Computers in Human Behavior 128 (2022) 107093

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

User interactions with chatbot interfaces vs. Menu-based interfaces: An empirical study

Quynh N. Nguyen [a,*], Anna Sidorova [b], Russell Torres [c]

[a] Stockton University, USA
[b] Professor, University of North Texas, USA
[c] Assistant Professor, University of North Texas, USA

ARTICLE INFO

Keywords:
Chatbot
Conversational agent
Human-computer interaction
User satisfaction
Self-determination theory

ABSTRACT

Rapid advances in Natural Language Processing (NLP) are transforming customer service by making it possible to create chatbot applications that can understand users' intents and response in a human-like manner. Chatbots promise to enhance customer experiences by creating more personal customer interactions than those afforded by traditional menu-based web applications. But are chatbots always superior to more traditional user interfaces (UI)? This study seeks to understand the differences in user satisfaction with a chatbot system vis-a-vis a menu-based interface system, and identify factors that influence user satisfaction. Grounded in the self-determination theory, the research model proposed here focuses on the effect of chatbot use on perceived autonomy, perceived competence, cognitive load, performance satisfaction, and system satisfaction. An experimental study was conducted, and data were analyzed using Partial Least Square Structural Equation Modeling. The findings indicate that chatbot systems lead to a lower level of perceived autonomy and higher cognitive load, compared with menu-based interface systems, resulting in a lower degree of user satisfaction. Implications of these findings for research and practice are discussed.

1. Introduction

Recent advances in AI have led to the emergence and wide adoption of AI-powered devices and applications. Chatbots, or conversational agents, which communicate with users via chat or speech interfaces and perform basic tasks such as search and question answering, are among the most popular AI applications today. Industry experts predict the global chatbot market size to reach USD 2485.7 million by 2028 (Grand View Research, 2021). It is forecasted that consumer retail spend via chatbots will reach $142 billion by 2024, and chatbot adoptions will cut business cost in healthcare, banking, and retail sectors $11 billion per annum by 2023 (Insider Intelligence, 2021).

AI-powered chatbots are often integrated into customer-facing applications and serve as easy and convenient conduits between companies and customers, helping customers to search for information, send feedback, file complaints, and even make purchases.

While chatbots can be considered extensions of traditional websites, the two differ in several important ways. Websites are commonly designed around menu-based user interfaces, which allow users to navigate multimedia content, including text, images, audios, videos, and interactive content. Companies use websites as digital platforms to promote their products and services and reach out to customers, and such websites vary in complexity, ranging from simple websites that provide basic information, to full-fledged e-commerce platforms. First-time users not familiar with the menu structure may have trouble navigating websites to find information, especially if they do not know how information is categorized. One of the potential advantages of chatbots is the fact that first-time users do not need to learn website-specific navigation structure as the chatbots typically provide a simple standard user interface that allows users to type in their questions using natural language. Users can chat with chatbots via familiar messenger platforms such as Facebook Messenger, WhatsApp, Skype, etc. Chatbots are often used in conjunction with ML-based recommendation agents to provide personalized answers, as well as product and service recommendations to users. In the global chatbot market, about 45% of end users prefer to use chatbots as the primary mode of communication to request and receive customer service (Grand View Research, 2017). But are chatbots always superior to traditional menu-based systems?

While human-computer interaction (HCI) has been extensively examined in Information Systems (IS) literature (Sidorova et al., 2013),

* Corresponding author. Office: C 104-a, 101 Vera King Farris Drive, Galloway, NJ, 08205-9441, USA.
E-mail addresses: Quynh.Nguyen@stockton.edu (Q.N. Nguyen), Anna.Sidorova@unt.edu (A. Sidorova), Russell.Torres@unt.edu (R. Torres).

https://doi.org/10.1016/j.chb.2021.107093
Received 14 May 2021; Received in revised form 27 September 2021; Accepted 5 November 2021
Available online 12 November 2021

the distinctive nature of human-chatbot interactions and its potential similarity to a human-to-human conversation warrants additional investigations chatbot usability factors. This research builds on self-determination theory (SDT) and seeks to examine the role of perceived autonomy, cognitive effort, and competence, as mediating factors in user satisfaction with chatbots vis-a-vis menu-based interfaces. Specifically, the study aims to address the following research questions:

Question 1: Do users experience different levels of autonomy and cognitive effort when completing a task using a chatbot as compared to a menu-based interface?

