

챗봇의 오류 유형과 오류 메시지 구조화 여부가 사용자 경험에 미치는 영향

The Effects of Chatbot's Error Types and Structures of Error Message on User Experience

이미진, 한광희
연세대학교 심리학과

Mi-Jin Lee(mijinleecc@gmail.com), Kwang-Hee Han(khan@yonsei.ac.kr)

요약

본 연구는 챗봇의 오류의 유형과 오류 메시지의 구조화 여부에 따른 태도, 행동의도, 사용성의 차이를 검증하고자 하였다. 챗봇의 오류 유형은 마음지각 이론을 기반으로 기대수준이 서로 다른 경험요소 오류와 기능요소 오류로 나누었으며 오류 메시지는 오류 사실만을 명시하는 비구조화 조건과 사과 표현, 오류의 원인, 개선 의지를 포함한 구조화된 조건으로 나누었다. 오류 유형에 따라 사용성 점수에 차이가 나타났으며, 구조화된 오류 메시지 조건에서 태도, 행동의도, 사용성 점수가 높게 나타났다. 경험요소 기대격차는 세 종속 변인을 예측하지 못했으나, 기능요소 기대격차는 세 종속 변인을 유의미하게 예측하였으며 기능요소 기대격차에 대하여 오류 유형과 오류 메시지 구조화 여부의 상호작용 경향성이 확인되었다. 본 연구는 구조화된 오류 메시지의 완충효과를 검증하였고 오류 유형에 따라 완충효과가 달라질 가능성을 확인하였다. 이 결과는 사용자 경험을 증진하는 오류대처전략설계에 적용가능할 것으로 기대한다.

■ 중심어 : | 챗봇 | 오류 | 사용자 경험 | 마음지각 | 기대위반 |

Abstract

The aim of this study is verifying the effects of chatbot's error types and structures of error message on attitude, behavior intention towards the chatbot and perceived usability of the chatbot. The error types of chatbot are divided into 'experience' error and 'agency' error, which set different expectancy level, according to mind perception theory. The structures of error message were either unstructured condition composed of error specification only or structured condition composed of apology, explanation and willingness of improvement. It was found that score of perceived usability was higher in experience error condition than agency error condition. Also, all three scores of dependent variables were higher in structured error message condition than unstructured error message condition. Furthermore, expectation gap of experience didn't predict the dependent variables but expectation gap of agency predicted all three dependent variables. Finally, the tendency of interaction effect between the error type and the structure of the error message on expectation gap of agency was observed. This study confirmed the mitigating effect of structured error messages and the possibility that these effects may vary by the type of error. The result is expected to be applicable to design of error coping strategies that enhance user experience.

■ keyword : | Chatbot | Error | User Experience | Mind Perception | Expectancy Violation |

* 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2019S1A5C2A03083499)

접수일자 : 2021년 02월 19일
수정일자 : 2021년 03월 09일

심사완료일 : 2021년 03월 10일
교신저자 : 한광희, e-mail : khan@yonsei.ac.kr

I. 서론

챗봇은 언어적인 대화를 통해 사용자와 상호작용하는 소프트웨어 에이전트다[1]. 초기의 챗봇은 등록된 키워드가 입력되면 정의된 응답을 출력하는 단순한 패턴 매칭 구조에 머물렀으나[2], 현재의 챗봇은 발달된 인공지능 기술을 기반으로 하여 다양한 자연어 처리가 가능하다. 챗봇은 고객 관리나 상담 그리고 예약과 같이 업무의 자동화를 위해 산업 분야에서 활용되고 있으며 챗봇의 활용과 역할의 범위는 점차 확장될 것이다. 현재 국내의 챗봇의 대부분은 검색을 통한 결과를 제공하는 단순한 형태의 소통이 가능한 1단계 수준과 사용자의 패턴과 상황을 고려하여 간단한 업무를 처리하는 2단계 수준에 머물러 있으나, 향후 감성인지 기술이 더해져 사용자와의 감정적 교류까지 가능한 3단계 수준으로 발전할 것으로 기대된다[3].

그러나 여전히 챗봇은 자연어로 이루어지는 복잡한 상호작용을 완벽하게 처리하지는 못한다. 사용자는 챗봇과의 상호작용에서 여러 오류를 경험할 수 있고, 오류의 경험은 사용자가 챗봇을 어떻게 지각하고 상호작용하는지에 큰 영향을 미친다[4]. 사용자는 오류가 발생하는 에이전트의 수행수준을 오류를 발생시키지 않는 에이전트의 수행수준보다 낮게 지각하게 된다. 수행수준이 기대수준보다 낮게 지각되면 부적 기대위반이 발생하고 이는 곧 에이전트에 대한 부정적인 태도 및 평가로 이루어진다. 이때, 오류로 인하여 발생하는 기대위반의 수준은 상호작용 이전에 사용자가 에이전트에게 가지는 기대수준에 따라 달라지는데[4][5], 기대감이 높을수록 기대위반 가능성이 높아지고 오류로 인한 부정적인 영향은 커지게 된다[7].

챗봇과의 상호작용에서 사용자의 기대감은 어떻게 형성될까? 사회적 행위자로서의 컴퓨터(Computer Are Social Actor; CASA) 패러다임에 따르면 사람은 에이전트도 사회적인 소통의 대상으로 인식한다[8]. 그러나 에이전트의 수행수준을 인간과 같은 수준으로 기대하는 것은 아니다. 마음지각이론에 따르면 에이전트의 기능적인 요소에 대한 기대는 높지만, 경험적인 요소에 대한 기대수준은 낮다[9]. 이러한 기대수준의 차이에 따라 해당 요소의 오류가 상호작용에 미치는 부정적

인 영향이 달라질 수 있다.

에이전트의 서로 다른 오류 유형이 사용자의 평가 및 태도에 미치는 영향을 이해하고 이에 대한 대처전략을 구성하는 것은 중요하다. 그러나, 대화형 에이전트에 있어서 여러 유형의 오류가 발생할 수 있음에도 불구하고 하나의 에이전트 내에서 발생 가능한 서로 다른 유형의 오류에 따른 사용자 경험의 차이에 관한 연구는 부족한 현실이다. 따라서 본 연구는 대화형 에이전트의 오류 유형을 마음지각 이론을 기반으로 경험요소 오류와 기능요소 오류로 나누어 기대수준이 서로 다른 요소의 오류가 에이전트에 대한 사용자의 태도와 행동의도 그리고 사용성에 영향을 미치는지 알아보고자 하였다. 또한, 오류에 후속되는 오류 메시지의 완충효과를 메시지의 구조화 여부에 따라 나누어 살펴보고자 하였다.

II. 이론적 배경 및 선행연구

1. 에이전트의 오류와 기대위반

인간은 타인과 상호작용을 하기 이전에 기대를 형성한다[10][11]. 기기나 에이전트와의 상호작용에서도 사용자는 정신모형(mental model)에 따라 기대수준을 설정하며 기대수준은 상호작용의 지각에 영향을 준다[6][7][12-14]. 상호작용 대상의 수행수준이 기대와 다를 때 기대위반이 발생하게 되는데 수행의 수준이 기대에 못 미치면 부적 기대 위반(negative violation)이 발생하고 수행의 수준이 기대를 능가하면 정적 기대 위반(positive violation)이 발생한다[11].

에이전트의 오류는 지각된 에이전트의 수행수준을 낮추어 부적 기대위반을 유발하고, 사용자 경험에 부정적인 영향을 미친다. 이때, 에이전트의 오류로 인한 부정적인 영향은 상호작용 이전에 설정된 기대수준에 따라 달라진다. Lee, Kiesler, Forlizzi, Srinivasa, & Rybski(2010)[13]의 연구에 따르면 서비스 로봇의 서비스 오류는 로봇과 로봇의 서비스에 대한 평가에 부정적인 영향을 미치며 서비스 오류가 평가에 미치는 부정적인 영향은 사전기대 수준에 따라 달라지는 것으로 나타났다. 상호작용 이전에 문구를 통해 로봇에게 해당 과업이 고난도라는 점과 오류가 발생할 가능성이 있다

는 점을 명시하여 수행수준에 대한 기대를 낮춘 조건에서 로봇의 오류로 인한 부정적인 영향이 감소하였다. Komatsu, Kurosawa, & Yamada(2012)[6]의 연구에서는 선택게임에서 로봇의 예상답안을 듣고 로봇의 의견을 따를지를 결정하는 실험을 진행하였다. 사전 연습 시행에서 에이전트의 적중률을 90% 혹은 10%로 설정하여 참가자의 기대수준을 조작하였다. 본 시행에서는 에이전트의 적중률을 33%로 동등하게 통제하고 참가자가 에이전트의 의견을 얼마나 수용하는지 확인하였다. 그 결과, 사전 시행에서 적중률을 10%로 설정하여 기대수준을 낮게 조작했던 집단이 사전 시행의 적중률을 90%로 높게 조작했던 집단보다 본 시행에서 로봇 의견의 수용 빈도가 2배 가까이 높은 것으로 나타났다. Go, & Sundar(2019)[14]의 연구에서는 챗봇과의 상호작용에서 대화가 진행되기 이전에 제시된 시각적인 단서(챗봇 혹은 사람)와 대화가 시작된 이후 밝혀진 대상의 정체성이 일치하지 않는 경우 대상에 대해 부정적인 태도를 보이며 더 낮은 행동의도를 나타냈다. 연구자들은 상호작용 이전에 정체성을 암시하는 시각적인 단서로 인해 형성된 정체성에 대한 기대위반을 부정적 태도의 원인으로 해석하였다.

