

## Study on the Impact of XAI Explanation Levels on Cognitive Load and User Satisfaction : Focusing on Risk Levels in Financial AI Systems

No-Ah Han\*, Yoo-Jin Hwang\*, Zoon-Ky Lee\*\*

\*Student, Graduate School of Information, Yonsei University, Seoul, Korea

\*\*Professor, Graduate School of Information, Yonsei University, Seoul, Korea

### [Abstract]

In this paper, we examine the impact of XAI explanations on user satisfaction and cognitive load according to the risk levels defined in the EU AI Act. XAI aims to make the internal processes of complex AI models understandable to humans and is widely used in both academia and industry. The importance and value of XAI are continuously rising; however, there has been little research determining the necessary level of explanation according to AI system risk levels. To address this gap, we designed an experiment with 120 participants, divided into 8 groups, each exposed to one of four levels of explainability(XAI) within low-risk and high-risk financial AI systems. A quantitative approach was used to measure cognitive load, user satisfaction, mental effort, and the clarity of the material design across the different AI system interfaces. The results indicate that the amount of information in explanations significantly affects cognitive load and user satisfaction, depending on the risk level. However, the impact of the level of explanation on user satisfaction was mediated by the material design, which determined how easily the information was understood. This research provides practical, regulatory, and academic contributions by offering guidelines for determining the necessary level of explanation based on AI system risk levels.

▶ **Key words:** XAI(Explainable AI), Right to Explanation, Cognitive Load, System Risk Levels

### [요 약]

본 연구에서는 EU AI 법안에서 정의한 AI 시스템의 위험 수준에 따라 XAI 설명이 사용자 만족도와 인지 부하에 미치는 영향을 살펴보았다. XAI는 복잡한 블랙박스 모델의 내부 처리 과정을 인간이 이해하기 쉽도록 설명하는 기술로, 학계와 산업계에서 널리 활용되고 있다. 그러나 AI 시스템의 리스크 수준에 따라 필요한 설명 수준을 결정하는 연구는 미흡하였다. 이에 본 연구는 120명의 참가자를 8개의 집단으로 나누어, 금융 AI 시스템의 리스크 수준에 따른 인지 부하와 사용자 만족도를 설문문을 통해 조사하였다. 사용자의 응답 결과는 요인 분석과 판별 타당성 검토 등을 포함한 정량적 통계 기법을 통해 분석되었으며, 그 결과 위험 수준에 따라 설명의 정보량이 인지 부하와 사용자 만족도에 미치는 영향이 통계적으로 유의미하게 차이를 확인하였다. 또한, 설명 수준이 사용자 만족도에 미치는 영향은 정보의 이해를 돕는 디자인에 의해 매개된다는 사실을 발견하였다. 본 연구는 AI 시스템의 리스크 수준에 따른 필요한 설명 수준을 결정하는 지침을 제공함으로써 실무적, 제도적, 학문적 기여를 할 수 있을 것으로 기대된다.

▶ **주제어:** XAI, 설명 요구권, 인지 부하, 시스템 위험도, 사용자 만족도

- First Author: No-Ah Han, Yoo-Jin Hwang, Corresponding Author: Zoon-Ky Lee
- \*No-Ah Han (gksshd8003@naver.com), Graduate School of Information, Yonsei University
- \*Yoo-Jin Hwang (yujinhwang00@yonsei.ac.kr), Graduate School of Information, Yonsei University
- \*\*Zoon-Ky Lee (zlee@yonsei.ac.kr), Graduate School of Information, Yonsei University
- Received: 2024. 08. 07, Revised: 2024. 08. 21, Accepted: 2024. 08. 27.

## I. Introduction

OpenAI가 2022년 11월 선보인 챗봇 서비스 Chat GPT는 출시 이후 단 5일 만에 사용자 100만 명을 돌파하며, 인공지능에 대한 대중의 인식을 크게 바꾸었다[1-2]. 이는 2016년 이세돌과 알파고가 펼친 바둑 대결과 유사한 수준으로 평가받고 있으나, 알파고는 바둑이라는 단일 영역에서 뛰어난 능력을 보인 것과 달리 Chat GPT는 기억력 및 반복 작업 등 보다 다양한 측면에서 인공지능이 인간을 앞선다는 것을 증명해 상대적으로 더욱 큰 충격을 안겼다[3-4]. 또한, Chat GPT의 근간이 되는 GPT(Generative pre-trained transformer) 기술은 2018년에 공개된 이후 꾸준히 발전됨에 따라, 이러한 인공지능 패러다임의 변화는 단순한 기술의 발전이 아닌 경험의 변화임을 짐작할 수 있다[5].

그러나, 인공지능의 진보와 함께 그에 따른 부작용 또한 여러 사회적 영역에서 지속적으로 보고되고 있다. 먼저 2017년 유명 할리우드 여배우 스칼렛 요한슨과 엠마 왓슨이 딥페이크 포르노에 노출된 것이 알려지며, 인공지능이 초래하는 윤리적 문제를 점화하였다[6]. 또한, 미국 최대의 전자상거래 기업 아마존은 AI 채용 분류 시스템이 학습 이후 성별로 채용 점수를 다르게 제시하는 시스템의 편향성을 발견하여 시스템을 폐기했다[7]. 2022년 러-우 전쟁 당시에는 젤렌스키 우크라이나 대통령이 항복하는 모습을 인공지능으로 조작한 가짜 영상이 유포되어 국제 사회 내 거대한 혼란을 불러일으키기도 하였다[8]. 최근 MIT Technology review에 따르면, AI가 객관적인 판단을 내릴 것이라는 일반적인 기대와 달리, 학습 데이터에 따라 특정한 정치 성향을 보이는 것으로 밝혀졌다[9]. 이와 같은 사례들은 인공지능의 편향성과 불투명성, 악용 가능성과 같은 위험을 반증한다.

그 결과 다양한 정책 기관에서는 인공지능의 위험성을 인지하고 이를 위한 정책을 마련하고 있다. 먼저 2016년 미국 백악관은 인공지능 활용과 발전을 위한 제도적 준비를 담은 'AI 미래를 위한 준비'라는 보고서를 발표했다[10]. 이를 필두로 2018년 EU 역시 개인정보 보호 강화, 의사결정 자동화 시스템의 설명 가능성(Explainability)을 설명하는 GDPR(General Data Protection Regulation)을 제시하였다[11]. 그리고 2019년에는 OECD, G20 등과 같은 국제단체에서도 인공지능 윤리 가이드를 발표했다. 국내에서도 2020년 정부가 인공지능 윤리기준을 발표하며, AI에 대한 제도적 제한을 가시화하였다.

이렇듯, 국내외를 막론하고 인공지능이 초래하는 사회적 위험성은 줄어들며 인공지능이 가지는 유용성을 더욱 활용하기 위한 노력이 이어지고 있다. 이 중에서도 최근에는 인공지능이 가지는 구조적 한계를 극복하기 위한 기술의 필요성이 증대되고 있다. 일반적으로 인공지능에는 내부 처리 과정이 복잡한 심층 학습(Deep Learning) 모형이 활용됨에 따라, 사용자가 개별 의사결정과정을 명확히 이해하는 것은 거의 불가능에 가깝다. 이러한 이유로 인공지능은 그 의사결정의 메커니즘을 일일이 파악할 수 없어 블랙박스(Black Box) 모형이라고 일컬어지기도 한다[12].

