

챗봇과의 대화에서 의도적인 부정확한 정보 제공에 대한 영향 요인 연구: 온라인 데이팅 서비스 이용 상황에서¹⁾

A Study of Influencing Factors for Intentional Inaccurate Information Provision in Conversations with Chatbots: In the Context of Online Dating Services

곽찬희 (Chanhee Kwak)

강남대학교 인공지능융합공학부 조교수²⁾

이준영 (Junyeong Lee)

충북대학교 경영정보학과 부교수³⁾

민진영 (Jinyoung Min)

중앙대학교 산업보안학과 부교수⁴⁾

최한별 (HanByeol Stella Choi)

명지대학교 경영정보학과 조교수⁵⁾

〈 국문초록 〉

챗봇은 편리한 기능뿐만 아니라 친근하고 인간적인 경험을 제공하는 대화형 소통 도구로서 활용 범위가 크게 확대되고 있다. 쌍방향 소통이 가능한 챗봇은 사용자와 정보를 주고받으며 다양한 과업을 수행할 수 있는데, 이 때 사용자는 의도적으로 부정확한 정보를 제공하기도 한다. 본 연구는 온라인 데이팅 서비스 이용 상황에서 대화형 에이전트에 대한 사회적 실재감, 개인 정보 제공에 대한 지각된 위험, 알고리즘에 대한 신뢰를 주요한 영향 요인으로 고려하여, 이들이 부정확한 정보 제공 의도에 미치는 영향을 파악하고, 그 효과가 에이전트 유형에 따라 변하는지를 함께 살펴보았다. 이를 위해, Amazon Mechanical Turk(MTurk)으로부터 데이터를 수집하고 구조방정식 모형을 이용하여 분석하였다. 분석 결과, 부정확한 정보 제공 의도에 관련된 요인들 간의 유의미한 관계를 확인하고 나아가 이들이 에이전트 유형에 따라 달라지는 것을 실증적으로 확인하였다. 본 연구 결과를 통해 온라인 환경에서 부정확한 정보 제공 행위에 대한 학술적인 시사점 및 이런 의도를 감소시키기 위한 챗봇 설계 방안 등의 실무적 시사점을 도출하였다. 또한, 온라인에서의 부정확한 정보로 인해 발생할 수 있는 결과에 대한 윤리적인 시사점도 논하였다.

주제어: 챗봇, 알고리즘에 대한 신뢰, 사회적 실재감, 지각된 위험, 부정확한 정보 제공

1) 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019S1A5A2A03041910)

2) 제1저자, chk@kangnam.ac.kr

3) 교신저자, junyeong.lee@cbnu.ac.kr

4) 공동저자, jymin@cau.ac.kr

5) 공동저자, hbschoi@mju.ac.kr

1. 서론

초창기 챗봇(Chatbot)은 주로 단순한 정보를 제공하는 고객 지원 서비스에 제한적으로 사용되었으나, 인공지능 기술의 발전과 디지털 전환 추세에 따라 그 활용 범위가 빠르게 확대되고 있다(Nguyen et al., 2022). 오늘날 챗봇은 사용자의 다양한 요청에 대응하고 개인화된 추천을 제공하며 레스토랑이나 호텔 예약을 지원하는 등 다양한 상황에 활용되고 있다(Park et al., 2022). 단순한 고객 응대 시스템과 비교하면, 챗봇은 인간과 동등한 수준의 자연어 구사 및 이해가 가능하며, 대화 맥락 이해 능력을 갖췄다는 특징이 있다. 아울러 단방향 소통이 아닌 쌍방향 소통을 지원한다는 점 또한 챗봇과 기존 자동화 고객 응대 도구를 가르는 큰 차이점이다. 이를 기반으로 챗봇은 단순히 문제 해결이나 도움 요청 대응에서 나아가, 사용자와 정보를 주고받으며 다양한 과업을 수행할 수 있다. 예를 들어, 레스토랑이나 호텔의 예약을 위해 일정과 개인 정보를 제공하면 챗봇이 예약을 확정하거나(Hsu et al., 2023), 항공권 예약 및 결제를 수행할 수 있다(Sundar & Kim, 2019).

그러나 챗봇 사용자들이 항상 정확하고 진실된 개인정보를 입력하는 것은 아니다. 웹사이트 가입 시 자신에 대해 거짓된 정보를 입력하거나 사실을 부풀리고 과장된 설명을 하는 등, 사용자들은 의도적으로 부정확한 정보를 제공할 수도 있다(Sharabi, 2021). 이렇게 부정확한 정보가 온라인 공간에 누적되면 온라인 정보의 진실성과 신뢰성에 대한 문제가 커질 수 있으며, 이는 과거와 현실의 호도 및 왜곡을 야기하거나 잘못된 정보에 기반한 부정확한 의사 결정, 나아가 사회 혼란까지도 일으킬 수 있다. 따라서 부정확한 정보 제공의 원인을 파악하고, 챗봇과 같은 디지털 상호작용이 이러한 현상에 어떤 영향을 미칠 수 있는지에 대한 연구가 필요하다. 하지만 부정확한 정보 제공 혹은 거짓 정보 제공에 대한 기존 연구들은

사용자가 해당 행동을 하는 요인들에 주로 집중하였고, 디지털 에이전트라는 정보 수신자에 대해 고려하지 못하였다(DePaulo & Kashy, 1998; Logsdon & Patterson, 2009). 반면, 소통 대상의 유형에 따라 정보 제공의 정확성 차이를 고려한 연구들은 대부분 유형(예. 컴퓨터 대 인간)의 효과에 집중하고 해당 효과가 나타나는 메커니즘은 고려하지 않았다(Lucas et al., 2014; Turner et al., 2005). 이들을 같이 고려함으로써, 본 연구는 디지털 전환 시대에 더욱 중요해질 챗봇과 같은 대화형 에이전트 이용 상황에서 부정확한 개인 정보를 제공하는 행동에 영향을 미치는 요인을 찾고, 대화를 나누는 상대가 인간인 경우와 비교하여 이에 대한 이해를 넓히는데 주요한 역할을 할 것이다. 이를 위해 본 연구는 다음과 같은 연구 문제를 제안한다:

RQ1: 사용자의 부정확한 정보 제공 의도에 영향을 미치는 요인은 무엇인가?

RQ2: 대화형 에이전트의 유형에 따라 사용자의 부정확한 정보 제공 의도에 미치는 요인들의 영향이 달라지는가?

위 연구문제에 답하기 위해, 본 연구는 온라인 데이팅 서비스 이용 상황에서 사용자의 의도적인 부정확한 정보 제공 행동에 영향을 미치는 요인을 살펴보고자 한다. 온라인 데이팅 서비스는 챗봇이 활용되며, 부정확한 정보 제공이 빈번하게 일어나는 온라인 환경이다(Hancock et al., 2004; Wang, 2023). 인터넷을 매개로 사람들을 연결해 파트너를 매칭해주는 온라인 데이팅 서비스 사용을 위해선 사용자의 자발적 정보 제공이 필수적이다. 이 때, 사용자는 자신을 더 매력적인 존재로 표현하기 위해 사실과 차이가 있는 정보를 제공하기도 한다(Hancock et al., 2004; Peng, 2020). 타인에게 더 매력적인 대상으로 보이고자 하는 개인의 동기와 사회적으로 더 나은 이미지를 가진 인물로 스스로를 표현하고자 하는 욕망이 자신을 원래 모습보다 더 낫게 표현하려 하고, 부정확한 정보 제공으로 이어진다(Sharabi & Caughlin, 2019). 온라인

인 데이팅 서비스의 부정확한 정보는 사소한 정보의 오류로 보일 수도 있으나, 앞서 언급한 대로 부정확한 정보가 온라인 공간에 누적되면 온라인 정보의 진실성과 신뢰성에 대한 문제가 확대될 수 있다. 이에 따라, 온라인 데이팅 서비스 이용 상황에서 부정확한 개인정보 제공에 영향을 미치는 요인들을 탐구하고, 나아가 이러한 행동과 대화형 에이전트의 유형 사이의 관계를 심층적으로 탐구할 것이다.

이를 위해, 대화형 챗봇 에이전트와 인간 에이전트 유형 각각과 대화하는 시나리오의 설문을 설계한 뒤 Amazon Mechanical Turk(MTurk)에서 데이터를 수집했다. 총 267개(대화형 챗봇 에이전트 180개, 인간 에이전트 87개)의 수집된 데이터를 바탕으로 PLS(Partial Least Square) 구조방정식 모형을 분석하였고, 분석 결과에 기반하여 다음과 같은 시사점을 도출하였다. 학술적 시사점으로 본 연구는 온라인 환경에서의 부정확한 정보 제공 행위, 대화형 에이전트와의 소통 시 개인의 행동 대한 연구 분야의 학술적인 이해를 넓혔다. 이에 더해, 실무적인 시사점으로 사회적 실재감을 줄이거나 의인화 정도 조절, 개인정보 제공이 갖는 위험 감소 방안 등 부정확한 정보 제공 의도를 감소시키기 위한 챗봇 설계 방안을 제안하고 서비스 상황에 맞는 대화형 에이전트 형태 설정의 필요성을 논하였다. 마지막으로, 온라인에서의 부정확한 정보로 인해 발생할 수 있는 결과에 대한 윤리적인 시사점도 논하였다.

2. 문헌 연구

2.1. 챗봇

챗봇은 자연어를 사용해 인간과 대화할 수 있는 소프트웨어 기반 에이전트로, 인간과 유사한 대화 스타일을

사용하여 사용자의 적극적인 소통 참여를 유도하고 다양한 기능을 제공함은 물론, 궁극적으로 사용자 만족도 향상을 목표로 하는 기술이다 (Diederich et al., 2022). 최근 자연어 처리 기술의 발달로 대화의 맥락 이해나 장기적인 관계 형성이 가능해지고, 다양한 챗봇 개발 프레임워크가 공개되며 챗봇의 활용 범위 또한 빠르게 확장되고 있다(Hsu et al., 2023; 박은영, 2024). 챗봇은 다채로운 과업에 활용될 수 있는데, 상품 구매나 회원가입 등을 위해 개인정보를 챗봇으로 제공받아 처리하거나, 사용자의 필요에 맞는 개인 맞춤형 서비스 제안 및 추천 시스템을 챗봇 형태로 구현할 수도 있다. 이에 더해 편리한 유지 보수, 인건비 절감, 기능 확장의 용이성 측면에서 챗봇은 매력적인 고객 응대 기술로 자리잡았다.

챗봇 연구들은 주로 챗봇과 사람의 소통 사이의 역학 관계에 집중하여 진행되었다. 챗봇과 사람 사이의 상호작용은 사람들이 컴퓨터를 매개로 한 대화에서 상대가 인간이 아니라는 것을 알면서도 사람들 사이에서 보편적으로 통용되는 상호작용과 관련된 사회적 기준을 기대한다고 가정한다(Nass et al., 1994). 이에, 많은 연구자들이 이러한 상호 작용에 영향을 끼치는 챗봇 요소들에 대해 탐구하였다. 예를 들어, 챗봇의 인간 유사성(Han, 2021), 시각적인 모습(Go & Sundar, 2019), 어투의 격조(Shams et al., 2024), 개인화 정도(Behera et al., 2024) 등이 챗봇 이용에 중요한 요소로 연구되었다.

하지만 사용자들이 컴퓨터를 매개로 한 소통에서 상대의 정체성과 무관하게 동일한 상호작용을 하는 것은 아니다. 즉, 챗봇의 다양한 요소에 더해 에이전트 유형은 특정한 효과를 일으키거나 챗봇 사용에 있어 중요한 역할을 맡을 수 있다. 예를 들어, 동일한 환경에서 같은 소통을 수행하더라도 대화 대상이 기계화된 소프트웨어인 경우와 인간인 경우 상호 작용 행태에 차이가 있을 수 있다(Park et al., 2022; 진평위, 이정, 2021). 이에 최근 챗봇 연구들은 동일한 커뮤니케이션 환경에서 에이전트

의 유형이 미치는 영향에 대해서 다양하게 탐구하였는데, 결제를 위한 개인정보 제공(Sundar & Kim, 2019), 설문 조사 수행(Kim et al., 2019), 일상 대화 교환(Hill et al., 2015) 등에서 에이전트에 따라 사용자의 행동이 달라짐을 보였다. 그러나, 기존 대부분의 연구는 일반적인 챗봇 이용 상황을 상정하여 연구를 진행하였다. 하지만, 실제 인터넷 환경에서 사용자들의 챗봇과 같은 AI 기반 서비스 활용은 다양한 형태로 나타나며, 특히 정보 교환 상황에서 항상 정확하고 진실된 정보제공이 보장되지는 않는다(Biener & Waerber, 2024). 따라서, 챗봇 이용 상황에서 부정확한 정보 제공과 에이전트의 형태가 어떤 역할 관계를 갖는지 밝히는 연구가 필요하다.