선행연구들을 종합하면 에이전트의 기대위반은 에이전트에 대한 태도와 평가에 부정적인 영향을 미치며, 에이전트의 의견을 수용하거나 상호작용을 지속할 의도를 낮춘다. 이때 오류가 미치는 부정적인 영향은 상호작용 이전에 설정된 기대수준에 따라 달라지는데, 높은 사전기대수준은 부적 기대위반이 발생할 가능성을 더욱 높이며 오류가 사용자 경험에 미치는 부정적인 영향을 더욱 크게 만든다.

2. 마음지각 이론에 따른 기대수준과 오류의 영향

인간은 에이전트를 사회적인 소통의 대상으로 인식하지만[8], 에이전트에 대한 기대수준은 인간에 대한 기대수준과는 다르게 설정된다[15].

마음지각이론에 따르면 마음은 경험(experience)과 기능(agency)의 2개의 서로 독립된 차원으로 지각된다. 경험 차원은 배고픔, 고통, 즐거움 등에 해당하고 기능 차원은 자아통제, 도덕성, 기억, 감정 이해, 계획 수립과 관련된 것이다[9]. 즉, 마음지각에서의 경험요소는

감정을 느낄 수 있는 능력을 의미하며 기능요소는 행동을 수행하는 능력을 의미한다. 로봇의 경험 차원과 기능 차원은 모두 성인 인간보다 낮은 수준으로 지각되는데 특히 로봇의 감정을 느끼고 표현하는 경험 차원이 자신을 통제하고 계획을 수립하는 기능 차원보다 현저히 낮게 지각된다[9]. 마음지각이론을 기반으로 한 HRI(Human Robot Interaction) 연구에 따르면 경험요소에 해당하는 감정을 느끼고 표현하는 요소에 대해서는 에이전트의 수행수준이 낮을 것으로 기대하는 반면, 에이전트 자신을 통제하거나 계획을 수립하는 등의 행동과 관련된 수준을 높을 것으로 기대한다 [16][17]. Appel, Izydorzcyk, Weber, Mara,& Lischetzke(2020)[16]의 연구에서는 로봇의 기대수준을 안내 문구로 조작하고 지각된 기능요소 수준과 경험요소 수준을 평가하도록 하였다. 이때 도구 조건, 기능요소 조건, 경험요소 조건으로 조작하였는데, 경험요소와 기능요소를 강조하지 않은 일반적인 도구 조건에서 기능요소의 수준을 경험요소의 수준보다 높게 평가하는 것으로 나타났다. Yam, Bigman, Ilies, De Cremer, Soh, & Gray(2020)[17]는 로봇의 얼굴 제시 여부에 따른 의인화가 마음지각의 정도를 높이는 것을 확인하였으며 얼굴 제시 여부와 관계없이 모든 조건에서 기능요소를 지각된 경험요소보다 높은 수준으로 평가했다.

마음지각이론을 기반으로한 선행연구들에 따르면 상호작용 이전에 로봇에 대한 마음지각수준은 경험요소보다 기능요소가 더 높다. 챗봇과의 상호작용에서도 마음지각 요소 간 기대수준의 차이가 존재할 수 있으며 이 경우 서로 다른 기대수준에 따라 해당 요소의 오류가 사용자 경험에 미치는 영향에 차이가 나타날 수 있다.

3. 에이전트의 오류와 구조화된 메시지의 완충효과

챗봇과 같은 에이전트를 설계하고 개발할 때 직면하는 가장 큰 도전 중 하나는 오류의 부정적인 효과를 경감시킬 수 있는 적합한 전략을 세우는 것이다[18]. 여러 선행연구에서는 에이전트의 오류로 인해 유발되는 부정적인 감정과 반응이 에이전트의 예의 바른 사과 행동을 통해 완화될 수 있음을 검증하였으며[19-22], 사용

자가 선호하는 오류의 메시지의 내용과 구조를 확인하였다[19][22][23].

Akgun, Cagiltay, & Zeyrek(2010)[19]은 e-러닝 환경에서 참가자에게 오류를 경험하게 한 뒤, 오류 상황에서 받기를 원하는 사과 메시지의 문장을 주관식으로 기록하게 하였다. 응답을 유형에 따라 나누어 최종적으로 결과 책임명시, 사과, 오류의 이유설명, 개선 약속, 대책 및 보상 제시의 5개의 유형으로 분류하였다. 이 중 책임명시와 사과가 가장 선호되는 요소였으며 이는 전체 참가자의 50% 이상이 언급하였다. 이로써 사람 간 상호작용에서 직접적인 사과 발언과 책임을 명시하는 것이 전반적인 상황에서 선호된다는 선행연구[24][25]가 인간-컴퓨터 상호작용에도 적용됨을 확인하였다. Park, MacDonald, & Khoo(2012)[20]의 연구에서는 전자도서관 및 e-러닝 시스템에서 오류 메시지를 사과유형, 중립 유형, 비 사과유형으로 나누어 제시하여 시스템에 대한 참가자의 평가가 달라지는지 알아보았다. 사과유형과 중립 유형 간 신뢰도의 차이는 없었으나, 사과유형이 중립 유형과 비 사과유형보다 오류로 인한 좌절감을 경감시키고 시스템에 대한 매력도와 사용성을 높이는 것으로 나타났다. Lewicki, Polin, & Lount Jr (2016)[23]는 사람 간의 대화에서 신뢰가 위반되는 상황을 설정하고 신뢰 회복에 가장 효과적인 사과의 구조를 연구하였다. 사과에 대한 표현, 문제의 이유에 대한 설명, 책임감 명시, 후회에 대한 표현, 복구 방안 제공, 용서 구하기와 같은 요소들의 개수와 조합을 달리하여 제시하고 사과 메시지에 대하여 평가하도록 하였다. 그 결과, 단일한 요소로 구성된 메시지보다 여러 요소를 복합적으로 포함하는 메시지가 신뢰 회복에 더욱 효과적인 것으로 나타났다.

선행연구를 종합하면 오류 메시지의 유형은 사과의 방식을 취하며 여러 요소를 포함하는 것이 선호된다. 특히, 오류 메시지를 구조화할 때, 직접적인 사과의 표현과 함께 오류 발생에 대한 책임과 원인을 명시하는 것이 오류로 인한 부정적인 영향을 감소시키는데 효과적일 수 있다.

III. 연구가설 및 연구문제

기술의 발달과 함께 챗봇이 수행하는 기능은 점차 다양해지고 있으며, 여러 산업 분야로 그 역할이 확장되고 있다. 따라서 챗봇의 오류가 사용자 경험에 미치는 영향에 대해 이해하고 오류에 대한 올바른 대처전략을 세우기 위한 노력이 필요하다. 에이전트의 오류 및 부적 기대위반이 사용자 경험에 미치는 영향을 연구한 여러 선행연구가 존재한다[5][6][13][14]. 그러나 하나의 에이전트 내에서 발생할 수 있는 여러 유형의 오류 및 부적 기대위반이 사용자 경험에 미치는 영향에 관한 연구는 충분하지 않으며, 특히 역할과 범위가 더욱 확장되고 있는 챗봇을 대상으로 한 연구는 더욱 부족한 현실이다. 따라서 본 연구는 하나의 챗봇이 일으킬 수 있는 오류 유형을 마음지각 이론을 기반으로 나누어 오류 유형에 따라 오류가 사용자 경험에 미치는 부정적인 영향에 차이가 있는지를 알아보려고 하였다.

사용자 경험은 챗봇에 대한 태도, 챗봇에 대한 행동의도 및 지각된 챗봇의 사용성으로 세분화하여 측정하였다. 태도는 대상에 대한 긍정적이거나 부정적인 개인의 느낌을 의미한다[26]. 행동의도는 대상을 사용하고 자 하는 개인의 주관적인 가능성을 의미하며[27], 사용성은 대상을 사용함으로써 개인의 수행이 향상된다고 믿는 정도를 의미한다[28]. 선행연구를 통해 챗봇을 비롯한 에이전트와의 상호작용에서 태도[14][29][30]와 행동의도[6][14]에 대한 오류의 부정적인 영향이 확인된 바 있고 오류에 대한 적절한 대처는 에이전트의 사용성을 높일 수 있는 것으로 나타났다[31]. 따라서 본 연구에서는 태도와 행동의도 및 사용성을 종속변인으로 하여 오류 유형과 오류 메시지 구조화 여부가 미치는 영향을 검증하고자 하였다.