이에 대한 해답으로 떠오른 XAI(eXplainable AI, 설명 가능한 AI) 기술은 인공지능의 의사결정과정을 인간이 이해할 수 있도록 설명하는 것을 목적으로 하며, 학계, 산업계에서 크게 주목받고 있다. 특히 최근에는 인공지능의 투명성을 확보하기 위해 알고리즘의 설명 요구권 보장이 강조되고 있으며, 그에 따라 XAI의 중요성이 더욱 높아지고 있다[13].

그러나, XAI는 사용자들이 부담하는 위험 정도에 따라 그 중요도가 달라질 수 있으나, 지금까지 AI 시스템의 리스크 수준에 따라 필요 설명 수준을 결정한 연구는 미흡하였다. 또한, 사용자의 인지 부하에 따라 AI 시스템의 결과에 대한 설명이 필요하다고 느끼는지의 정도도 충분히 달라질 수 있으나, 관련 연구는 대부분 AI 시스템의 설명 수준 변화에 따른 신뢰도, 만족도, 작업 부하 등을 살펴보는 것에서 그쳤다.

따라서, 본 연구는 AI 시스템에 대한 XAI 설명 정도가 인지 부하를 매개하여 사용자 만족도에 미치는 영향을 설문조사 기법을 통해 살펴본다. 또한, 소비자들이 부담하는 위험 수준에 따라 설명 정도의 영향력이 달라지는지 비교하고자 한다. 본 연구의 세부 연구 질문은 Table 1과 같다.

Table 1. Research Questions

1	Do XAI explanation levels affect the cognitive load perceived by users?
2	Does user satisfaction vary according to the cognitive load perceived by users?
3	Do XAI explanation levels affect user satisfaction through cognitive load perceived by users?
4	Do RQ1 to RQ3 vary depending on the level of risk borne by consumers?

## II. Preliminaries

### 1. Related works

#### 1.1 XAI Explanation Level

XAI 기술의 유형은 설명 대상이 특정 결과인지, 모형 전반인지에 따라 전역 설명(Global Explanation)과 지역 설명(Local Explanation)으로 구분된다. 전역 설명은 다시 대리 모형(Surrogate Model)을 이용한 방법과 PFI(Permutation Feature Importance)와 같이 변수들의 중요도로 모형을 이해하는 방법으로 나눌 수 있다. 또한, 지역 설명에는 LIME, SHAP과 같이 머신러닝 알고리즘을 설명하는 기술로 활용도가 높은 기법들이 있다[14].

선행 연구의 결과를 종합하여 XAI 설명 수준을 구분하면 Table 2과 같이 나타낼 수 있다. 이 중 설명 수준 4는 결과 설명에 비교군 정보를 추가로 제공한 형태로, 금융업에서 빈번히 활용되는 ‘벤치마킹 가격 전략’을 XAI 설명에 활용한 것이다.

Table 2. Categories of XAI Explanation levels

Explanation levels		Characteristics
1	No XAI Explanation	Only AI system results exposed (No XAI Explanation)
2	Model Explanation (PFI, Surrogate Model)	Features used by the AI system and their importance exposed (Global XAI)
3	Result Explanation (LIME, SHAP)	Feature impact on the AI system's output exposed (Local XAI)
4	Result Explanation + Relative Comparison	Provide comparative group information at Explanation Level 3 (XAI + Additional Information)

#### 1.2 Human Factors in XAI

XAI 인간 요인은 인공지능 모형에 대한 ‘설명’에 따라 달라지는 인간의 의사결정, 신뢰도, 작업 수행 능력 등을 일컫는 용어이다. 인간 요인 연구는 주로 설명 유형, 여부, 복잡도 등에 따라 사용자 만족도, 인지 부하, 유용성, 신뢰성 등이 어떻게 변하는지 관찰하며, 주요 연구는 설명 유형과 복잡도에 따라 사용자가 체감하는 인지 부하가 다르게 나타난 결과를 제시하였다[15-19]. 또한, 모형에 AI 문해력 및 작업 친숙도, 인구통계 요인, 인지 필요도 등과 같은 사용자 요인을 조절 변수 및 통제 변수로 활용한 연구도 활발하게 수행되었다[18,20].

#### 1.3 Cognitive Load

인지 부하(Cognitive Load)는 사용자가 특정 작업을 수행하며 겪는 정신적, 육체적 스트레스 또는 부담을 의미한다[21]. 이러한 인지 부하는 내재적 부하, 외재적 부하, 관

련 부하로 나뉜다. 먼저 내재적 부하(Intrinsic Load)는 과제 복잡도에 따라 발생하는 부하로, 어렵거나 생소한 개념을 이해하려고 하는 경우 높아질 수 있다. 반면 외재적 부하(Extraneous Load)는 이해해야 할 대상의 디자인이나 제시된 방식에 따라 발생하는 부하로, 주로 복잡한 도표나 설명을 접할 때 높아질 수 있다. 마지막으로 관련 부하(Germane Load)는 기존 지식과 새로운 지식을 통합할 때 발생하는 부하이다[21]. 외재적 부하는 설명이나 표현 방법을 수정하여 낮출 수 있어, 사용자 만족도와 깊이 연관되어 있다[19]. 또한 외재적 부하는 디자인을 간결하게 구성하거나 분할 원칙 등을 이용하여 낮출 수 있는 것으로 알려져 있다[22].

이러한 인지 부하는 설문을 이용한 주관적 평가로 측정할 수 있다. 주관적 평가에서 인지 부하는 주로 다섯 가지 하위 변수로 측정한다. 첫째, 정신적 노력(Mental Effort)은 작업 수행 중 사용자가 느끼는 정신적인 노력을 의미한다[23]. 둘째, 작업 난이도(Task Difficulty)는 사용자가 작업 수행 중 느낀 작업 난이도를 측정한다[24]. 셋째, 학습 과정(Learning Process) 변수는 학습 과정 중 학습자의 경험과 인식을 평가한다[25]. 넷째, 학습 효과(Learning Outcome)는 학습 과정 중 학습자의 경험과 인식을 평가하는 변수로, 학습을 통해 학습자 얼마나 많은 것을 배우고 느꼈는지를 측정한다[26]. 다섯째, 자신감(Confidence) 또는 자기 평가(Self-evaluation)는 학습 활동이 학습 결과에 미친 영향을 평가한다.

마지막으로 인지 부하를 직접적으로 측정하는 변수는 아니나, 인터페이스 내에 디자인이 포함된 경우 ‘디자인(Material Design)’ 변수가 인지 부하 측정에 유용하게 활용될 수 있다. 이 변수는 해당 시스템 디자인이 사용자 이해에 미치는 영향도를 측정하며, 표현된 설명 내용 구조와 배치 등과 같은 요소들이 평가에 활용된다. XAI 관련 연구에서는 이 중 주로 ‘정신적 노력’과 ‘디자인’이 채택되어 사용되고 있다.

