

# 메신저 기반의 모바일 챗봇 서비스 사용자 경험 평가 -구글(Allo)과 페이스북(M messenger)을 중심으로-

강희주<sup>1</sup>, 김승인<sup>2\*</sup>

<sup>1</sup>홍익대학교 국제디자인전문대학원 디지털미디어디자인전공

<sup>2</sup>홍익대학교 국제디자인전문대학원 교수

## Evaluation on the Usability of Chatbot Intelligent Messenger Mobile Services -Focusing on Google(Allo) and Facebook(M messenger)

Hee Ju Kang<sup>1</sup>, Seung In Kim<sup>2\*</sup>

<sup>1</sup>Digital Media Design, International Design School for Advanced Studies, Hongik University

<sup>2</sup>Prof. of Digital Media Design, International Design School for Advanced Studies, Hongik University

**요약** 본 연구는 해외에서 대표적으로 인식되는 메신저 기반의 어시스턴트 챗봇 서비스인 구글(Allo)과 페이스북 M을 대상으로 사용성 경험 평가를 하고, 연구결과를 바탕으로 국내 챗봇 서비스의 사용성을 개선하고 발전 방향을 모색하는데 목적이 있다. 우선, 문헌연구를 하여 인공지능(AI) 챗봇 서비스에 대한 전반적인 이해와 챗봇의 특성 및 국내 메신저 기반 어시스턴트 챗봇 서비스의 현황 및 전망을 정리 및 분석하였다. 사용자 평가를 실시하기 위해 피터 모빌의 허니콤 모델을 재구성하여 사용자들에게 심층 인터뷰를 실시하였다. 그 결과 다음과 같은 개선 사항들이 도출되었다. 첫째, 사용자가 챗봇의 기능들을 쉽게 이해할 수 있도록 직관적인 요소들을 사용해야 한다. 둘째, 사용성에 방해가 되는 챗봇의 질문을 개선해야 한다. 셋째, 사용자의 만족도 향상을 위해 답변 정확도를 높여야 한다. 향후 본 연구를 바탕으로 사용자들과 챗봇의 사용성 평가가 지속해서 이루어진다면, 해당 서비스에 대한 구체적인 연구와 가이드라인을 제시할 수 있기를 기대한다.

• **주제어** : 챗봇 서비스, 메신저, 모바일 서비스, 애플리케이션 서비스, 사용자 경험 디자인

**Abstract** This project has been conducted to improve the usability of Chatbot Services such as Google(Allo) and Facebook M(Messenger). Based on the evaluation, this study aims to suggest the solutions to improve the usability of domestic Chatbot services and future directions for their development. It provides the overall understanding of the AI Chatbot service and the feature of Chatbot service through literature search. Furthermore, we summarized the current standing and the prospect of domestic messenger-based assistant Chatbot services. For conducting user evaluation, Peter Morville's honeycomb model is applied to in-depth user interviews. The followings are elements that could be amended to improve the service. The service should be incorporated by intuitive elements for users' understanding its functions and eliminate any elements that interfere with usability. The accuracy should be increased to improve the user satisfaction. This research will provide the future guidelines to improve the usability of Chatbot services through continuous evaluation by users.

• **Key Words** : Chatbot Services, Intelligent messenger, Mobile Service, App Service, User Experience

\*Corresponding Author : 김승인(r2d2kim@naver.com)

Received June 2, 2017

Accepted September 20, 2017

Revised August 31, 2017

Published September 28, 2017

## 1. 서론

### 1.1 연구의 배경 및 목적

전 세계적으로 사람들이 모바일 메신저를 소셜 미디어인 페이스북, 트위터 등 보다 더 많이 쓰기 시작하였다. 메신저 기반 SNS가 편리하게 사용됨에 따라 챗봇에 대한 영향력이 커져, 모바일 메신저는 이제 단순한 메시지 기능을 넘어 플랫폼 영역을 확장하고 있다. 인공지능 기계와 채팅을 통해 원하는 서비스나 정보를 얻는 ‘챗봇’은 사용자가 별도로 웹 사이트나 애플리케이션을 따로 실행하지 않고도 대화하듯 정보를 얻을 수 있고, 메시지만 보내면 원하는 서비스를 받을 수 있다.