에이전트와의 상호작용에서 감정을 느끼고 표현하는 경험요소에 대한 기대치보다 계획을 수립하고 수행하는 기능요소에 대한 기대치가 더욱 높다. 따라서 하나의 챗봇에 대한 기대수준은 경험요소보다 기능요소가 높을 것이다. 그리고 기대수준이 높은 기능요소에서의 오류가 기대수준이 낮은 경험요소의 오류보다 사용자 경험에 더욱 큰 부정적인 영향을 줄 것으로 예상된다. 따라서 연구가설 1과 연구가설 2는 다음과 같다.

연구가설 1: 상호작용 이전 챗봇에 대한 기대수준은 경험요소보다 기능요소가 높을 것이다.

연구가설 2: 오류 유형에 따라 오류가 태도(a)와 행동의도(b) 및 사용성(c)에 차이가 있을 것이다. (태도(a)와 행동의도(b) 및 사용성(c)은 경험요소 오류조건보다 기능요소 오류조건에서 더 낮을 것이다.)

또한, 본 연구에서는 챗봇과의 상호작용에서 구조화된 오류 메시지의 완충효과를 확인하고자 하였다. 앞서 살펴본 선행연구에 따르면 사과와 양식을 따르고 복합적인 요소를 포함한 구조화된 메시지가 오류의 부정적인 영향을 감소시킨다. 챗봇과의 상호작용에서도 구조화된 오류 메시지는 오류의 부정적인 영향을 감소시킬 것으로 기대할 수 있다. 따라서 연구가설 3은 다음과 같다.

연구가설 3: 오류 메시지의 구조화 여부에 따라 태도(a)와 행동의도(b) 및 사용성(c)에 차이가 있을 것이다. (태도(a)와 행동의도(b) 및 사용성(c)은 구조화한 오류 메시지 조건보다 구조화하지 않은 오류 메시지 조건에서 더 낮을 것이다.)

오류는 지각된 수행수준을 낮추어 부적 기대위반을 발생시키며 이때 기대수준과 지각된 수행수준 간 기대격차가 나타나게 된다. 본 연구에서는 오류 유형 및 오류 메시지 구조화 여부에 따른 태도와 행동의도 및 사용성의 차이가 기대격차에 기인한 것인지 알아보려고 하였다. 연구문제는 다음과 같다.

연구문제 1: 각 요소의 부적 기대격차가 사용자 경험에 미치는 정도는 어떠한 것인가?

연구문제 2: 구조화된 오류 메시지가 부적 기대격차를 감소시킬 것인가?

IV. 연구방법

1. 연구설계 및 대상

오류가 발생한 요소가 경험요소인지 기능요소인지와 오류 메시지의 구조화 여부에 따라 챗봇에 대한 태도와 행동의도 그리고 사용성에 차이가 있는지를 검증하고자 하였다. 2(오류 유형:기능/경험) X 2(구조화 여부:구조화/비구조화)의 참가자 간 설계(between-subjects design)로 실험을 구성하였다. 독립변인은 오류 유형과 오류메시지의 구조화 여부이며, 종속 변인으로는 에이

진트에 대한 태도, 행동의도 그리고 사용성을 측정하였다.

켈트릭스를 통해 온라인 설문으로 진행되었으며 2020년 10월 8월 4일부터 2020년 10월 17일까지 응답을 수집했다. 참가자는 연세대학교 심리학과 학부생이었으며, 보상으로 1 크레딧을 지급하였다. 전체 참가자 131명 중, 설문을 완료하지 않은 7개의 응답과 응답시간이 60분을 초과한 1개의 응답은 분석에서 제외하였다. 참가자의 연령은 만18세에서 24세로 평균 연령은 20.43세($SD = 1.68$)였다. 성별은 41명이 남성(33.3%), 82명이 여성(66.7%)이었다.

2. 실험절차

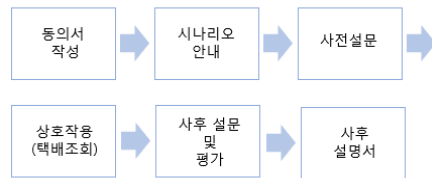


그림 1. 실험절차

실험절차는 [그림 1]과 같다. 참가자가 온라인 설문 링크에 접속하여 참가동의서를 작성하는 절차로 실험을 시작하였다. 대화형 택배 예약 및 조회 프로그램 '딜리버'를 통하여 상세택배 조회를 하게 된다는 안내문을 제시하였는데, 이때 '딜리버'는 대화형 에이전트로서 감정을 이해하고 표현하도록 학습되었으며 신규 예약을 진행과 접수된 택배의 현황을 조회하고 관리할 수 있음을 명시하였다. 안내문을 읽고 나면 딜리버에 대한 경험요소 및 기능요소 수행에 대한 기대감의 수준 설문에 응답하도록 하였다. 다음으로 택배의 배송상태 상세조회 시나리오를 제시하였다. 참가자가 수행할 과제는 챗봇과의 상호작용을 통해 배송이 지연되고 있는 택배의 배송상황을 상세조회하고 택배의 정확한 도착 시점을 확인하는 것이었다. 참가자는 오류 유형과 사과 메시지 구조화 여부에 따른 4가지 조건 중 하나의 상호작용 시나리오에 무선적으로 할당되었다. 모든 조건의 상호작용 시나리오는 감정표현을 포함한 인사, 택배 배송 현황 조회, 택배 배송 현황 상세조회 장면으로 구성하였다. 상호작용은 피험자가 응답 선택지를 마우스를 통해

클릭하는 선택형 방식으로 진행하였으며, 사용자의 선택에 따라 사전에 정해진 응답목록 중에 적합한 응답이 제시되도록 설계하였다. 또한 실제 서비스를 제공하는 택배 챗봇인 ‘롯데택배 로다’를 모사하여 응답목록과 화면의 문구를 설정하였다. 예를 들어 참가자가 ‘택배조회’를 선택하면 ‘운송장 직접입력’과 ‘최근 배송 내역’이 제시되도록 하였으며, 택배 배송 현황 안내 문구 역시 실제 택배 조회 시 제시되는 양식에 따라 제시될 수 있도록 하였다. 상호작용을 모두 마친 이후, ‘딜리버’에 대한 태도, 행동의도 및 사용성에 대한 설문과 경험요소와 기능요소의 지각된 수행수준 설문에 응답하도록 하였다. 참가자가 모든 설문에 응답한 이후에는 사후설명문을 제시하여 연구의 목적을 상세하게 안내하였다.

3. 실험처치와 처치물

실험에서 참가자가 경험요소와 기능요소 중 하나의 요소에서의 오류를 경험하도록 구성하였다. 경험요소 오류조건에서는 참가자의 정서표현에 대해 ‘딜리버’가 적합한 정서표현을 하지 못하는 오류 상황을 제시하였다. 오류가 발생하는 정서표현은 배고픔, 기쁨, 고통 중 참가자의 선택에 대응하여 제시하였다. 기능요소 오류조건에서는 택배 간략 조회 성공 이후 이어지는 상세조회에 실패하는 오류 상황을 제시하였다. 하나의 요소에서 오류가 발생하면 다른 요소는 성공하도록 하였다. 경험요소 오류조건에서는 기능요소인 상세조회가 성공적으로 이루어지도록 하였으며, 기능요소 오류조건에서는 경험요소인 정서표현에 있어서 적절한 응답이 제시되도록 하였다.

오류 메시지의 구조화 여부는 오류 상황을 안내하는 메시지로 조작하였다. 구조화되지 않은 조건에서의 오류 메시지는 챗봇의 실패를 명시하는 단독 문장으로 구성하였다. 구조화된 조건에서의 오류 메시지는 직접적인 사과 표현과 오류 발생의 원인 및 보완 가능성을 명시하는 복합 문장으로 구성하였다. 각 조건에 사용된 문장의 예시는 아래 [그림 2]와 같다.

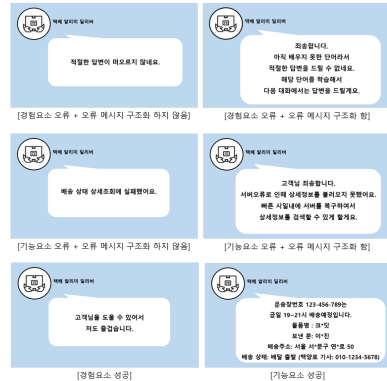


그림 2. 각 조건에 따른 상호작용 자극 예시

4. 종속변인의 측정

각 종속 변인은 설문 문장을 제시하고 해당 표현에 대해 얼마나 동의하는지 7점 척도(1점=매우 동의하지 않음, 7점=매우 동의함)로 측정하였다.