Question 2: What factors influence users' performance satisfaction and system satisfaction when completing a task using a chatbot as compared to a menu-based interface?

The study is expected to contribute to HCI research by highlighting the role of autonomy, competence, and cognitive effort in explaining the difference in user experiences with NLP based chatbots vis-à-vis traditional menu-based interface. The rest of this paper proceeds with a brief review of chatbot literature and relevant theoretical perspectives, a set of testable hypotheses, and a description of the research methodology, data collection and analysis. We conclude the paper with discussions of our findings and their implications for theory and practice.

2. Literature review

2.1. AI chatbots

For the purposes of this study, we define a chatbot as *an AI application that uses NLP to understand and enable conversation between a human and a machine*. Chatbots are also called conversational agents or dialog systems. Innovations in AI, ML, and NLP in recent years have helped advance chatbot technology, and current AI chatbots can accomplish many tasks, such as checking the weather, ordering food, booking travel and hotel, providing product info, or automating enterprise IT help desks. AI chatbots continue to evolve as they interact with customers, thus enabling improved interactions with customers over time. Fueled by the wide adoption and use of messaging apps (Følstad & Brandtzæg, 2017), chatbots offer businesses a new way to connect and communicate with their customers. Virtual agents embedded in an online store to assist older users with search and navigation have been credited with increased perceived social support, trust, and patronage intention for online store (Chattaraman et al., 2012). Chatbots have been used in financial sector to enhance customer experience and improve operational efficiency (Jang et al., 2021).

Extant research sheds light on the factors that make user interactions with chatbots more successful. An study of user interactions with a social adaptive virtual agent in a job interview context found that the interviewees reacted socially to intelligent virtual agents when the agents perceived users' behavior and adapted their emotions and social attitude accordingly (Youssef et al., 2015). Experienced users have been found to be more satisfied with conversational agents as compared to novice users on all the evaluated dimensions including impression, control, effectiveness, navigability, learnability etc. (Semeraro et al., 2008). Users were found to exhibit different personality traits and communication attributes during their initial interactions with a chatbot, as compared to interacting with humans (Mou & Xu, 2017). Users were also found to adapt their language when communicating with AI chatbots: human-chatbot communication duration was longer, but users relied on shorter messages with limited vocabulary, compared to their communication with fellow humans (Hill et al., 2015). Chatbots design cues such as human-like language or name influence user perceptions of a chatbot as being more or less human-like (Araujo, 2018). Users have been shown to respond differently to different types of chatbot interfaces, e.g. simple text chatbots and chatbots with anthropomorphic design features (Ciechanowski et al., 2019; Sheehan, 2018). While users enjoyed interactions with both a simple text chatbot and a complex animated avatar chatbot, they had more pleasant experience with the simple text chatbot (Ciechanowski et al., 2019). When interacting with the avatar chatbot, users in the study experienced greater uncanny valley effect (Mori et al., 2012), meaning they had greater negative feelings and discomfort toward the human-like chatbot. Users exhibited preference for chatbots with similar personality traits to those of the user, with matching personality resulting in higher on user-chatbot engagement and leading to higher sales (Shumanov & Johnson, 2021). But what is the effect of chatbot interaction on user satisfaction with their own performance on the task and their motivation to engage with the task in the future? Self-determination theory identifies factors that drive intrinsic motivation and thus are important in the context of chatbot use and task performance.

2.2. Self-determination theory

Grounded in the social psychology literature which examines the impact of social environments on human motivation, attitude, and behavior, Self-Determination Theory (SDT) is a macro theory of human motivation and personality (Ryan & Deci, 2000b). SDT examines factors that energize human action and the degree to which individual behavior is self-motivated and self-determined. The key constructs of SDT include intrinsic and extrinsic motivation, and the basic psychological needs that drive motivation. Intrinsic motivation refers to the motivation based on internal rewards, and is marked by action initiated due to enjoyment, interest, and satisfaction (Ryan & Deci, 2000a). Numerous studies highlight the importance of intrinsic motivation in the IS use context, including a finding that over two-thirds of users use the Internet for personal enjoyment (Fallows, 2005). Extrinsic motivation refers to the motivation that is driven by external rewards, such as fame, money, praise, etc. An extrinsically motivated individual performs an action or behavior because of an external reward, even though he or she may not enjoy the activity (Ryan & Deci, 2000a). According to SDT, people are active organisms that have growth tendencies, i.e. they repeatedly show effort, master challenges, and integrate new experiences (Ryan & Deci, 2000b). These tendencies develop naturally but require social nutriment and support. SDT identifies three basic psychological needs, which, if satisfied, allow healthy development and function and foster motivation for human activity. These needs are competence, relatedness, and autonomy (Ryan & Deci, 2000b).