최근 들어 인공지능 기술이 발전하면서 컴퓨터가 인간의 대화를 기계적으로 학습하여 자연스러운 인간의 언어로 질문이나 명령을 해도 맥락을 파악하여 응답할 수 있게 되었다[1]. 마치 사람과 대화를 하는 것처럼 대화를 통해 정보 검색을 돕고, 딥러닝 기술을 이용하면서 대화가 축적될수록 스스로 학습을 통해 상황에 따라 고객이 원하는 것을 알아서 추천할 수 있는 기반이 마련됐다.

현재 챗봇 시장은 인공지능이 접목되면서 소수의 관련 기술 중심의 글로벌 기업에 의해 주도되고 있는 양상을 띠고 있다[2]. 시장조사 기관인 TMR(Transparency Market Research)에 따르면, 2015년 구글, 페이스북, 마이크로소프트 등 글로벌 기업이 챗봇 시장이 97.5%를 점유하고 있다고 발표했다.

세계 최대 모바일 메신저인 왓츠앱과 페이스북 메신저를 운영하는 페이스북은 인공지능을 적용한 챗봇 베타 버전을 발표했다. 채팅을 통해 사용자가 원하는 정보를 검색해주거나 구매, 예약을 도와주는 인공지능 가상비서 ‘M’을 활용한 서비스다. 그뿐만 아니라, 자사 메신저 플랫폼에 기업과 개발자들이 쉽게 챗봇을 구축해 소비자들이 음식 주문, 날씨 안내, 온라인 쇼핑, 결제, 자동 상담 등 챗봇 서비스를 활용할 수 있도록 하였다[3]. 2016년 9월 21일 구글은 머신러닝을 적용한 구글 AI 가상비서인 ‘Allo’라는 애플리케이션을 공개했다. 이것 역시 ‘@google’이라고 입력하면 필요한 정보를 찾아주는 가상비서 메신저다.

국내 모바일 메신저 대표 서비스인 카카오톡과 네이버 라인도 국내 AI 시장 선점을 위해 인공지능 개발에 힘을 쏟고 있다고 한다[4]. 위와 같은 해외 챗봇 서비스를 비교하는 것은 국내 모바일 메신저 서비스 사용성을 개선한다는 점에서 의미 있는 연구가 될 것으로 사료가 된

다. 이에 본 연구는 해외 대표적 서비스인 구글(Allo)과 페이스북(M)의 사용성 경험 평가를 비교분석으로 향후 국내 메신저 챗봇 서비스 사용성 개선을 위한 필요 사항을 도출하고 제언하는 데 목적이 있다.

### 1.2 연구 범위 및 방법

본 연구는 위와 같은 목적에 따라 1차로 문헌연구를 통해 모바일 메신저기반 챗봇 어시스턴트 서비스의 정의, 특성, 현황, 전망 등을 고찰하였다. 2차로 구글 알로와 페이스북 M의 사용성 평가 실험을 하기 위해 해당 서비스의 사용경험이 없는 이들을 대상으로 사용성 평가(usability test)를 실시하였다. 또한, 피터모빌의 허니콤 모델을 재구성하여 각 서비스에 대한 의견을 도출하고, 해당 서비스의 사용경험이 있는 사용자를 대상으로 심층 면접 인터뷰를 진행함으로써 각 서비스들의 사용성을 평가하고 분석하였다.

## 2. 메신저기반 어시스턴스

### 챗봇 서비스의 이해

#### 2.1 메신저 기반 어시스턴스 챗봇 서비스 정의

‘채팅’과 ‘로봇’을 합성한 용어인 ‘챗봇’은 사람과의 문자 대화를 통해 질문에 알맞은 답이나 각종 연관 정보를 제공하는 ‘인공지능(AI)기반의 커뮤니케이션 소프트웨어’를 칭한다[5]. 방대한 데이터와 기계학습 방식인 딥러닝을 기반으로 사람 도움 없이도 질문에 답해주고 대화하는 알고리즘이다. 챗봇은 봇(bot)의 일종으로, 현재 봇 종류에는 챗봇, 음성인식봇, 가상 개인 비서가 있다.

가상 개인 서비스란 머신러닝, 음성인식, 문장분석, 상황인지 등 인공지능 기술과 첨단 기술의 결합으로 사용자의 언어를 이해하고 사용자가 원하는 지시사항을 수행하는 소프트웨어 애플리케이션으로 정의한다[6].