인지 부하는 교육학에서 오랫동안 활발하게 연구되어온 개념이나, AI 시스템에서 사용자가 느끼는 인지 부하에 대한 연구는 상대적으로 부족했다. 인지 부하는 XAI 시스템 사용자의 인지 부담을 세부 항목으로 나누어 측정할 수 있으며, 각 세부 영향 요인 따라 시스템 설계를 달리 할 수 있다는 점에서 유용성이 높다. 이에, 인지 부하 이론을 활용할 경우 시스템 위험 수준에 보다 적합한 XAI 시스템을 구축할 수 있다. 특히 XAI와 같이 사용자에게 AI 시스템 결과를 납득시키는 것이 주요한 목표인 경우, ‘이해’의 걸림돌이 되는 인지 부하를 낮추는 방향으로 시스템을 설계

하는 것이 유용한 대안이 될 수 있다.

또한, 인지 부하는 AI 시스템이 내린 결과에 대한 근거가 부족할 때 높아질 수 있으며, 반대로 결과에 대한 설명이 필요 이상으로 복잡할 때도 높아질 수 있다는 특성이 있어, 연구의 필요성이 높다. 그러나, 관련 연구 대부분은 AI 시스템의 설명 수준 변화에 따른 신뢰도, 만족도, 작업 부하 등을 살펴보는 것에 그쳤다[16-19]. 따라서 본 연구는 AI 시스템에 대한 설명의 정보량이 사용자 만족도에 미치는 영향을 인지 부하를 매개하여 살펴보고자 한다.

### 1.4 Risk Level

XAI는 사용자들이 부담하는 위험 정도에 따라 그 중요도가 달라질 수 있으나, 지금까지 AI 시스템의 리스크 수준에 따라 필요 설명 수준을 결정한 연구는 미흡하였다. 이에, 본 연구는 소비자들이 부담하는 위험 수준에 따라 XAI 설명 정도의 영향력에 차이가 있는지 살펴보고자 한다. 이에, 위험 수준을 구분하기 위한 명확한 기준을 정립하기 위해, 2024년 EU가 발간한 ‘Artificial Intelligence Act: Council and Parliament strike a deal on comprehensive AI rules’ 보고서를 활용하였다.

본 보고서가 정의한 AI 시스템의 위험 수준은 Table 3와 같이 구분할 수 있는데, 해당 분류 기준에 따르면 대출 심사, 신용 평가 등을 제공하는 AI 시스템은 High-Risk 시스템에 속하며, 스팸 필터, 상품 추천 등과 같은 AI 시스템은 ‘Minimal-Risk’ 시스템으로 분류할 수 있다[27]. 본 연구에서는 금융 산업의 대표적인 High-Risk AI 시스템인 대출 심사와 대표적인 Minimal-Risk 시스템인 상품 추천을 각각 소비자들이 느끼는 높은 위험 수준(Risk-High), 낮은 위험 수준(Risk-Low)으로 설정했다[28].

Table 3. Risk Levels of AI Systems

Risk Level	Definition
Unacceptable	AI Systems that pose clear threats to safety, livelihood, and human rights
High	AI Systems that pose significant threats and must meet stringent pre-market validation before release
Limited	AI systems with specific transparency obligations
Minimal	AI systems that pose little to no risk to citizens' rights or safety

## III. The Proposed Scheme

### 1. Research Framework

연구는 AI 시스템에 대한 XAI 설명 정도가 인지 부하를

매개해 사용자 만족도에 미치는 영향을 소비자들이 부담하는 위험 수준에 따라 비교·분석하기 위하여, 선행연구를 참고해 Fig. 1과 같은 연구모형을 구성하였다.

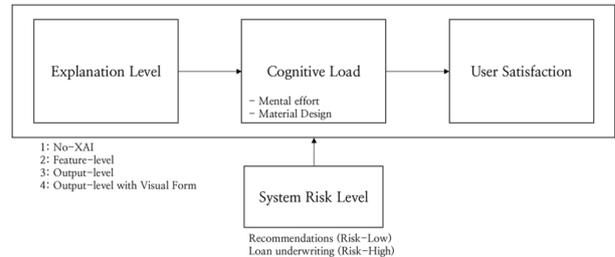


Fig. 1. Research Framework

## 2. Research Hypothesis

### 2.1 Level of XAI

연구에서 XAI 설명 수준은 설명에 포함된 정보량과 비례한다. 따라서 XAI 설명 수준의 변화는 설명 내용을 이해하기 위한 정신적 노력과, 설명 표현을 이해하기 위한 디자인 부하에 영향을 미칠 것이다. 이에 따라 설정된 가설 1,2는 다음과 같다.

- H1: XAI의 설명 수준에 따라 정신적 노력이 달라질 것이다.
- H2: XAI의 설명 수준에 따라 디자인 부하가 달라질 것이다.

### 2.2 Cognitive Load

인지 부하가 높다는 것은 고객이 결과를 이해하는 데 더 큰 노력이 필요하다는 뜻이다. 특히 상품 추천에서 고객이 결과를 이해하기 어렵다면 만족도가 떨어질 수 있으나, 가독성 높은 디자인으로 결과를 쉽게 이해할 수 있을 경우 만족도는 높아질 수 있다. 이에 따른 가설 H3, H4는 다음과 같다.

- H3: 정신적 노력은 고객 만족도에 부(-)의 영향을 미친다.
- H4: 디자인은 고객 만족도에 정(+)의 영향을 미친다.

### 2.3 Risk Level

XAI의 설명 수준과 인지 부하, 그리고 인지 부하와 고객 만족도 간의 영향은 소비자가 부담하는 위험 수준에 따라 다를 수 있다. 예를 들어, 대출 심사와 같은 고위험 AI 시스템의 경우 시스템 결과가 사용자에게 미치는 영향이 상대적으로 크다. 비록 가상 시나리오지만, 목돈이 필요한 상황에 대출 거절은 소비자에게 민감하게 작용하기 때문이다. 이에 대출 심사 결과에 대한 상세한 설명이 없는 경우, 특히 거절을 당한 경우 심사 결과만으로는 쉽게 결과를 받아들이기 어려울 수 있다. 이에 심사 결과의 이유를 추론하는 과정에서 정신적 인지 부하가 높아질 수 있다.

반면, 상품 추천과 같이 시스템 위험 수준 자체가 낮은 경우 사용자가 상품이 추천된 이유를 명확히 알지 못해도 어떠한 비용도 발생하지 않는다. 하지만 설명에 포함된 정보량이 많은 경우 사용자는 설명을 이해하는 데에 많은 인지 자원을 소비하기 때문에 인지 부하가 높아질 수 있다. 따라서 가설 H5와 H6은 다음과 같다.

H5: XAI 설명 수준이 정신적 노력에 미치는 영향은 시스템 위험 수준에 따라 다를 것이다.

