집단에서는 해당 경로의 부(-)의 영향이 유의하며 그 영향력 또한 크게 나타났다(p < 0.05). 반면, 저빈도 사용자 집단에서는 해당 경로계수가 상대적으로 낮게 나타났으며(β = -0.132), 그 영향력 또한 상대적으로 약한 것으로 나타났다. 이러한 결과는 인간 지원 불가성이 지각된 유용성에 미치는 부정적 영향이 고빈도 사용자 집단에서 더 크게 나타남을 의미하며, 사용 빈도에 따라 해당 관계에 대한 사용자 민감도가 달라질 수 있음을 시사한다.

다만, 사용 빈도가 사용자 경험 수준 및 서비스 기대 수준과 어떠한 관계를 가지는지에 대해서는 본 연구의 측정 범위에서는 직접 검증되지 않았으며, 관련 해석은 5.3절에서 보다 상세히 논의한다.

Table 11. Bootstrapped MGA results across length-of-stay groups

Path	β (Low frequency)	β (High frequency)	Two-tailed *p*-value	Decision
PHSU → PU	-0.132	-0.430	0.014	Significant

5. Discussion

5.1 Discussion of findings

본 연구 결과는 AI 서비스 실패 상황에서도 사용자 행동 의사결정이 여전히 지각된 사용용이성(PEOU)과 지각된 유용성(PU)을 중심으로 한 인지적 평가에 기반하여 이루어질 수 있음을 보여준다. 이는 서비스 실패가 사용자 행동을 직접적으로 변화시키기보다는, 시스템에 대한 인지적 평가를 통해 간접적으로 지속사용의도(CI)와 전환 의도(SI)에 영향을 미친다는 것을 의미한다. 이러한 결과는 기술수용모형(TAM)의 기본 논리를 지지할 뿐만 아니라, 그 적용 범위를 일반적인 기술 사용 상황에서 서비스 실패라는 부정적 맥락으로 확장한다는 점에서 중요한 이론적 의의를 가진다(Davis, 1989; Venkatesh and Davis, 2000).

추가 분석 결과, 서비스 실패 상황에서는 지각된 유용성(PU)보다 지각된 사용용이성(PEOU)이 사용자 행동 의도와 보다 밀접한 관련성을 가지는 것으로 나타났다. 이는 사용자가 실패 상황에서 상호작용 과정에서 요구되는 노력 비용과 인지적 부담에 보다 민감하게

반응함을 시사한다. 즉, 시스템이 추가적인 조작이나 이해 비용을 요구할 경우, 사용자는 보다 쉽게 부정적 평가를 형성하고 사용을 중단하거나 대안을 모색할 가능성이 높다.

서비스 실패 유형별 분석 결과, 지각된 맥락 연속성 단절(PCD)이 PEOU에 가장 큰 부정적 영향을 미치는 것으로 나타났다. 이는 과정적 실패(process failure)가 결과적 실패(outcome failure)보다 서비스 과정 평가에 더 큰 영향을 미친다는 기존 연구와 일관된다(Smith et al., 1999). 본 연구는 이러한 메커니즘을 AI 대화형 서비스 맥락으로 확장하여, 대화의 연속성과 맥락 유지 능력이 사용자 인지 부담을 감소시키고 PEOU 인식 형성에 중요한 역할을 한다는 점을 보여준다. 특히 맥락 단절이 발생할 경우 사용자는 동일 정보를 반복 입력해야 하며, 이는 인지적 부담을 증가시키고 시스템의 사용용이성 평가를 저하시킬 수 있다.