#### 2.2 메신저 기반 어시스턴스 챗봇 서비스 특성

본 연구의 대상인 Google(Allo)과 Facebook(M)은 대표적인 모바일 메신저이며 챗봇 형태 서비스를 제공하며 다음과 같은 특징을 지닌다.

첫째, 대화형 인터페이스상에서 규칙 또는 인공지능으로 사용자와 소통하는 서비스이다[7]. 대화형 인터페이스는 사용자가 채팅하는 방법을 새로 배울 필요가 없으며,

어느 환경에서든 대화의 맥락을 놓치지 않고 연속된 경험을 할 수 있어 사용자를 포함한 사람들, 브랜드, 서비스, 봇 등과 소통할 수 있다[8].

둘째, 접근이 가능하며 개인에 맞춘 가상비서 서비스 경험을 제공한다. 인공지능 기술을 기반으로 메시지 인식을 통해 사용자가 원하는 최적의 답변을 찾아낼 수 있다[9].

셋째, 문자(Text) 형식 인간의 언어와 감성을 컴퓨터가 이해하고 학습하기 위한 데이터 분석 기술인 자연어 처리(Natural Language processing), 텍스트 마이닝(Text Mining), 패턴인식, 상황인지(Context-Awareness) 등의 기술이 포함된다[10].

넷째, 빅데이터 분석 기반으로 개인별 맞춤 서비스를 받는다.

다섯째, 애플리케이션을 앱 스토어에서 내려받을 수 있다.

여섯째, 두 서비스의 대화 형태는 열린 구조이며, 봇이 질문의 의도를 파악하고 의도에 해당하는 답변을 제공한다.

## 2.3 메신저 기반 어시스턴스 챗봇 서비스 현황 및 전망

최근 인공지능이 미래의 유망기술로 떠오르며 가상 개인비서 시장은 2024년 412억 달러 규모로 증가할 것으로 전망되었다[11]. 국내에서도 챗봇을 대화형 커머스 플랫폼으로써 유통 및 금융산업 등 다양하게 활용되고 있다. 모바일 메신저를 가진 카카오톡과 네이버는 인공지능 경쟁에서 뒤처지면 포털-메시지 시장에서의 위상이 흔들 것이라는 위기감이 팽배하여 인공지능 플랫폼 사업에 속도를 내고 있다[12].

카카오톡은 국내 메신저 점유율이 94%이며 30만 개 기업 아이디를 확보하고 있다. 최근 카카오프레인 설립하여 챗봇 시스템을 고도화하고, AI를 접목하여 다기능 개인비서를 만들 계획이라고 밝혔다[13]. 네이버는 대화형 인공지능 시스템 라운을 개발하여 네이버 앱, 네이버 톡톡, 유니버 앱 서비스에 활용되고 있다[14]. 국내 금융권에서 챗봇을 도입한 곳은 우리은행, 기업은행, 신한카드, 카카오뱅크, NH 농협은행, K뱅크가 있으며, 유통업계에서 도입한 곳은 인터파크, 11번가, GS홈쇼핑, CJ 오쇼핑, 배달 앱 숙박 앱 등이 있다[15].

모바일 대화형 커머스 서비스 구축과 강화를 통해 고

객 확보와 추가구매 및 재방문을 유지할 수 있으며, 챗봇 서비스는 다른 기기 간의 연계를 통하여 다양한 분야에 서 발전 활용될 전망이다.

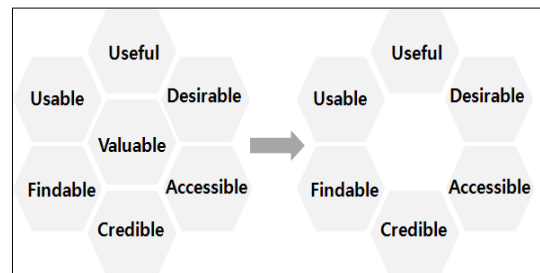
## 3. 실험 방법 및 분석 결과

### 3.1 실험 대상 선정

본 연구의 사용성 평가 대상은 모바일 가상비서 챗봇 서비스인 구글(Allo)과 Facebook(M messenger)이다. 피실험자 대상은 모바일 메신저를 자주 이용하지만, 구글(알로)와 페이스북(M 메신저) 서비스 사용 경험이 없는 20~30대 대한민국 남녀를 대상으로 실험자를 선정하였으며, 각각 3명씩 총 6명으로 구성하였다.