H6: XAI 설명 수준이 디자인에 미치는 영향은 시스템 위험 수준에 따라 다를 것이다.

### 2.4 User Satisfaction

앞서 언급했듯, 인지 부하가 사용자 만족도에 미치는 영향은 사용자가 부담하는 위험 수준에 따라 다를 수 있다. 고위험 시스템은 설명으로 결과를 납득하기 어려운 경우 정신적 부하가 높아질 수 있으며, 손실에 민감한 인간 심리 특성으로 시스템 만족도와 더욱 직접 연관될 수 있다. 즉, 고위험 시스템 사용자의 만족도는 결과와 설명 간 타당성에 기초하여 결정될 것이다. 반면, 저위험 시스템 사용자는 결과에 대한 설명이 사용자에게 미치는 영향이 적기 때문에, 설명을 이해하고자 하는 유인 또한 낮을 수 있다. 이에 따라 사용자 만족도는 설명 타당성보다는 설명이 얼마나 이해하기 쉽게 표현되어 있는가를 나타내는 디자인에 더 큰 영향을 받을 수 있다. 이에 따라 설정된 가설 H7과 H8은 다음과 같다.

H7: 정신적 노력은 시스템 위험 수준에 따라 사용자 만족도에 미치는 영향이 다를 것이다.

H8: 디자인은 시스템 위험 수준에 따라 사용자 만족도에 미치는 영향이 다를 것이다.

### 3. Situational Questions

본 연구는 선행 연구 검토 결과를 바탕으로 연구 맥락과 목적에 맞게 설문지를 일부 변형하여 사용했다. 선행 연구 중 일부는 영문으로 문항이 구성되어 있어 번역을 수행하였고, 사전 설문 및 응답자와 인터뷰를 통해 혼란을 가져오는 문항은 제외하였다. 최종적으로 구성된 측정 항목 및 설문 문항은 Table 4에 제시되었으며, 각 항목은 5점 리커트 척도로 측정하였다.

Table 4. Situational Questions

Variables		Situation Question	Ref
User satisfaction	USS1	이 AI 시스템은 나의 요구를 적절히 충족시킬 수 있다고 생각한다.	[29] [30] [31]
	USS2	이 AI 시스템은 나의 요구를 효율적으로 충족시킬 수 있을 것이다.	
	USS3	전반적으로 나는 이 AI 시스템에 만족할 것이다.	
	USS4	AI 시스템 결과를 이해한 이후 만족감을 느꼈다.	
Mental effort	MEF1	AI 시스템의 결과를 이해하는 것은 인지적으로 까다로운 작업이었다.	[31] [32]
	MEF2	AI 시스템 결과의 목적에 따라 그 내용을 이해하는 것이 어려웠다.	
	MEF3	AI 시스템 결과는 이해 난이도가 높았다.	
	MEF4	AI 시스템 결과에 나와있는 개념을 이해하는데 어려움을 느꼈다.	
Material Design	MTD1	AI 시스템 결과의 화면구성은 읽기에 편했다.	[33]
	MTD2	AI 시스템 결과 화면은 사용하기 쉽도록 만들어졌다.	
	MTD3	AI 시스템 결과 화면은 내용의 핵심을 이해하기 쉽도록 구성되었다.	
	MTD4	AI 시스템 결과의 내용 구조를 쉽게 파악할 수 있었다.	

## 4. Experiment Interface

### 4.1 Scenario Description

Table 5는 각 화면 상단의 시나리오 설명을 나타낸다.

Table 5. Scenario Description

Work	Interface
Recommendations (Risk-Low)	AI 예금상품 추천 응답자는 현재 대한은행에 예금상품을 가입하기 위해 AI 예금상품 추천 시스템을 이용하였으며, 다음과 같은 추천상품을 안내받은 상황이라 가정합니다. 화면을 잘 보고, 다음 질문에 응답해주세요.
Loan underwriting (Risk-High)	AI 대출심사 안내 응답자는 현재 대한은행에 대출심사를 신청한 상태이며, 다음과 같은 심사결과를 안내받은 상황이라 가정합니다. 화면을 잘 보고, 다음 질문에 응답해주세요.

### 4.2 User Interface

Table 6은 실험을 위한 전체 상세 인터페이스를 나타낸다. 먼저 설명 수준이 같은 경우 정확한 비교를 수행하기 위해 같은 인터페이스로 구성하였다. 또한 설명 수준이 높아짐에 따라 노출 정보가 많아져, 설명 수준 3부터는 '결과'와 '설명'을 나누어 표현했다.

인터페이스는 AI 시스템의 설명 수준이 높아짐에 따라 결과에 대한 정보가 추가되도록 구성했다. 먼저 Level 1은 XAI가 적용되지 않은 경우로, 결과 외에 인터페이스상 포함된 정보는 없도록 하였다. Level 2의 경우, 어떤 정보가 중요하게 쓰였는지에 대한 설명이 포함되어 있다. 설명 예시를 살펴보면, 사용자가 최근에 조회했던 상품의 조회 횟수 정보와 우대금리 등의 정보를 활용했음을 명시하고 있다. 이를 통해 사용자는 ‘행복 플러스 예금’ 상품이 어떤 근거로 추천되었는지에 대한 이유를 어느 정도 파악할 수 있을 것이다.

Level 3은 어떤 요인이 얼마나 영향을 미쳤는가에 대한 정보를 1,000점 만점의 점수를 이용하여 표현했다. 해당 인터페이스는 신용평가나 FDS 시스템 등에서 주로 사용되는 방식이며, LIME, SHAP과 같은 지역 설명 모형을 이용할 경우에도 활용할 수 있다. 마지막으로 Level 4의 경우, 비교군을 두어 해당 점수가 얼마나 높은지 혹은 얼마나 낮은지를 체감할 수 있도록 인터페이스를 제작하였다.

