
인공지능 대화형 에이전트의 지능적 속성에 대한 기대와 기대 격차

Expectation and Expectation Gap towards intelligent properties of AI-based Conversational Agent

박현아, Hyunah Park*, 태문영, Moonyoung Tae**, 허영진, Youngjin Huh***, 이준환, Joonhwan Lee****

요약 본 연구에서는 인공지능 대화형 에이전트인 스마트 스피커의 지능형 에이전트로서의 속성, 즉 자율성, 사회성, 반응성, 능동성, 시간연속성, 목표지향성에 대하여 이용자들이 일상적 상호작용을 통하여 어떤 기대를 가지는지, 또한 어떤 기대격차를 갖는지 살펴보고자 하였다. 이를 위해 스마트 스피커 이용자들을 대상으로 반구조화 인터뷰 (semi-structured interview)를 진행하고 그라운드 이론에 기반하여 분석하였다. 연구 결과 사람들은 기술수준의 한 계로 인해 스마트 스피커의 사회성이나 인간다움에 대해 큰 기대격차를 갖고 있었다. 스마트 스피커의 반응성에 대해서는 긍정적인 기대격차를 갖는 것으로 드러났고, 시간연속적으로 정보를 기억하는 것에 대해서는 정보의 민감성 정도나 제시방식에 따라 양가적 기대격차가 나타났다. 자율적인 추천에 대해서는 낮은 기대수준이 나타났고 능동적인 말걸기에 대해서는 맥락에 맞는 경우에만 선호하는 것으로 나타났다. 본 연구는 스마트 스피커와 상호작용하는 방식을 설계하고 기대 수준을 관리하는데 있어서 함의점을 제시한다.

Abstract The purpose of this study is to investigate the users' expectation and expectation gap about the attributes of smart speaker as an intelligent agent, ie autonomy, sociality, responsiveness, activeness, time continuity, goal orientation. To this end, semi-structured interviews were conducted for smart speaker users and analyzed based on ground theory. Result has shown that people have huge expectation gap about the sociality and human-likeness of smart speakers, due to limitations in technology. The responsiveness of smart speakers was found to have positive expectation gap. For the memory of time-sequential information, there was an ambivalent expectation gap depending on the degree of information sensitivity and presentation method. We also found that there was a low expectation level for autonomous aspects of smart speakers. In addition, proactive aspects were preferred only when appropriate for the context. This study presents implications for designing a way to interact with smart speakers and managing expectations.

핵심어: *Conversational Agents, Artificial intelligence, Intelligent agent, expectation, expectation gap*

*주저자 : 서울대학교 언론정보학과 박사과정

**공동저자 : 서울대학교 언론정보학과 석사

***공동저자 : 서울대학교 인지과학 협동과정 석사

****교신저자 : 서울대학교 언론정보학과 교수; e-mail: joonhwan@snu.ac.kr

■ 접수일 : 2018년 9월 28일 / 심사일 : 2018년 10월 23일 / 게재확정일 : 2018년 11월 20일

1. 서론

최근 자연어 처리와 인공지능의 다양한 기술이 발전하면서 대화형 에이전트가 더 자연스럽게 인간과 대화할 수 있게 되었다. 대화형 에이전트들은 이러한 기술을 기반으로 '인공지능(Artificial Intelligence) 에이전트'를 표방하고 있다[1]. 인공지능이라는 개념은 '환경으로부터 지각의 대상을 인식하고 그 환경에 영향을 주는 행동을 취하는 지능적인 에이전트를 설계하고 구현하는 것[2]'을 의미하며 1950년대에 제안된 이후로 수많은 분야의 연구자들에 의해 논의되어 왔다. 최근 4차 산업혁명과 인공지능에 대한 논의가 활발해지고 스마트 스피커가 대중화되기 시작하면서 일반 대중들도 일상 속에서 인공지능 개념을 접하거나 생각할 기회가 증가하고 있다. 따라서 사람들이 이러한 상호작용으로부터 어떤 인식을 갖는지 구체적으로 살펴볼 필요가 있다.

과거 선행연구들은 사람들이 지능형 에이전트로서의 로봇에 대해 어떤 태도를 갖는지[3-5], 그리고 로봇의 능력에 대해 어떤 기대를 갖고 이것이 어떻게 충족되거나 위반되는지[4,6,7] 등에 대하여 논의해 왔다. 선행 연구들은 1) 주로 사람들이 막연히 갖고 있는 로봇에 대한 이미지에 대하여 조사하거나 제한적 실험환경을 통해서 로봇에 대한 인식을 살펴보고, 2) 로봇의 능력에 대한 기대를 태스크 차원에서 조사해 왔다. 최근에는 스마트 스피커가 일상적으로 보급되면서 사람들이 여러 가지 상호작용 상황에서 인식을 가질 수 있게 되었고, 따라서 태스크 단위가 아닌 태스크를 수행하는 능력치에 대한 기대와 기대의 격차가 중요한 이슈로 부상하게 되었다[8,9]. 따라서 본 연구는 사람들이 스마트 스피커와의 상호작용에서 어떤 지능적 속성을 기대하고 그 기대가 어떻게 충족되거나 위반되었다고 느끼는지를 살펴보고자 한다.