In IS research, SDT constructs have been used in the context of user's motivation to continue e-learning technology (Roca & Gagné, 2008; Sørebø et al., 2009). An SDT-based extension of the Technology Acceptance Model (TAM) was used to investigate e-learning continuance intention in the workplace, suggesting that users who feel autonomous and competent are more likely to continue using IT. As these basic needs driving intrinsic and extrinsic motivation are satisfied, users report increased perceived usefulness and perceived playfulness, and stronger intention to use the IT (Roca & Gagné, 2008). Perceived relatedness also positively affects perceived usefulness and perceived playfulness. A study of associations between SDT and UTAUT constructs showed strong correlations between performance expectancy and extrinsic motivation and between perceived enjoyment and intrinsic motivation (Lee et al., 2015). In business setting, a research study on business process configuration found that business process rules complexity significantly reduces competence, and perceived autonomy and perceived competency has an impact on user motivation (Torres & Sidorova, 2015).

3. Research model and hypotheses

In both chatbots vs. menu-based systems, the interaction process starts when a user formulates a specific task goal, such booking a trip to a destination on a desired travel date and forms an intention to use a system to complete the task. In the next steps, the user will execute a sequence of actions to complete the task. These steps are different between the two interfaces, thus it is expected to lead to different user experiences.

3.1. Perceived autonomy

Perceived autonomy is defined as the extent to which a person feels he/she is able to control how he/she performs a task within the system (Nikou & Economides, 2017). Studies in marketing literature show that consumers have different information control levels vary across different information systems (Ariely, 2000). When interacting with menu-based website, users have control over translating their intents into system commands through menu selections. By browsing the website, examining different input controls and menu options, they can familiarize themselves with the "input language" of the system, and prepare themselves for the articulation task. In the NLP-based interaction, users delegate the task of understanding their intents and translating it into system commands to the chatbot agent. However, users have no control over how the chatbot performs the translation. A typical chatbot is designed to understand a pre-defined set of user intents and expects to receive certain inputs from the user. However, the realm of permissible intents and required inputs are not displayed to the user at the onset of the interaction. Therefore, it is possible that original user input does not match an intent understood by the chatbot or does not contain the inputs required for task execution. In such a case, the output from the chatbot will include a request for clarification or additional inputs, thus increasing the number of interaction cycles necessary for the user to achieve a task goal.[1] Because the user does not have sufficient information to anticipate what additional inputs will be requested by the chatbot, or whether a given input will be correctly translated into the desired intent, the user is likely to perceive less control in chatbot-based interactions, as compared to menu-base interactions. Hence, we hypothesize:

Hypothesis 1. Chatbots with NLP interface will lead to a lower level of perceived autonomy than with menu-based interface.

3.2. Cognitive effort

Cognitive effort refers to the amount of attentional capacity allocated to obtain and process information in order to perform a task (Hong et al., 2004). In information processing theory, cognitive effort is related to the limited information processing capacity possessed by humans as well as the speed at which information can be processed (Kahneman, 1973). Webpage designs with many elements can impact user's visual processing, thereby increasing cognitive load (Harper et al., 2009). Chatbot users see fewer visual elements thus requiring less cognitive effort from the user. Furthermore, in order to perform the articulation task using a menu-based interface, a user is required to understand existing menus and input requirements and make decisions about the appropriate menu choices. In the case of a chatbot-based interface, the user can express his/her task in a natural language and delegate the task of parsing the text input into system commands and structured inputs to the chatbot agent. The chatbot interface is designed to provide a conversational experience, so user interaction can be more natural, like chatting with a friend rather than browsing a website. Hence, we expect that users expend less cognitive effort interacting with a chatbot, as compared with menu-based interfaces.