태도와 행동의도는 선행연구에 사용된 설문을 번안하여 재구성하였다[25][26]. 에이전트에 대한 태도를 측정하기 위해 8개의 문장을 제시하였으며 제시된 표현은 “흥미롭다”, “우수하다”, “도움이 된다”, “매력적이다”, “재미있다”, “기분을 좋게 한다”, “호감이 간다”, “품질이 좋다”였다. 에이전트에 대한 행동 의사는 3개의 문장을 제시하였으며, 제시된 표현은 “더 탐색하고 싶다”, “다시 이용하고 싶다”, “친구나 지인에게 추천하고 싶다”였다.

사용성은 PSSUQ (Post-Study System Usability Questionnaire) version3 를 번안하여 설문 문항을 구성하였다[27]. 16개의 문장을 제시하였으며 “전반적인 시스템은 사용하기 쉽다.” “화면에 제시된 정보의 내용은 명확하다.” “이 시스템은 내가 기대했던 기능과 능력을 모두 갖추고 있다.”와 같은 문장이 포함되었다.

상호작용 이전 경험요소와 기능요소의 기대수준을 측정하기 위해서 설문을 추가로 구성하였다. 경험요소는 “정서표현에 공감할 수 있다.” “정서를 표현할 수 있다.” 2개의 문장으로 설문을 구성하였고, 기능요소는 “판단하고 제어할 수 있다” “계획을 기억하고 관리할 수 있다” 2개의 문장으로 구성하였다. 상호작용 이전에 4개의 문항을 제시하여 기대수준이 어떠한 지 7점 척도(1점 = 매우 낮음, 7점 = 매우 높음)로 응답하도록 하

였다.

상호작용 이후 경험요소와 기능요소의 지각된 수행 수준을 측정하기 위해서 기대수준 측정에 사용된 4문장을 상호작용 이후에 다시 제시하고 지각된 수행수준이 어떠한지 7점 척도로 응답하도록 하였다.

각 설문 문항의 신뢰도(Cronbach' α)는 태도(.92), 행동의도(.88), 사용성(.93), 경험요소 기대수준(.74) 사전 기능요소 기대수준(.68), 경험요소 수행수준(.87), 기능요소 수행수준(.75)로 모든 항목에 대하여 .60이상으로 신뢰성이 검증되었다.

또한, 분석을 위해 기대수준 점수와 지각된 수행수준의 점수의 차이 값을 기대격차 점수로 사용하였다. 기대수준 점수에서 지각된 수행수준 점수를 감하여 기대격차를 계산하였다. 기대수준과 수행수준이 같은 경우 기대격차는 0이며, 수행이 기대에 못 미치는 부적 기대위반의 기대격차 점수는 음수(-), 수행이 기대를 넘어서는 정적 기대위반의 기대격차 점수는 양수(+)로 나타난다.

V. 결과

본 연구는 오류 유형과 오류 메시지의 구조화 여부에 따라 챗봇에 대한 태도와 행동의도 및 사용성에 차이가 있는가와 그 차이가 기대격차에 기인하는가를 분석하고자 하였다. 모든 분석은 IBM SPSS Statistics 25를 사용하였다.

1. 기대수준 및 수행수준

1.1 조작의 점검

각 조건에 따라 상호작용 이전에 에이전트에게 가지는 기대수준에 차이가 있다면 이는 기대격차의 크기에 영향을 줄 수 있다. 집단 간 사전 기대수준에 차이가 없음을 확인하기 위하여 일원분산분석(One-way ANOVA)를 실시하였다. 그 결과, 집단 간 에이전트의 경험요소에 대한 기대수준($M = 3.46, SD = 1.30$)에 차이가 없었으며, $F(3, 119) = 0.38, p = .764$, 기능요소에 대한 기대수준($M = 5.66, SD = 0.80$)에도 차이가 없었다, $F(3, 119) = 1.25, p = .296$.

시나리오를 통해 오류 상황이 적절하게 조작되었다면 오류 요소에 대하여 상호작용 이후에 지각된 수행수준은 낮을 것이다. 조작 점검을 위해 오류 유형에 따라 지각된 수행수준을 독립표본 t검정 (independent sample t-test)을 통해 확인하였다. 그 결과, 오류 유형에 따라 지각된 경험요소 수행수준의 차이가 유의미했다. 경험요소오류를 경험한 집단의 지각된 경험요소 수행 점수($M = 2.65, SD = 1.23$)가 기능요소 오류를 경험한 집단($M = 3.67, SD = 1.37$)보다 유의미하게 낮았다, $t(121) = -4.37, p < .001$. 또한, 오류 유형에 따라 지각된 기능요소 수행수준의 차이가 유의미했다. 경험요소오류를 경험한 집단($M = 5.44, SD = 1.07$)보다 기능요소오류를 경험한 집단($M = 4.90, SD = 1.18$)이 기능요소 수행수준을 더욱 낮게 지각했다, $t(121) = 2.64, p = .009$. 따라서 시나리오를 통한 오류의 조작이 적절하였음을 확인하였다.

1.2 마음요소에 따른 기대수준

상호작용 이전 챗봇에 대한 기대수준은 경험요소보다 기능요소가 높을 것이라는 가설 1을 검증하기 위해 대응표본 t검정(paired t-test)을 실시하였다. 그 결과, [그림 3]에 제시된 바와 같이 사전 기능요소에 대한 기대 점수 평균($M = 5.66, SD = 0.80$)이 사전 경험요소에 대한 기대 점수 평균($M = 3.46, SD = 1.30$)보다 높았으며, 그 차이가 유의미하였다, $t(122) = -15.24, p < .001$. 따라서 가설 1이 채택되었다.

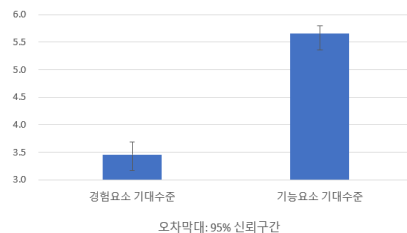


그림 3. 마음지각 요소에 따른 기대수준

2. 오류 유형과 오류 메시지 구조화 여부에 따른 사용자 경험

오류 유형과 오류 메시지 구조화 여부에 따른 태도,

행동의도 및 사용성의 차이를 알아보기 위하여 이원분산분석(two-way ANOVA)을 실시하였다.

2.1 태도

태도에 대하여 오류 유형의 주효과는 나타나지 않았다. 경험요소오류를 경험한 집단($M = 4.15, SD = 1.27$)의 태도보다 기능요소오류를 경험한 집단($M = 3.88, SD = 1.12$)의 태도 점수가 더 낮은 것으로 나타났으나 이 차이는 통계적으로 유의미하지 않았다, $F(1, 119) = 1.63, p = .204$. 오류 메시지 구조화 여부에 따른 주효과는 유의미한 것으로 확인되었다. 구조화된 오류 메시지가 제시된 집단($M = 4.23, SD = 1.21$) 이 구조화되지 않은 오류 메시지가 제시된 집단($M = 3.80, SD = 1.15$)보다 태도 점수가 더 높았다, $F(1, 119) = 4.19, p = .043$. 오류 유형과 구조화 여부의 상호작용 효과는 나타나지 않았다, $F(1, 119) = 0.07, p = .796$. 조건별 태도 점수에 대한 기술통계치는 표 1에 제시하였으며, 구조화 여부에 따른 태도 점수의 차이는 그림 4와 같다. 따라서 가설 2a는 기각되고 가설 3a는 채택되었다.