2.2. 온라인에서 자기 공개 행동(self-disclosure) 및 부정확한 정보 제공 행동

자기 공개 행동(self-disclosure)은 일반적으로 개인이 자신에 대해 타인에게 드러내는 정도로 정의되며(Kim & Dindia, 2011), 친구 및 파트너 간의 관계부터 고용주 및 조직과의 관계에 이르기까지 다양한 관계를 맺을 때 기본이 되는 행동이다(Cozby, 1973; Sprecher & Hendrick, 2004). 자기 공개와 같은 개인정보 공개 정도를 결정할 때 소통의 맥락은 큰 역할을 하는데(Acquisti et al., 2015; John et al., 2011; Ozdemir et al., 2017), 커뮤니케이션의 일부인 사회적 요인과 맥락이 자기 공개 정도를 결정하는 주요 요인이기 때문이다(Barasch & Berger, 2014). 예를 들어, 자기 공개는 온라인 커뮤니티 결성을 촉진시키고 사회적 수용도를 향상시키는 역할을 수행한다(Berg & McQuinn, 1989; Posey et al., 2010). 또한 사회적 의사소통의 욕구 중 하나로 자기 공개를 활용하기도 한다(Wilson et al., 2012). 이러한 자기 공개를 이해하는 주요한 사회적 요인으로 사회적 실재감이 있다. 사회적 실재감은 컴퓨터를 매개로 하는 커뮤니케이션 환경에서 사람

들의 경험을 촉진하는 핵심 요인 중 하나로, 기존 연구에서 자기 공개와 밀접하게 같이 다루어져 왔다(e.g., Kim & Song, 2016; Lee et al., 2024).

온라인 환경에서는 또한 개인정보가 처음 공유된 후에도 오랫동안 지속되거나 누적될 수 있기에, 이에 따른 개인정보 보호 위험도 비례하여 증가할 수 있다(Xu & Dinev, 2022). 처음에는 무해하거나 유익해 보이는 정보도 의도하지 않은 사람이 접근하거나 활용 및 오용하는 경우도 있고, 시간이 지남에 따라 사람들이 예상하지 못했거나 동의하지 않은 방식으로 집계 및 분석되어 향후 예상치 못한 부정적인 결과를 초래할 수 있다(Kwak et al., 2022; Nicol et al., 2022; Xie & Kang, 2015). 이를 막기 위해 사람들은 자기 공개 행동을 제한하거나 거짓으로 부정확한 정보를 제공하기도 한다(Sannon et al., 2018).

거짓 행동과 관한 연구들을 살펴보면, 개인은 거짓말을 성공적으로 숨길 가능성이 높을 때 거짓말을 할 가능성이 더 높으며(Logsdon & Patterson, 2009), 사회적 실재감이 높을수록 거짓말과 관련된 심리적 비용이 증가하고 거짓말을 하려는 의지가 감소한다(DePaulo & Kashy, 1998; Van Zant & Kray, 2014). 이처럼, 부정확한 정보 제공 혹은 거짓 정보 제공에 대한 기존 연구들은 사용자가 해당 행동을 하는 요인들에 주로 집중하여 연구를 진행해 왔지만, 디지털 에이전트라는 정보 수신자를 고려하지 못하였다.

Short et al. (1976)은 사회적 실재감이 단순히 커뮤니케이션 매체의 특징이 아니라, 커뮤니케이션 파트너에 따라 사회적 실재감의 차이가 발생할 수 있다고 주장했다. 이에 따라, 거짓 행동에 관한 일부 연구들은 상대방이 인간 혹은 컴퓨터인지에 따라 다른 행동을 보일 수 있다는 점을 밝혔다. 예를 들어, Lucas et al. (2014)는 개인은 진실을 말하는 것이 오히려 부정적으로 비춰질 때 컴퓨터보다 사람에게 거짓말을 더 많이 한다는 사실을

밝혔고, Turner et al. (2005)은 컴퓨터 소프트웨어 면접관이 아닌 사람이 질문할 때 개인이 약물 사용에 대해 거짓 말을 할 가능성이 더 높다는 사실을 발견했다. 하지만, 이러한 연구들도 대부분 유형(예. 컴퓨터 대 인간)의 효과에 집중하고 해당 효과가 나타나는 메커니즘은 고려하지 않았다(Lucas et al., 2014; Turner et al., 2005). 본 연구는 이러한 두 가지 기존 연구의 한계를 같이 고려하여 살펴보고자 한다.

2.3. 온라인 데이팅 서비스에서 자기 공개 행동 및 부정확한 정보 제공

온라인 데이팅은 사용자들의 이상형, 관심사, 지역, 나이, 호감도 등의 개인정보를 제공받아 이를 바탕으로 데이팅 상대를 매칭해주는 서비스를 제공한다(박윤주, 2012; 안효진, 김승인, 2017). 따라서 개인 정보 제공 및 공유는 온라인 데이팅 서비스 이용 시 필수적인 부분이다. 온라인 데이팅 서비스에서 개인 정보 제공은 다른 개인 정보 제공과 구별되는 몇 가지 특징을 갖는다. 먼저, 정보 비대칭성 및 자기표현의 자유성이다. 개인의 자기 정보 공개는 타인 혹은 서비스 제공자에 의해 간접적으로 평가되지만, 제공된 정보의 진실성을 확인하기 쉽지 않다(Simmons & Lee, 2020). 따라서 정보의 진실성은 사용자의 선의에 따라 결정되고, 사용자는 자기 표현에 대한 절대적인 자유도를 갖는다(He et al., 2024). 이로 인해 온라인 데이팅 서비스에서 개인에 대한 다소간의 과장 혹은 부정확한 정보가 빈번히 등장한다(Hancock et al., 2004). 이는 타인에게 더 매력적인 상대로 보이게 하는 개인의 동기 또한 반영된다(Peng, 2020).

이에 더해, 온라인 데이팅 서비스 이용 시엔 개인 정보 제공에 대한 우려 또한 존재한다. 불특정 다수에게 세부적인 개인 정보를 공개하는 행위는 필연적으로 개인 정보 제공의 우려를 야기한다(Kelly & Teevan, 2003; Nichols,

1997; Oard & Kim, 1998). 이런 위험을 회피하기 위해 사용자들은 자세하고 정확한 정보 제공 대신, 사실과는 다소 거리가 있는 정보를 제공하기도 한다. 예를 들어, 자신의 사는 곳을 의도적으로 다르게 기입하거나, 타인의 사진을 제공하는 경우가 있다(Willis et al., 2023). 이에 따라, 본 연구는 온라인 데이팅 서비스 이용 상황에서 부정확한 정보 제공 행동에 대한 메커니즘을 이해하고, 에이전트의 형태에 따라 이들이 어떻게 달라지는지 알아보고자 한다.

3. 가설

3.1. 알고리즘에 대한 신뢰

점점 더 디지털화되는 세상에서 알고리즘은 많은 서비스에 내장되어 있으며 검색 엔진 결과부터 쇼핑이나 스트리밍 플랫폼의 맞춤형 제품 추천까지 일상의 많은 부분과 밀접하게 관련되어 있다. 대부분의 사람들에게 '알고리즘'이라는 용어는 더 이상 낯설지 않으며, 사람들은 알고리즘이 처리하는 프로세스에 대해 특정한 믿음과 기대를 가지게 되었다(김경애, 류방란, 2019). 따라서 사람들은 알고리즘 기반 시스템이 정확한 프로세스를 수행하고, 개인화된 경험 및 서비스 제공하며, 과업의 효율성을 극대화할 것으로 기대한다.

알고리즘에 대한 신뢰는 불확실하고 취약한 상황에서 알고리즘 에이전트에 대한 개인의 태도라고 정의할 수 있다(Lee, 2018). 개인이 기술 플랫폼의 정확성, 일관성, 투명성을 얼마나 신뢰하는지는 개인이 다양한 서비스를 이용하는 방식 및 그들의 행동에 영향을 미칠 수 있는 결정적인 요소다(Lee & See, 2004). 알고리즘의 기능을 신뢰하는 사용자는 쇼핑 사이트의 상품 추천이나 데이팅 플랫폼의 잠재적 매칭 등 시스템이 제공하는 추천 결과

를 세심한 계산의 산물로 여기고 자신의 필요나 선호에 부합할 것이라고 기대할 수 있다(Sharabi, 2021). 또한 알고리즘의 기능에 대한 믿음은 사용자가 제공하고자 하는 정보의 양과 질에도 영향을 미칠 수 있는데, 사용자가 데이터를 효과적으로 활용하여 더 나은 결과를 생성하는 알고리즘의 능력에 대한 믿음이 있으면 더 정확하고 상세한 정보를 공유하려는 경향이 높아질 수 있다(Sundar & Kim, 2019). 이처럼, 챗봇의 활용성이 높은 온라인 데이팅 서비스를(Wang, 2023) 이해하기 위해, 알고리즘에 대한 신뢰를 주요 요인으로 추가 고려하였고, 이는 챗봇과 인간이라는 두 가지 대화형 에이전트에 따른 부정확한 정보 제공 행동에 미치는 영향을 비교하는데 기여할 것이다.

3.2. 대화형 에이전트의 사회적 실재감(Social Presence)

기존 연구들은 개인이 디지털화된 에이전트와 사회적 상호작용을 할 때, 음성 에이전트(Kim et al., 2022), 자동화된 메신저(Tzeng, 2004), 챗봇(Ho et al., 2018; Sundar & Kim, 2019)을 포함한 다양한 에이전트에 대해 사회적 실재감을 느낄 수 있음을 밝혔다. 알고리즘에 대한 신뢰는 사용자가 대상의 사회적 실재감을 인지하는 방식에 다양한 영향을 미치므로, 대화형 에이전트의 실재감에 대한 인식을 형성하는데 중요한 역할을 할 수 있다. 문헌에 따르면 상대에 대한 신뢰는 상대방의 유능함, 선의, 성실성을 결정하며 (McKnight et al., 2002), 이러한 대화 에이전트의 자질은 궁극적으로 사회적 실재감을 향상시킬 수 있다. 사람들이 사용 중인 챗봇 뒤에 의식이 있다고 인식할 때, 마치 함께 있는 것처럼 느낄 가능성이 높으며, 이러한 사회적 실재감은 챗봇과 친밀감을 느끼는 정도에 영향을 미친다(Lee et al., 2020). 또한, 사용자가 알고리즘의 응답에 대한 논리와 근거를 이해하면 에이전트를 더 사려 깊

고 인간답다(human-like)고 인지할 수 있다(Ribeiro et al., 2016). 이에 더해 알고리즘에 대한 신뢰는 풍성한 상호작용을 바탕으로 대화 에이전트가 더 사회적으로 실재감 있게 보이도록 할 수 있으며(Giles, 2002), 보안과 신뢰감을 조성하여 사용자가 플랫폼을 실제 또는 진정성 있는 사회적 실체로 간주하도록 유도할 수 있다(Shin & Park, 2019). 예를 들어, 사용자가 대화 에이전트의 알고리즘과 소프트웨어가 자신의 질문을 정확하게 이해하고 응답한다고 신뢰하면 챗봇을 더 인간과 비슷하다고 인식할 수 있다(Lankton et al., 2015). 이는 인지된 사회적 실재감을 높이고 더 긴밀한 상호작용으로 연결될 수 있다. 반면에 사용자의 알고리즘과 소프트웨어에 대한 신뢰도가 낮으면 플랫폼을 비인간적이거나 인위적인 것으로 간주하여 상호 작용이 감소할 수 있다. 따라서 문헌들을 종합하여 다음과 같은 가설을 세울 수 있다:

H1a: 알고리즘에 대한 개인의 신뢰는 온라인 환경에서 개인 정보 제공을 요청하는 대화형 에이전트의 사회적 실재감을 향상시킬 것이다.