경로 분석 결과를 종합하면, PEOU는 SI와 CI 모두에 유의한 영향을 미치는 반면, PU는 주로 CI와의 관련성이 두드러지는 것으로 나타났다. 이는 사용자가 인지하는 상호작용 과정의 부담과 서비스 결과의 가치 평가가 서로 다른 방식으로 행동 의도에 연결될 수 있음을 시사한다. 구체적으로, 상호작용 과정에서의 인지적 부담은 전환 행동을 보다 직접적으로 유발하는 반면, 시스템의 문제 해결 능력에 대한 평가는 지속사용의도와 보다 밀접하게 관련된다. 이러한 결과는 시스템의 기능적 효용이 장기적 사용 결정에 중요한 역할을 한다는 기존 연구와도 일관된다(Bhattacherjee, 2001). 종합하면, 서비스 실패 상황에서 사용자 행동은 단일 요인에 의해 결정되기보다는, 상호작용 비용과 결과 가치가 서로 다른 경로를 통해 복합적으로 작용하는 것으로 이해할 수 있다.

본 연구는 전환 의도(SI)와 지속사용의도(CI)를 서로 대립되는 단일 선택이 아닌, 병렬적인 행동 결과 변수로 설정하였다. SI는 서비스 실패 상황에서의 즉각적인 대응 행동을 반영하는 반면, CI는 시스템에 대한 장기적 사용 의향을 반영한다. 따라서 두 변수는 동일한 연속선상에 위치한 개념이 아니라, 서로 다른 수준의 의사결정을 나타내는 변수로 해석될 수 있다(Bansal et al., 2005; Bhattacherjee, 2001). 이러한 결과는 SI와 CI가 대립적 관계라기보다는, 상황에 따라 달라지는 행동 반응임을 보여준다. 또한 두 변수 간에는 일정 수준의 관련성이 존재할 수 있으며, 특정 상황에서는 잠재적인 상충 관계(trade-off relationship)가 나타날 가능성도 있다. 향후 연구에서는 두 변수 간의 직접 경로를 포함하여 사용자 행동 의사결정 메커니즘을 보다 정교하게 분석할 필요가 있다.

마지막으로, 본 연구는 사용자 경험 수준에 따라 행동 의사결정 메커니즘이 달라질 수 있음을 보여준다. 저빈도 사용자는 주로 사용용이성에 기반하여 판단하는 경향이 있는 반면, 고빈도 사용자는 문제 해결 능력과 서비스 회복 메커니즘, 특히 인간 지원의 접근 가능성을 보다 중요하게 고려하는 경향을 보인다. 이러한 결과는 기술수용과 서비스 실패 간의 상호작용을 보다 동적인 관점에서 이해할 필요성을 시사한다.

또한 본 연구의 내생 변수 설명력(R^2 = 0.198~0.316)이 중간 수준에 머문 것은 사용자 행동 의사결정이 인지적 평가 외에도 감정 반응과 책임 귀인 등 다양한 요인의 영향을 받을 가능성을 시사한다(Smith and Bolton, 2002). 본 연구가 감정 및 귀인 변수를 실증 모형에 포함하지 않은 것은 TAM 기반 인지 평가 경로의 기초적 안정성을 우선적으로 검증하기 위한 이론적 선택이었으며, 이는 향후 '인지–감정–행동' 통합 모형 구축을 위한 실증적 기반을 마련하였다는 점에서 의의를 가진다.

5.2 External validity

본 연구는 중국 사용자를 연구 표본으로 활용하였으며, 연구 결과는 이론적 차원에서 일정 수준의 외적 타당성을 가진다고 볼 수 있다. 다만, 연구 결과의 타 문화 및 서비스 맥락으로의 일반화 가능성에 대해서는 신중한 해석이 필요하다. 중국 시장은 AI 고객 서비스의 상용화 수준과 사용자 경험 축적 측면에서 비교적 높은 수준을 보이며, 서비스 실패 상황에서 다양한 문제 유형을 관찰하기에 적절한 연구 맥락을 제공한다.

그러나 본 연구의 결과는 주로 고빈도 인간–AI 상호작용이 이루어지고, 문제 유형이 비교적 표준화된 서비스 환경에서 보다 높은 설명력을 가질 가능성이 있다. 반면, 의료·법률과 같은 고위험 또는 고도로 개인화된 서비스 환경에서는 신뢰 지각 및 위험 평가와 같은 요인이 보다 중요한 역할을 할 수 있으며, 이 경우 인지 평가 중심의 메커니즘은 상이하게 나타날 가능성이 있다.