표 1. 조건별 태도 점수

요인	집단	n	M	SD
오류 유형	경험요소	60	4.15	1.27
	기능요소	63	3.88	1.12
구조화 여부	구조화 X	62	3.80	1.15
	구조화 O	61	4.23	1.21
전체		123	4.01	1.20

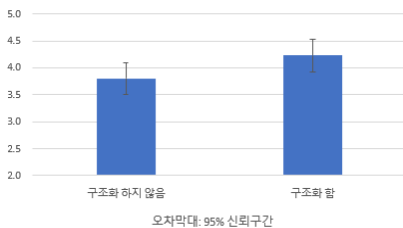


그림 4. 오류 메시지 구조화 여부에 따른 태도

2.2 행동의도

행동의도에 대한 오류 유형의 주효과는 나타나지 않

았다. 경험요소의 오류를 경험한 집단($M = 4.14, SD = 1.33$)의 행동의도가 기능요소 오류를 경험한 집단($M = 3.85, SD = 1.38$)의 행동의도 보다 높게 나타났으나 이 차이는 통계적으로 유의하지 않았다, $F(1, 119) = 1.44, p = .233$. 행동의도에 대한 메시지 구조화 여부의 주효과는 유의미한 것으로 나타났다. 구조화된 메시지를 제시받은 집단($M = 4.33, SD = 1.35$)의 행동의도가 구조화되지 않은 메시지를 제시받은 집단($M = 3.65, SD = 1.30$)의 행동의도보다 유의미하게 높았다, $F(1, 119) = 8.03, p = .005$. 행동의도에 대한 오류 유형과 메시지 구조화 여부의 상호작용 효과는 유의미하지 않았다, $F(1, 119) = 1.67, p = .199$. 조건별 행동의도 점수에 대한 기술통계치는 [표 2]에 제시하였으며, 구조화 여부에 따른 태도 점수의 차이는 [그림 5]와 같다. 따라서 가설 2b는 기각되고 가설 3b는 채택되었다.

표 2. 조건별 행동의도 점수

요인	집단	n	M	SD
오류 유형	경험요소	60	4.14	1.33
	기능요소	63	3.85	1.38
구조화 여부	구조화 X	62	3.65	1.30
	구조화 O	61	4.33	1.35
전체		123	3.99	1.36

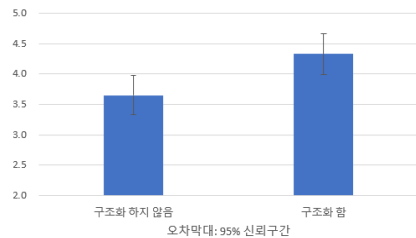


그림 5. 오류 메시지 구조화 여부에 따른 행동의도

2.3 사용성

사용성에 대한 오류 유형의 주효과는 나타났다. 경험요소 오류를 경험한 집단($M = 5.02, SD = 0.86$)보다 기능요소 오류를 경험한 집단($M = 4.65, SD = 0.92$)의 사용성이 유의미하게 낮은 것으로 나타났다, $F(1, 119) = 5.32, p = .023$. 또한, 사용성에 대한 오류 메시지 구

구조화 여부의 주효과도 유의미한 것으로 나타났다. 구조화되지 않은 오류 메시지를 제시받은 집단($M = 4.67$, $SD = 0.88$)의 사용성보다 구조화된 오류 메시지를 제시받은 집단($M = 4.99$, $SD = 0.91$)의 사용성이 유의미하게 높았다, $F(1, 119) = 4.07$, $p = .046$. 그러나 오류 유형과 메시지 구조화 여부의 상호작용 효과는 나타나지 않았다, $F(1, 119) = 0.01$, $p = .925$. 조건별 사용성 점수에 대한 기술통계치는 [표 3]에 제시하였으며, 구조화 여부에 따른 태도 점수의 차이는 [그림 6]과 같다. 따라서 가설 2c와 가설 3c 모두 채택되었다.

표 3. 조건별 사용성 점수

요인	집단	n	M	SD
오류 유형	경험요소	60	5.02	0.86
	기능요소	63	4.65	0.92
구조화 여부	구조화 X	62	4.67	0.88
	구조화 O	61	4.99	0.91
전체		123	4.83	0.91

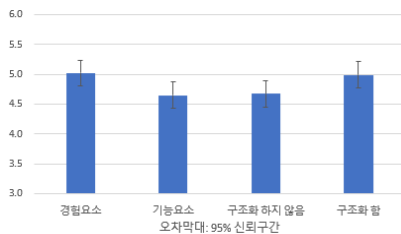


그림 6. 오류 유형과 오류 메시지 구조화 여부에 따른 사용성

3. 기대격차가 사용자 경험에 미치는 영향

마음지각의 요소에 따른 기대격차 점수가 사용자 경험에 미치는 영향을 알아보기 위해서 경험요소 기대격차 점수와 기능요소 기대격차 점수를 독립변인으로 하여 태도와 행동의도 그리고 사용성 각각에 대한 다중회귀분석을 실시하였으며, 결과는 [표 4]에 제시하였다.

먼저, 두 기대격차가 태도에 미치는 영향을 살펴보았다. 회귀식은 태도 변량의 21.3%를 설명하는 것으로 나타났다, $F(2, 120) = 16.20$, $p < .001$. 각 독립변인을 세부적으로 살펴본 결과, 경험요소 기대격차의 영향은 유의미하지 않은 것으로 나타났으며($p = .089$), 기능요소 기대격차 점수의 독립적인 영향력은 유의미했다(β

$= .431$).

다음으로 두 기대격차가 행동의도에 미치는 영향을 살펴보았다. 두 기대격차는 행동의도 변량의 13.7%를 설명하는 것으로 나타났다, $F(2, 120) = 9.56$, $p < .001$. 각 독립변인을 세부적으로 살펴본 결과, 경험요소 기대격차의 영향은 유의미하지 않았으나($p = .245$), 기능요소 기대격차 점수의 독립적인 영향력은 유의미했다($\beta = .351$).

마지막으로 두 기대격차가 사용성에 미치는 영향을 살펴보았다. 두 기대격차가 사용성 변량에 대해서는 26.6%를 설명하는 것으로 나타났다, $F(2, 120) = 21.77$, $p < .001$. 각 독립변인을 세부적으로 살펴본 결과, 사용성에 대한 경험요소 기대격차의 영향은 유의미하지 않았으나($p = .928$), 기능요소 기대격차 점수의 독립적인 영향력은 유의미하였다($\beta = .516$).

표 4. 태도, 행동의도 및 사용성에 대한 다중회귀분석 결과

종속변수	독립변수	B	β	t
태도	(상수)	4.270		40.06***
	경험요소기대격차	.116	.139	1.71
	기능요소기대격차	.456	.431	5.32***
	$R^2 = .213$ Durbin-Watson=2.01			
행동의도	(상수)	4.224		33.27***
	경험요소기대격차	.094	.099	1.17
	기능요소기대격차	.422	.351	4.214***
	$R^2 = .137$ Durbin-Watson=1.97			
사용성	(상수)	5.032		64.57***
	경험요소기대격차	-.004	-.007	-0.09
	기능요소기대격차	.413	.516	6.59***
	$R^2 = .266$, Durbin-Watson=2.24			

*** $p < .001$

4. 오류 유형과 오류 메시지 구조화 여부에 따른 기대격차

오류의 유형과 오류 메시지 구조화의 여부에 따른 기대격차 점수에 차이를 확인하기 위해서 기대격차를 종속변인으로 하는 이원분산분석을 실시하였다. 구체적으로는 오류 메시지의 구조화가 부적 기대격차를 완화하는가를 알아보려고 하였다.

4.1 전체 기대격차

전체 기대격차에 대한 오류 유형의 주효과는 나타나

지 않았다. 경험요소오류를 경험한 집단($M = -0.95$, $SD = 1.58$)의 전체 기대격차점수가 기능요소 오류를 경험한 집단($M = -0.61$, $SD = 2.12$)보다 부적 기대격차가 더욱 큰 것으로 나타났으나, 이 차이는 통계적으로 유의미하지 않았다, $F(1, 119) = 1.03$, $p = .312$. 기대격차에 대한 오류 메시지 구조화 여부의 주효과의 경향성을 확인하였다. 통계적으로 유의미한 수준에 미치지 못하는 못했으나, 구조화된 오류 메시지가 제시된 집단($M = -0.47$, $SD = 1.74$)이 구조화되지 않은 오류 메시지가 제시된 집단($M = -1.08$, $SD = 1.98$)보다 부적 기대격차가 더 작은 것으로 나타났다, $F(1, 119) = 3.43$, $p = .066$. 오류 유형과 구조화 여부의 상호작용효과는 나타나지 않았다, $F(1, 119) = 0.68$, $p = .412$. 이 결과는 [그림 7]과 같으며, 기술통계치는 [표 5]에 제시하였다.