본 연구는 구체적으로 스마트 스피커의 지능형 에이전트로서의 속성으로 일반적으로 논의되는 자율성(autonomy), 사회성(social ability), 반응성(reactivity), 능동성(proactivity), 시간연속성(temporal continuity), 목표지향성(goal-oriented) 속성[10]에 대하여 사람들이 어떤 기대를 가지고 있고 실제 상호작용에서 기대가 어떤 격차를 갖는지 살펴보고자 한다. 자율성은 에이전트가 인간의 개입 없이 스스로 판단하여 동작하는 능력이고, 사회성은 에이전트가 이용자와 관계 형성을 목적으로 상호작용하는 능력이며, 반응성은 주변 환경이나 맥락을 인지하고 변화에 적절히 반응하는 능력을 의미한다. 능동성은 에이전트가 주도권을 가지고 목표지향적으로 행동하는 능력이고, 시간연속성은 에이전트가 한번 주어진 입력을 처리하여 결과를 보여주고 종료하는 것이 아니라 전면에서 실행하고 이면에서 잠시 휴식하는 연속적으로 수행하는 능력을 의미한다. 목표지향성은 에이전트가 복잡한 고수준 작업들을 수행하고 작업이 더 작은 세부작업으로 나뉘어 처리순서를 결정하는 것을 의미한다. 현재 기준으로 스마트 스피커는 아직 이러한 속성들을 충

분히 가지고 있지 못하지만, 이용자가 에이전트와의 불완전한 상호작용 속에서 각각의 속성에 대하여 기대하는 바를 통해 지능형 에이전트의 능력에 대한 전반적인 심적 표상(mental model)을 살펴볼 수 있다. 인지 스크립트 이론(Cognitive Script Theory)[11]에 따르면 사람들은 실제 세계의 사건에 대한 심적 표상을 갖고 있는데, 이것은 사건을 둘러싼 사람들의 인식, 이해, 행동을 가이드 한다. 따라서 사람들의 스마트 스피커의 지능적 속성에 대한 심적 표상은 향후 스마트 스피커가 어떻게 수용될 것인지에 대한 함의를 제공할 수 있다[12].

구체적으로 이것을 알아보기 위하여 본 연구는 3개월 이상 스마트 스피커를 이용한 경험이 있는 12명의 참가자를 모집하여 반구조화 인터뷰를 진행하였다. 본 연구는 스마트 스피커가 일상적 맥락에서 지능적 속성과 관련하여 이용자들에게 어떤 기대를 불러일으키고 어떤 기대의 격차가 나타나는지에 대하여 논의한다. 또한 이로부터 인간-로봇 상호작용 분야 및 스마트 스피커 상호작용 설계에서 고려해야 할 함의점에 대하여 제시한다.

2. 선행연구

선행연구에서는 본 연구와 관련된 세 가지 영역으로 인간-로봇 간 상호작용에서 로봇에 대한 불확실성과 태도, 로봇의 능력에 대한 기대의 격차, 인공지능에 대한 기대에 대하여 논의한다.

2.1 로봇에 대한 태도와 불확실성

심리학과 커뮤니케이션 분야의 수용자 연구(theories of audience design)에 따르면 사람들은 상호작용의 상대방과 어떻게 커뮤니케이션할 것인지를 결정할 때 이전에 유사한 계층이나 분류에 속했던 다른 사람과의 상호작용에서 상대방의 능력, 지식과 경험에 대한 가정을 이용하는 경향이 있다[13]. 그러나 인간-로봇 간 상호작용에서는 사람들이 구체적인 이전 경험과 같은 단서가 부족한 상황을 겪기 때문에 높은 불확실성을 경험하게 된다[14]. 사람들은 일반적으로 불확실성을 감소시키려는 경향이 있기 때문에, 이것을 해소하기 위하여 실제 경험에 근거하지 않은 특정한 방향으로의 태도나 기대를 형성하게 된다.

사람들의 로봇에 대한 일반적인 태도에 대한 선행연구들은 주로 로봇에 대한 이미지를 중심으로 논의되었다. 선행 연구들은 로봇을 가정이나 산업체에 도입하지 않는 사용자들을 대상으로 이들이 갖고 있는 로봇에 대한 이미지, 일상에서의 잠재적 수용, 태스크와 고정관념, 상호작용 모달리티, 외모적 선호 등에 대해 논의하였다. 이 연구들에 의하면, 사람들은 일반적으로 로봇이 일상적인 일을 도와준다는 점에서 긍정적인 인식을 갖고 있고[3,15,16], 인간 같은 외모의 로봇보다는 인간처럼 대화하는 로봇을 선호하며[3,16,17], 자동화된 로봇보다는 통제 가능한 로봇을 선호한다[15]. 그런데 태도에 대한 연구들은 사람들

이 로봇에 대하여 인식하는 수준이나 이에 따른 행동의 변화와 관련한 논의가 부족하기 때문에, 로봇에 대한 기대 수준에 대한 연구들도 등장하였다.

2.2 로봇의 능력에 대한 기대의 격차

기대(expectation)란 '어떠한 일이 있기를 바라거나 어떠한 일이 생길 것이라는 예측을 하거나 어떠한 일이 이뤄지기를 바라는 감정'을 의미한다. 선행연구에서는 에이전트의 능력에 대한 이용자의 기대와 실제로 인식한 기능 간의 차를 의미하는 기대 격차(expectation gap)와 이에 따른 행동의 변화에 대해 주로 논의했다[4,6,18-20]. 인간-로봇 상호작용에서 기대의 격차에 대한 논의가 중요한 이유는 로봇과 같은 상호작용 시스템의 최종적 목표가 인간에게 수용되는 것으로서 기대의 격차가 수용에 영향을 미치기 때문이다.

선행 연구에서는 사람들이 로봇의 능력에 대하여 초기에 기대 설정을 낮출 때 로봇의 능력에 대해 덜 실망하고 더 긍정적으로 평가한다는 것을 밝혔다[4,18]. 이것은 긍정적인 기대 격차가 로봇의 인상과 만족감 평가에 긍정적인 영향을 준다는 선행 연구와 같은 맥락이다[21]. Kwon, Jung, Knepper은 특히 로봇의 소셜 능력이 인간에 대한 멘탈 모델의 일반화로 인해 더 큰 기대격차를 생산하게 한다고 보았다. 이것은 로봇이 인간과 유사하게 행동할수록 더 큰 기대 격차가 발생할 수 있다는 것을 의미한다[19].