Table 6. User Interface

Level		User Interface															
Risk	XAI																
Low	1	-AI 예금상품 추천 결과 안내- Description AI가 고객님의 행복 플러스 예금 상품을 추천합니다.															
	2	-AI 예금상품 추천 결과 안내- Description AI가 고객님의 행복 플러스 예금 상품을 추천합니다. (설명: 이 상품추천에는 고객님의 관련 상품 조회 횟수, 우대금리 충족 여부 등의 정보가 중요하게 쓰였습니다)															
	3	-AI 예금상품 추천 결과 안내- (결과) AI 시스템이 고객님의 행복 플러스 예금 상품을 추천합니다. (설명) <table border="1"> <tr> <td>평가 활용 항목</td> <td>평가 점수*</td> </tr> <tr> <td>관련 상품 조회 3회</td> <td>+ 200점</td> </tr> <tr> <td>관련 상품군 문의 1회</td> <td>+ 150점</td> </tr> <tr> <td>최근 예금 상품 만기 도래</td> <td>+ 180점</td> </tr> </table>	평가 활용 항목	평가 점수*	관련 상품 조회 3회	+ 200점	관련 상품군 문의 1회	+ 150점	최근 예금 상품 만기 도래	+ 180점							
	평가 활용 항목	평가 점수*															
관련 상품 조회 3회	+ 200점																
관련 상품군 문의 1회	+ 150점																
최근 예금 상품 만기 도래	+ 180점																
4	-AI 예금상품 추천 결과 안내- (결과) AI 시스템이 고객님의 행복 플러스 예금 상품을 추천합니다. (설명) <table border="1"> <tr> <td>평가 활용 항목</td> <td>비교 그룹</td> <td>내의 영향</td> <td>평가 점수*</td> </tr> <tr> <td>관련 상품 조회 3회</td> <td>0.5회</td> <td>3회</td> <td>+ 200점</td> </tr> <tr> <td>관련 상품군 문의 1회</td> <td>0회</td> <td>1회</td> <td>+ 150점</td> </tr> <tr> <td>최근 예금 상품 만기 도래</td> <td>1도래</td> <td>도래</td> <td>+ 180점</td> </tr> </table>	평가 활용 항목	비교 그룹	내의 영향	평가 점수*	관련 상품 조회 3회	0.5회	3회	+ 200점	관련 상품군 문의 1회	0회	1회	+ 150점	최근 예금 상품 만기 도래	1도래	도래	+ 180점
평가 활용 항목	비교 그룹	내의 영향	평가 점수*														
관련 상품 조회 3회	0.5회	3회	+ 200점														
관련 상품군 문의 1회	0회	1회	+ 150점														
최근 예금 상품 만기 도래	1도래	도래	+ 180점														

High	1	-AI 대출심사 결과 안내- Description AI로 추정된 신용점수는 780점으로, 심사 결과 대출 거절되었습니다.															
	2	-AI 대출심사 결과 안내- Description AI로 추정된 신용점수는 780점으로, 심사 결과 대출 거절되었습니다. (설명: 대출 심사에는 연체 이력, 제2금융권 대출이력 등이 중요하게 쓰였습니다)															
	3	-AI 대출심사 결과 안내- (결과) AI로 추정된 신용점수는 830점으로, 심사 결과 대출 거절되었습니다. (설명) <table border="1"> <tr> <td>평가 활용 항목</td> <td>평가 점수*</td> </tr> <tr> <td>최근 3년 연체 횟수</td> <td>-100점</td> </tr> <tr> <td>최근 3년 제2금융 대출 횟수</td> <td>-50점</td> </tr> <tr> <td>연체 적정 단위 기간</td> <td>-20점</td> </tr> </table>	평가 활용 항목	평가 점수*	최근 3년 연체 횟수	-100점	최근 3년 제2금융 대출 횟수	-50점	연체 적정 단위 기간	-20점							
	평가 활용 항목	평가 점수*															
최근 3년 연체 횟수	-100점																
최근 3년 제2금융 대출 횟수	-50점																
연체 적정 단위 기간	-20점																
4	-AI 대출심사 결과 안내- (결과) AI로 추정된 신용점수는 830점으로, 심사 결과 대출 거절되었습니다. (설명) <table border="1"> <tr> <td>평가 활용 항목</td> <td>비교 그룹</td> <td>내의 영향</td> <td>평가 점수*</td> </tr> <tr> <td>최근 3년 연체 횟수</td> <td>0.1회</td> <td>2회</td> <td>-100점</td> </tr> <tr> <td>최근 3년 제2금융 대출 횟수</td> <td>0.5회</td> <td>3회</td> <td>-50점</td> </tr> <tr> <td>연체 적정 단위 기간</td> <td>1년</td> <td>3년</td> <td>-20점</td> </tr> </table>	평가 활용 항목	비교 그룹	내의 영향	평가 점수*	최근 3년 연체 횟수	0.1회	2회	-100점	최근 3년 제2금융 대출 횟수	0.5회	3회	-50점	연체 적정 단위 기간	1년	3년	-20점
평가 활용 항목	비교 그룹	내의 영향	평가 점수*														
최근 3년 연체 횟수	0.1회	2회	-100점														
최근 3년 제2금융 대출 횟수	0.5회	3회	-50점														
연체 적정 단위 기간	1년	3년	-20점														

5. Procedure

앞서 구성한 측정 항목과 인터페이스에 기초하여 8개에 독립된 설문 링크를 구성하고, 인지 부하 측정을 위한 정량 연구 방법의 하나인 설문조사를 실시하였다. 설문은 약 2주에 걸쳐서 수집되었으며, 응답자는 10~50대까지 다양한 연령대에서 총 120개 표본을 획득했다. 표본은 8개의 인터페이스별로 15명씩 배치되었다. 응답자에 대한 설명은 Table 7로 제시하였다.

Table 7. Participants Information

Demographic Variables		
Age Group	10-19	0.8%
	20-29	40.0%
	30-39	45.8%
	40-49	6.7%
	50 >	6.7%
Gender	Male	45.0%
	Female	55.0%
Control Variables		
XAI Explanation Level	1: No-XAI	25.0%
	2: Feature-level	25.0%
	3: Output-level	25.0%
	4: Output-level with Visual Form	25.0%
System Task Type	Recommendations (Risk-Low)	50.0%
	Loan underwriting (Risk-High)	50.0%

## IV. Results

### 1. Exploratory Factor Analysis

측정 항목의 탐색적 요인분석은 SPSS 29를 이용하여 수행했다. 탐색적 요인분석의 회전 방법은 Kaiser 정규화가 있는 베리맥스를 이용했다. 또한 주성분 분석의 경우, 고정된 요인을 3개로 설정하고 수행하였다. 탐색적 요인분석 수행 결과는 Table 8과 같다. 탐색적 요인분석 대상 요인에서 '적재치(Factor Loading Score) > 0.6'을, 다른 요인에서는 '적재치 < 0.4'를 기준으로 수행했다[34]. 그 결과 12개 측정 항목 중 'MTD4'가 다른 항목에서 요인 적재치가 0.46이고, 또한 가장 높은 요인의 요인 적재치 역시 0.59로 기준을 만족하지 못해 제거하였다.

Table 8. Exploratory Factor Analysis Results

	Factor 1	Factor 2	Factor 3
USS1	0.90	-0.11	0.21
USS2	0.91	-0.15	0.12
USS3	0.91	-0.13	0.18
USS4	0.80	-0.07	0.18
MEF1	-0.21	0.84	-0.23
MEF2	-0.20	0.84	-0.24
MEF3	-0.04	0.88	-0.10
MEF4	-0.06	0.86	-0.32
MTD1	0.12	-0.14	0.89
MTD2	0.18	-0.26	0.85
MTD3	0.26	-0.36	0.75
MTD4	0.46	-0.28	0.59

### 2. Convergent Validity Analysis

요인분석으로는 선택된 측정 항목이 측정 대상 변수와 개념적으로 일치하는 정도를 확인하는 수렴 타당성 검증을 수행하였다. 수렴 타당성 검증의 기준은 각 측정 항목의 표준화 적재치(Standard Loading)가 0.7 이상인지, 각 변수의 AVE(Average Variance Extracted)가 0.5 이상인지로 설정하였다[35]. 수렴 타당성 결과는 Table 9에 제시되었다.