[1] It is likely that there is a strong relationship between the fulfillment of the psychological need of autonomy and the dialog control afforded by a particular system. The dialog control construct is actively used in HCI and consists of two components, dialog initiative, i.e. who has control over the dialog (user-directed, system, directed or mixed) and dialog flow (McTear, 2010). While the examination of the specific role played by dialog control in the fulfillment of the psychological need for autonomy is outside the scope of this study it is a potentially useful direction for future research.

Hypothesis 2. Chatbots with NLP interface will lead to a lower level of cognitive effort than with a menu-based interface.

3.3. Perceived competence

Perceived competence is defined as the extent to which a person feels confident about their ability to complete a task, and it reflects the user's perception of whether they can interact with a system effectively to perform a task and gain value outcomes (Nikou & Economides, 2017). According to SDT, competence is a basic psychological need and, when satisfied, is an essential antecedent of the motivation required to engage in, and complete, a task (Ryan & Deci, 2002). Individuals experiencing low levels of perceived competence may judge task difficulty as excessive and therefore abandon the task prior to completion. Previous IS research found that individual competence in the use of computers is related to self-efficacy significantly (Munro et al., 1997).

Importantly, perceived competence is likely to be influenced by the psychological need for autonomy. When a user finds difficulty in controlling an information system, they will be less confident in their performance due to a lack of control over the process and outcomes (Igbaria & Iivari, 1995; Marakas et al., 1998). Therefore, a user will experience diminished feelings related to their ability to successfully execute the underlying task. As perceived autonomy increases and the individual gains control over the task and associated tools, the user is reassured that he or she has both the skills and the capability to meet the challenge of completing the task, rather than the frustration and loss of confidence they would experience in its absence. We propose that higher levels of perceived autonomy are associated with greater perceived competence.

The cognitive effort involved in information system use is also likely to be a highly influential factor related to perceived competence. When a user expends less cognitive effort on a task, the task is perceived as less difficult and perceived competence for the given task is high. On the other hand, increased cognitive effort necessary to perform a task using an information system is likely to have a negative effect on a user's perceived competence related to task completion. Hence, we propose the following:

Hypothesis 3. Perceived autonomy has a positive effect on perceived competence.

Hypothesis 4. Cognitive effort has a negative effect on perceived competence.

3.4. Satisfaction

Performance satisfaction is the extent to which a person is pleased with their performance in completing a task. After a user executing a task via an information system, a user can evaluate his/her performance in various ways, such as completion time, information adequacy, error minimization, etc. In the philosophical literature, a person's responsibility for something is determined by whether he/she has control over that thing (Smith, 2008). Hence, we expect that when a user feels in control of his/her interaction with the system, he/she would perceive more responsibility for the outcomes of the interaction. On the other hand, when a user does not feel in control of the system operation, he/she is likely to shift the responsibility for the outcomes to the system. This would extend to both positive and negative outcomes. The user who does not feel in control of the process of performing the task is expected to blame the system for the errors. Similarly, when the system delivers good outcomes, he/she would not feel that such outcomes are the results of his or her work. Studies in satisfaction judgments show that when a person feels he/she is responsible for a decision, and that decision leads to positive outcomes, performance satisfaction is maximized (Oliver & Desarbo, 1988). Hence, we propose that higher perceived autonomy would lead to higher performance satisfaction. System satisfaction refers to an extent an individual feel pleased with respect to the system and the mechanics of interaction (Wixom & Todd, 2005). When a user perceived he/she has control over the system and can direct the system to execute the task as he/she wants, he/she would be more satisfied with the system. Because information systems are generally used to attain some goal, the degree to which the system's capabilities support task completion is highly related to effective system use (Burton-jones & Grange, 2012). Therefore, we propose the following hypotheses:

Hypothesis 5. Perceived autonomy has a positive effect on performance satisfaction.

Hypothesis 6. Perceived autonomy has a positive effect on system satisfaction.