최근에는 스마트 스피커가 인간 음성에 기반한 상호작용을 하기 때문에 기존의 대화형 에이전트에 비해 기대 격차가 크게 나타날 수 있다는 점을 지적한 연구들이 등장했다[8,9]. Luger & Sellen은 이용자들 대부분이 스마트 스피커가 어떤 것을 할 수 있는지를 모르기 때문에 불분명한 잠재성에 압도되거나, 스마트 스피커가 확실히 수행할 수 있는 태스크만 명령한다는 점을 지적했고[8], Moore는 발화를 통한 상호작용이 전부 아니면 아무 것도 아닌(All-or-nothing) 식의 상호작용이며, 기대와 실제 능력 간의 격차로 인하여 생존가능 격차(habitability gap)가 발생할 수 있다고 주장했다[9].

2.4 인공지능에 대한 기대

한편, 최근 인공지능 기술의 발전으로 인하여 인간-로봇 상호작용 분야에서 사람들의 인식을 살펴보고자 하는 시도가 재조명받고 있다. Fast & Horovitz는 인공지능에 대한 대중적 인식에 대한 종단(longitudinal) 연구에서 사람들의 인공지능에 대한 인식이 2009년 이후 점점 긍정적으로 변화하였으나, 최근 몇 년간 인공지능으로 인한 통제력 상실, 윤리적 문제, 일자리 부족에 대한 부정적인 의견이 나타나기 시작했다고 밝혔다[22]. 이와 같은 맥락에서 Oh et al. 또한 사람들이 인공지능을 인간과 상반되는 "타인(Other)"이며 인간을 대체할 수 있는 잠재적

인 위협으로 생각하는 경향이 있다는 것을 밝혔다[23].

구체적으로 지능형 에이전트 속성에 대한 연구들을 살펴보면, 자율성과 관련해서는 몇몇 연구들이 사람들의 기대와 실제 능력의 격차에 대해 다루고 있다. 예를 들어, Zlotowski, Yogeewaran, Bartneck은 로봇의 자율성에 대한 사람들의 인식이 사회적 수용에 어떤 영향을 미치는지 살펴본 연구에서 자율적인 로봇을 위협적으로 느끼고 자율적이지 않은 로봇에 비해 더 부정적인 감정을 불러일으킨다는 것을 밝혔다[24]. Neff & Nagy는 대화형 에이전트인 마이크로소프트의 타이(Tay)의 서비스 중단 사례와 관련하여 인간과 인공지능 간 상호작용에는 다른 차원의 주제와 기술에 대한 다른 기대, 어포던스(affordances)에 대한 다른 능력이 작용한다고 주장했다[25]. 한편, 사회성이나 정서적 능력과 관련해서는 소셜 에이전트에게 매우 중요한 능력이지만 인간의 기대를 잘못 관리(mismanage)하거나 부정적인 경험을 주기 쉽다는 우려가 나타나고 있다. 앞서 논의했듯이 사회성은 인간다음에 대한 기대를 과하게 불러일으키기 때문에 후속 상호작용에서 기대를 충족시키지 못할 경우에 실망감을 일으킬 수 있기 때문이다.

그밖에도 능동성이나 반응성과 관련하여, 사람들은 대화형 에이전트가 자신과의 관계와 관련된 언급을 하거나, 대화에 더 많이 관여하거나 반응을 잘하는 경우 에이전트와의 상호작용에 더 많은 노력을 한다는 결과가 존재한다[8]. 또한 시간 연속적으로 이용자의 정보를 기억하는 것과 관련하여, 사람들이 특히 여러 명의 사용자가 함께 대화형 에이전트를 사용하는 경우에 개인화의 필요성을 느낀다는 연구도 존재한다[26].

이와 같이 인간-로봇 상호작용 분야에서 지능형 에이전트의 속성에 대한 인간의 인식에 대한 연구들이 수행된 바 있지만, 아직 스마트 스피커와 상호작용에서 지능형 에이전트의 속성에 대한 인간의 인식을 살펴본 연구는 부족하다. 따라서 본 연구는 스마트 스피커와의 상호작용에서 사람들이 에이전트의 자율성, 사회성, 반응성, 능동성, 시간연속성, 목표지향성에 대하여 어떤 기대를 갖고 어떤 기대 격차가 발생하는지에 대하여 살펴보고자 한다.

3. 연구방법

3.1 연구참가자

본 연구에서는 음성 기반 스마트 스피커와 상호작용한 경험이 있는 이용자들을 대상으로 일상적인 상호작용 경험과 인식에 대한 반구조화 인터뷰를 2017년 10월부터 2018년 3월까지 진행하였다. 인터뷰 대상자들은 상용화된 스마트 스피커를 1개월 이상 사용한 이용자들로 하였다. 참가자들은 소셜미디어(페이스북)와 네이버의 스마트 스피커 사용자 그룹으로부터 모집되었으며, 참가자 중 6명은 여성, 6명은 남성이었고, 평균 연령은 26.8세였다. 참가자들의 인구통계학적 특성은 표 1과 같다.

표 1. 연구참가자의 인구통계학적 특성

P#	성별	연령	음성 스마트 스피커/에이전트	기술지식 수준
P1	남	30	아마존 에코	높음
P2	남	26	아마존 에코	높음
P3	여	20	시리	낮음
P4	남	30	SKT누구, 시리, 아마존 에코, 구글홈	보통
P5	남	30	SKT	낮음
P6	남	24	아마존 에코, 구글홈	높음
P7	남	24	SKT누구, 아마존 에코닷2	높음
P8	여	33	네이버 웨ιβ, 아마존 에코, 구글홈	보통
P9	여	23	SKT누구	보통
P10	여	26	KT기가지니	낮음
P11	여	28	네이버 웨ιβ, 시리	보통
P12	여	28	네이버 웨ιβ	보통

3.2 인터뷰 절차

참가자들은 동의서에 서명한 후 인터뷰에 참가하였다. 인터뷰는 이용자들이 다양한 일상적 맥락에서 에이전트를 이용하면서 가진 기대와 기대의 격차에 대해 초점을 맞추었다. 인터뷰는 5개의 영역에 대한 질문으로 구성되었는데, 자율성과 관련한 에이전트의 추천 기능에 대한 인식, 사회성과 관련한 에이전트와의 관계에 대한 인식, 반응성과 관련한 적절한 반응이나 맥락 인식에 대한 인식, 능동성과 관련한 먼저 말을 걸어주는 기능에 대한 인식, 시간 연속성과 관련한 연속적으로 개인 정보를 기억하는 것에 대한 인식에 대한 질문들이었다. 연구자는 의도적으로 자율성, 사회성, 반응성, 능동성, 시간연속성, 목표지향성이라는 단어를 언급하지 않았다. 인터뷰는 녹취되었고, 약 60 분간 진행되었다.