모든 측정 항목의 표준화 적재치는 0.7보다 큰 것으로 나타났고, t-value 역시 일반적으로 활용되는 기준인 1.96 (유의수준 5%)을 크게 웃돌아 통계학적으로 유의한 것을 확인하였다. 아울러 모든 변수에서 CR과 Cronbach's Alpha 값이 0.7을 웃돌았으며, AVE 값도 기준치인 0.5를 넘었다.

Table 9. Convergent Validity Analysis Results

Item	Std. Loading	t-value	AVE	CR	Cronbach's α
USS1	0.93	69.08	0.83	0.95	0.93
USS2	0.93	47.96			
USS3	0.94	84.11			
USS4	0.83	19.58			
MEF1	0.92	30.96	0.80	0.94	0.92
MEF2	0.92	24.66			
MEF3	0.84	11.57			
MEF4	0.89	14.86			
MTD1	0.89	30.32	0.81	0.93	0.88
MTD2	0.92	48.85			
MTD3	0.89	32.73			

### 3. Discriminant Validity Analysis

판별 타당성은 서로 다른 개념을 설명하는 측정 항목 간의 개념이 잘 구분되는지를 의미한다. 일반적으로 판별 타당성은 상관계수를 이용해 변수 간의 상관계수가 0.6 미만이고, AVE의 제곱근 값이 0.7 이상이면서 동시에 AVE 값이 다른 변수와의 상관계수보다 큰 경우 만족한다고 본다. Table 10은 본 연구에서 활용된 측정 변수들에 대한 AVE, 상관계수를 산출한 결과이다.

모든 변수의 상관계수가 0.6 미만으로 기준치보다 낮았으며, 각 변수의 AVE 제곱근 값은 0.89~0.91로 앞서 설명한 기준을 모두 만족하는 것을 알 수 있다.

Table 10. Discriminant Validity Analysis Results

	Mean(S.D.)	MEF	MTD	USS
MEF	2.35(1.07)	0.89		
MTD	3.58(0.94)	-0.55**	0.90	
USS	3.14(1.02)	-0.31**	0.50**	0.91

### 4. Hypothesis Testing

#### 4.1 Low-Risk

저위험 그룹에 대한 가설 검정 결과는 Fig. 2와 같다.

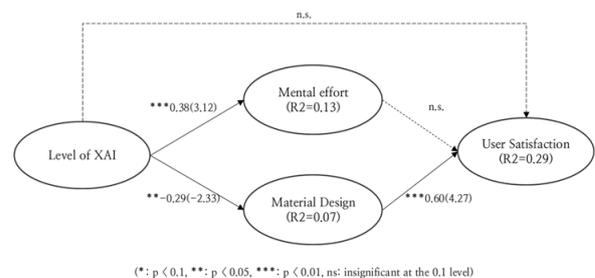


Fig. 2. Low-Risk Group Model Results

결과를 살펴보면, 설명 수준은 인지 부하(정신적 노력, 디자인) 모두에 유의수준 5%에서 유의한 영향을 미쳤다. 또한 디자인의 경우 사용자 만족도에 유의한 영향을 미치는 것으로 나타났다. 따라서 'H1: XAI의 설명 수준에 따라 정신적 노력이 달라질 것이다, H2: XAI의 설명 수준에 따라 디자인 부하가 달라질 것이다, H4: 자료 디자인은 고객 만족도에 정(+)의 영향을 미친다'는 모두 유의수준 5%에서 채택되었으며, 'H3: 정신적 노력은 고객 만족도에 부(-)의 영향을 미친다'는 기각되었다.

금융 상품 추천 시스템의 경우 결과에 대한 XAI 설명 수준(정보량)이 높아짐에 따라 정신적 부하가 높아지는 경향을 보였다. 또한 정신적 부하의 증감은 사용자 만족도와 유의한 관계를 보이지 않았다. 반면, 시스템 디자인은 XAI 설명 수준과 사용자 만족도를 매개했다. XAI의 설명 수준이 높아질수록 사용자는 디자인을 복잡하고 어렵게 느꼈다는 의미로, 설명의 정보량이 많아지면서 발생하는 자연스러운 현상으로 해석할 수 있다. 또한 사용자는 디자인 요인을 이해하기 쉽다고 느꼈을 때 만족도가 높아지는 경향을 보였다.

이 결과를 통해 얻을 수 있는 시사점은 크게 두 가지이다. 첫째, 저위험 AI 시스템에서 사용자는 원인과 결과의 논리 구조를 중요하게 생각하지 않는다는 점이다. XAI 설명 수준이 높아질수록 사용자는 AI 시스템의 인과를 보다 명확히 파악할 수 있다. 하지만 저위험 시스템에 풍부한 설명은 사용자가 디자인 요인을 복잡하다고 느끼게 만들어 사용자 만족도를 낮췄을 뿐, 사용자가 시스템 결과의 논리구조를 정확히 이해함으로써 얻는 만족감은 없었다.

4.2 High-Risk

고위험 그룹에 대한 가설 검정 결과는 Fig. 3과 같다.

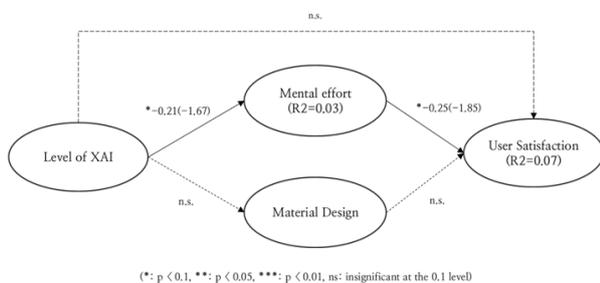


Fig. 3. High-Risk Group Model Results

먼저 가설 'H1: XAI의 설명 수준에 따라 정신적 노력이 달라질 것이다'는 유의수준 10%에서 채택되었다. 분석 결과 유의수준 5%에서 설명 수준이 높아짐에 따라 사용자가

느끼는 정신적 부하는 감소하는 경향을 보였다(H3). 반면 정신적 노력이 높아짐에 따라 사용자 만족도는 유의하게 감소하는 경향을 보였다.

고위험 AI 시스템의 경우 XAI 설명 수준이 높아짐에 따라 정신적 노력이 유의하게 감소했다. 이 현상은 설명이 없는 경우, 사용자가 결과만으로 원인을 추론하기 위해 더 많은 인지 자원을 필요로 했기 때문으로 보인다. 고위험 시스템의 결과는 사용자 재무 상황에 직접 영향을 미칠 수 있다. 이에 따라 설명 수준이 높아질수록 사용자는 심사 결과에 대한 보다 많은 정보를 얻을 수 있고, 어떤 원인으로 이 결과가 도출되었는지를 이해할 수 있어 인지 부하가 점점 낮아지는 것으로 분석된다. 이 결과는 저위험 시스템과 상반되는데, 고위험 시스템에서 사용자는 더욱 정교한 설명 제공할 때 인지 부하가 낮아지며 그에 따른 사용자 만족도도 높아지는 것으로 추정되었다.