인간과 유사한 대화 스타일을 모방하는 챗봇을 서비스함으로써 조직은 사용자 참여를 유도하고 다양한 기능을 제공하며 궁극적으로 사용자 만족도를 높일 수 있다(Nguyen et al., 2022). 알고리즘으로 구현된 챗봇의 특성상, 챗봇의 사람의 언어를 이해하는 능력, 자동화 정도, 사람과 유사한 응답을 구성하는 정도는 알고리즘에 의해 크게 좌우된다. 따라서, 챗봇에 대한 인식은 사용자가 가진 알고리즘에 대한 신뢰에 따라 크게 변할 수 있으며, 이는 대화형 에이전트의 사회적 실재감 강화에 기여할 수 있다(Shin, 2022). 반면 대화형 에이전트가 인간인 경우, 알고리즘과 인간 에이전트와의 긴밀한 관계가 챗봇에 비해 제한적일 것이므로, 알고리즘에 대한 사용자의 신뢰가 사회적 실재감에 미치는 정의 효과 또한 챗봇에 비해 낮을 것이다. 따라서 다음의 가설을 세울 수 있다.

H1b: 알고리즘에 대한 신뢰가 사회적 실재감에 미치는 정 (+)의 영향은 대화형 에이전트가 인간인 경우보다 챗봇인 경우에 더 크다.

3.3. 개인 정보 제공에 대한 인지된 위험

개인 정보 공개에 대한 위험을 인지하는 정도는 정보 공개 여부와 범위 결정에 중요한 역할을 한다(Cheng et al., 2021). 서비스 이용 시 불확실하거나 잠재적으로 의심스러운 개인 데이터 제공 요청을 받으면, 사용자는 자연스럽게 위험에 대한 노출을 최소화하는 경향이 있다(Rossmannek et al., 2024). 이에 따라 요청을 회피하거나 부정확한 정보를 제공하는 등 개인 정보 보호를 위한 행동을 취할 수 있으며, 이는 자신의 프라이버시를 보호하고 잠재적인 개인 정보 오용을 방지하기 위함이다(Dinev & Hart, 2006).

이 때, 알고리즘에 대한 신뢰는 개인 정보 공유와 같이 잠재적 위험이 있는 상황과 관련해 위험을 인지하는 정도를 크게 완화할 수 있다. 알고리즘에 대한 신뢰도가 높은 사용자는 알고리즘이 데이터를 정확하고 투명하게 처리할 것으로 예상하며, 개인정보 공개와 관련된 위험을 낮게 인식할 수 있다(Shin, 2020; Shin et al., 2022). 즉, 개인이 알고리즘 작동 방식을 이해하면 이를 신뢰할 수 있고, 신뢰할 수 있는 알고리즘으로 운영되는 소프트웨어 제품은 데이터의 무결성을 보호할 것이라고 믿는 경향이 높아질 수 있다(Shin, 2022). 신뢰할 수 있는 소프트웨어 제품은 엄격한 데이터 보호 조치를 준수하여 사용자에게 데이터의 안전성을 보장하고, 보다 정직하고 개방적인 정보 공유를 장려할 것으로 기대된다. 따라서 대화형 에이전트에 대해 사용자가 느끼는 위험이 완화될 수 있다(Martin et al., 2017). 마찬가지로 알고리즘의 공정성과 보안성을 신뢰하는 사용자는 이러한 시스템과 더 가까이 소통할 의향이 있을 수 있다(Sundar & Kim, 2019). 반대로 알고리즘에 대한 신뢰도가 낮은 사용자는 데이터

유출이나 개인정보 오용 가능성을 우려하여 위험을 더 높게 인지할 수 있다(Rohden & Zeferino, 2023). 따라서 다음과 같은 가설을 세울 수 있다:

H2a: 알고리즘에 대한 개인의 신뢰는 개인 정보 제공에 대한 지각된 위험을 줄일 수 있다.

알고리즘에 대한 인식이 인지된 위험에 미치는 영향은 대화 상대에 따라 달라질 수 있다. 예를 들어, 인간 상담원은 어느 정도의 재량권과 예측 불가능성을 가지고 있는 것으로 인식되지만, 알고리즘을 기반으로 하는 챗봇은 기 설계된 규칙에 기인하여 행동을 수행할 것이기에 더 높은 수준의 일관성과 신뢰성을 제공할 수 있다(Sundar & Kim, 2019). 즉, 알고리즘에 대한 신뢰는 특히 챗봇과 상호 작용할 때 인지된 위험을 줄이는 데 중요한 요소가 되고, 인간과 상호작용할 때는 그 효과가 줄어들 것으로 예상된다. 따라서 다음과 같은 가설을 세울 수 있다:

H2b: 알고리즘에 대한 신뢰가 지각된 위험에 미치는 부(-)의 영향은 대화형 에이전트가 인간인 경우보다 챗봇인 경우에 더 크다.

3.4. 부정확한 정보 제공 의도

개인이 대화형 에이전트에게 강한 사회적 실재감을 느낄 때, 해당 대상과의 상호작용에 사회적 규범과 규제가 적용될 수 있다(Park et al., 2022). 이러한 사회적 실재감은 사회적 책임감과 신뢰감을 이끌어낼 수 있으며, 이는 사용자 응답의 정직성과 투명성을 높일 수 있다. 또한 기존 연구에 따르면, 개인은 다른 사람이 자신을 감시하거나 판단하고 있다고 믿을 때 거짓말을 할 가능성이 줄어든다(Cohn et al., 2022). 따라서 대화형 에이전트의 사회적 실재감을 높게 인식하면 사용자는 본인의 행동을 사회적 규범 내에서 맞추려 하고, 또한 자신이 주시되고 있다고 느껴 개인 정보에 대해 부정확한 정보를 제공할 가능성이 낮아질 수 있다. 따라서 다음과 같은 가설을 세울

수 있다:

H3a: 대화형 에이전트의 사회적 실재감을 높게 인지하면 해당 에이전트가 요청하는 개인 정보에 대해 부정확한 정보를 제공할 의도가 줄어든다.

챗봇이 사회적 실재감을 가지고 있더라도 사회적 실재감으로 인한 부정직한 행동의 억제 효과는 인간과 상호작용할 때보다 낮을 것이다. 챗봇의 사회적 실재감이 높으면 사람과 대화하는 것처럼 느껴져 상호작용이 더욱 친밀하고 매력적으로 느껴질 수 있다. 그러나 그렇다고 하더라도 이는 동일한 소통 환경에서 사람과 사람 간의 상호작용과는 다르다(Song et al., 2022). 인간 상담원과 상호작용할 때는 상대의 개인에 대한 주관적 판단, 보복 또는 평판 손상에 대한 두려움이 강력한 억제 요인으로 작용할 수 있는 반면, 챗봇과 상호작용할 때는 사회적 실재감이 높은 챗봇이라도 이러한 사회적, 윤리적 억제력이 같은 정도로 적용되지 않을 수 있다(Schuetzler et al., 2019). 따라서 다음과 같은 가설을 세울 수 있다:

H3b: 대화형 에이전트의 사회적 실재감이 개인 정보에 대해 부정확한 정보 제공 의도에 미치는 부(-)의 영향은 대화형 에이전트가 인간인 경우보다 챗봇인 경우 더 약하게 나타난다.

개인 입장에서는 자신의 정보가 스스로 원하는 용도로 정확하게 사용되는지 정확히 알 수 없으므로 이에 대한 위험과 불확실성을 항상 인지하게 되며, 이에 따라 부정확한 정보를 제공할 의도가 존재한다. 기존 연구에 따르면, 개인은 온라인에서 나이, 성별, 관심사 등의 개인 정보를 허위로 표현하는 경향이 더 강하다(Drouin et al., 2016). 특히, 개인정보 오남용 사례가 언론에 자주 등장하면서 개인은 허위 정보를 제공하여 위험은 줄이고 이익만 얻고자 하는 유혹을 받을 수 있다. 잘못된 개인정보를 제공하면서도 원하는 상품이나 서비스를 이용할 수 있다면 그렇게 하는 것이 정보 제공자에게 매력적인 선

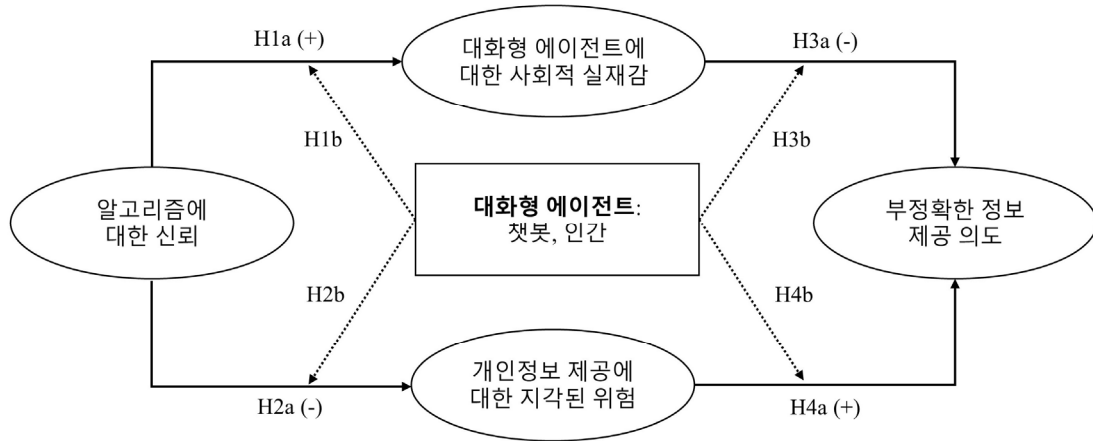
택지가 될 수 있기 때문이다(LaMothe & Bobek, 2020). 결과적으로, 개인이 대화형 에이전트에게 개인정보를 제공할 때 위험을 인식하면 부정확하거나 허위 정보를 제공하는 경향이 증가할 수 있다. 따라서 다음과 같은 가설을 세울 수 있다:

H4a: 개인정보 제공에 대한 지각된 위험은 대화형 에이전트가 요청하는 개인 정보에 대해 부정확한 정보를 제공할 의도를 높인다.

개인 정보 제공에 따른 위험이 갖는 부정확한 정보 제공 의도에 대한 영향의 효과는 대화형 에이전트의 종류에 따라 다르게 나타날 수 있다. 이는 사용자가 인간 에이전트와의 상호작용에서 느끼는 사회적 규범이나 부담감으로부터 발생하는 개인 정보 제공에 수반한 위험이, 챗봇과의 상호작용에서는 상대적으로 낮기에 보다 솔직하고 성실한 정보를 제공할 가능성이 있기 때문이다(Sundar & Kim, 2019). 또한, 챗봇 에이전트는 감정이 없는 소프트웨어로 개인 정보를 사사roi 평가하거나 획득한 정보를 오남용할 위험이 상대적으로 낮다고 인식될 수 있다(Zhang et al., 2021). 나아가, 인간과의 소통에서 느낄 수 있는 당혹감이나 창피함은 챗봇 에이전트를 상대할 때 고려 대상에서 제외된다. 반면, 인간과의 소통에서는 위의 위험들이 맥락적 요소로 중요하게 고려된다. 또한, 개인이 에이전트와 물리적으로 상호작용하지 않더라도, 인간 에이전트와 대화하는 과정에서는 창피하고 불편하며 바람직하지 않은 콘텐츠를 숨기는 자기 검열이 발생할 수 있다(Lau et al., 2018). 따라서 다음과 같은 가설을 세울 수 있다.

H4b: 개인정보 제공에 대한 지각된 위험이 개인 정보에 대한 부정확한 정보 제공 의도에 미치는 정(+)의 영향은 대화형 에이전트가 인간인 경우보다 챗봇인 경우 더 작게 나타난다.

연구 모형은 <그림 1>에 설명되어 있다.



〈그림 1〉 제안된 연구 모형

4. 연구방법

4.1. 데이터 수집

본 연구는 Amazon MTurk에서 횡단면 설문조사를 실시하여 데이터를 수집했다. 데이터 수집을 위해 실험실과 같은 환경(lab-experiment environment)을 설계하였다. 참가자는 무작위로 두 가지 대화형 에이전트(챗봇 또는 인간) 중 하나에 배정되었고, 이 무작위 배정은 설문조사 내 문항(manipulation check)으로 재차 확인하였다. 참가자들은 대화형 에이전트가 온라인 데이트 서비스 이용을 위해 개인 정보를 요청하는 상황에 노출된 뒤, 설문조사를 진행하였다. 부록 B에 실험 환경으로 설정된 시나리오가 제시되어 있다.