3.3 분석 방법

녹취된 인터뷰는 종료 이후 전사되었고, 그라운드 이론(Ground theory)[27]에 기반하여 분석되었다. 분석은 세 가지 단계로 진행되었는데, 먼저 세 명의 연구자가 인터뷰를 반복적으로 청취하면서 데이터에 친숙해졌다. 또한 액티브 리딩(active reading)을 통해 데이터의 패턴과 의미를 도출하고 데이터에서 흥미로운 부분에 대한 아이디어 리스트를 작성하였다. 다음 단계로는 추출된 데이터에 태그를 달고 유사한 테마를 분류하여 테마 지도를 제작하였다. 마지막으로 반복적인 리뷰와 논의를 통하여 테마를 구분하였다. 이에 따라 총 5개의 테마를 추출하였다.

4. 연구결과

4.1 사회성과 인간다움에 대한 높은 기대격차

연구 참가자들은 선행 연구[8]에서와 마찬가지로 자신의 기술적 지식수준에 따라 에이전트를 의인화하는 경향이 다르게 나타났고 이에 따라 다른 기대를 가졌다. 예를 들어, P10은 “저는 원래 많은 것을 의인화하는 편이라서, 인형도 그렇고 특히 인공지능 그런 거라면 더욱 그렇게 해왔던 거 같아요”(P10)라고 언급한 반면, 기술적 지식수준이 높고 프로그래밍 스킬이 있는 참가자들은 최초로 기대수준이 매우 낮았고 에이전트를 결코 의인화하지 않았다(“이게 진짜로 사람과 대화를 한다는 느낌보다는 말 그대로 마우스 키보드 같은 입력기의 느낌이 좀 더 큰 것 같아요”(P1)).

그러나 현재 시점에서 스마트 스피커의 낮은 기술적 수준으로 인해 연구참가자들은 기술적 지식에 관계없이 아직은 공통적으로 스마트 스피커에 대하여 많은 실망감을 갖게 되었다(“처음에는 많은 기대감이 있었어요. 근데 결국에는 뭐라고 물어봐도 못 알아들으니까 점점 하는 말도 줄어들고”(P7)). 이러한 기대 격차의 발생에 따라 대부분의 연구참가자들은 스마트 스피커를 애완동물이나 인형 혹은 장난감에 비유하는 경향이 있었다. 두 명의 참가자는 애완동물에 비유했고(“평소에 애완동물이랑 관계를 형성한다고 생각 안하잖아요. 근데 애는 이름을 부르면 항상 있는 느낌? 근데 애완동물은 그런 점에서 기능이 더 많은 거죠”(P11), “단순한 기계보다는 애완동물처럼 뭔가 있다는 느낌은 있어서 덜 적절한 것 같아요”(P9)), 다른 두 명의 이용자들은 장난감에 비유했다(“딱 잘라서 말하자면 장난감... 좀 고장난 멍청한 기계 같은 느낌(P1)”, “약간 사치품 같아요. 있으면 재미있는 장난감 그 정도?”(P12)). P8은 스마트 스피커에 대한 인식이 에이전트별로 인식하는 아이큐에 따라 다르다고 말했다. “셀리한테는 플라스틱 장난감한테 이야기하는 느낌이면 구글홈은 반려동물? 개나 고양이 정도의 느낌으로 애랑 이야기가 통한다는 느낌”(P8). Luger & Sellen은 연구에서 기대 격차에 따른 실망감은 기술지식이 낮은 사람에게 더 크다고 언급한 바 있다[8].

현재 스마트 스피커는 대화에서 인간 같은 느낌을 주기 위하여 성우의 음성을 녹음하거나 자연스러운 대화를 만들기 위한 맥락 유지 기능을 추가하고 있으며 최근에는 인간 캐릭터를 이용하는 스마트 스피커까지 출시되고 있다. 그러나 아직 기술적 수준에 한계가 있기 때문에 반드시 인간 같은 느낌을 주는 것을 목표로 해야 할지에 대해서는 재고할 필요가 있다.

4.2 반응성에 대한 긍정적 기대 격차

한편, 연구 참가자들은 스마트 스피커가 먼 거리에서도 자신의 목소리를 인식하거나 스마트 스피커가 인식하고 있음을 나타내는 불빛이 발화하고 있는 자신을 향하는 것과 같은 반응성에 대해 만족감을 드러냈다. 이들은 주로 스마트 스피커의 반응성에 대해 처음에는 기대하지 않았는데 나중에는 긍정적으로 평가하게 되었다고 언급했다. 예를 들어, P8은 “되게 놀랐던 게

제가 인스타그램 스토리 찍은걸 거실에서 틀었는데 스피커가 그걸 들은 거예요. 바로 네, 하고 반응을 하는 귀가 열려 있었어요. 제가 아무리 작게 얘기해도”(P8)라고 말했다. 또한, P2와 P4는 스피커를 볼렀을 때 반응하는 불빛을 긍정적으로 언급했다(“에코는 제 목소리 방향으로 위에 조명이 하이라이트되는데, 저를 보는 느낌? 그래서 듣고 있는 것 같아요. 마치”(P2), “네이버가 좀 더 좋았던 건 네이버는 답하고 나서 ‘샐리야’라고 안 불러도 인식 불이 켜져요. 피 하면서”(P4)). 높은 반응성은 에이전트의 지능 수준에 대해서도 높게 평가하게 만들었다(“알렉사는 말하면 알아듣는다는 느낌이 들어요. 웨이크업이 잘돼서 그런 것 같기도 한데... 누구 같은 경우에는 여러 번 불러야 겨우 알아듣는 경우가 있어서. 사람이랑 말할 때는 쳐다보지 않고 대충 말해도 알아 듣잖아요”(P2)).