표 5. 조건별 전체 기대격차 점수

요인	집단	n	M	SD
오류 유형	경험요소	60	-0.95	1.58
	기능요소	63	-0.61	2.12
구조화 여부	구조화 X	62	-1.08	1.98
	구조화 O	61	-0.47	1.74
전체		123	-0.78	1.88

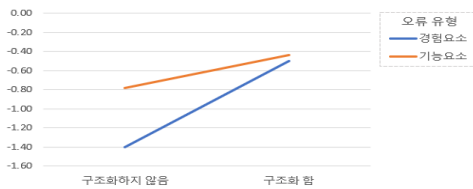


그림 7. 오류 유형과 구조화 여부에 따른 전체기대격차

4.2 경험요소 기대격차

경험요소 기대격차점수에 대한 오류 유형의 주효과가 확인되었다. 경험요소오류를 경험한 집단($M = -0.73$, $SD = 1.31$)의 경험요소 기대격차점수가 기능요소 오류를 경험한 집단($M = 0.14$, $SD = 1.42$)보다 부적 기대격차가 더욱 컸으며 기능요소 오류를 경험한 집단의 경험요소 기대격차점수는 0보다 커서 지각된 수행수준이 기대수준보다 높은 것으로 나타났다, $F(1, 119) = 12.66$, $p = .001$. 기대격차에 대한 오류 메시지

구조화 여부의 주효과는 나타나지 않았다. 구조화된 오류 메시지가 제시된 집단($M = -0.11$, $SD = 1.49$)이 구조화되지 않은 오류 메시지가 제시된 집단($M = -0.46$, $SD = 1.36$)보다 부적 기대격차가 더 작은 것으로 나타났으나 통계적으로 유의하지 않았다, $F(1, 119) = 2.11$, $p = .149$. 오류 유형과 구조화 여부의 상호작용효과는 나타나지 않았다, $F(1, 119) = 0.56$, $p = .813$. 이 결과는 [그림 8]과 같으며, 기술통계치는 [표 6]에 제시하였다.

표 6. 조건별 경험요소 기대격차 점수

요인	집단	n	M	SD
오류 유형	경험요소	60	-0.73	1.31
	기능요소	63	0.14	1.42
구조화 여부	구조화 X	62	-0.46	1.36
	구조화 O	61	-0.11	1.49
전체		123	-0.28	1.43

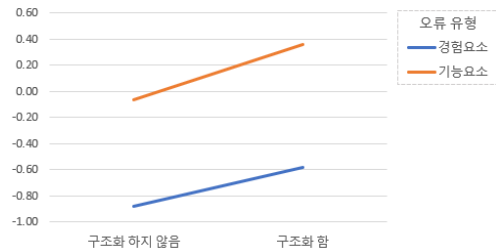


그림 8. 오류 유형과 구조화 여부에 따른 경험요소기대격차

4.3 기능요소 기대격차

기능요소 기대격차에 대한 오류 유형의 주효과가 확인되었다. 기능요소의 부적 기대격차는 경험요소오류를 경험한 집단($M = -0.22$, $SD = 1.01$)보다 기능요소 오류를 경험한 집단($M = -0.75$, $SD = 1.18$)이 더 큰 것으로 나타났다, $F(1, 119) = 7.44$, $p = .007$. 기대격차에 대한 오류 메시지 구조화 여부의 주효과는 나타나지 않았다. 구조화된 오류 메시지가 제시된 집단($M = -0.36$, $SD = 1.08$)이 구조화되지 않은 오류 메시지가 제시된 집단($M = -0.62$, $SD = 1.18$)보다 부적 기대격차가 더 작은 것으로 나타났으나 통계적으로 유의하지 않았다, $F(1, 119) = 1.80$, $p = .183$. 오류 유형과 구조

화 여부의 상호작용효과는 경향성이 확인되었다. $F(1, 119) = 2.90, p = .091$. 경험요소 오류를 경험한 집단 중 구조화된 오류 메시지를 받은 집단($M = 0.08, SD = 0.88$)의 기능요소 기대격차 점수는 0보다 커서 기능요소에 대한 수행수준을 기대보다 높게 지각하였고, 구조화하지 않은 오류 메시지를 받은 집단($M = -0.52, SD = 1.06$)의 기능요소 기대격차는 0보다 작아 부적 기대위반이 나타났다. 반면, 기능요소 오류를 경험한 집단은 구조화된 오류 메시지를 받은 집단($M = -0.79, SD = 1.09$)의 기능요소 부적 기대격차점수가 구조화되지 않은 오류 메시지를 받은 집단($M = -0.72, SD = 1.29$)보다 구조화된 오류 메시지를 받은 집단($M = -0.79, SD = 1.09$) 더 작은 것으로 나타났다. 이 결과는 [그림 9와 같으며, 기술통계치는 [표 7]에 제시하였다.

표 7. 조건별 기능요소 기대격차 점수

요인	집단	n	M	SD
오류 유형	경험요소	60	-0.22	1.01
	기능요소	63	-0.75	1.18
구조화 여부	구조화 X	62	-0.62	1.18
	구조화 O	61	-0.36	1.08
전체		123	-0.49	1.13

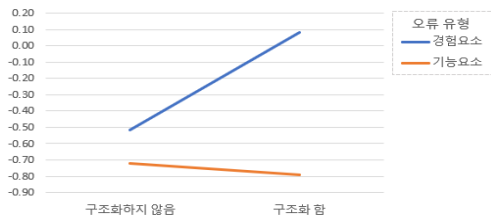


그림 9. 오류 유형과 구조화 여부에 따른 기능요소기대격차

VI. 논의

본 연구는 오류의 유형과 오류 메시지의 구조화 여부에 따라 태도와 행동의도 그리고 사용성에 차이가 있는지를 검증하였다. 오류의 유형은 마음지각 이론에 근거하여 마음지각의 축인 경험요소와 기능요소를 기준으로 나누었다. 또한, 구조화된 오류 메시지가 오류로 인한 부정적인 영향에 대해 완충작용을 하는지 검증하고

자 하였다. 비구조화된 오류 메시지는 단순하게 오류를 명시하는 문장으로 구성하였고 구조화된 오류 메시지는 사과 표현과 오류의 원인 및 개선의 의지를 포함한 복합적인 문장으로 구성하였다. 각 조건에 따른 이원분산분석의 결과를 종합하면 [표 8]과 같다.

표 8. 오류유형과 메시지구조화 여부에 따른 이원분산분석 결과의 종합

종속 변인	요인	집단	M	SD	F	p
태도	오류유형	경험요소	4.15	1.27	1.63	.204
		기능요소	3.88	1.12		
	구조화 여부	구조화 X	3.80	1.15	4.19	
		구조화 O	4.23	1.21		
상호작용	전체	4.01	1.20	0.07	.796	
행동의도	오류유형	경험요소	4.14	1.33	1.44	.233
		기능요소	3.85	1.38		
	구조화 여부	구조화 X	3.65	1.30	8.03	
		구조화 O	4.33	1.35		
상호작용	전체	3.99	1.36	1.67	.199	
사용성	오류유형	경험요소	5.02	0.86	5.32	.023
		기능요소	4.65	0.92		
	구조화 여부	구조화 X	4.67	0.88	4.07	
		구조화 O	4.99	0.91		
상호작용	전체	4.83	0.91	0.01	.925	
전체 기대격차	오류유형	경험요소	-0.95	1.58	1.03	.312
		기능요소	-0.61	2.12		
	구조화 여부	구조화 X	-1.08	1.98	3.43	
		구조화 O	-0.47	1.74		
상호작용	전체	-0.78	1.88	0.68	.412	
경험요소 기대격차	오류유형	경험요소	-0.73	1.31	12.66	.001
		기능요소	0.14	1.42		
	구조화 여부	구조화 X	-0.46	1.36	2.11	
		구조화 O	-0.11	1.49		
상호작용	전체	-0.28	1.43	0.56	.813	
기능요소 기대격차	오류유형	경험요소	-0.22	1.01	7.44	.007
		기능요소	-0.75	1.18		
	구조화 여부	구조화 X	-0.62	1.18	1.80	
		구조화 O	-0.36	1.08		
상호작용	전체	-0.49	1.13	2.90	.091	

또한, 오류 상황에서 발생하는 부적 기대격차가 태도와 행동의도 그리고 사용성을 유의미하게 예측하는지와 구조화된 오류 메시지가 부적 기대격차를 완화하는지도 살펴보았다. 결과를 종합하면 다음과 같다.

먼저, 챗봇에 대한 사전 기대감은 경험요소보다 기능요소가 더욱 높은 것으로 나타났다. 이를 통해 마음지각을 기반으로 진행된 여러 HCI 선행연구에서 확인된 에이전트에 대한 기대수준의 차이가 챗봇에도 같은 양

상으로 나타남을 확인하였다.

오류의 유형을 기대수준이 서로 다른 경험요소와 기능요소로 나누어 오류 유형에 따라 태도와 행동의도 및 사용성에 차이가 있는지 검증하였다. 그 결과, 경험요소 오류조건보다 기능요소오류 조건에서 태도와 행동의도 그리고 사용성을 더 낮게 평가하였으나 사용성에 대해서만 그 차이가 통계적으로 유의미한 것으로 나타났다. 세 종속 변인에 대한 경험요소 및 기능요소 기대격차의 회귀분석 결과, 전체 기대격차의 사용성에 대한 예측력 변량의 26.6%로 태도(21.3%)나 행동의도(13.7%)에 대한 예측력보다 높은 것으로 나타났다. 전체 기대격차는 오류 유형에 따라 서로 다른 기대수준을 기반으로 한다. 따라서 전체 기대격차가 종속 변인에 미치는 영향이 가장 높았던 사용성에 대해서는 오류 유형에 따른 차이가 유의미했던 반면 사용성보다 기대격차가 미치는 영향력이 낮은 태도와 행동의도에서는 차이가 통계적으로 유의미하지 않았을 수 있다.