총 407명의 응답자로부터 챗봇 시나리오와 인간 시나리오에 대한 응답을 받았다. 분석 수행 전에 다음 기준에 따라 신뢰도가 낮은 데이터를 폐기하였다: (a) 총 응답 시간이 5분 미만인 경우, (b) 모든 질문에 동일한 답변을 제공한 경우, (c) 응답자가 지시를 따르지 않은 경우 (e.g., '아니오' 응답이 필요한 질문에 다르게 답한 경우), (d) 대화형 에이전트 유형을 잘못 식별한 경우 (e.g., 인간

에이전트에 배정되었으나 챗봇 에이전트라 응답한 경우). 그 결과, 추가 분석에 활용하기에 적합하다고 판단된 응답은 챗봇과 인간 시나리오 각각 180개와 87개였다. <표 1>에는 응답자의 인구통계학적 통계를 요약하였으며, 이후 분석은 해당 데이터를 활용하여 진행하였다.

4.2. 측정 항목

본 연구에 활용된 설문 항목들은 기존 연구에서 채택하여, 연구 맥락에 맞게 문구 수정 후 활용하였다. 알고리즘에 대한 신뢰도는 Sundar and Kim (2019)의 5개 항목을 사용했다. 정보 공개에 대한 지각된 위험은 Xu et al. (2011)의 3개 항목을 사용했다. 대화 에이전트에 대한 사회적 실재감의 4개 항목은 Pavlou et al. (2007)의 항목을 활용하였다. 부정확한 개인 정보 제공 의도(개인 정보에 대한 왜곡 의도)는 Lwin et al. (2007)의 3개 항목을 사용하였다. 이 때, 설문 참여자의 응답 피로도를 줄이기 위해 개인 정보를 키/몸무게로 정하여 질문하였다. 데이트 상대의 매력을 평가할 때 키와 몸무게는 주요한 정보가 되며, 또한 거짓으로 제공하는 경우가 많기 때문이다(Sharabi & Caughlin, 2019). 앞서 설명한 모든 변수

〈표 1〉 응답자의 인구통계학적 특성

		챗봇 (n=180)		인간 (n=87)	
		빈도	백분율	빈도	백분율
성별	남	117	65.0	48	55.2
	여	63	35.0	39	44.8
연령	20대	21	11.7	10	11.5
	30대	111	61.7	56	64.4
	40대	29	16.1	12	13.8
	50대 이상	19	10.6	9	10.3
최종 학력	전문대 이하	59	32.8	26	29.9
	대졸	68	37.8	34	39.1
	석사 이상	53	29.4	27	31.0
연소득수준	~\$40,000	48	26.7	24	27.6
	\$40,000~\$60,000	49	27.2	28	32.2
	\$60,000~\$80,000	52	28.9	19	21.8
	\$80,000~	31	17.2	16	18.4

는 '전혀 아니다'에서 '매우 그렇다'까지의 7점 리커트 척도를 사용하여 측정했다. 추가적으로, 개인이 온라인 환경에서 개인 정보 공개 요청을 얼마나 자주 받는지, 그리고 해당 요청에 얼마나 적극적으로 응하는지에 따라 개인 정보 제공 의도가 달라질 수 있으므로(Sundar et al., 2013; Sundar et al., 2020), 인터넷에 이미 공개된 개인 정보의 항목 수를 조사하여 본 연구의 통제 변수로 활용하였다(Punj, 2019). 이 외에 나이, 성별, 교육 수준, 직업도 통제 변수로 활용하였다.

5. 연구 결과

5.1. 측정 모형

측정 모형을 검증하기 위해 먼저 모든 구성요소에 대한 Cronbach's alpha 및 합성 신뢰도(composite reliability)를 확인했다. <표 2>에서 확인할 수 있듯, 모든 경우에서 Cronbach's alpha 및 합성 신뢰도는 0.8 이상으로 높은 신

뢰성을 보임을 확인하였다. 그 뒤, Varimax 로테이션을 사용해 탐색적 요인 분석(Exploratory Factor Analysis)을 수행하였다. 요인 분석 과정에서 알고리즘에 대한 신뢰에 대한 항목 하나가 삭제되었다. 분석에 사용된 최종 문항들은 부록 A에 제시되어 있다. 부록 C에서 볼 수 있듯이, 모든 요인 적재값이 0.6보다 크고, 해당 변수에 대한 모든 항목의 적재값이 다른 변수에 대한 적재값보다 높아 항목들의 전반적인 수렴 및 판별 타당도가 충족되었음을 보였다(Hair et al., 2006). 또한 평균 분산 추출(Average Variance Extracted: AVE)을 조사하여, 각 변수의 AVE의 제공근과 다른 변수와의 상관관계를 비교하였다. <표 2>에 표시된 바와 같이, 모든 경우에서 AVE 값이 0.5보다 높고, 대각선에 위치한 AVE 값의 제공근이 변수 간 상관관계보다 크므로 수렴 타당도와 판별 타당도가 만족스러운 것으로 나타났다(Fornell & Larcker, 1981). 마지막으로 동일 방법 편향(common method bias)의 존재 가능성을 확인하기 위해 두 가지 방법을 활용하였다. 첫째, Harman의 단일 요인 테스트(Podsakoff & Organ, 1986)을 분석한 결과, 단일 요인은

〈표 2〉 타당성 및 신뢰성 분석 결과

타당성 및 신뢰성 분석 결과 (챗봇, n=180)								
	신뢰	실재감	위험	부정확	공개정보	Cronbach' Alpha	합성 신뢰도	AVE
알고리즘에 대한 신뢰	0.824					0.843	0.894	0.68
대화형 에이전트에 대한 사회적 실재감	0.502**	0.902				0.923	0.945	0.813
개인정보 제공에 대한 지각된 위험	-0.065	0.308**	0.895			0.881	0.926	0.807
부정확한 정보제공 의도	0.203	0.423**	0.614**	0.925		0.915	0.946	0.855
기 공개된 개인 정보 수준	-0.190*	-0.224**	-0.245**	-0.252**	-	-	-	-
타당성 및 신뢰성 분석 결과 (인간, n=87)								
	신뢰	실재감	위험	부정확	공개정보	Cronbach' Alpha	합성 신뢰도	AVE
알고리즘에 대한 신뢰	0.85					0.872	0.912	0.722
대화형 에이전트에 대한 사회적 실재감	0.484**	0.887				0.91	0.937	0.787
개인정보 제공에 대한 지각된 위험	0.331**	0.288**	0.885			0.862	0.916	0.784
부정확한 정보제공 의도	0.351**	0.344**	0.669**	0.937		0.931	0.956	0.879
기 공개된 개인 정보 수준	-0.037	-0.029	-0.328**	-0.242*	-	-	-	-

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

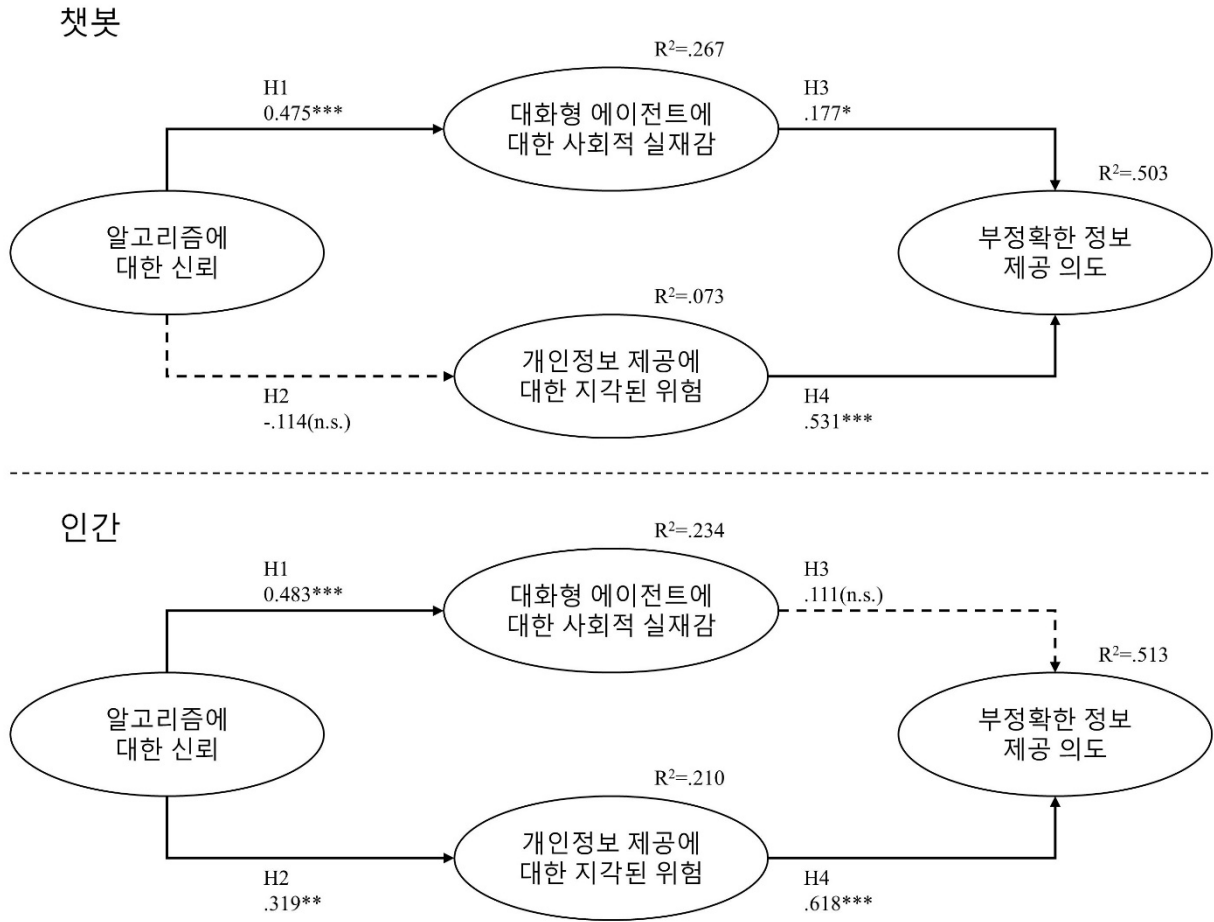
전체 분산의 각각 30.70%(챗봇)와 34.24%(인간)을 설명하여 우세한 메소드 요인이 나타나지 않았다. 둘째, Liang et al. (2007)의 제안에 따라, 메소드 요인과 본래 변수에 의해 설명되는 분산을 분해하였다. 두 대화형 에이전트에 대해, 지표의 본래 설명 분산의 평균은 각각 0.787(챗봇), 0.790(인간)인 반면, 메소드 분산의 평균은 각각 0.007(챗봇), 0.003(인간)(부록 D)이어서, 그 비율이 각각 113.57(챗봇), 269.63(인간)에 가까운 것을 확인하였다. 따라서, 동일 방법 편이의 존재가 본 연구에 심각한 문제가 되지 않는 것을 확인했다.

5.2. 구조 방정식 모형

본 연구 모형에 있는 경로 분석을 테스트하기 위해 부분 최소제곱법(Partial Least Squares, PLS)을 사용하였다. PLS 방법은 최소 제곱을 사용하여 예측 오차를 최소화하는 구조 방정식 모델링 방법(Gefen & Straub, 2005)으로, 적은 표본을 테스트하는 데 적합하다(Chin, 1998). 본 연

구의 표본 크기가 상대적으로 작기 때문에(각각 n=180, 87), PLS를 활용하여 모형을 분석하였다. 연구 모형의 분석 결과는 <그림 2>에 나와 있다. <그림 2>에서, 통제 변수는 제외되었다.