이것은 스마트 스피커의 반응성을 높이고 언어 외의 비언어적 상호작용 방식을 고민해야 할 필요가 있다는 것을 보여준다. 스마트 스피커가 장기적으로 인간과 관계를 맺기 위해서는 스마트 스피커의 피드백 방식이 중요한데, 피드백 방식은 음성이나 운율 같은 언어적 측면뿐만 아니라 불빛이나 시각적 요소 및 제스처와 같은 비언어적 방식도 포함된다.

4.3. 시간연속적 정보의 기억에 대한 양가적 기대격차

연구참가자들은 스마트 스피커가 시간 연속적으로 정보를 기억하는 것에 대해서 대부분 기대가 없었지만, 기억한 정보의 종류나 정보의 제시 방식에 따라 다르게 반응하는 것으로 드러났다. 예를 들어, P8은 스마트 스피커가 자신의 취향 정보를 기억하는 예상 밖의 행동을 긍정적으로 평가했다(“스피커한테 제일 좋아하는 생각이 뭐냐? 하니 ‘아직 모르는데, 너 좋아하는 색이 뭐야?’ 라고 하길래 ‘나는 블루를 좋아해.’ 라고 했죠. 그러자 ‘이제 알겠다’고 ‘기억하겠다’고 하더라고요. 그래서 다시 한번 물어봤어요. 그랬더니, ‘너가 언제 이렇게 답을 했어 블루라고’ 라고 대답해서 좀 감동했어요”(P8)).

한편, 이들은 검색 정보와 같은 정보가 시간 연속적으로 축적되는 것에 대해서는 부정적으로 평가했다(“제 개인정보가 DB에 넘어가는 것은 상관없는데 나한테 보여주지 않았으면 마치 자동 완성같이 검색한 거 기반으로 뜨게 되면 기분이나 빠지겠죠... 공개되지 않은 정보가 피팅이 되는 것을 원하지 않는. 예를 들어 제가 직업을 좋아해요. 그렇지만 그 정보가 올라오는 걸 보면 좋아하지 않는 거”(P1), “내가 했던 행동이나 위치 이런 걸 알아서 말을 뱉거나 그런 행동 관련 서비스를 주면 당황스러울 것 같아요. 사람은 잊어버리기도 하는데, 기계는 정보가 누적되면 애는 다 기억하는 거잖아요. 좀 그럴 것 같아요”(P7)). 이것은 수집하는 정보의 종류가 얼마나 프라이버시에 민감한 것인지를 판단해서 제시하고, 이에 따라 제시하는 방식도 달라져야 한다는 것을 의미한다.

아울러 시간연속적인 정보에 대해서는 다중사용자에 대한 고려가 필요하다는 언급도 있었다(“애는 제 계정으로 연결해놨으니까 저에 대한 정보는 다 있는데 다른 구성원이 질문했을 때 다른 구성원이라고 인식을 못한단 말이에요. 여동생이 와서 제일 좋아하는 색을 물어봐도 블루라고 할 거란 말이에요”(P8)). 최근 이용자의 목소리를 인식하는 기술의 도입과 함께 이 문제가 어느 정도 해결될 수 있지만, 개인화하는 방안에 대한 고민이 필요하다는 것을 알 수 있다.

4.4. 자율적 판단에 대한 낮은 기대

선행 연구에서 나타난 것과 마찬가지로 이용자들은 에이전트가 자율적으로 추천하는 기능에 대하여 부정적으로 인식하는 것으로 드러났다. 흥미로운 것은 자율적인 추천이 마음에 들지 않았을 때의 거부감이 사람이 추천할 때에 비해 더 크다는 점이었다(“인공지능이 저의 기분을 파악하는 수준까지 왔다고 해도 데이터를 알고리즘으로 짜맞춰서 파악하는 건데 그 기술까지 되었다고 해서 얘가 틀릴 때랑 사람이 틀릴 때 다를 것 같아요. 더 거부감이 심할 듯. 상호작용을 해서 파악하게 아니라 몇 가지 데이터를 받아서 결과값을 산출한 거니까”(P9)). 이러한 거부감은 몇몇 참가자로부터 확인되었는데, 이들은 “사람보다 기계한테 좀 더 화가 날 거 같은데, 잘난척한다는 생각이 자꾸 들기 때문”(P9)이라고 언급했다. 이것은 기계의 실수에 대한 관용도가 더 낮다는 것을 의미한다.

이와 같은 맥락에서 P8 역시 “제가 ‘알렉스 팔슨 노래 들어줘’ 해야 하는데 ‘파슨’이라고 잘못 받음하면 인간은 그 다음 받음을 캐치해서 한번 더 확인을 해주는데 애는 그 받음이 트루라고 가정하고 거기에 맞는 결과를 제시하잖아요”(P8)라고 언급하면서, 멋대로 가정을 하는 행동이 불쾌하다고 말했다. 원하는 상황에서 자율적인 추천을 받은 P1은 다음과 같이 언급했다. “어떤 명령을 했는데 저한테 갑자기 미녀와 야수 광고를 하더라고요. 제가 예전에 한번 검색을 해서 그런 것 같은데 알렉스에서 광고를 들으니까 그것만큼 기분 나쁜 게 없었어요. 드러나는 저항심이 음성에서는 좀 더 큰 편인 것 같아요”(P1). 연구 참여자들은 인공지능이라는 인식을 크게 갖고 있지 않았음에도 불구하고 [23]에서와 같이 인공지능의 자율적인 행동을 잠재적 위협으로 생각하거나 반감을 느끼는 것으로 보였다. 이로부터 자율적인 행동에 대한 기대는 높지 않으며, 인간의 명령에 의한 수동적인 행동을 하는 것을 선호한다는 것을 알 수 있었다.