또한 본 연구에서 지각된 맥락 연속성 단절(PCD)이 상대적으로 높은 설명력을 보인 결과는 중국의 고맥락 문화(high-context culture)

특성과 관련이 있을 가능성이 있다. 즉, 사용자들은 상호작용 과정에서 맥락의 연속성과 의도 이해에 대해 높은 기대를 가지며, 이러한 기대가 충족되지 않을 경우 인지적 부담을 보다 크게 인식할 수 있다.

따라서 불확실성 회피 성향(uncertainty avoidance)이나 개인주의–집단주의(individualism-collectivism)와 같은 문화적 차이에 따라 각 경로의 효과 크기에는 차이가 존재할 수 있다. 향후 연구에서는 다양한 문화권의 표본을 활용하여 본 연구 모형의 구조적 안정성과 경로 차이를 비교·검증할 필요가 있다.

5.3 Theoretical contributions and practical implications

본 연구의 핵심적 기여는 AI 서비스 실패 상황에서 사용자 행동을 단순한 이탈 또는 지속사용이라는 단일 결과 변수로 환원하지 않고, '지속사용–인간 지원 전환'이라는 이중 경로의사결정 문제로 재구성하였다는 데 있다. 이를 통해 본 연구는 실제 서비스 실패 상황에서 사용자가 직면하는 행동 선택의 복잡성을 보다 현실적으로 반영하며, 기존 연구가 주로 단일 행동 결과에 집중해온 한계를 보완한다.

본 연구의 이론적 의의는 기술수용모형(TAM)의 적용 범위를 전통적인 긍정적 사용 상황에서 서비스 실패라는 부정적 상황으로 확장하였다는 데 있다. 구체적으로 다음과 같은 이론적 기여를 제시할 수 있다.

첫째, TAM의 적용 범위를 부정적 서비스 맥락으로 확장하였다는 점에서 이론적 기여를 가진다. 본 연구는 서비스 실패라는 비이상적 상황에서도 '실패 경험–인지적 평가–행동 의사결정'의 심리적 경로가 유효함을 확인하였다.

둘째, 서비스 실패 유형에 따른 차별적 영향 메커니즘을 규명하였다. 연구 결과, 서로 다른 실패 유형은 서로 다른 인지 경로(PEOU 및 PU)를 통해 사용자 행동에 영향을 미치는 것으로 나타났다. 이러한 결과는 AI 서비스 실패의 작동 메커니즘에 대한 이해를 심화시키고, 향후 연구에서 보다 세분화된 실패 유형 분석의 필요성을 제시한다.

셋째, 실제 사용자 불만 데이터를 기반으로 한 서비스 실패 유형 도출 방법을 제시하였다. 본 연구는 LDA 토픽 모델링을 활용하여 실제 사용자 불만 데이터를 분석하고 서비스 실패 유형을 구성하였다. 이는 연구의 실증적 기반을 강화하였을 뿐만 아니라, 향후 AI 서비스 품질 및 서비스 실패 연구에서 보다 객관적인 변수 구성 방법을 제시하였다는 점에서 의의가 있다.