추가로 대화형 에이전트의 유형에 따른 효과의 차이를 확인하기 위해, 구조 방정식 모델에서의 해당 경로 계수를 통계적으로 비교하고 그 차이를 확인하는 분석을 수행하였다. 이를 위해, 먼저 Henseler et al. (2016) 이 제안한 MICOM(measurement invariance of composite models) 방식을 따라 다중집단분석의 선결조건인 측정 동일성을 확인하였다. 동일한 측정문항, 데이터처리, 모델 알고리즘을 사용하였기에 형태적 동일성을 확보하였고, PLS 알고리즘을 활용한 순열 검정을 통해 두 집단 간 변수의 평균의 상관관계가 1에 가까움을 확인하여 구성적 동일성을 확보하였다(Chin & Dibbern, 2009; Henseler et al., 2016). 마지막으로, 같은 순열 검정을 통해 집단 간 변수의 평균과 분산의 차이를 검정하는 평균 및 분산 균등성을 확인한 결과 이들 또한 확보되어, 최종적으로 완전 측



〈그림 2〉 구조방정식 모형 결과

정 동일성을 확보하였고 이후 집단 간 비교 분석을 수행하였다. 집단 간 비교는 Keil et al. (2000)이 제안한 다중 그룹 분석 수행 절차에 따라 진행하였으며, 각 그룹의 표본 수(챗봇 및 인간 그룹에 대하여 각각 m, n), 구조 방정식 모형의 경로 표준 오차(두 그룹 각각 $SE_{chatbot}$, SE_{human}), 및 분산에 대한 풀링된 추정(pooled estimators; S_p)를 사용하여 $m+n-2$ 자유도를 갖는 t-분포에 따라 t-검정을 계산하여 그 차이를 확인했다(아래 식 참조)(Sanchez-Franco, 2006). 본 연구에서 그룹은 각 대화형 에이전트의 상황을 관찰한 그룹이 되며, 두 그룹 간의 계수 차이를 비교하여 확인하였다. 분석 결과는 <표 3>에 정리했다.

$$t = \frac{\beta_{chatbot} - \beta_{human}}{S_p \times \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

$$S_p = \sqrt{\frac{(m-1)}{(m+n-2)} \times SE_{chatbot}^2 + \frac{(n-1)}{(m+n-2)} \times SE_{human}^2}$$

분석 결과를 정리하면 다음과 같다. 먼저 알고리즘에 대한 신뢰는 대화 에이전트의 사회적 실재감에 긍정적인 영향을 미치는 것으로 나타나 가설 H1a은 지지되었다. 하지만, 대화형 에이전트의 유형이 챗봇일 때와 인간일 때의 차이는 유의하지 않은 것으로 나타나, H1b의 경우 기각되었다.

또한, 알고리즘에 대한 신뢰가 개인정보 제공에 대한

〈표 3〉 다중 그룹 분석의 T-검정 결과

	표준 오차 (SE)		Sp	$\beta_{\text{챗봇}} - \beta_{\text{인간}}$	T-value	결과
	챗봇	인간				
H1b	0.071	0.102	0.082	-0.008	-0.745	n.s.
H2b	0.094	0.095	0.094	-0.433	-35.188***	챗봇(n.s.)<인간
H3b	0.079	0.106	0.089	0.066	5.708***	챗봇>인간(n.s.)
H4b	0.075	0.083	0.078	-0.087	-8.585***	챗봇<인간

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, n.s.=not significant

지각된 위험을 감소시키는 효과는 관찰되지 않아 H2a는 기각되었다. 각 대화형 에이전트 유형에 대해 살펴보면, 챗봇의 경우 예상된 방향으로 나왔으나 유의성이 충분하지 않았으며, 오히려 인간의 경우 지각된 위험을 높이는 방향으로 유의하게 나왔다. 하지만, 다중 그룹 분석의 수행 결과에서 인간인 경우의 계수 값이 정(+)으로 더 크게 나타났기에 H2b는 지지되었다.

대화형 에이전트에 대한 사회적 실재감이 부정확한 정보 제공 의도에 미치는 영향에 대한 H3a는 기각되었는데, 가설과 달리 챗봇의 경우엔 정(+)의 영향을 주었고 인간의 경우에는 유의하지 않은 결과를 보였다. 다중 그룹 분석의 수행 결과, 챗봇일 때 그 효과가 더 크게 보였기에 H3b의 경우 기각되었다.

개인정보 제공에 대한 지각된 위험이 부정확한 정보 제공 의도에 정(+)의 영향을 미치는 것으로 나타나 해당

가설인 H4a는 지지되었다. 추가로, 다중 그룹 분석의 수행 결과, 인간인 경우의 계수 값이 정(+)으로 더 크게 나타났기에 H4b도 지지되었다.

통제 변수의 경우, 기 공개된 개인 정보 수준의 크기가 큰 경우(온라인에 이미 개인 정보를 많이 남긴 경우), 지각된 위험을 적게 느끼는 것으로 나타나 정보 공개에 대한 위협성을 덜 인식하고 있음을 확인했다(챗봇: -0.267***, 인간: -0.317***). 또한, 사용자의 기 공개된 개인 정보 수준의 크기는 챗봇의 경우에만 사회적 실재감에 부정적인 영향을 미쳤으며 (챗봇: -0.134*), 인간의 경우엔 유의하지 않았다. 아울러, 교육 수준과 직업은 챗봇 상황에서 부정확한 정보 제공 의도에 유의미한 영향을 미쳤다 (각각 0.161**, -0.189*). 가설의 결과를 요약하면 아래 <표 4>와 같다.

〈표 4〉 가설 검정 결과 요약

가설	유의성		지지 여부	
	챗봇	인간	챗봇	인간
H1a 알고리즘에 대한 신뢰 → 대화형 에이전트에 대한 사회적 실재감 (+)	$p < 0.001$	$p < 0.001$	채택	채택
H1b 알고리즘에 대한 신뢰 → 대화형 에이전트에 대한 사회적 실재감: 챗봇 > 인간	n.s.		기각(n.s.)	
H2a 알고리즘에 대한 신뢰 → 개인정보 제공에 대한 지각된 위험 (-)	n.s.	$p < 0.01(+)$	기각	기각
H2b 알고리즘에 대한 신뢰 → 개인정보 제공에 대한 지각된 위험: 챗봇 < 인간	$p < 0.001$		채택(챗봇<인간)	
H3a 대화형 에이전트에 대한 사회적 실재감 → 부정확한 정보 제공 의도 (-)	$p < 0.01(+)$	n.s.	기각	기각
H3b 대화형 에이전트에 대한 사회적 실재감 → 부정확한 정보 제공 의도: 챗봇 < 인간	$p < 0.001$		기각(챗봇<인간)	
H4a 개인정보 제공에 대한 지각된 위험 → 부정확한 정보 제공 의도 (+)	$p < 0.001$	$p < 0.001$	채택	채택
H4b 개인정보 제공에 대한 지각된 위험 → 부정확한 정보 제공 의도: 챗봇 < 인간	$p < 0.001$		채택(챗봇<인간)	

6. 토론

6.1. 연구 결과

본 연구는 온라인 데이팅 서비스 이용 상황에서 부정확한 정보 제공 의도를 일으키는 영향 요인에 대해 알아 보았다. 이를 위해, 대화형 에이전트를 챗봇과 인간으로 구분하고, 각 에이전트를 대하는 상황에서 알고리즘에 대한 신뢰와 사회적 실재감, 개인정보 제공에 대한 위험이 어떻게 부정확한 정보 제공 의도에 영향을 주는지 실증적으로 확인하였다. 분석 결과, 알고리즘에 대한 신뢰는 대화형 에이전트의 사회적 실재감에 정의 영향을 미치는 것을 확인하였다. 다만, 해당 효과는 대화형 에이전트의 유형에 따라 다르지 않은 것으로 나타났는데, 이는 온라인 환경에서 대화로 사회적 실재감을 인지하는 맥락상 챗봇과 인간의 차이를 크게 느끼기는 어렵기 때문으로 해석된다.

알고리즘에 대한 신뢰는 개인정보 제공에 대한 지각된 위험을 줄이지 않았고, 오히려 인간의 경우엔 위험을 높이는 결과가 나왔다. 에이전트의 유형에 따른 상대적인 크기는 예상과 같았다는 것을 고려하면, 알고리즘에 대한 신뢰가 높은 사람들이 오히려 인간이 개인정보를 다루고 처리하는 상황을 더 우려하여 발생한 것으로도 볼 수 있다. 다만, 챗봇의 경우엔 위험에 영향을 주지 않았는데, 이는 처리 과정에 대한 위험을 줄일 수 있어도(Li, 2023), 개인 정보 제공 자체에 따른 위험이 존재하여(Ho et al., 2018) 두 영향이 상쇄되어 유의하지 않게 나온 것으로 해석할 수 있다.

대화형 에이전트에 대한 사회적 실재감이 부정확한 정보 제공 의도에 미치는 영향 또한 예상과는 다르게 챗봇에서만 긍정적으로 나타났다. 사회적 실재감이 오히려 챗봇에 대한 인지를 높이고, 대화 상대가 챗봇이라는 것을 인지한 사용자들이 거짓말 행동을 보인 것이라고 설

명할 수 있다. 온라인 데이팅 서비스 이용 상황의 특성상 사회적 실재감이 클수록 본인을 더 잘 보이게 하기 위해 부정확한 정보를 제공한 기존 연구들의 결과와 일관된다고 볼 수 있다(Hancock et al., 2004; Peng, 2020; Sharabi & Caughlin, 2019). 에이전트의 유형에 대한 상대적인 효과 또한 예상과 달리 챗봇일 때 더 컸는데, 이는 위 설명과 연결되어, 디지털 상황에서 사용자들이 커뮤니케이션 상대가 챗봇임을 인지할 때 거짓 행동을 더 많이 보이기 때문이라 해석할 수 있다(Croes & Antheunis, 2021).

개인정보 제공에 대한 위험이 부정확한 정보 제공 행동을 높일 거라는 가설은 두 에이전트 유형 모두에서 일관되게 확인할 수 있었고, 그 효과 또한 예상한 대로 에이전트가 인간일 때 더 큰 것을 확인할 수 있었다.

6.2. 학문적 시사점

본 연구 결과를 통해 다음의 학문적인 시사점을 제공할 수 있다. 먼저, 온라인에서의 의도적인 부정확한 정보 제공 행위에 대한 연구를 진행함으로써, 해당 행동의 중요성을 강조하고 이에 대한 이해를 넓혔다. 부정확한 정보 제공 행동은 개인의 서비스 제공에도 부정적인 영향을 줄 뿐만 아니라, 온라인 정보의 진실성과 신뢰성을 줄이고, 사회적 관계 형성을 저해함으로써 전반적인 온라인 환경에도 부정적인 영향을 줄 수 있다(Menard & Bott, 2024). 해당 행동에 영향을 미치는 다양한 요인을 확인함으로써, 부정확한 정보 제공 행동을 줄이며 더 바람직한 인터넷 사용 환경 구축에 기여할 수 있다. 예를 들어, 개인정보를 개별 정보 차원이 아니라 개개의 개인정보가 모인 정보의 집합이라고 이해한다면, 부정확한 정보가 갖는 의미와 영향력을 활용하여 디지털 공간에서 개인정보 집합의 의미에 대한 시사점을 제공할 수 있다. Floridi (2014)는 온라인 정보 공간을 인포스피어(Infosphere)로

정의하고 개인의 정보의 총체가 온라인의 하나의 개체를 이루는 인포그(Inforg) 개념을 주장하였다. 즉, 개인이 온라인에서 활동하며 개인에 관한 다양한 정보가 디지털화됨에 따라 이런 정보의 누적만으로 온라인에서 특정 개인에 대한 정보 개체를 구현할 수 있게 된다는 것이다. 소셜미디어 정보 등 개인에 의해 온라인에서 자유롭게 생성되는 정보는 그 진위 여부를 판단하기 어렵다. 따라서 부정확한 정보가 포함되어 구현되고 이해되는 인포그는 그 대상이 되는 현실의 개체와는 괴리가 있을 수 있으며, 따라서 온라인 공간을 현실과 괴리가 있는 왜곡된 공간으로 만들 수 있다. 본 연구는 이러한 관점에서 부정확한 정보, 특히 의도적으로 부정확한 정보를 만들어내는 데 영향을 미치는 요인들을 연구하여, 온라인 공간에서 발생할 수 있는 현실과의 괴리가 생겨나는 원인과 환경을 살펴보려 하였다. 본 연구의 이러한 시도는 온라인 공간의 부정확한 개인정보를 이해하는 정보 생성자 관점 연구들에게도 도움이 될 수 있을 것이다.

둘째, 챗봇과 인간이라는 대화형 에이전트 종류에 따라 요인들의 효과가 달라지는 것을 실증적으로 확인하였다. 기존 연구에서는 커뮤니케이션 파트너에 따라 거짓말 행동이 달라짐을 직접 효과로써 확인하였다면(Lucas et al., 2014; Turner et al., 2005), 본 연구에서는 이를 조절효과로 고려하여 각 관계의 효과를 구체적으로 살펴보았다. 예를 들어, 알고리즘에 대한 신뢰가 개인정보 제공에 따른 인지된 위험에 미치는 영향은 인간과 챗봇 에이전트 두 경우에 대해 다르게 발현되었다. 이처럼 같은 대화형 에이전트라 하더라도 대화 상대의 정체성에 따라 요인들의 상대적인 효과를 살펴볼 수 있기에 챗봇과 부정확한 정보 제공에 대한 이해를 더욱 높이는데 기여할 수 있다.