오류 메시지의 구조화 여부에 따른 태도와 행동의도 및 사용성의 차이를 검증한 결과, 오류 메시지의 구조화 여부에 따라 태도와 행동의도 및 사용성에 유의미한 차이가 나타났으며, 오류 메시지가 구조화된 조건에서 태도와 행동의도 그리고 사용성이 모두 유의미하게 높아지는 것을 확인할 수 있었다. 즉, 오류 후 제시되는 오류 메시지가 단순히 오류 상황을 명시하기보다 사과와 오류 발생의 원인 그리고 보완의 가능성을 포함하였을 때 에이전트에 대한 사용자의 태도가 더 긍정적이며 해당 에이전트와 더 많이 상호작용하고 하고자 하며, 에이전트의 사용성을 더 높게 평가했다. 본 연구에서는 이를 구조화된 오류 메시지가 오류로 인해 발생하는 불확실성을 완화하기 때문에 해석하고자 한다. 불확실성감소이론(Uncertainty Reduction Theory)[32]에 의하면 상호작용의 목적은 대상에 대한 불확실성을 감소시키는 것이다. 에이전트의 기대위반은 에이전트가 가지는 불확실성을 증가시키는데[33], 구조화된 오류 메시지는 오류 상황에 대한 복합적인 정보를 제시한다. 따라서 오류 및 기대위반으로 인한 불확실성을 감소시켜서 오류가 태도와 행동의도 그리고 사용성에 미치는 부정적인 영향을 완화하는 결과를 나타낸 것으로 해석할 수 있다.

기대격차가 태도와 행동의도 및 사용성을 예측하는 지 검증한 결과, 전체 기대격차는 세 종속 변인을 유의미하게 예측하였다. 부적 기대격차가 감소할수록 태도와 행동의도 및 사용성이 높아지는 것으로 나타났다. 이때, 경험요소 기대격차 독립적으로는 세 종속 변인을 예측하지 못하나 기능요소 기대격차는 유의미하게 예측하는 것으로 나타났다. 또한, 메시지 구조화 여부에 따라서 기대격차가 감소하는지 알아본 결과, 전체 부적 기대격차는 구조화되면 더 감소하는 경향성이 나타났다. 유형별로 나누어 살펴보면 경험요소 부적 기대격차는 구조화할 때 감소하였지만 기능요소 기대격차는 상호작용효과의 경향성이 나타났다. 경험요소 오류조건에서는 구조화된 메시지가 기능요소기대격차를 감소시킨 반면, 기능요소 오류조건에서는 구조화된 메시지가 오히려 기능요소기대격차를 증가시키는 양상이 나타난 것이다. 이는 해당 챗봇과의 상호작용이 목표로 하는 과업이 분명한 효용 지향적 상황이었기 때문으로 보인다. 오류 상황에서 오류 복구 전략에 대한 반응은 해당 서비스에 대한 사용자의 도식이나 정신모형에 따라 달라진다[34]. Lee, Kiesler, Forlizzi, Srinivasa,& Rybski (2010)[13]의 연구에 따르면 관계 지향적인 사용자와 실용 지향적인 사용자는 서비스 로봇의 오류 대처전략 중 서로 다른 전략을 선호했다. 사용자의 관계 지향성이 강해질수록 사과를 선호하였으나, 실용 지향성이 강해질수록 보상을 선호하며 사과보다도 아무 대처전략을 하지 않는 것을 선호하는 것으로 나타났다. 상세조회라는 분명한 과업 성취를 목표로 하는 본 실험의 시나리오에서 참가자가 지향성은 관계적이기보다 실용적이었을 가능성이 더욱 크다. 또한, 기능요소의 오류의 경우 최종과업의 실패로 이어지게 되는데, 사과방식의 구조화된 오류 메시지가 이러한 과업의 실패를 재명시함으로써 기능요소에 대한 지각된 수행수준을 더 낮추고 기대격차를 키웠을 것으로 해석할 수 있다.

VII. 연구의 시사점 및 한계점

1. 본 연구의 시사점

본 연구의 시사점은 다음과 같다.

먼저, 챗봇과의 상호작용에서 사과와 문제 발생의 원인 및 보완 가능성을 포함하여 복합적으로 구성된 오류 메시지의 완충효과를 재검증하였다. 사과유형의 메시지가 사과유형이 아닌 메시지보다 더욱 긍정적으로 평가된다는 선행 연구들[19-22]과 단일 요소가 아닌 복합적인 요소로 구성된 오류 메시지를 사용자가 더욱 선호한다는 선행연구들[19][22][23]의 결과가 챗봇에도 적용됨을 확인하였다. 웹페이지를 배경에서 진행된 선행 연구들을 통해 검증된 구조화된 메시지의 완충효과가 챗봇과의 상호작용에서도 유의미하게 나타난 것이다. 따라서 챗봇 오류의 부정적인 영향을 감소시키기 위한 전략으로써 오류 메시지를 설계할 때, 사과의 유형을 따르고 오류에 대한 복합적인 정보로 구조화된 구성이 단순하게 오류의 사실만을 명시하는 비구조화된 구성보다 적합할 것이다.

둘째, 구조화된 오류 메시지의 효과는 오류의 유형에 따라 달라질 수 있음을 확인하였다. 기능요소가대격차는 태도와 행동의도 및 사용성을 유의미하게 예측하였으며 기능요소가대격차에 대한 오류 유형과 오류 메시지 구조화 여부에 따른 상호작용효과의 경향성이 나타났다. 따라서 오류 메시지의 구조화와 하면서 오류 유형에 따라 차별화된 세부 전략을 설계하는 것이 사용자 경험에 긍정적인 영향을 줄 수 있다. 특히 기능요소의 오류 상황에서 해당 오류가 과업의 성취와 관련된 경우, 오류에 대한 구체적인 정보만을 전달하는 구조화된 오류 메시지는 챗봇의 실패 사실을 부각할 가능성이 있다. 이때에는 실용 지향성이 강해질수록 보상에 대한 선호가 강해진다는 선행연구[13]에 따라 보상이나 오류에 대한 즉각적 수정 방안을 제시하는 추가적인 대처전략을 고려할 필요가 있다.

셋째, 과업을 위한 맥락에서는 기능요소의 부적 기대 위반 수준이 낮아질수록 태도와 행동의도 및 사용성이 높아지는 것을 확인하였다. 그러나 경험요소의 기대격차는 세 종속 변인을 유의미하게 예측하지 못했다. 즉, 오류 발생 상황에서 에이전트의 계획 및 관리 그리고 판단 및 제어 수행수준에 대한 기대격차를 최소화할 수 있는 대처전략을 수립하는 것이 사용자의 긍정적인 사용 경험 설계에 효과적일 것이다. 그러나 이는 맥락에 따라서 달라질 수 있다. 사용자의 사전 기대감을 해당

서비스가 얼마나 충족시켰는가는 서비스에 대한 평가에 큰 영향을 주는 요소이다[35][36]. 따라서 경험요소에 대한 기대수준이 기능요소에 대한 기대수준보다 더 높은 정서적 사용맥락에서는 이와 같은 결과가 반전될 가능성이 존재한다.

또한, 기대수준을 낮추는 방안을 기대격차를 낮추는 전략으로 고려해볼 수 있다. 상호작용 이전에 기대수준이 낮은 경우, 오류로 인해 수행수준이 낮게 지각되더라도 부적 기대격차의 크기는 기대수준이 높을 때보다 낮아진다. 기대격차의 크기는 사용자 경험을 유의미하게 예측하므로 수행수준이 향상되기 어려운 조건에서는 오류의 가능성을 명시하는 등의 방식으로 사전 기대수준을 낮게 조작하는 것 역시 전략이 될 수 있다. 그러나 기대수준이 너무 낮은 경우, 사용자가 상호작용을 할 필요성을 잃게 되므로[6], 이를 고려하여 제한적으로 적용해야 할 것이다.

2. 본 연구의 한계점

본 연구의 한계점은 다음과 같다.

먼저, 본 연구에 사용된 자극은 챗봇형 에이전트의 양식을 따랐으나 선택형으로 구성하여 실제의 상호작용과는 차이가 있을 수 있다. 현재 챗봇은 여러 응답 중 사용자가 하나를 택하는 룰 베이스(Rule-base)형과 사용자가 입력한 텍스트를 입력하여 열린 대화 형태에서 상호작용이 진행되는 지능형 방식이 혼용되어 있다. 그러나 본 연구에서는 참가자가 시나리오에 따라 상호작용을 하도록 설계하기 위하여 룰 베이스형을 적용하였기에 실제의 챗봇과는 차이가 있다.