4.3 Group Comparison

소비자가 부담하는 위험 수준에 따라 XAI 설명이 인지 부하와 인지 부하를 매개하여 사용자 만족도에 미치는 영향이 서로 다름을 확인하기 위해 고위험 시스템과 저위험 시스템 간 그룹 비교를 수행하였으며, 결과는 Table 11로 제시하였다.

Table 11. Group Comparison Analysis Results

Hypothesis	t-value	Results
H5: XAI Explanation Level → Mental Effort	3.67	Accept
H6: XAI Explanation Level → Design	-1.37	Reject
H7: Mental Effort → User Satisfaction	1.61	Reject
H8: Design → User Satisfaction	2.41	Accept

고위험과 저위험 AI 시스템 간의 그룹 비교는 각 경로 계수에 대한 t-검정을 통해 수행했다. 그 결과 가설 'H5: XAI 설명 수준 → 정신적 노력(인지 부하), H8: 디자인(인지 부하) → 사용자 만족도'는 유의수준 5%에서 채택되었으며, 나머지 두 가설 H6과 H7은 기각되었다. 따라서 XAI 설명 수준이 정신 노력에 미치는 영향과 디자인 요소가 사용자 만족도에 미치는 영향은 AI 시스템의 위험 수준에 따라 유의하게 다른 것으로 나타났다.

V. Conclusions

본 연구는 AI 시스템에 대한 XAI 설명 정도가 인지 부하를 매개하여 사용자 만족도에 미치는 영향을 파악하였

으며, 소비자가 부담하는 위험 수준에 따라 설명 정도의 영향력이 달라지는지 비교하였다.

연구 결과, 위험 수준에 따라 설명의 정보량이 인지 부하와 사용자 만족도에 미치는 영향이 통계적으로 유의하게 차이를 알 수 있었다. 즉, 사용자는 AI 시스템 리스크 수준에 따라 결과에 대한 설명을 다르게 인식했다. 구체적으로는, 상품 추천과 같은 저위험 AI 시스템은 설명 수준이 높아질수록 사용자가 느끼는 인지 부하가 비례하여 증가했다. 하지만 설명 수준이 사용자 만족도에 미치는 영향은 표현 방법이 얼마나 이해하기 쉬운가를 나타내는 디자인을 매개하여 결정되었다.

또한, 사용자는 저위험 AI 시스템에서 이해하기 쉬운 표현에 만족감을 얻는다는 사실을 발견하였다. 이는 금융 상품 추천과 같은 저위험 AI 시스템에서는 설명에 포함된 정보량을 높이는 방식 대신, 간단한 설명과 이해하기 쉬운 디자인이 사용자 만족도를 더욱 높일 수 있다는 것으로 해석된다. 저위험 시스템의 경우, 사용자가 '설명'이 많은 것을 선호하지 않아 발생한 현상으로 추정되며, 그에 따라 설명 수준이 높아짐에 따라 '과잉 설명'으로 인한 불만이 높아진 것으로도 볼 수 있다. 반면, 대출 심사와 같은 고위험 금융 AI 시스템은 더 구체적이고 이해할 수 있는 충분한 설명이 사용자 만족도를 높였다.

이 연구는 AI 시스템 위험도에 따른 설명 수준과 증점 사항을 제공하여 실무와 학술 양 측면에서 유용하다. 실무적으로는 금융 AI 시스템의 위험 수준에 따라 필요한 XAI 기술을 보다 최적화된 방법으로 적용할 수 있도록 도울 수 있다. 정책적으로는 최근 금융혁신기획단 등에서 추진하고 있는 금융 분야 AI 적용 방안에 대한 유용한 참고 자료로 활용되어, 위험 수준에 따른 적절한 설명 수준을 확인함으로써 무분별한 규제나 방지로 인한 정책 부작용을 예방하는 데 도움을 줄 수 있다.

학술적으로는 금융 도메인에서 XAI 시스템 설명하고자 하는 AI 시스템의 위험 수준에 따라 인지 부하가 사용자 만족도에 미치는 영향이 차이를 밝혔다는 점에서 이전에 없던 연구로 학술적 기여가 기대된다.

그러나, 본 연구는 다음과 같은 한계를 가진다. 첫째, 데이터 수집 방식의 한계이다. 연구는 실험 참가자 개개인에 대한 XAI 결과가 아닌 가상 시나리오에 기초한 화면을 이용하여 수행되었다. 그 결과 응답자 본인의 결과가 아니기 때문에 실제 대출 심사나 상품 추천에서 느꼈을 감정이나 생각을 충분히 담지 못했을 수 있다. 이러한 이유로 제시한 인터페이스를 제대로 확인하지 않고 응답했을 가능성

도 존재한다.

둘째, 고위험 AI 시스템에 대한 연구 모형은 '정신적 노력 → 사용자 만족도'로 가는 경로가 통계적으로 유의했으나 설명력( $R^2=0.07$ )이 낮았다. 따라서 고위험 시스템의 경우, 다양한 추가 변수를 문헌 연구를 통해 찾아 추가 고려하는 과정이 필요해 보인다. 셋째, 디자인 요인에 대한 검토가 부족했다. 연구에서 사용한 인터페이스 구조는 주로 금융권에서 사용하는 구조를 차용했다. 하지만, 이 구조가 충분한 학술 검토를 통해 구성한 것이 아님에 따라, 다소 편향이 존재할 가능성이 있다.

## ACKNOWLEDGEMENT

This work was supported by ICONS(Institute of Convergence Science), Yonsei University.

## REFERENCES

- [1] Johnson, A., Here's What To Know About OpenAI's ChatGPT-What It's Disrupting And How To Use It., <https://www.forbes.com/sites/ariannajohnson/2022/12/07/heres-what-to-know-about-openais-chatgpt-what-its-disrupting-and-how-to-use-it/>
- [2] Mollick, E., Chat-GPT is a tipping point for AI., <https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai>
- [3] Zastrow, M., South Korea Trumpets \$860-million AI Fund after AlphaGo Shock., <https://www.nature.com/articles/nature.2016.19595>
- [4] Taecharunroj, V., "What Can Chat-GPT Do? Analyzing Early Reactions to the Innovative AI Chatbot on Twitter." *Big Data and Cognitive Computing*, Vol. 7, No. 1, pp 35, February 2023. DOI 10.3390/bdcc7010035
- [5] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I., "Improving language understanding by generative pre-training.", 2018
- [6] Hern, A., AI Used to Face-swap Hollywood Starts into Pornography Films, <https://www.theguardian.com/technology/2018/jan/25/ai-face-swap-pornography-emma-watson-scarlett-johansson-taylor-swift-daisy-ridley-sophie-turner-maisie-williams>
- [7] Dastin, J., Amazon scraps secret AI recruiting tool that showed bias against women, <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/>
- [8] Filimowicz, M. (Ed.), "Deep Fakes: Algorithms and Society.