추가로, 알고리즘에 대한 신뢰가 대화형 에이전트 상황에서 가지는 효과에 대해서도 확인하였다. 알고리즘에 대한 신뢰는 에이전트의 종류에 관계없이 소통 상대의

사회적 실재감을 강화하는 요인으로 작용하였다. 이는 대화형 에이전트라는 대상 자체의 기술적인 환경 요인이 대상의 정체성보다 강하게 작용한 것으로 해석할 수 있다(Pataranutaporn et al., 2023). 또한, 사용자의 기 공개된 개인 정보 수준은 지각된 위험 사이에 부(-)의 영향을 보 이면서, 사용자가 온라인 데이터 공유에 익숙한 것이 챗봇 상호작용과 관련된 위험 수준을 결정하는 데 중요한 요인이 될 수 있음을 제시했다. 따라서, 대화형 에이전트나 정보 제공 행동에 대한 이해 등 프라이버시 관련 연구 수행 시 알고리즘에 대한 신뢰와 사람들이 이미 온라인에 공개한 정보 수준은 주요 요인으로 다뤄져야 한다. 또한, 에이전트의 유형에 따라 이들의 영향이 달라지기에 이에 대한 추가 탐구가 필요하다.

마지막으로, 앞서 이야기한대로 온라인 정보는 개별 정보로서의 용도를 넘어 정보의 집합으로서 다양하게 사용되고 있다. 특히 AI의 활용이 보편화된 시대에 이러한 데이터가 AI의 학습 데이터로 사용되면 윤리적인 문제를 낳을 수 있다. 부정확한 정보로 학습된 AI는 현실에 대한 편향된 시각을 보여줄 수 있을 것이기 때문이다. 온라인 테이팅 서비스 이용 상황에서와 같이 상대방에게 잘 보이기 위한 약간의 거짓말이 현실에서는 허세 정도로 이해될 수 있지만, AI가 학습하는 온라인 데이터에서는 허세와 같은 개념으로 부정확한 정보가 이해되기 보다는 그 데이터가 이야기하는 그대로의 정보가 AI가 이해하는 현실이 될 것이다. 따라서 부정확한 정보는 AI의 편향과 연결되고 윤리적 문제로 연결될 수 있다. 이러한 측면에서, 본 연구는 AI의 윤리 문제를 낳을 수 있는 편향된 데이터 생성 환경에 대한 이해를 제공할 수 있을 것이다.

6.3. 실무적 시사점

본 연구 결과의 실무적 함의는 다음과 같다. 첫째로, 온라인에서 발생하는 부정확한 정보 제공 행동을 줄이기 위한

챗봇 설계 요소를 제안할 수 있다. 예를 들어, 챗봇과 대화할 때 사회적 실재감이 부정적인 행동을 높이고 있으므로, 챗봇 설계 시 사회적 실재감을 줄이는 형태로 의인화 정도를 고려하여 설계 및 구축할 수 있다. 또한, 개인정보 제공에 대한 위험은 부정적인 행동을 줄이는 과정에서 중요한 역할을 수행하기에, 이를 줄일 수 있는 챗봇 설계 및 보안 요소, 기술적인 지원, 제도적인 장치 등의 고려가 필요하다.

또한, 대화형 에이전트의 유형에 따라 앞선 효과들이 달라질 수 있기에, 서비스 제공자들은 본인들의 서비스와 상황에 적합한 에이전트 형태를 선택할 필요가 있다. 같은 메시지를 전달하더라도 에이전트의 유형에 따라 사용자들의 인식이 달라지며 그 효과가 상이해지기에, 전략적인 에이전트 유형 선택이 필요하다. 예를 들어, 개인의 알고리즘에 대한 신뢰가 부정확한 정보 제공 의도에 미치는 영향의 경로는 챗봇일 경우 사회적 실재감을 통하고, 인간일 경우 대화 에이전트의 지각된 위험을 통하는 등 에이전트 유형에 따라 달라진다. 이에 따라, 에이전트 유형 선택에 따라 사용자의 부정확한 정보 제공 의도를 줄이기 위한 노력이 달라져야 하기에 이를 고려할 필요가 있다.

마지막으로, 본 연구는 AI를 사용하는 사용자 윤리와 관련한 시사점을 갖는다. 챗봇과 같은 AI 도구들은 수집된 데이터를 학습해 서비스를 발전시키는 특성을 가지므로, 잘못된 데이터가 미치는 파장이 기존의 디지털 도구에 비해 더 크다. 체리피킹 혹은 단기간의 이익 극대화를 위한 부정확한 정보 제공이 지속적으로 발생한다면, 데이터의 왜곡이 심화되고 챗봇을 포함한 다양한 AI 도구들은 잘못된 정보를 학습하여 궁극적으로 서비스의 질과 신뢰성은 물론이고 사용자 경험 또한 타격을 받을 것이다. 따라서 챗봇을 도입하고자 하는 기업은 이용자들에게 부정확한 정보 제공이 단순히 이익을 챙기는 행위가 아닌, 정보 교환 상대를 교란시키고 망가뜨리는 행위임을 고지해야 한다(Guzman & Lewis, 2020). 또한 사용자의 윤리적인 챗봇 활용을 장려하기 위해 알고리즘의 투

명성과 신뢰성을 강화하고 정보 제공에 동반된 여러 위험 요소를 최소화하는 것이 필수적이다(McGuire et al., 2023; Murtarelli et al., 2021). 아울러 기술적으로는 사용자의 부정확한 정보 제공을 탐지하고 이를 대응할 수 있는 체계 및 기술의 개발이 필요하다.

6.4. 한계 및 향후 연구 방향

본 연구는 몇 가지 한계점을 갖는다. 먼저, 본 연구는 설문 응답 환경을 실험 형태로 구축했지만, 응답자에게 실제 챗봇 애플리케이션을 제공하지 않았다. 챗봇 구현과 연구 환경 제어의 어려움이 이유였으나, 실제 환경과 다소간의 차이가 있다. 향후 연구에서는 이를 고려하여 실제 개발된 챗봇을 활용해 연구를 진행한다면, 일반화나 확장성을 확보하기 용이할 것으로 예상된다. 나아가 챗봇의 다양한 특성들(예. 사진 등 시각 요소, 텍스트에 포함된 감정, 상호관계성 등)을 고려하여 챗봇과 인간 상황을 더욱 구체적이고 다채롭게 제시한다면, 챗봇 구현 및 설계에 대하여 더욱 현실적이며 실무적인 시사점을 제공할 수 있을 것이다. 또한, 본 연구에서는 알고리즘에 대한 신뢰, 대화형 에이전트에 대한 사회적 실재감, 개인정보 제공에 대한 위험을 주요 요인으로 상정하였는데, 매칭 추천 시나리오와 같은 온라인 데이팅 플랫폼의 다양한 요소와 상황을 고려하여 연구를 수행한다면 부정확한 정보 제공에 대한 이해를 더욱 높일 수 있을 것이다. 마지막으로, 연구에 쓰인 샘플의 숫자를 고려한 분석을 수행하였지만, 그 숫자가 크지 않아 추후 연구에서는 더 많은 샘플을 확보하여 연구를 진행한다면 더 신뢰할 수 있는 결과를 제공할 수 있을 것이다.

〈참고문헌〉

[국내 문헌]

1. 김경애, 류방란 (2019). 교육에서의 4 차 산업혁명 기술 활용에 대한 기대와 우려. *교육과학연구*, 50(3), 55-79.
2. 박윤주 (2012). 고객의 암묵적 이상형을 반영하여 배우자 선택기준을 동적으로 조정하는 온라인 매칭 시스템: 의사결정나무의 활용을 중심으로. *Information Systems Review*, 14(3), 115-129.
3. 박은영 (2024). 고객의 조절조점 성향과 생성형 AI 기반 챗봇에 대한 친숙도가 개인정보 제공의도에 미치는 영향: 프라이버시 계산이론을 중심으로. *지식경영연구*, 25(2), 49-68.
4. 안효진, 김승인 (2017). 국내 모바일 애플리케이션 소셜 데이팅 서비스에 대한 사용자 경험 연구—이음과 정오의 데이트를 중심으로. *Journal of Digital Convergence*, 15(3), 335-341.
5. 진평위, 이정 (2021). 전자상거래 라이브채팅의 유형이 소비자가 지각하는 판매자에 대한 사회적 실재감과 신뢰에 미치는 영향. *지식경영연구*, 22(1), 287-308.
6. Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221), 509-514.
7. Barasch, A., & Berger, J. (2014). Broadcasting and narrowcasting: How audience size affects what people share. *Journal of Marketing Research*, 51(3), 286-299.
8. Behera, R. K., Bala, P. K., & Ray, A. (2024). Cognitive Chatbot for personalised contextual customer service: Behind the scene and beyond the hype. *Information Systems Frontiers*, 26(3), 899-919.
9. Berg, J. H., & McQuinn, R. D. (1989). Loneliness and aspects of social support networks. *Journal of Social and Personal Relationships*, 6(3), 359-372.
10. Biener, C., & Waeber, A. (2024). Would I lie to you? How interaction with chatbots induces dishonesty. *Journal of Behavioral and Experimental Economics*, 102279.
11. Cheng, X., Hou, T., & Mou, J. (2021). Investigating perceived risks and benefits of information privacy disclosure in IT-enabled ride-sharing. *Information & Management*, 58(6), 103450.
12. Chin, W. W. (1998). The Partial Least Squares Approach for Structural Equation Modeling. In G. A. Marcoulides (Ed.), *Modern Methods for Business Research* (pp. 295-336). Lawrence Erlbaum Associates.
13. Chin, W. W., & Dibbern, J. (2009). An introduction to a permutation based procedure for multi-group PLS analysis. In *Handbook of partial least squares: Concepts, methods and applications* (pp. 171-193). Springer.
14. Cohn, A., Gesche, T., & Maréchal, M. A. (2022). Honesty in the digital age. *Management Science*, 68(2), 827-845.
15. Cozby, P. C. (1973). Self-disclosure: A literature review. *Psychological Bulletin*, 79(2), 73-91.
16. Croes, E. A., & Antheunis, M. L. (2021). 36 questions to loving a Chatbot: Are people willing to self-disclose to a chatbot? In *Chatbot Research and Design: 4th International Workshop, Conversations 2020, Virtual Event*, November 23-24, 2020, Revised Selected Papers 4.
17. DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology*, 74(1), 63-79.
18. Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. (2022). On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems*, 23(1), 96-138.
19. Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, 17(1), 61-80.
20. Drouin, M., Miller, D., Wehle, S. M., & Hernandez, E. (2016). Why do people lie online? "Because everyone lies on the internet." *Computers in Human Behavior*, 64, 134-142.
21. Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford.
22. Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.
23. Gefen, D., & Straub, D. (2005). A practical guide to

- factorial validity using PLS-Graph: Tutorial and annotated example. *Communications of the Association for Information Systems*, *16*(1), 91-109.
24. Giles, D. C. (2002). Parasocial interaction: A review of the literature and a model for future research. *Media Psychology*, *4*(3), 279-305.
 25. Go, E., & Sundar, S. S. (2019). Humanizing Chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, *97*, 304-316.
 26. Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human-machine communication research agenda. *New Media & Society*, *22*(1), 70-86.
 27. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis*. Prentice Hall.
 28. Han, M. C. (2021). The impact of anthropomorphism on consumers' purchase decision in Chatbot commerce. *Journal of Internet Commerce*, *20*(1), 46-65.
 29. Hancock, J. T., Thom-Santelli, J., & Ritchie, T. (2004). Deception and design: The impact of communication technology on lying behavior. *Proceedings of the SIGCHI conference on Human factors in computing systems*, (129-134). New York: ACM.
 30. He, Y., Xu, X., Huang, N., Hong, Y., & Liu, D. (2024). Enhancing user privacy through ephemeral sharing design: Experimental evidence from online dating. *Information Systems Research*. <https://doi.org/10.1287/sre.2021.0379>
 31. Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review*, *33*(3), 405-431.
 32. Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human-human online conversations and human-Chatbot conversations. *Computers in Human Behavior*, *49*, 245-250.
 33. Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a Chatbot. *Journal of Communication*, *68*(4), 712-733.
 34. Hsu, P. F., Nguyen, T., Wang, C. Y., & Huang, P. J. (2023). Chatbot commerce—How contextual factors affect Chatbot effectiveness. *Electronic Markets*, *33*(1), Article 14.
 35. John, L. K., Acquisti, A., & Loewenstein, G. (2011). Strangers on a plane: Context-dependent willingness to divulge sensitive information. *Journal of Consumer Research*, *37*(5), 858-873.
 36. Keil, M., Tan, B. C., Wei, K.-K., Saarinen, T., Tuunainen, V., & Wassenaar, A. (2000). A cross-cultural study on escalation of commitment behavior in software projects. *MIS Quarterly*, *24*(2), 299-325.
 37. Kelly, D., & Teevan, J. (2003). Implicit feedback for inferring user preference: A bibliography. *ACM Sigir Forum*, *37*, 18-28.
 38. Kim, J., & Dindia, K. (2011). Online self-disclosure: A review of research. In K. B. Wright & L. M. Webb (Eds.), *Computer-mediated communication in personal relationships* (pp. 156-180). Peter Lang Publishing.
 39. Kim, J., Merrill Jr, K., Xu, K., & Kelly, S. (2022). Perceived credibility of an AI instructor in online education: The role of social presence and voice features. *Computers in Human Behavior*, *136*, 107383.
 40. Kim, J., & Song, H. (2016). Celebrity's self-disclosure on Twitter and parasocial relationships: A mediating role of social presence. *Computers in Human Behavior*, *62*, 570-577.
 41. Kim, S., Lee, J., & Gweon, G. (2019). Comparing data from Chatbot and web surveys: Effects of platform and conversational style on survey response quality. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, *19*, 1-12.
 42. Kwak, C., Lee, J., & Lee, H. (2022). Could you ever forget me? Why people want to be forgotten online. *Journal of Business Ethics*, *179*, 25-42.
 43. LaMothe, E., & Bobek, D. (2020). Are individuals more willing to lie to a computer or a human? Evidence from a tax compliance setting. *Journal of Business Ethics*, *167*(2), 157-180.
 44. Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information*