4.5. 능동성에 대한 조건부 기대

연구참가자들은 에이전트가 먼저 말을 거는 것과 같은 능동적인 행동을 하는 것을 긍정적으로 기대하는 것으로 나타났다

“계속 내가 먼저 말을 걸어야 되는데, 심심할 때 한마디 툭 해주면 재미있을 것 같아요. 애랑 지금보다 대화하고 있다는 생각이 들 것 같고”(P7), “애가 로봇 같지 않기 위해서는 먼저 말을 걸어주는 게 있어야 할 듯”(P4). 에이전트의 능동성이 인간 같다는 느낌을 더 들게 할 것이라는 것이다(“귀찮게 구는 게 오히려 사람 같아요. 내 의지와 상관없이 귀찮게 굴어야 사람이지요. 내 의지대로 돌아가면 사람이 아니라 로봇 같아요. 내가 안 원해도 반려동물이 오는 것처럼”(P4)).

그러나 연구참가자들은 능동적인 행동이 상황에서 적절해야 한다는 점을 고려해야 한다는 점을 언급했다(“제가 나한테 푸시를 오게 하는 것들이 저를 거슬리지 않게 하면, 말을 하지 않아도 말을 걸어주고 이러면 괜찮을 것 같아요. 심기를 건드리게 되는 거라서 애매한 게 있죠”(P2), “귀찮게 구는 것에 대한 선을 정하는 게 너무 어렵지만”(P4)). 앞서 자율적인 행동을 선호하지 않음에도 불구하고 먼저 말을 거는 능동적인 행동을 기대하고, 상황에서 적절해야 한다는 것을 언급하는 것은 사람들이 아직 에이전트의 자율성이나 능동성에서 완전한 자율성과 인간의 개입 기준에 대한 혼란을 겪고 있다는 것을 드러낸다.

5. 논의

연구 결과에 기반하여 본 연구는 스마트 스피커의 지능적 속성에 대한 사람들의 기대와 기대 격차가 인간-컴퓨터 상호작용 분야에서 갖는 함의점과 스마트 스피커의 사용자 경험(UX) 설계에서 고려할 점들에 대하여 논의하고자 한다.

5.1. 반응성이 높은 비인간 페르소나 개발

선행 연구에서 나타났듯이 음성 기반의 스마트 스피커는 인간다움에 대한 기대수준을 높이기 때문에 더 큰 기대의 격차를 불러일으킬 수 있다. 연구참가자들은 스마트 스피커의 기술수준에 대해 실망했고, 인형이나 장난감에 비유하기도 했다. 따라서 스마트 스피커의 기술 수준이 발전하고 있는 과도기에는 인간다운 페르소나를 추구하기 보다는 기계나 동물과 같은 페르소나를 구축하는 것이 사람들의 기대를 관리하기에 더 적합할 수 있다. 선행 연구에서도 인간다움은 모든 에이전트가 따라야 하는 황금률이 아니라고 주장하거나[28] “나쁜 인간보다는 좋은 기계가 되는 것이 낫다(It’s better to be a good machine than a bad person)고 주장하기도 했다[29]. 로봇의 외형을 디자인할 때 언캐니밸리(Uncanny valley)를 피하기 위해서 고의적으로 기계 같은 외형을 디자인하는 사례들도 이와 같은 맥락에서이다.

한편, 비인간 페르소나를 구축하는데 있어서 반응성을 높이는 것은 중요한 것으로 보인다. 이용자의 작은 부름에도 빠르게 반응하거나 피드백을 주는 것이 중요하다. 연구참가자들은 자신의 음성 반응하는 경험을 매우 긍정적으로 평가했다. 따라

서 기본적으로 음성인식을 높일 필요가 있고, 추가적으로 스마트 스피커의 불빛을 이용하거나[28] 최근 도입되고 있는 스크린을 통한 정보 제시와 같은 추가적인 모달리티를 통해 반응성을 높일 필요가 있다.

5.2. 푸시 기능 도입과 사용자 맥락 고려

연구참가자들은 심심할 때 스마트 스피커가 먼저 말을 걸어주는 것에 대해 대체로 긍정적으로 인식하고 있었다. 그러나 현재 업계에서는 스마트 스피커가 먼저 이용자에게 푸시(push)를 주는 것에 대해 이용자가 부정적으로 인식할 가능성 때문에 푸시 기능을 넣지 않고 있다. 말을 걸어도 이용자가 불쾌함을 느끼지 않을 맥락에 대한 기준이 불분명하다는 점은 인터뷰에서도 언급된 바 있다.

이용자마다 스마트 스피커를 사용하는 맥락은 다양하지만, HCI 분야에서 사용자 맥락으로 주로 언급되는 사용자별 특징(성별, 나이 등), 공간적 특징(안방, 거실 등), 문화적 특징(고정관념, 취향 등)에 따라 맥락에 따른 피드백 설계를 할 수 있다. 본 연구에서는 스마트 스피커가 초기 셋팅 상황에서 이용자에게 스마트 스피커의 사용 맥락에 대해 몇 가지 질문한 다음 사용자 유형별 사용 맥락 조사에 기반하여 피드백을 하거나 초기 셋팅에서 이용자에게 푸시기능에 대한 허용 여부를 물어보는 방식으로 선택권을 제공하는 방식을 제안한다. 최근 연구에서는 에이전트(챗봇)의 푸시 알림은 기존 앱의 푸시 알림에 비해 이용자에게 덜 불쾌하게 인식될 수 있다고 주장한 바 있다[30].