### 3.2 실험 방법

사용자 경험 측정을 위하여 [Fig. 1]의 피터 모빌의 허니콤 모델 7가지 사용성 원칙을 이용하였다. 모바일 애플리케이션의 특성상 가치성에 대한 항목의 결과가 적합하지 않다고 판단하여 피터 모빌의 허니콤 모형 7가지 원칙 중 가치성을 제외하여 재구성하였다. [16]



[Fig. 1] Reconstructing the principle of Honeycomb Model of Peter Morville

#### 3.2.1 태스크 수행 평가

아래 <Table 1>과 같이 실제 서비스의 수행 절차대로 태스크를 진행하였다. 태스크를 실행하면서 피실험자에게 태스크 단계마다 심층 인터뷰를 하여 피실험자의 수행을 분석하였다.

6가지 사용성 평가를 진행하기 위해 주요기능에 관한 태스크를 수행한 실험자들에게 추가로 실험을 진행하였다. 허니콤 항목 분류에 따른 설문 문항 설계는 <Table 2>와 같다. 주요 내용을 비롯해 총 30문항에 대한 심층 면접 인터뷰를 진행하였다.

〈Table 1〉 Usability Evaluation Test

Task	Type	Contents
Task1	Call a friend	Talk to your opponent
Task2	Appointment Time	Set lunch time
Task3	Place Reservation	Get recommendations on restaurants
Task4	Add an event	Add an event at 12
Task5	Send a voice msg	Voice Message
Task 6	Search past conversations	Search for "Place"
Task7	Places to Travel	Recommendations on Honeymoon
Task8	Interlocking service	Flight Booking
Task9	Subscribe	Follow News/Shopping
Task10	Emotional Expression	A happy sign
Task11	Gift Recommendation	Recommendation on earring

〈Table 2〉 Questionnaire design based on Honeycomb Model classification

Item	Contents
Accessible	<ul style="list-style-type: none"> <li>Is it easy to talk to a virtual assistant chatbot?</li> <li>Is the service what you expected?</li> <li>Do you think there is enough guidance information to use the service?</li> </ul>
Desirable	<ul style="list-style-type: none"> <li>Does the service provide differentiated functionality?</li> <li>Do you think it provides the right graphical response?</li> <li>Is menu name, icon, and button provided by the service easy to understand?</li> <li>Do you feel pleasure in design of sticker / emoticon?</li> <li>Did you feel the brand identity when you saw the screen design?</li> </ul>
Credible	<ul style="list-style-type: none"> <li>Do you trust the result of the virtual assistant's answer?</li> <li>Do you think the service is responding to you properly?</li> <li>Is it easier for you to ask questions again when you make a mistake?</li> </ul>
Findable	<ul style="list-style-type: none"> <li>Is it easy to find the word you were talking about last time?</li> <li>Is it easy to find the content you are looking for?</li> <li>Is it easy to see what each menu functions?</li> <li>Is the questioning menu convenient?</li> </ul>
Useful	<ul style="list-style-type: none"> <li>Do you think the functions provided by the service are useful to you?</li> <li>Does it provide the right features for forming a bond with the acquaintance?</li> <li>Are the services presented differently depending on the question?</li> <li>Is interoperability with other services appropriate?</li> <li>Have you experienced a screen interruption due to a virtual assistant service pop-up window or notification window?</li> </ul>
Usable	<ul style="list-style-type: none"> <li>Is the user's guide easy to recognize?</li> <li>Is it easy to share various information?</li> <li>Is it possible to check and reply to the other person easily and quickly?</li> <li>Does it help you to build a bond relationship?</li> </ul>

### 3.4 실험 결과 분석

평가자에게 재구성된 허니콤 모델의 6가지 항목에 대해 5점 척도로 평가하게 하였다. 접근성, 가시성, 신뢰성, 검색성, 유용성, 사용성에 대한 평가 내용의 점수가 높을수록 긍정을 나타내며, 그 결과를 6가지 항목에 따른 점수로 나타내었다.

먼저 Allo는 사용성, 검색성부분에서 우수했으며, M은 가시성과 신뢰성 점수가 높았다.

그러나 평균적으로 보통, 혹은 그 이하의 점수를 나타내고 있으면 모든 항목에서의 개선이 필요함을 확인할 수 있었다. 항목별 세부 평가 내용의 점수를 살펴보면, <Table 3>과 같다.