넷째, 사용 경험에 따른 사용자 행동 의사결정 메커니즘의 잠재적 차이를 탐색하였다. 본 연구의 다집단 분석(MGA) 결과는 사용 빈도에 따라 PHSU → PU 경로의 효과 크기에 유의한 집단 간 차이가 존재함을 확인하였다. 이러한 결과는 사용 빈도가 사용자 행동 의사결정 과정에서 차별적인 역할을 할 가능성을 시사한다. 다만, 본 연구는 사용자의 숙련도와 기대 수준을 직접 측정하지 않았으므로, 사용 빈도와 이러한 잠재 변수 간의 관계는 추론적 해석에 그치며, 향후 연구에서는 이를 직접 측정하여 검증할 필요가 있다. 본 연구의 실증 결과에 한정하여 잠정적으로 해석할 때, 고빈도 사용자에게서 PHSU의 부정적 영향이 더 크게 나타난 점은, 고빈도 사용자가 인간 지원의 부재(PHSU)를 PU 평가의 중요한 요인으로 인식할 가능성을 보여준다. 이러한 결과는 TAM을 지속사용 모형(Bhattacherjee, 2001)과 결합한 연구 흐름을 확장하였다는 점에서 의의를 가진다.

실무적 측면에서 본 연구는 다음과 같은 시사점을 제시한다.

첫째, 대화의 연속성과 맥락 기억 능력을 우선적으로 개선할 필요가 있다. 연구 결과, 맥락 단절은 지각된 사용용이성에 가장 큰 부정적 영향을 미치는 요인으로 나타났으며, 이는 사용자 상호작용 부담을 증가시키는 주요 원인으로 작용한다. 따라서 기업은 다중 턴 대화 상황에서의 맥락 이해 능력과 대화 연속성 유지 기능을 우선적으로 개선할 필요가 있다.

둘째, AI + 인간 상담원의 심리스(Seamless)한 혼합 회복 전략을 구축할 필요가 있다. 사용자가 AI 서비스 실패를 경험했을 때, 이를 서비스 이탈이 아닌 보완적 경로로 유도하는 것이 중요하다. 구체적으로 맥락 인식형 핸드오버(Hand-over) 시스템을 구축하여, AI가 문제를 해결하지 못해 인간 상담원으로 전환될 때 이전 대화 기록과 사용자 의도가 실시간으로 전달되도록 함으로써 사용자의 반복

설명에 따른 인지적 부하를 줄여야 한다. 원활한 상담원 전환은 단기적 전환(SI)이 장기적 이탈(CI 저하)로 이어지는 것을 방지하며, 오히려 서비스 실패를 극복하는 긍정적인 서비스 회복 도구로 기능하게 한다.

셋째, 사용자 경험 수준별 차별화된 맞춤형 리텐션(Retention) 전략이 필요하다. 사용자의 서비스 이용 빈도에 따라 실패를 인식하는 민감도가 다르므로 차별화된 접근이 필요하다. 고빈도 사용자는 인간 지원의 부재(PHSU)를 서비스 가치 자체를 부정하는 결과적 실패로 인식하며 유용성(PU) 평가가 급격히 하락하는 경향이 있으므로, 즉각적인 인간 상담원 연결옵션을 우선 제공하여 리텐션을 방어해야 한다. 저빈도 사용자는 주로 사용용이성(PEOU)에 기반하여 판단하므로, 인터페이스 내에서 조작을 돕는 가이드를 강화하여 서비스 이용 과정에서의 인지적 비용을 낮추는 데 집중해야 한다.

6. Conclusion

6.1 Summary of findings

본 연구는 AI 고객 서비스 실패 상황에서 '서비스 실패–인지적 평가–행동 의도'의 구조적 관계를 실증적으로 검증하였다. 분석 결과, 서비스 실패는 PEOU와 PU를 매개로 행동 의도에 간접적으로 영향을 미치며, PEOU의 영향력이 PU보다 강한 것으로 나타났다. 또한 사용 빈도에 따라 행동 메커니즘이 구조적으로 변화하며, 고빈도 사용자일수록 인간 지원의 부재(PHSU)를 핵심 실패 요인으로 인식하는 것으로 확인되었다.

6.2 Limitations and future research

본 연구는 AI 고객 서비스 실패 상황에서 사용자 행동 메커니즘을 분석하였다는 점에서 일정한 학문적 의의를 가지지만, 다음과 같은 연구의 한계가 존재한다. 특히, 본 연구는 서비스 실패 상황에서 작용할 수 있는 다양한 심리적 메커니즘 중 일부에 초점을 두고 있다는 점에서 이론적 범위가 일정 부분 제한될 수 있음을 우선적으로 시사한다.