- Systems*, *16*(10), 880–918.
45. Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on human-computer interaction*, *2*(CSCW), 1–31.
 46. Lee, J., Lee, D., & Lee, J. G. (2024). Influence of rapport and social presence with an AI psychotherapy chatbot on users' self-disclosure. *International Journal of Human-Computer Interaction*, *40*(7), 1620–1631.
 47. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80.
 48. Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 2053951718756684.
 49. Lee, Y. C., Yamashita, N., Huang, Y., & Fu, W. (2020). "I hear you, I feel you": Encouraging deep self-disclosure through a chatbot. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12.
 50. Li, J. (2023). Security implications of AI Chatbots in health care. *Journal of Medical Internet Research*, *25*, e47551.
 51. Liang, H., Saraf, N., Hu, Q., & Xue, Y. (2007). Assimilation of enterprise systems: The effect of institutional pressures and the mediating role of top management. *MIS Quarterly*, *31*(1), 59–87.
 52. Logsdon, J. M., & Patterson, K. D. (2009). Deception in business networks: Is it easier to lie online? *Journal of Business Ethics*, *90*, 537–549.
 53. Lucas, G. M., Gratch, J., King, A., & Morency, L. P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100.
 54. Lwin, M., Wirtz, J., & Williams, J. D. (2007). Consumer online privacy concerns and responses: A power-responsibility equilibrium perspective. *Journal of the Academy of Marketing Science*, *35*(4), 572–585.
 55. Martin, K. D., Borah, A., & Palmatier, R. W. (2017). Data privacy: Effects on customer and firm performance. *Journal of Marketing*, *81*(1), 36–58.
 56. McGuire, J., De Cremer, D., Hesselbarth, Y., De Schutter, L., Mai, K. M., & Van Hiel, A. (2023). The reputational and ethical consequences of deceptive Chatbot use. *Scientific Reports*, *13*(1), 16246.
 57. McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334–359.
 58. Menard, P., & Bott, G. J. (2024). Artificial intelligence misuse and concern for information privacy: New construct validation and future directions. *Information Systems Journal*, 1–46.
 59. Murtarelli, G., Gregory, A., & Romenti, S. (2021). A conversation-based perspective for shaping ethical human-machine interactions: The particular challenge of Chatbots. *Journal of Business Research*, *129*, 927–935.
 60. Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72–78.
 61. Nguyen, Q. N., Sidorova, A., & Torres, R. (2022). User interactions with Chatbot interfaces vs. Menu-based interfaces: An empirical study. *Computers in Human Behavior*, *128*, 107093.
 62. Nichols, D. M. (1997). Implicit rating and filtering. *Proceedings of the 5th DELOS Workshop on Filtering and Collaborative Filtering*, 31–36.
 63. Nicol, E., Briggs, J., Moncur, W., Htait, A., Carey, D. P., Azzopardi, L., & Schafer, B. (2022). Revealing cumulative risks in online personal information: A data narrative study. *Proceedings of the ACM on Human-Computer Interaction*, *6*, 1–25.
 64. Oard, D. W., & Kim, J. (1998). Implicit feedback for recommender systems. *Proceedings of the AAAI Workshop on Recommender Systems*, 31–36.
 65. Ozdemir, Z. D., Jeff Smith, H., & Benamati, J. H. (2017). Antecedents and outcomes of information privacy concerns in a peer context: An exploratory study. *European Journal of Information Systems*, *26*(6), 642–660.
 66. Park, G., Yim, M. C., Chung, J., & Lee, S. (2022). Effect of AI Chatbot empathy and identity disclosure on

- willingness to donate: The mediation of humanness and social presence. *Behaviour & Information Technology*, *42*(12), 1998–2010.
67. Pataranutaporn, P., Liu, R., Finn, E., & Maes, P. (2023). Influencing human-AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness. *Nature Machine Intelligence*, *5*(10), 1076–1086.
 68. Pavlou, P. A., Liang, H., & Xue, Y. (2007). Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective. *MIS Quarterly*, *31*, 105–136.
 69. Peng, K. (2020). To be attractive or to be authentic? How two competing motivations influence self-presentation in online dating. *Internet Research*, *30*(4), 1143–1165.
 70. Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, *12*(4), 531–544.
 71. Posey, C., Lowry, P. B., Roberts, T. L., & Ellis, T. S. (2010). Proposing the online community self-disclosure model: The case of working professionals in France and the UK who use online communities. *European Journal of Information Systems*, *19*(2), 181–195.
 72. Punj, G. N. (2019). Understanding individuals' intentions to limit online personal information disclosures to protect their privacy: Implications for organizations and public policy. *Information Technology and Management*, *20*, 139–151.
 73. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.
 74. Rohden, S. F., & Zeferino, D. G. (2023). Recommendation agents: An analysis of consumers' risk perceptions toward artificial intelligence. *Electronic Commerce Research*, *23*(4), 2035–2050.
 75. Rossmannek, O., David, N. A., Schramm-Klein, H., & Van der Borgh, M. (2024). Customer misbehavior and service providers' risk perception in the sharing economy. *Journal of Business Research*, *170*, 114340.
 76. Sanchez-Franco, M. J. (2006). Exploring the influence of gender on the web usage via partial least squares. *Behaviour & Information Technology*, *25*(1), 19–36.
 77. Sannon, S., Bazarova, N. N., & Cosley, D. (2018). Privacy lies: Understanding how, when, and why people lie to protect their privacy in multiple online contexts. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, Canada* (pp.1–13).
 78. Schuetzler, R. M., Grimes, G. M., & Giboney, J. S. (2019). The effect of conversational agent skill on user behavior during deception. *Computers in Human Behavior*, *97*, 250–259.
 79. Shams, G., Kim, K. K., & Kim, K. (2024). Enhancing service recovery satisfaction with Chatbots: The role of humor and informal language. *International Journal of Hospitality Management*, *120*, 103782.
 80. Sharabi, L. L. (2021). Exploring how beliefs about algorithms shape (offline) success in online dating: A two-wave longitudinal investigation. *Communication Research*, *48*(7), 931–952.
 81. Sharabi, L. L., & Caughlin, J. P. (2019). Deception in online dating: Significance and implications for the first offline date. *New Media & Society*, *21*(1), 229–247.
 82. Shin, D. (2020). How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance. *Computers in Human Behavior*, *109*, 106344.
 83. Shin, D. (2022). Expanding the role of trust in the experience of algorithmic journalism: User sensemaking of algorithmic heuristics in Korean users. *Journalism Practice*, *16*(6), 1168–1191.
 84. Shin, D., Kee, K. F., & Shin, E. Y. (2022). Algorithm awareness: Why user awareness is critical for personal privacy in the adoption of algorithmic platforms? *International Journal of Information Management*, *65*, 102494.
 85. Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, *98*, 277–284.
 86. Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Wiley.
 87. Simmons, M., & Lee, J. S. (2020). Catfishing: A look

- into online dating and impersonation. In *International Conference on Human-Computer Interaction*. Springer, 349–358.
88. Song, M., Xing, X., Duan, Y., Cohen, J., & Mou, J. (2022). Will artificial intelligence replace human customer service? The impact of communication quality and privacy risks on adoption intention. *Journal of Retailing and Consumer Services*, *66*, 102900.
 89. Sprecher, S., & Hendrick, S. S. (2004). Self-disclosure in intimate relationships: Associations with individual and relationship characteristics over time. *Journal of Social and Clinical Psychology*, *23*(6), 857–877.
 90. Sundar, S. S., Kang, H., Wu, M., Go, E., & Zhang, B. (2013). Unlocking the privacy paradox: Do cognitive heuristics hold the key? In *CHI'13 extended abstracts on Human Factors in Computing Systems* (pp.811–816).
 91. Sundar, S. S., & Kim, J. (2019). Machine heuristic: When we trust computers more than humans with our personal information. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–9.
 92. Sundar, S. S., Kim, J., Rosson, M. B., & Molina, M. D. (2020). Online privacy heuristics that predict information disclosure. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*.
 93. Turner, C. F., Villarroel, M. A., Rogers, S. M., Eggleston, E., Ganapathi, L., Roman, A. M., & Al-Tayyib, A. (2005). Reducing bias in telephone survey estimates of the prevalence of drug use: A randomized trial of telephone audio-CASI. *Addiction*, *100*(10), 1432–1444.
 94. Tzeng, J. Y. (2004). Toward a more civilized design: studying the effects of computers that apologize. *International Journal of Human-Computer Studies*, *61*(3), 319–345.
 95. Van Zant, A. B., & Kray, L. J. (2014). “I can’t lie to your face”: Minimal face-to-face interaction promotes honesty. *Journal of Experimental Social Psychology*, *55*, 234–238.
 96. Wang, H. (2023). Algorithmic colonization of love: The ethical challenges of dating App algorithms in the age of AI. *Techne: Research in Philosophy & Technology*, *27*(2), 260–280.
 97. Willis, M. L., Oliver, E., & March, E. (2023). Dating in the dark: Vulnerable narcissism predicts inauthentic self-presentation in online dating. *Telematics and Informatics*, *81*, 101985.
 98. Wilson, R. E., Gosling, S. D., & Graham, L. T. (2012). A review of Facebook research in the social sciences. *Perspectives on Psychological Science*, *7*(3), 203–220.
 99. Xie, W., & Kang, C. (2015). See you, see me: Teenagers’ self-disclosure and regret of posting on social network site. *Computers in Human Behavior*, *52*, 398–407.
 100. Xu, H., & Dinev, T. (2022). Guest editorial: Reflections on the 2021 impact award: Why privacy still matters. *MIS Quarterly*, *46*(4), 20–32.
 101. Xu, H., Luo, X. R., Carroll, J. M., & Rosson, M. B. (2011). The personalization privacy paradox: An exploratory study of decision making process for location-aware marketing. *Decision Support Systems*, *51*(1), 42–52.
 102. Zhang, L., Pentina, I., & Fan, Y. (2021). Who do you choose? Comparing perceptions of Human vs Robo-advisor in the context of financial services. *Journal of Services Marketing*, *35*(5), 634–646.