또한, 최종과업의 성취 여부가 결과에 영향을 주었을 가능성이 있다. 기능요소는 최종과업의 성취 여부와 직접적으로 연관되어있다. 따라서, 유의미하게 나타난 사용성에 대한 오류 유형의 주효과가 오류의 유형이 아닌 최종과업의 성취 여부에 따른 효과였을 가능성이 존재한다. 따라서 후속 연구에서는 사용성에 대한 오류 유형의 주효과를 재검증을 하기 위해 최종과업의 성취 여부가 통제된 시나리오를 활용해야 할 것이다.

마지막으로 오류 메시지 구조화 여부에 따른 조건별 문장 길이를 통제하지 못하였다. 오류 메시지가 구조화 되지 않은 조건보다 구조화된 조건에서 제시된 오류 메

시지의 문장 길이가 길었으며 조건 간 서로 다른 메시지 길이가 상호작용에 영향을 주었을 가능성이 존재한다. 따라서 문장의 길이를 통제된 후속 연구가 이루어져야 할 것이다.

참 고 문 헌

- [1] A. Følstad and P. B. Brandtzæg, "Chatbots and the new world of HCI," *Interactions*, Vol.24, No.4, pp.38-42, 2017.
- [2] J. Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, Vol.9, No.1, pp.36-45, 1966.
- [3] 문용식, *인공지능 기반의 챗봇 서비스의 국내외 동향 분석 및 발전 전망*, 한국정보문화진흥원, 2018.
- [4] Z. Ashktorab, M. Jain, Q. V. Liao, and J. D. Weisz, "Resilient chatbots: repair strategy preferences for conversational breakdowns," In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp.1-12, 2019.
- [5] T. Komatsu and S. Yamada, "Adaptation gap hypothesis: How differences between users' expected and perceived agent functions affect their subjective impression," *Journal of Systemics, Cybernetics and Informatics*, Vol.9, No.1, pp.67-74, 2011.
- [6] T. Komatsu, R. Kurosawa, and S. Yamada, "How does the difference between users' expectations and perceptions about a robotic agent affect their behavior?," *International Journal of Social Robotics*, Vol.4, No.2, pp.109-116, 2012.
- [7] A. C. Horstmann and N. C. Krämer, "When a Robot Violates Expectations: The Influence of Reward Valence and Expectancy Violation on People's Evaluation of a Social Robot," In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp.254-256, 2020.
- [8] C. Nass, J. Steuer, and E. R. Tauber, "Computers are social actors," In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp.72-78, 1994.
- [9] H. M. Gray, K. Gray, and D. M. Wegner, "Dimensions of mind perception," *Science*, Vol.315 No.5812, pp.619-619, 2007.
- [10] J. K. Burgoon, "A communication model of personal space violations: Explication and an initial test," *Human communication research*, Vol.4, No.2, pp.129-142, 1978.
- [11] J. K. Burgoon, "Interpersonal expectations, expectancy violations, and emotional communication," *Journal of Language and Social Psychology*, Vol.12, No.1-2, pp.30-48, 1993.
- [12] T. Komatsu and S. Yamada, "How do robotic agents' appearances affect people's interpretations of the agents' attitudes?," In *CHI'07 Extended Abstracts on Human Factors in Computing Systems*, pp.2519-2524, 2007.
- [13] M. K. Lee, S. Kiesler, J. Forlizzi, S. Srinivasa, and P. Rybski, "Gracefully mitigating breakdowns in robotic services," In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.203-210, 2010.
- [14] E. Go and S. S. Sundar, "Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions," *Computers in Human Behavior*, Vol.97, pp.304-316, 2019.
- [15] C. Edwards, A. Edwards, P. R. Spence, and D. Westerman, "Initial interaction expectations with robots: Testing the human-to-human interaction script," *Communication Studies*, Vol.67, No.2, pp.227-238, 2016.
- [16] M. Appel, D. Izydorczyk, S. Weber, M. Mara, and T. Lischetzke, "The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers," *Computers in Human*

- Behavior, Vol.102, pp.274-286, 2020.
- [17] K. C. Yam, Y. E. Bigman, P. M. Tang, R. Ilies, D. De Cremer, H. Soh, and K. Gray, "Robots at work: People prefer—And forgive—Service robots with perceived feelings," *The Journal of Applied Psychology*, 2020.
retrived from <https://doi.org/10.1037/apl0000834>
- [18] Z. Ashktorab, M. Jain, Q. V. Liao, and J. D. Weisz, "Resilient chatbots: Repair strategy preferences for conversational breakdowns," In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp.1-12, 2019.
- [19] M. Akgun, K. Cagiltay, and D. Zeyrek, "The effect of apologetic error messages and mood states on computer users' self-appraisal of performance," *Journal of Pragmatics*, Vol.42, No.9, pp.2430-2448, 2010.
- [20] S. J. Park, C. M. MacDonald, and M. Khoo, "Do you care if a computer says sorry? User experience design through affective messages," In *Proceedings of the designing interactive systems conference*, pp.731-740, 2012.
- [21] J. Y. Tzeng, "Toward a more civilized design: studying the effects of computers that apologize," *International Journal of Human-Computer Studies*, Vol.61, No.3, pp.319-345, 2004.
- [22] S. Yuan, B. Brüggemeier, S. Hillmann, and T. Michael, "User Preference and Categories for Error Responses in Conversational User Interfaces," In *Proceedings of the 2nd Conference on Conversational User Interfaces*, pp.1-8, 2020.
- [23] R. J. Lewicki, B. Polin, and R. B. Lount Jr, "An exploration of the structure of effective apologies," *Negotiation and Conflict Management Research*, Vol.9, No.2, pp.177-196, 2016.
- [24] E. Olshtain, "Apologies across languages," *Cross-cultural pragmatics: Requests and apologies*, Vol.155173, 1989.
- [25] A. Cohen and E. Olshtain, "Apology: A speech act set," *Sociolinguistics and language acquisition*, Vol.18, p.35, 1983
- [26] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "User acceptance of computer technology: A comparison of two theoretical models," *Management science*, Vol.35, No.8, pp.982-1003, 1989.
- [27] M. Fishbein and I. Ajzen, "Belief, attitude, intention, and behavior: An introduction to theory and research," *The Journal of Philosophy and Rhetoric*, Vol.10, No.2, pp.130-132, 1977.
- [28] F. D. Davis, "A technology acceptance model for empirically testing new end-user information systems: Theory and results," *Massachusetts Institute of Technology*, 1985.
- [29] D. J. Brooks, M. Begum, and H. A. Yanco, "Analysis of reactions towards failures and recovery strategies for autonomous robots," In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication*, pp.487-492, 2016.
- [30] M. Salem, F. Eyssel, K. Rohlfing, S. Kopp, and F. Joublin, "To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability," *International Journal of Social Robotics*, Vol.5, No.3, pp.313-323, 2013.
- [31] S. Honig and T. Oron-Gilad, "Understanding and resolving failures in human-robot interaction: Literature review and model development," *Frontiers in psychology*, Vol.9, p.861, 2018.
- [32] C. R. Berger and R. J. Calabrese, "Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication," *Human communication research*, Vol.1, No.2, pp.99-112, 1974.
- [33] P. R. Spence, D. Westerman, C. Edwards, and A. Edwards, "Welcoming our robot overlords: Initial expectations about interaction with a robot," *Communication Research Reports*,

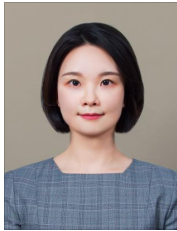
Vol.31, No.3, pp.272-280, 2014.

- [34] T. Ringberg, G. Odekerken-Schröder, and G. L. Christensen, "A cultural models approach to service recovery," *Journal of Marketing*, Vol.71, No.3, pp.194-214, 2007.
- [35] R. L. Oliver, "A cognitive model of the antecedents and consequences of satisfaction decisions," *Journal of marketing research*, Vol.17, No.4, pp.460-469, 1980.
- [36] R. A. Spreng, G. D. Harrell, and R. D. Mackoy, "Service recovery: impact on satisfaction and intentions," *Journal of services marketing*, 1995.

저 자 소 개

이 미 진(Mi-Jin Lee)

준회원



- 2013년 8월 : 성신여자대학교 생활문화소비자학과(생활과학사)
- 2019년 8월 ~ 현재 : 연세대학교 심리학과 석사과정

〈관심분야〉 : HCI, UX

한 광 희(Kwang-Hee Han)

정회원



- 1994년 ~ 현재 : 연세대학교 심리학과 교수

〈관심분야〉 : 인공지능, HCI, 인지심리학