〈Table 3〉 Comparison by usability evaluation result

	Contents	Allo	M messenger
Accessible	Steps	4.4	2
	Usage Forecast	3	1.4
	Information	3.2	2.2
	Chatting process	4.2	2.2
	Service Process	4.2	2.8
Desirable	Function	4	3
	Graphics	4.2	3.8
	Menu Configuration and UI Elements	4.6	4.8
	Sticker/Emoticon	3	2.2
	Brand Identity	4	1.8
Credible	Answer output	2.8	2.6
	Answer Flow	3.6	2.8
	Cancellation	3.4	1
	Credibility	4	3
	Mistakes	2.6	3.8
Findable	Keywords	3.8	1.4
	Search Results	3.4	2.6
	Contents Recognition	4.2	1.6
	Menu Functions	4	2
	Menu	2.6	1.8
Useful	Functionality	3.8	3.2
	Bond formation	3.4	2
	Differentiated Service	4.2	4
	Interlocking	4	2.8
	Interferes	4.5	2.6
Usable	Guideline	3.8	2.8
	Sharing	4.6	3
	Alarm	3.8	3
	Bonding Relationship	3.8	3.4
	Service	4.2	3

평가결과에 의한 비교표를 확인해보면 두 애플리케이션의 사용성에 대해 우수한 면을 갖고 있다. 챗봇이 채팅

형 인터페이스이므로 새롭게 배우지 않고, 쉽게 이용할 수 있어 편리하다. Allo의 경우 검색 엔진을 통해 챗봇이 답을 해주기 때문에 실시간으로 구체적이고 다양한 정보 제공이 가능하였지만, M의 경우 페이스북 'Extended App'을 통해 사용자가 직접 메신저 내에서 정보 제공 기능을 클릭하여야지만 받을 수 있었다. 이러한 이유로 챗봇 서비스와 깊은 유대감을 형성하기에는 Allo의 기능이 다소 편리하지만, Search Engine 외 서비스를 받기 적절치 않았고, M의 경우 사용자가 다른 사용자와 대화를 하는 도중, M이 나와서 자연스럽게 '일정 잡기, 콜택시 예약' 등 애플리케이션 밖의 서비스가 가능하였다.

신뢰성과 유용성 부분에서 이슈는 질문이 1-0, 2-0 순서에는 챗봇이 명확하게 대답을 할 수 있지만 1-0, 1-1, 1-2 한 주제로 파생되는 질문에는 연결을 짓지 못하여 다시 처음부터 물어봐야 하므로 어려움을 느낀다고 응답하였다.

#### 4. 결론

최근 전 세계적으로 사람들이 모바일 메신저를 소셜 미디어보다 더 많이 쓰기 시작하여 메신저 기반 SNS가 편리하게 사용됨에 따라 챗봇에 대한 영향력이 커져, 국내 모바일 메신저 어시스턴스 챗봇은 발전 중이다. 이에 본 연구는 해외 대표적 서비스인 구글(Allo)과 페이스북(M)의 사용성 경험 평가를 비교분석으로 향후 국내 메신저 챗봇 서비스 사용성 개선을 위한 필요 사항을 도출하고 제안하는 데 목적이 있다. 사용성 평가에 대한 이론적 고찰을 함으로써 모바일 메신저 챗봇 서비스와 사용성 평가에 대한 이해를 넓히고, 피터 모빌의 허니콤 모델을 재구성하여 실험을 진행하였다. 본 연구를 통한 결과는 다음과 같다.

첫째, 메신저 서비스 특성상 인터페이스를 습득할 필요가 없어 편리하지만, 사용자가 기능들을 쉽게 접근할 수 있도록 접근방법의 제공이 필요하다. 두 애플리케이션에서 가이드라인이 없는 기능들이 다소 많아, 사용자가 어떤 기능들이 있는지 모르기 때문에 쉽게 학습할 수 있도록 행동 유도성을 높여야 한다.

둘째, 사용자의 질문이 아닌 챗봇이 임의로 보여주는 연관성이 없는 질문들로 인해 사용자가 화면 방해 경험을 느낄 수 있으므로 정확도를 개선해야 한다. 질문을 통해 자연스럽게 더 많은 대화를 이어나가면서 가상비서

챗봇과 유대감을 형성할 수 있지만 대화의 흐름을 방해할 수 있다.