Users with high perceived competence are more engaged in a task and more confident in their ability to complete the task using an information system (Zylka et al., 2015). As users are more engaged, they are expected to be more satisfied with their performance. Previous studies have shown that perceived competence has direct effects on intention or motivation to use technology systems, such as mobile learning application (Shroff & Keyes, 2017), or online learning (Chen & Jang, 2010). Perceived competence also increases perceived ease of use and perceived usefulness (Sørebø et al., 2009; Teo et al., 2009), which in turn affect user attitudes toward information system and their associated user satisfaction. We propose that a user with higher perceived competence would be more satisfied with his or her performance. Hence, we present the following hypotheses:

Hypothesis 7. Perceived competence has a positive effect on performance satisfaction.

Hypothesis 8. Perceived competence has a positive effect on system satisfaction.

Research suggests that the amount of effort a user expends to complete the task affects how they evaluate the system. In other words, user satisfaction with an information system is a cost-benefit analysis in which the user evaluates the effort required to complete the task in relation to the benefit of task completion. (Au et al., 2008). Previous research suggests that the higher the effort users exert, the less satisfied they become (Al-Maskari & Sanderson, 2010), and that cognitive load has a greater direct effect on satisfaction than performance outcomes (Hu et al., 2017). Hence, we predict that cognitive effort has a direct impact on performance satisfaction. When a user has a high cognitive effort on the task, he or she would be less likely to be satisfied with his or her performance on the task. Furthermore, such an increase in cognitive effort will also negatively impact system satisfaction, causing the user to calculate that the cost of using the system outweighs the associated benefit. Thus, we propose Hypothesis 9 and 10:

Hypothesis 9. Cognitive effort has a negative effect on performance satisfaction.

Hypothesis 10. Cognitive effort has a negative effect on system satisfaction.

Fig. 1summarizes the research hypotheses.

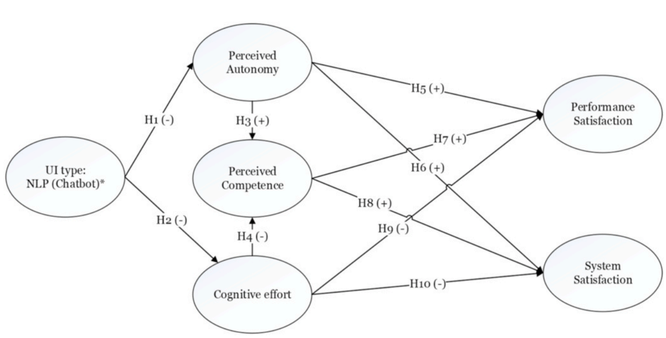


Fig. 1. Research model.
Note. *As opposed to menu-based interface.

4. Research methodology

We conducted an experimental study to test our research model. For the purpose of the study, we sought to identify a context with which subjects would be relatively familiar, and which could be amenable to creating relatively simple experimental tasks. After reviewing several chatbots on the market, we focused on chatbots deployed in the online travel industry. To attract more customers and gain market shares, travel e-commerce companies expand their services to other online platforms such as mobile websites, apps, and chatbots. Hipmunk is a consumer-oriented online travel company that offers comprehensive travel search, from flights to hotels and vacation rentals. The company launched Hello Hipmunk chatbot, an AI-powered virtual travel-planning assistant that not only helps search for flight and accommodation options but also gives advice and recommendations (Sawers, 2016). The chatbot is merged with popular messaging platforms such as Facebook, Slack, and Skype. We decided to choose Hipmunk website and chatbot for our study because (1) Hipmunk chatbot is an AI-powered bot so it meets our study goal, and (2) travel searching requires cognitive effort with specific goals in mind.

The Hipmunk website has a menu-based user interface. Users click on the menu to choose what they want to find (flights, hotels, cars, packages). The website then displays search results for users to examine. Users can also sort results and filter the results. The Hello Hipmunk chatbot starts by asking a user's current location so it can start giving travel advice. Users can ask something like "I want a non-stop flight from New York to Miami leaving 08/30 and returning 09/15" or "I'd like a hotel in San Diego on Jun 17–21". They can also set specific preferences such as "I prefer American Airlines" or "I'd like the cheapest flight". The search results will be displayed on the chat screen.