〈부록 A〉 설문 문항

측정개념	변수	문항	참고문헌							
알고리즘에 대한 신뢰	신뢰1 (생략됨)	(1) 기계가 작업을 수행하면, 사람이 같은 작업을 수행할 때보다 결과가 더 객관적이다.	Sundar and Kim (2019)							
	신뢰2	(2) 기계는 안전한 방식으로 정보를 처리할 수 있으므로, 내 개인 정보를 기계에 공개해도 괜찮다.								
	신뢰3	(3) 기계는 정밀도가 높기 때문에, 내 개인 정보를 안전하게 처리할 것이다.								
	신뢰4	(4) 기계는 험담을 하지 않으므로, 다른 사람과 내 개인 정보를 공유하지 않을 것이다.								
	신뢰5	(5) 사람보다 기계에게 개인정보를 공개하는 것이 더 안전하다.								
대화형 에이전트에 대한 사회적 실재감	실재감1	(1) 이 에이전트에게 사람과 접촉하는 느낌이 있다.	Pavlou et al. (2007)							
	실재감2	(2) 이 에이전트에게 개인적인 친밀감이 느껴진다.								
	실재감3	(3) 이 에이전트에게 인간적인 따뜻함이 느껴진다.								
	실재감4	(4) 이 에이전트에게 인간적인 감성이 느껴진다.								
대화형 에이전트를 통한 개인정보 제공의 지각된 위험	위험1	(1) 데이팅 서비스를 위해 이 에이전트에 개인정보를 제공하는 것은 예상치 못한 다양한 문제를 발생시킬 수 있다.	Xu et al. (2011)							
	위험2	(2) 이 에이전트를 통해 데이팅 서비스 제공업체에 개인정보를 공개하는 것은 위험할 수 있다.								
	위험3	(3) 이 에이전트를 통해 데이팅 서비스 제공업체에 개인정보를 공개할 경우 손실 가능성이 높다.								
부정확한 정보 제공 의도	부정확1	(1) 나는 이 에이전트에게 본인의 키와 몸무게를 제공하지 않기 위해 가상의 답변을 구성하는 것을 고려할 것이다.	Lwin et al. (2007)							
	부정확2	(2) 나는 이 에이전트에게 키와 몸무게를 제공할 때 다른 이름이나 웹/이메일 주소를 사용하여 실제 신원을 밝히지 않고 등록된 사용자로서 모든 액세스 권한과 혜택을 누릴 수 있도록 할 것이다.								
	부정확3	(3) 나는 이 에이전트에게 키와 몸무게를 제공할 때 부분적으로만 데이터를 입력할 것이다.								
기 공개된 개인정보 수준	공개정보1	아래 항목 중 인터넷에서 검색할 수 있는 모든 항목에 체크하세요. 즉, 해당 항목을 검색할 수 있나요? (누가 정보를 공개했는지는 중요하지 않다.)	Punj (2019)							
		<table border="0"> <tr> <td>a. 귀하의 이메일 주소</td> <td>f. 귀하의 정당 또는 소속 정당</td> </tr> <tr> <td>b. 집 주소</td> <td>g. 귀하의 이름이 적혀있는 것들</td> </tr> <tr> <td>c. 집 전화번호</td> <td>h. 귀하의 사진</td> </tr> <tr> <td>d. 휴대폰 번호</td> <td>i. 귀하의 동영상</td> </tr> <tr> <td>e. 귀하의 고용주 또는 귀하가 근무하는 회사</td> <td>j. 귀하가 속한 그룹 또는 조직</td> </tr> </table>		a. 귀하의 이메일 주소	f. 귀하의 정당 또는 소속 정당	b. 집 주소	g. 귀하의 이름이 적혀있는 것들	c. 집 전화번호	h. 귀하의 사진	d. 휴대폰 번호
a. 귀하의 이메일 주소	f. 귀하의 정당 또는 소속 정당									
b. 집 주소	g. 귀하의 이름이 적혀있는 것들									
c. 집 전화번호	h. 귀하의 사진									
d. 휴대폰 번호	i. 귀하의 동영상									
e. 귀하의 고용주 또는 귀하가 근무하는 회사	j. 귀하가 속한 그룹 또는 조직									

〈부록 B〉 온라인 데이팅 서비스 시나리오 예시

시나리오 1: 챗봇	
챗봇:	안녕, 친구! 저는 소개팅 챗봇 알렉스입니다. 마음에 드는 이성을 찾고 있나요? (안녕하세요, 소개팅 챗봇입니다. 함께 데이트할 상대를 추천해도 될까요?)
사용자:	네.
챗봇:	좋아요! 데이트를 성사시키기 위해 몇 가지 질문을 할게요. 계속할까요? (상대를 찾으려면 개인 정보를 제공해야 합니다. 계속하시겠습니까?)
사용자:	물론이죠.
시나리오 2: 인간	
카운슬러:	안녕하세요! 저는 여러분의 개인 매칭 카운슬러 알렉스입니다. 마음에 드는 이성을 찾고 있나요? (안녕하세요, 저는 여러분의 개인 매칭 카운슬러입니다. 함께 데이트할 상대를 추천해도 될까요?)
사용자:	네.
카운슬러:	좋아요! 데이트를 성사시키기 위해 몇 가지 질문을 할게요. 계속할까요? (상대를 찾으려면 개인 정보를 제공해야 합니다. 계속하시겠습니까?)
사용자:	물론이죠.

〈부록 C〉 탐색적 요인 분석 결과

변수	(챗봇, n=180)					(인간, n=87)				
	1	2	3	4	5	1	2	3	4	5
신뢰2	.622	.265	-.296	.225	.040	.740	.173	.222	.105	.146
신뢰3	.714	.276	-.116	.209	.103	.822	.194	.041	.056	-.021
신뢰4	.819	.099	.038	-.105	-.062	.715	.147	.393	.159	-.046
신뢰5	.806	.185	.033	-.004	.123	.861	.226	-.127	.130	-.089
실재감1	.199	.816	.138	.135	.108	.148	.809	.056	.078	-.025
실재감2	.174	.815	.072	.108	.005	.182	.795	.121	.152	-.034
실재감3	.146	.836	.112	.221	.050	.193	.839	.115	.077	.006
실재감4	.172	.844	.077	.124	.040	.185	.831	.074	.135	.023
위험1	-.070	.153	.796	.231	.089	.116	.098	.635	.294	-.181
위험2	-.058	.148	.673	.332	-.015	.081	.153	.766	.337	-.112
위험3	-.035	.195	.650	.369	.101	.200	.227	.639	.301	-.084
부정확1	.041	.178	.250	.773	.032	.134	.176	.231	.821	-.112
부정확2	.022	.258	.152	.850	.076	.074	.175	.335	.835	-.123
부정확3	.071	.117	.218	.861	.064	.190	.093	.149	.864	.024
공개정보1	-.103	-.115	-.093	-.113	-.956	-.015	-.010	-.174	-.124	.952

〈부록 D〉 동일 방법 편의 분석 결과

동일 방법 편의 분석 결과 (챗봇, n=180)					
측정개념	변수	Factor loading on construct (R1)	R1 ²	Factor loading on method (R2)	R2 ²
알고리즘에 대한 신뢰	신뢰2	0.731 ^{***}	0.534	0.136 [*]	0.018
	신뢰3	0.840 ^{***}	0.706	0.027	0.001
	신뢰4	0.866 ^{***}	0.750	-0.151	0.023
	신뢰5	0.869 ^{***}	0.755	-0.024	0.001
대화형 에이전트에 대한 사회적 실재감	실재감1	0.830 ^{***}	0.689	0.008	0.000
	실재감2	0.946 ^{***}	0.895	-0.023	0.001
	실재감3	0.921 ^{***}	0.848	-0.045	0.002
	실재감4	0.916 ^{***}	0.839	0.070	0.005
개인정보 제공에 대한 지각된 위험	위험1	0.919 ^{***}	0.845	0.105	0.011
	위험2	0.924 ^{***}	0.854	-0.110	0.012
	위험3	0.852 ^{***}	0.726	-0.013 [*]	0.000
부정확한 정보제공 의도	부정확1	0.864 ^{***}	0.746	0.071	0.005
	부정확2	0.898 ^{***}	0.806	0.050	0.003
	부정확3	1.012 ^{***}	1.024	-0.122 ^{**}	0.015
기 공개된 개인 정보 수준	공개정보1	1		0	
평균			0.787		0.007
동일 방법 편의 분석 결과 (인간, n=87)					
측정개념	변수	Factor loading on construct (R1)	R1 ²	Factor loading on method (R2)	R2 ²
알고리즘에 대한 신뢰	신뢰2	0.817 ^{***}	0.667	0.049	0.002
	신뢰3	0.896 ^{***}	0.803	-0.019	0.000
	신뢰4	0.767 ^{***}	0.588	0.077	0.006
	신뢰5	0.920 ^{***}	0.846	-0.108	0.012
대화형 에이전트에 대한 사회적 실재감	실재감1	0.864 ^{***}	0.746	0.032	0.001
	실재감2	0.846 ^{***}	0.716	0.026	0.001
	실재감3	0.902 ^{***}	0.814	-0.018	0.000
	실재감4	0.936 ^{***}	0.876	-0.038	0.001
개인정보 제공에 대한 지각된 위험	위험1	0.890 ^{***}	0.792	-0.001	0.000
	위험2	0.917 ^{***}	0.841	-0.026	0.001
	위험3	0.847 ^{***}	0.717	0.028	0.001
부정확한 정보제공 의도	부정확1	0.906 ^{***}	0.821	0.062	0.004
	부정확2	0.911 ^{***}	0.830	0.038	0.001
	부정확3	0.999 ^{***}	0.998	-0.104	0.011
기 공개된 개인 정보 수준	공개정보1	1		0	
평균			0.790		0.003

저 자 소 개



곽 찬 희 (Chanhee Kwak)

강남대학교 인공지능융합공학부 데이터사이언스전공 조교수로 재직 중이다. KAIST 경영대학에서 박사학위를 취득하였다. 주요 관심분야는 Data Analytics, Information Privacy 등이다. Journal of Management Information Systems, Journal of Business Ethics, International Journal of Information Management, 등 학술지에 논문을 발표하였다.



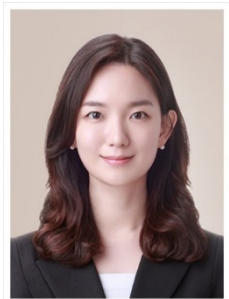
이 준 영 (Junyeong Lee)

충북대학교 경영정보학과 부교수로 재직 중이다. KAIST 경영대학에서 박사학위 취득 후 중국과학기술대학(USTC) 및 한국기술교육대학교 조교수를 역임하였다. 주요 관심분야는 collective dynamics and human behavior in IS이다. Journal of Management Information Systems, Journal of the Association for Information Systems, Journal of Business Ethics, International Journal of Information Management, International Journal of Electronic Commerce, Communications of the ACM 등의 학술지에 논문을 발표하였다.



민 진 영 (Jinyoung Min)

중앙대학교 경영경제대학 산업보안학과에서 부교수로 재직 중이며 KAIST 경영대학에서 경영정보시스템 전공으로 박사학위를 취득하였다. 주요 관심 분야는 Privacy, Algorithm Automation and Platform 등이다. Computers in Human Behavior, International Journal of Information Management, Journal of the Association for Information Science and Technology, Communication of the ACM, 경영학연구, 경영정보학연구 등에 논문을 게재한 바 있다.



최 한 별 (HanByeol Stella Choi)

명지대학교 경영대학 경영정보학과 조교수로 재직 중이다. KAIST 경영대학 박사학위 취득 후 KAIST 경영대학 디지털혁신연구센터 연구조교수로 재직하였다. 주요 관심분야는 프라이버시, 정보 보안, 공유 경제, 정보 시스템의 사회적 효과 및 비즈니스 애널리틱스 등이다. 지금까지 Journal of Management Information Systems, Decision Support System, Security Journal 등 주요 학술지에 논문을 발표하였다.

〈 Abstract 〉

A Study of Influencing Factors for Intentional Inaccurate Information Provision in Conversations with Chatbots: In the Context of Online Dating Services

Chanhee Kwak^{*}, Junyeong Lee^{**}, Jinyoung Min^{***}, HanByeol Stella Choi^{****}

Chatbots are becoming increasingly popular as interactive communication tools that provide not only convenience but also a friendly and humanized experience. Due to the interactive nature of chatbots, they can exchange information with users to perform various tasks, and users sometimes intentionally provide inaccurate information. Considering social presence of conversational agents, perceived risk of providing personal information, and trust in algorithms as key influencing factors, this study explores the effects of those factors on the intention to provide inaccurate information in the context of online dating services and examine whether these effects vary across types of conversational agents. We conducted an analysis of structural equation model using data collected from Amazon Mechanical Turk (MTurk). The analysis results showed significant relationships between factors related to the intention to provide inaccurate information and empirically confirmed that those relationships vary by types of conversational agents. Our findings have academic implications for the behavior of providing inaccurate information in online environments and practical implications for designing chatbots to reduce such intentions. We also discuss the ethical implications of the consequences of inaccurate information online.

Key words: Chatbot, Trust in algorithms, Social presence, Perceived risk, Inaccurate information provision

* Assistant Professor, Department of Artificial Intelligence Convergence, Kangnam University

** Associate Professor, Department of Management Information Systems, Chungbuk National University

*** Associate Professor, Department of Industrial Security, Chung-Ang University

**** Assistant Professor, Department of Management Information Systems, Myongji University