셋째, 사용자의 만족도 향상을 위해 답변 어려움을 감소시켜야 한다. 사용자가 가상의 서비스라고 판단하여 질문의 답변을 못 하였을 때 관용적인 태도를 보였으나, 사용자가 재질문을 하므로 서비스의 사용성을 떨어뜨리는 요인으로 작용하였다. 대화의 흐름이 자연스럽게 진행될 수 있도록, 서비스를 개선해야 할 것이다.

본 연구는 기존에 연구되지 않았던 메신저 챗봇인 구글(Allo)과 페이스북(M)의 사용자 평가 비교연구로써 사용성을 개선하는 데 필요한 사항들을 도출하고 발전 방향을 제시하였다는 것에 의의가 있다. 그러나 서비스가 해외 메신저 서비스인 만큼 영어로 실험이 진행되었기 때문에 언어적인 부분에 한계가 있음을 밝힌다. 향후 본 연구를 바탕으로 모바일 메신저 지능형 챗봇의 사용성 평가가 지속적으로 이루어진다면, 해당서비스에 대한 구체적인 방안과 가이드라인을 제시할 수 있을 것으로 기대한다.

#### REFERENCES

- [1] <http://www.newstomato.com/ReadNews.aspx?no=671231>
- [2] Chatbots and Artificial Intelligence: Market Assessment, Application Analysis, and Forecasts 2017 - 2022
- [3] J. H. Jang, "The emergence and development trend of Chatbot Service based on Artificial Intelligence", ICT Issues and Trends", NIA, 2016
- [4] [http://biz.chosun.com/site/data/html\\_dir/2017/02/14/2017021402818.html](http://biz.chosun.com/site/data/html_dir/2017/02/14/2017021402818.html)
- [5] S. H. Lee, Internet/IT, Hanhwa Investment Industry, 2017.
- [6] K. B. Research, Evolution of Artificial Intelligence Technology and Evolution of Virtual Private Secretary Service Research, No. 16-53, 2016.
- [7] M. S. Lim, "Smart Paradox : An Effect of Mobile Instnt Messengers", J.of the Korea Convergence

Society, Vol. 3, No. 4, pp. 7-13, 2012.

- [8] Y. Yoon, "Development of Smart Senior Classification Model based on Activity Profile Using Machine Learning Method", J. of the Korea Convergence Society, Vol. 8, No. 1, pp. 25-34, 2017.
- [9] S. Cho, "Development of User Oriented Vulnerability Analysis Application on Smart Phone", J. of the Korea Convergence Society, Vol. 3, No. 2, pp. 7-12, 2012.
- [10] K. H. KIM, "Artificial Intelligence", KISTI Market report, Korea Institute of Science and technology information, 2016.
- [11] "Artificial Intelligence Change, Power Report, KIISA, 2015.
- [12] <http://m.post.naver.com/viewer/postView.nhn?volumeNo=7611619&memberNo=10391613>
- [13] <http://www.edaily.co.kr/news/NewsRead.edy?DCD=A00504&SCD=JE41&newsid=04185286615828224>
- [14] <http://www.newspim.com/news/view/20160321000321>
- [15] <http://www.bikorea.net/news/articleView.html?idxno=15422>
- [16] M. Yeom, "Analysis of Cosmetics App Smart UI convergence Design in Mobile Environments", J. of the Korea Convergence Society, Vol. 7, No. 2, pp. 13-17, 2016.

## 저자소개

### 강 희 주(Hee Ju Kang)

[정회원]



- 2010년 2월 ~ 2013년 2월 : 시드니대학교 Visual Art 전공 (학사)
- 2014년 9월 ~ 현재 : 홍익대학교 디자인대학원 디지털 미디어 디자인과
- 2015년 5월 ~ 현재 : 한국 IBM Consultant

<관심분야> : 사용자경험디자인, 휴먼 컴퓨터 인터랙션, 인공지능

### 김 승 인(Seung-In Kim)

[정회원]



- 2001년 3월 : 홍익대학교 국제디자인전문대학교 교수
- 2006년 3월 : 홍익대학교 국제디자인전문대학원 디자인혁신센터 센터장
- 2017년 1월 : 사단법인 한국서비스디자인학회 회장

<관심분야> : 사용자경험디자인, 서비스디자인, 시각디자인