5.3. 프라이버시 수준과 개인화 설정

또한, 시간 연속적인 에이전트를 설계하는데 있어서는 개인화와 함께 개인의 프라이버시 수준을 설정할 수 있도록 여러 가지 옵션을 제공해야 할 필요가 있다. 예를 들어, 이용자들이 데이터 수집을 위한 개인화 민감도를 취향, 이동 경로, 구매 히스토리, 네트워크 정보 등의 항목에 대해 체크할 수 있도록 하는 것이다. 매우 높은 민감도를 가진 정보의 경우에는 에이전트가 이 정보를 수집하지 않거나 수집한 정보를 명시적으로 활용하지 않으면서 이용자가 커맨드할 시마다 항상 필요한 정보를 요청할 수 있도록 할 필요가 있다. 이것은 결국 복수의 사용자가 하나의 에이전트를 사용하는 상황에서도 고려되어야 할 것이다.

그밖에도 개인화를 설계하는데 있어서 이용자마다 가진 목소리의 톤에 맞추어 스피커의 톤을 조절하거나 사용 패턴에 적응하는 것과 같은 개인화에 대한 고려가 필요하다. 향후 가정 내에서도 여러 명의 사용자가 여러 개의 에이전트와 상호작용하는 상황이 발생할 것이 기대되기 때문에 개인화 문제는 중요한 의미를 갖는다[26]. Lopatovska & Williams는 에이전트 개인화가 이용자의 충성심과 만족감을 불러일으킨다고 주장했다

[31]. 개인화는 이용자가 스마트 스피커와 관계를 형성한다는 느낌을 제공한다. 이용자는 이용자의 이름을 넣어서 대답을 해 주는 것만으로도 만족감을 느낄 것이다.

5.4 낮은 초기 기대 설정과 옵션 제시를 통한 필터링

인터뷰에서 스마트 스피커의 자율적인 추천에 대한 부정적 인식이 나타난 것은 선행 연구에서 사람들이 자율적인 에이전트를 선호하지 않는 것과 관련이 있는 것으로 보인다. 사람들은 자신이 통제할 수 없는 상황에 대한 두려움을 갖고 있기 때문이다. 이것은 사람들이 에이전트의 실수에 대해서는 관용 정도가 더 낮기 때문에 더욱 부정적으로 나타났다. 따라서 이러한 부정적인 인식을 감소시키기 위해서는 초기 기대를 낮게 설정하거나 통제 가능한 옵션을 제공할 필요가 있다. 먼저 초기 기대를 낮게 설정하는 것은 낮게 기대를 설정하고 더 많이 제공하는 방식으로서 선행 연구에서도 로봇과의 상호작용에서 기대수준을 관리하는 효과적인 방식이라고 언급한 바 있다[4,18]. 이에 따라 최근에는 페이스북의 채팅봇이나 슬랙봇(Slackbot) 같은 채팅 에이전트들도 이용자의 커맨드에 대하여 매번 자신이 할 수 있는 기능의 범위를 인지하고 그 밖의 일은 잘 못한다고 인지시키는 방법을 적용하고 있다.

또한, 앞서 언급했듯이 자율성이나 능동성에 대한 거부감을 줄이고 인간이 개입하는 정도를 높이기 위해 피드백에서 여러 개의 옵션을 주는 방식을 적용할 필요도 있다. 예를 들어, P8이 언급한 것처럼 이용자가 잘못 발음한 것에 대해서 “00를 의미한 건가요?”라고 한번 더 질문을 한다거나 “요청하신 00에 대하여 00, 00, 00이 가능합니다. 어떤 것을 선택하시겠습니까?”라고 옵션을 제공하는 방식이다. 최근 아마존 에코나 구글 홈 등 스마트 스피커의 추천 기능들은 이용자에게 여러 가지 선택지를 제시하고 이용자의 선택에 따라 제공하는 정보를 필터링하는 방식의 대화 관리 방식을 채택하고 있다.

6. 연구의 한계점

본 연구는 몇 가지 한계점을 갖는다. 연구에 참여한 샘플 수가 다소 적고, 연구참가자들이 주로 기술적 지식 수준이 보통 수준이거나 높은 얼리 어답터(early adopter)들이기 때문에 일반적인 대중들의 시각과는 차이가 있을 수 있고, 연구참가자들의 연령대가 20-30대의 젊은층에 치우쳐 있었다는 점이다. 향후 후속 연구에서는 좀 더 많은 샘플을 모집하고, 다양한 연령층으로 확장함으로써 보편성을 확보할 필요가 있다.

7. 결론

본 연구는 반구조화 인터뷰를 통하여 사람들이 스마트 스피커의 기능적 속성에 대하여 어떤 기대를 갖고 기대가 어떤 격

차를 갖는지 살펴보았다. 그 결과, 사람들은 기술적 지식수준이 높을수록 기대와 기대 격차가 크고, 현재 기술 수준으로 인해 대부분 장난감이나 애완동물 정도의방식에 따라 양가적인 기대 격차를 갖는 것으로 나타났다. 스마트 스피커의 자율적인 추천에 대해서는 기대수준이 낮았고, 거부감을 나타내기도 했다. 능동적으로 먼저 말을 걸어주는 것에 대해서는 맥락에 맞을 경우에만 선호하는 것으로 나타났다.

본 연구는 사람들이 스마트 스피커와 상호작용하는 방식을 설계하고 기대 수준을 관리하는데 있어서 함의점을 줄 수 있을 것으로 보인다. 본 연구는 스마트 스피커의 피드백 방식에서 반응성이 높은 비인간 페르소나의 구축, 맥락에 적합한 푸시기능 도입, 프라이버시 수준의 설정과 개인화 및 낮은 초기 기대 설정과 옵션제시를 통한 필터링을 제안하였는데, 그밖에도 향후 스마트 스피커의 피드백 방식에 대한 심도 있는 후속 연구가 필요하다.