첫째, 방법론적 한계이다. 본 연구는 횡단면 설문 데이터에 기반하고 있어 특정 시점에서의 사용자 인지와 행동 의도만을 반영하며, 변수 간 인과 관계를 직접적으로 검증하기에는 한계가 있다. 또한 측정된 전환 의도(SI)와 지속사용의도(CI)는 자기보고식 행동 의향으로, 실제 행동과의 괴리가 존재할 수 있다(Ajzen, 1991). 향후 연구에서는 종단 연구 설계(longitudinal design)와 행동 로그 데이터를 결합하여, 반복적인 실패 경험 상황에서의 사용자 인지·행동 변화 및 의도–행동 일치도를 보다 체계적으로 검증할 필요가 있다.

둘째, 심리적 메커니즘 통합의 한계이다. 본 연구는 TAM을 중심으로 지각된 사용용이성과 지각된 유용성이라는 인지적 평가 요인에 초점을 맞추었으나, 이는 복합적인 심리 변수들의 영향을 통제한 상태에서 인지적 평가 메커니즘의 기초적 작동 방식을 우선적으로 확인하고자 한 연구 설계상의 선택이었으며, 동시에 이론적 설명력의 범위를 일정 부분 제한하는 결과로 이어졌음을 시사한다. 서비스 실패 상황에서는 인지적 평가뿐만 아니라 감정 반응, 귀인 판단, 신뢰 훼손 등이 동시에 작용할 가능성이 있으며, 이러한 요인들이 상호작용하며 사용자 행동 의사결정에 영향을 미칠 수 있다. 또한 본 연구에서 내생 변수의 설명력이 중간 수준(R^2 = 0.198~0.316)에 머문 것은 이러한 추가 변수들이 행동 의도 설명에 기여할 가능성을 시사한다(Smith and Bolton, 2002). 따라서 향후 연구에서는 감정 변수와 귀인 메커니즘을 통합하여 '인지–감정–행동'의 통합적 분석 틀을 보다 정교하게 구축할 필요가 있다.

셋째, 실제 실패(actual failure)와 지각된 실패(perceived failure)의 구분 문제이다. 본 연구에서는 사용자 주관적 인식을 기준으로 서비스 실패를 정의하였으나, 객관적인 시스템 오류로 인한 실패(actual failure)와 기대 불일치로 인해 발생하는 주관적 실패 인식(perceived failure)은 서로 다른 경로를 통해 사용자 행동에 영향을 미칠 가능성이 있다. 따라서 향후 연구에서는 실험 연구 방법을 활용하여 서로 다른 실패 상황을 통제하고, 각 실패 유형이 사용자 인지 및 행동에 미치는 차별적 영향을 보다 명확히 규명할 필요가 있다.

넷째, 외적 타당성의 한계이다. 본 연구의 표본은 주로 중국 사용자로 구성되어 있으며, 관련 논의는 5.2절에서 상세히 다루었다. 향후 연구에서는 다양한 문화적 맥락에서 본 연구 모형을 검증하고, 불확실성 회피나 개인주의–집단주의와 같은 문화적 변수의 조절 효과를 추가적으로 분석할 필요가 있다.

종합하면, 향후 연구에서는 본 연구에서 제시한 '서비스 실패–인지적 평가–행동 의사결정' 프레임워크를 기반으로 귀인 변수와 감정 변수를 통합하고, 종단 연구와 실험 연구 방법을 결합하여 AI 서비스 실패 상황에서의 사용자 행동 메커니즘을 보다 체계적이고 심층적으로 검증할 필요가 있다.

References

Ajzen, I., The theory of planned behavior, *Organizational Behavior and Human Decision Processes*, 50(2), 179-211, 1991.