- Routledge.”, Routledge, 2022
- [9] Melissa Heikkilä., AI language models are rife with different political biases, <https://www.technologyreview.com/2023/08/07/1077324/ai-language-models-are-rife-with-political-biases/>
- [10] National Science and Technology Council Committee on Technology., PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE., [https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/NSTC/preparing\\_f\\_or\\_the\\_future\\_of\\_ai.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_f_or_the_future_of_ai.pdf)
- [11] Goodman, B., & Flaxman, S., “European Union regulations on algorithmic decision-making and a right to explanation.”, *AI magazine*, Vol. 38, No. 3, pp 50-57, October 2017, DOI 10.1609/aimag.v38i3.2741
- [12] Von Eschenbach, W. J., “Transparency and the black box problem: Why we do not trust AI.”, *Philosophy & Technology*, Vol. 34, No. 4, pp 1607-1622, September 2021, DOI 10.1007/s13347-021-00477-0
- [13] Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D., “Expanding explainability: Towards social transparency in ai systems.”, In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp 1-19, Yokohama, Japan
- [14] Arya, V., Bellamy, R. K., Chen, P. Y., Dhurandhar, A., Hind, M., Hoffman, S. C., ... & Zhang, Y., “One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques.” *arXiv*, September 2019, DOI 10.48550/arXiv.1909.03012
- [15] Nimmo, R., Constantinides, M., Zhou, K., Quercia, D., & Stumpf, S., User Characteristics in Explainable AI: The Rabbit Hole of Personalization?, In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp 1-13, Honolulu, USA
- [16] Herm, L. V., “Impact of explainable ai on cognitive load: Insights from an empirical study.”, *arXiv*, April 2023, DOI 10.48550/arXiv.2304.08861
- [17] Sanneman, L., Tucker, M., & Shah, J. A., An Information Bottleneck Characterization of the Understanding-Workload Tradeoff in Human-Centered Explainable AI, In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, pp 2175-2198, Rio de Janeiro, Brazil
- [18] Jansen, A., Leborgne, F., Wang, Q., & Zhang, C., Contextualizing the “Why”: The Potential of Using Visual Map As a Novel XAI Method for Users with Low AI-literacy, In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp 1-7, New York, USA
- [19] Hu, P. J. H., Hu, H. F., & Fang, X., “Examining the mediating roles of cognitive load and performance outcomes in user satisfaction with a website.” *Mis Quarterly*, Vol. 41, No. 3, pp 975-988, September 2017
- [20] Panigutti, C., Beretta, A., Giannotti, F., & Pedreschi, D., Understanding the impact of explanations on advice-taking: a user study for AI-based clinical Decision Support Systems, In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp 1-9, New Orleans, USA
- [21] Sweller, J., “Cognitive load during problem solving: Effects on learning.” *Cognitive science*, Vol. 12, No. 2, pp 257-285, December 2003, DOI 10.1016/0364-0213(88)90023-7
- [22] Mayer, R. E., & Moreno, R., “Nine ways to reduce cognitive load in multimedia learning”, *Educational psychologist*, Vol. 38, No. 1, pp 43-52, June 2010, DOI 10.1207/S15326985EP3801\_6
- [23] Paas, F. G.M “Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach”, *Journal of educational psychology*, Vol. 84, No. 4, pp 429, 1992, DOI 10.1037/0022-0663.84.4.429
- [24] Ayres, P, “Using subjective measures to detect variations of intrinsic cognitive load within problems.”, *Learning and Instruction*, Vol. 16, No. 5, pp 389-400, October 2006, DOI 10.1016/j.learninstruc.2006.09.001
- [25] Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G., “Development of an instrument for measuring different types of cognitive load.”, *Behavior Research Methods*, Vol. 45, No. 4, pp 1058-1072, April 2013
- [26] Kalyuga, S., “The expertise reversal effect. In *Managing cognitive load in adaptive multimedia learning*”, IGI Global, pp 58-80, DOI 10.4018/978-1-60566-048-6.ch003
- [27] EY Asset, AI Act Political Agreement Overview, [https://www.ey.com/en\\_be/tax/tax-alerts/2023/eu-artificial-intelligence-act](https://www.ey.com/en_be/tax/tax-alerts/2023/eu-artificial-intelligence-act)
- [28] Financial Services Commission, Guide to AI Development and Utilization in the Financial Sector, <https://www.fsc.go.kr/comm/getFile?srvcId=BBSTY1&upperNo=78235&fileTy=ATTACH&fileNo=8>
- [29] Rana, N. P., Dwivedi, Y. K., Williams, M. D., & Weerakkody, V., “Investigating success of an e-government initiative: Validation of an integrated IS success model”, *Information systems frontiers*, Vol. 17, pp 127-142, May 2014
- [30] Han, C. W., Ham, S. N., & Moon, H. Y., “Influence of the characteristics of the O2O food delivery platform service on the trust, satisfaction and reuse intention”, *Journal of Hospitality and Tourism Studies*, Vol. 21, No. 1, pp 115-128, 2019
- [31] Wilkinson, D. L., “Effects of prior knowledge and spatial ability on learning outcomes and cognitive load associated with rich and lean multimedia presentation”, University of Kansas, 2004
- [32] Miller, C. D., “Demystifying Aesthetics: An Examination of the Relationships and Effects of Emotional Design on Learner Cognitive Load and Task Performances.”, University of Minnesota, 2007
- [33] Whelan, R. R., “The multimedia mind: Measuring cognitive load in multimedia learning.”, New York University, 2006
- [34] Chulmo Koo, Hee-Woong Kim, Youhee Joun, “A Study on the Explorative or Exploitive Use of Smartphone : Focusing on IS Continuance and Habit Theory”, *The Journal of Society for*

e-Business Studies, Vol. 16, No. 3, pp 67-91, Aug 2011, DOI  
10.7838/jsebs.2011.16.3.067

- [35] Seddon, P. B., "A respecification and extension of the DeLone  
and McLean model of IS success.", Information systems research,  
Vol. 8, No. 3, pp 240-253, Sep 1997, DOI 10.1287/isre.8.3.240

## Authors



No-Ah Han received a B.S. degree in Information Statistics from Yonsei University in Wonju, Korea in 2016. He is currently pursuing an M.S. degree at Graduate School of Information, Yonsei University, Korea.

His current interests encompass LLM, data augmentation and HCI.



Yoo-Jin Hwang received a B.S. degree in Economics and Business Administration from Konkuk University in Seoul, Korea in 2023. She is currently pursuing her M.S. degree at Graduate School of Information, Yonsei

University, Korea. Her current interests encompass data analytics and machine learning in finance.



Zoon-Ky Lee received the B.S. degree in Computer Science from Seoul National University, Korea in 1985, M.S. in Social Psychology from Carnegie Mellon University, USA in 1991, and Ph.D in Management

Informatics from Southern California University, USA in 1999. He is currently a Professor in the Graduate School of Information at Yonsei University, Korea. Zoon-ky Lee is interested in Big Data Analytics, Digital Transformation, and Open Collaboration.