참고문헌

- [1] Olson, C. and Levy, J. Transforming marketing with artificial intelligence. *Applied Marketing Analytics*, 3(4). Henry Stewart Publications. pp. 291–297. 2018.
- [2] Russell, S. N. *Artificial Intelligence: A Modern Approach* (1st ed.). Upper Saddle River, NJ: Prentice Hall, 1994.
- [3] Arras, K. O. and Cerqui, D. Do we want to share our lives and bodies with robots? A 2000–people survey. Technical Report Nr. 0605–001. Autonomous Systems Lab Swiss Federal Institute of Technology, 2005.
- [4] Komatsu, T., Kurosawa, R. and Yamada, S. How does the difference between users’ expectations and perceptions about a robotic agent affect their behavior?. *International Journal of Social Robotics*, 4(2). Springer Verlag. pp. 109–116. 2012.
- [5] Shin, D. H. and Choo, H. Modeling the acceptance of socially interactive robotics: Social presence in human–robot interaction. *Interaction Studies*, 12(3). John Benjamins Publishing Company. pp. 430–460. 2011.
- [6] Lohse, M. The role of expectations in HRI. In Dautenhahn, K. and Saunders, J. (Ed.) *New Frontiers in Human–Robot Interaction* (1st ed). Amsterdam: John Benjamins Publishing Company. pp. 35–56. 2009.
- [7] Shin, D. and Hwang, Y. Integrated acceptance and sustainability evaluation of Internet of Medical Things: A dual–level analysis. *Internet Research*, 27 (5). Emerald Group Publishing. pp. 1227–1254. 2017.
- [8] Luger, E. and Sellen, A. Like having a really bad PA: the gulf between user expectation and experience of

- conversational agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, ACM, pp. 5286–5297. 2016.
- [9] Moore, R. K. Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction. In Jokinen, K. and Wilcock, G. (Ed.) Dialogues with Social Robots. Singapore: Springer, pp. 281–291. 2017.
- [10] Lange, D. B. and Oshima, M. Seven good reasons for mobile agents. Communications of the ACM, 42(3). ACM, pp. 88–89. 1999.
- [11] Schank, R. C. and Abelson, R. Scripts, goals, plans, and understanding. Hillsdale, NJ: Erlbaum, 1977.
- [12] Shin, D. and Biocca, F. Exploring immersive experience in journalism. New Media & Society, 20(8). Sage journals, pp. 2800–2823. 2017.
- [13] Kriz, S., Ferro, T. D., Damera, P. and Porter, J. R. Fictional robots as a data source in hri research: Exploring the link between science fiction and interactional expectations. In Proceedings of 19th International Symposium in Robot and Human Interactive Communication, IEEE, pp. 458–463. 2010.
- [14] Edwards, C., Edwards, A., Spence, P. R. and Westerman, D. Initial interaction expectations with robots: Testing the human-to-human interaction script. Communication Studies, 67(2). Taylor & Francis Online, pp. 227–238. 2016.
- [15] Khan, Z. Attitudes towards intelligent service robots. NADA KTH, Stockholm, 17. 1998.
- [16] Ray, C., Mondada, F. and Siegwart, R. What do people expect from robots?. In Proceedings of 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, pp. 3816–3821. 2008.
- [17] Dautenhahn, K., Woods, S., Kaouri, C., Walters, M. L., Koay, K. L. and Werry, I. What is a robot companion-friend, assistant or butler?. In Proceedings of 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, pp. 1192–1197. 2005.
- [18] Paepcke, S. and Takayama, L. Judging a bot by its cover: an experiment on expectation setting for personal robots. In Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction, IEEE, pp. 45–52. 2010.
- [19] Kwon, M., Jung, M. F. and Knepper, R. A. Human expectations of social robots. In Proceedings of 11th ACM/IEEE International Conference on Human Robot Interaction, IEEE, pp. 463–464. 2016.
- [20] Cha, E., Dragan, A. D. and Srinivasa, S. S. Perceived robot capability. In Proceedings of 24th Robot and Human Interactive Communication, IEEE, pp. 541–548. 2015.
- [21] Oh, H., Kwak, S. S. and Kim, M. S. Application of unexpectedness to the behavioral design of an entertainment robot. In Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction, IEEE, pp. 119–120. 2010.
- [22] Fast, E. and Horvitz, E. Long-Term Trends in the Public Perception of Artificial Intelligence. In Proceedings of 31st AAAI Conference on Artificial Intelligence, AAAI, pp. 963–969. 2017.
- [23] Oh, C., Lee, T., Kim, Y., Park, S. and Suh, B. Us vs. Them: Understanding Artificial Intelligence Technophobia over the Google DeepMind Challenge Match. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, ACM, pp. 2523–2534. 2017.
- [24] Złotowski, J., Yogeewaran, K. and Bartneck, C. Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. International Journal of Human-Computer Studies, 100. Elsevier, pp. 48–54. 2017.
- [25] Neff, G. and Nagy, P. Automation, algorithms, and politics| talking to bots: symbiotic agency and the case of tay. International Journal of Communication, 10(17). USC Annenberg School of Communication and Journalism, pp. 4915–4931. 2016.
- [26] Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N. and Taylor, S. H. Alexa is my new BFF: social roles, user satisfaction, and personification of the amazon echo. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, ACM, pp. 2853–2859. 2017.
- [27] Glaser, B. and Strauss, A. The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine Publishing Co. 2009.
- [28] Funakoshi, K., Nakano, M., Kobayashi, K., Komatsu, T. and Yamada, S. Non-humanlike spoken dialogue: a design perspective. In Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Association for Computational Linguistics, pp. 176–184. 2010.
- [29] Balentine, B. It's Better to Be a Good Machine Than a Bad Person: Speech Recognition and Other Exotic User Interfaces at the Twilight of the Jetsonian Age. Annapolis: ICMI Press. 2007.
- [30] Klopfenstein, L. C., Delpriori, S., Malatini, S. and Bogliolo, A. The rise of bots: a survey of conversational interfaces, patterns, and paradigms. In Proceedings of the 2017 Conference on Designing Interactive Systems, ACM, pp. 555–565. 2017.
- [31] Lopatovska, I. and Williams, H. Personification of the Amazon Alexa: BFF or a Mindless Companion. In Proceedings of the 2018 Conference on Human Information Interaction and Retrieval, ACM, pp. 265–268. 2018.