Bansal, H.S., Taylor, S.F. and James, Y.S., Migrating to new service providers: Toward a unifying framework of consumers' switching behaviors, *Journal of the Academy of Marketing Science*, 33(1), 96-115, 2005.

Bhattacherjee, A., Understanding information systems continuance: An expectation-confirmation model, *MIS Quarterly*, 25(3), 351-370, 2001.

Bitner, M.J., Evaluating Service Encounters: The Effects of Physical Surroundings and Employee Responses, *Journal of Marketing*, 54(2), 69-82, 1990.

Blei, D.M., Ng, A.Y. and Jordan, M.I., Latent Dirichlet Allocation, *Journal of Machine Learning Research*, 3, 993-1022, 2003.

Borsci, S., Malizia, A., Schmettow, M., van der Velde, F., Tariverdiyeva, G., Balaji, D. and Chamberlain, A., The chatbot usability scale: the design and pilot of a usability scale for interaction with AI-based conversational agents, *Personal and Ubiquitous Computing*, 26(1), 95-119, 2022. doi:10.1007/s00779-021-01582-9

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. and Blei, D., Reading tea leaves: How humans interpret topic models, *Advances in Neural Information Processing Systems*, 22, 2009.

Chebat, J.C. and Slusarczyk, W., How emotions mediate the effects of perceived justice on loyalty in service recovery situations: an empirical study, *Journal of Business Research*, 58(5), 664-673, 2005.

Chen, Q., Gong, Y., Lu, Y. and Tang, J., Classifying and measuring the service quality of AI chatbot in frontline service, *Journal of Business Research*, 145, 552-568, 2022.

Cho, S.R. and Jung, S.C., The Effect of AI Chatbot Service Failure on User Emotions and Switching Intention, *Journal of Korea Big Data Society*, 10(2), 129-141, 2025. doi:10.36498/kbigdt.2025.10.2.129

Cohen, J., *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.

Davis, F.D., Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly*, 13(3), 319-340, 1989.

Falk, R.F. and Miller, N.B., *A Primer for Soft Modeling, Akron*, OH: University of Akron Press, 1992.

Folkes, V.S., Consumer Reactions to Product Failure: An Attributional Approach, *Journal of Consumer Research*, 10(4), 398-409, 1984. doi:10.1086/208978

Gu, C., Zhang, Y. and Zeng, L., Exploring the mechanism of sustained consumer trust in AI chatbots after service failures: a perspective based on attribution and CASA theories, *Humanities and Social Sciences Communications*, 11, 1400, 2024. doi: 10.1057/s41599-024-03879-5

Hair, J.F., Risher, J.J., Sarstedt, M. and Ringle, C.M., When to use and how to report the results of PLS-SEM, *European Business Review*, 31(1), 2-24, 2019. doi:10.1108/EBR-11-2018-0203

Henseler, J., Ringle, C.M. and Sarstedt, M., A new criterion for assessing discriminant validity in variance-based structural equation modeling, *Journal of the Academy of Marketing Science*, 43, 115-135, 2015. doi:10.1007/s11747-014-0403-8

Henseler, J., Ringle, C.M. and Sarstedt, M., Testing measurement invariance of composites using partial least squares, *International Marketing Review*, 33(3), 405-431, 2016. doi:10.1108/IMR-09-2014-0304

Huang, M.H. and Rust, R.T., Artificial Intelligence in Service, *Journal of Service Research*, 21(2), 155-172, 2018.

Huang, Y.S. and Dootson, P., Chatbots and service failure: When does it lead to customer aggression, *Journal of Retailing and Consumer Services*, 68, Article 103044, 2022. doi:10.1016/j.jretconser.2022.103044

Hwang, J.H. and Sung, Y.J., Effects of Chatbot Anthropomorphism Level and Message Appeal Type on Purchase Intention, *Korean Journal of Consumer and Advertising Psychology*, 24(3), 245-266, 2023.

Izadi, S. and Forouzanfar, M., Error correction and adaptation in conversational AI: a review of techniques and applications in chatbots, *AI,* 5(2), 803-841, 2024. doi:10.3390/ai5020041

Jeon, S.Y. and Yoo, H.J., Analysis of the effect of chatbot service quality on chatbot use intention and comparative study between Industries, *Journal of Service Management*, 24(3), 395-418, 2023.

Kock, N., Common method bias in PLS-SEM: a full collinearity assessment approach, *International Journal of e-Collaboration*, 11(4), 1-10, 2015. https://doi.org/10.4018/IJeC.2015100101

Leo, X. and Huh, Y.E., Who gets the blame for service failures? Attribution of responsibility toward robot versus human service providers and service firms, *Computers in Human Behavior*, 113, 106520, 2020.

Liu, D., Li, C., Zhang, J. and Huang, W., Robot service failure and recovery: Literature review and future directions*, International Journal of Advanced Robotic Systems*, 20(4), 2023. doi:10.1177/17298806231191606

Ma, M. and Park, C., Effects of AI Chatbot Characteristics on Perceived Benefits and Continuous Use Intention: A Comparison of Korea and China. *The Academy of Customer Satisfaction Management*, 26(3), 41-65, 2024. 10.34183/KCSMA.26.3.3

Ozuem, W., Ranfagni, S., Willis, M., Salvietti, G. and Howell, K., Exploring the relationship between chatbots, service failure recovery and customer loyalty: A frustration-aggression perspective, *Psychology & Marketing*, 41(10), 2253-2273, 2024. doi:10.1002/mar.22051

Parasuraman, A., Zeithaml, V.A. and Berry, L.L., A conceptual model of service quality and its implications for future research, *Journal of Marketing,* 49(4), 41-50, 1985. doi:10.2307/1251430

Röder, M., Both, A. and Hinneburg, A., "Exploring the Space of Topic Coherence Measures", *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399-408, 2015.

Smith, A.K., Bolton, R.N. and Wagner, J., A model of customer satisfaction with service encounters involving failure and recovery, *Journal of Marketing Research,* 36(3), 356-372, 1999.

Smith, A.K. and Bolton, R.N., The effect of customers' emotional responses to service failures on their recovery effort evaluations and satisfaction judgments, *Journal of the Academy of Marketing Science,* 30(1), 5-23, 2002. doi:10.1177/03079450094298

Venkatesh, V. and Davis, F.D., A theoretical extension of the technology acceptance model: four longitudinal field studies, *Management Science,* 46(2), 186-204, 2000.

Venkatesh, V., Morris, M.G., Davis, G.B. and Davis, F.D., User acceptance of information technology: toward a unified view, *MIS Quarterly,* 27(3), 425-478, 2003.

Weiner, B., An attributional theory of achievement motivation and emotion, *Psychological Review,* 92(4), 548-573, 1985. doi:10.1037/0033-295X.92.4.548

Wirtz, J. and Zeithaml, V., Cost-effective service excellence, *Journal of the Academy of Marketing Science,* 46(1), 59-80, 2018. doi:10.1007/s11747-017-0560-7

Yang, C.G., A study on satisfaction and dissatisfaction of AI chatbot service, *Asia-Pacific Journal of Business Venturing and Entrepreneurship,* 17(2), 167-177, 2022.

Author listings

GUO MENGLING: iukkoi@kookmin.ac.kr
Highest degree: Master Student, Department of Smart Experience Design, Kookmin University
Position title: Graduate Student, Kookmin University
Areas of interest: Human-Computer Interaction, User Experience, AI Service, Service Failure, User Behavior

Jinho Yim: hci.yim@kookmin.ac.kr
Highest degree: PhD, Industrial Systems and Information Engineering, Korea University
Position title: Professor, Dept. of Smart Experience Design, Graduate School of Techno Design, Kookmin University
Areas of interest: Aging UX, Medical & Healthcare UX, AI-Augmented Edutainment & Smart Learning, Human Factors & Inclusive Design, Leisure Experience Design