4.1. Research procedure

We collected data using an online experiment involving students from a public university as participants. The study deployed a $2 \times 2 \times 2$ factorial experimental design including 8 different treatments (system interface, task, scenario). With respect to our main variable of interest, the type of the system interface, we employed a counter-balanced design for our experiment as we were interested in studying how each participant performed the search task on both website and chatbot systems, which allowed us to examine both the within-subject and between-subject differences. The counter-balanced design enhances experimental control by entering all respondents into all treatments in a randomized order (Campbell and Stanley, 1963). The students were asked to use the Hipmunk website and chatbot to complete two information search tasks, including flight ticket and hotel room search. The participants were randomly assigned to one of the eight treatments, each containing two scenarios in the same category and two sets of user instructions (menu-based website vs chatbot). The orders of the systems, the scenarios and instructions were randomized. In category 1, participants were asked to work as an administrative coordinator for a company, and they had to book a round trip flight ticket and hotel room for their manager. We provided them departure and destination airport names, and budget amounts. Category 1 included 2 scenarios with 2 different business trip destinations and on different dates. In category 2, participants were asked to plan for their personal vacation, and they could decide their destinations and budgets. Category 2 included 2 scenarios: spring trip and summer trip. Participants were given short instructions on how to use the Hipmunk website and chatbot for travel search.

After completing each search task, participants were asked to provide their search results and fill out a questionnaire about their experience during the task, including their perceived autonomy and competence, their cognitive effort when completing the task, their performance satisfaction, process satisfaction, and system satisfaction. Thus, the participants filled out a questionnaire immediately after each system interaction: one for the chatbot and one for the menu-based system. In addition, the survey also captured control variables such as demographic information and the frequency of travel information search via travel website and chatbot.

4.2. Measurement

Performance satisfaction and system satisfaction were measured by items developed by the authors. The choice to develop measurement scales for these two constructs in the presence of multiple existing scale was dictated by our need to have a short simple scale that captured system satisfaction and performance satisfaction separately. Extant scales that we have reviewed were either too long, or appear to measure constructs other than satisfaction, or contained items that were not relevant to the context of the experiment. For example, the System Usability Scale (SUS) is both longer and contains items that measure constructs other than satisfaction, such as ease of use and self-efficacy (Bangor et al., 2009). UEQ+ is both long and appears to measure the various aspects of the system itself rather than the psychological feeling of satisfaction directed at the system (Schrepp et al., 2021). Perceived autonomy and perceived competence were measured using scales adapted from Deci and Ryan (2000) and Gagné (2003) (Gagné, 2003; Ryan & Deci, 2000b). The cognitive effort scale was adapted from Hu et al., 2017 (Hu et al., 2017). We opted for the use of the cognitive effort scale rather than other scale used to measure similar construct such as the NASA Task Load Index[2] (NASA TLI) as it focusses on the cognitive aspects of the task load, whereas NASA TLI incorporates both physical, temporal and emotional (frustration) aspects. All items were measured using a 5-point Likert scale, with 1 being strongly disagree and 5 being strongly agree. The items used in different measurement scales are provided in the Appendix.

4.3. Data collection

We recruited participants through announcements in their classes at a public university in the US. Student participants were offered extra credits for participation in the study. The invitation to participate was distributed to 597 undergraduate students. Sampling took place over the spring and summer semesters of 2018. We got a total of 419 responses, resulting in a response rate of 70.18%. We excluded 103 invalid and incomplete responses. In the end, we retained 316 useable responses. Table 1 displays demographic information.

Table 1
Demographic profile.

Variable	Value	Sample	Percentage
Gender	Male	153	48.42%
	Female	161	50.95%
	Other	2	0.63%
Age (years)	18–21	156	49.37%
	22–25	109	34.49%
	26–30	34	10.76%
	31–40	15	4.75%
	41–50	1	0.32%
	51 and above	1	0.32%
Frequency of flight search on website	Never	39	12.34%
	Rarely	104	32.91%
	Sometimes	80	25.32%
	Often	55	17.41%
	Always	38	12.03%
Frequency of flight search on chatbot	Never	242	76.58%
	Rarely	52	16.46%
	Sometimes	14	4.43%
	Often	7	2.22%
	Always	1	0.32%
Frequency of hotel search on website	Never	52	16.46%
	Rarely	93	29.43%
	Sometimes	82	25.95%
	Often	53	16.77%
	Always	36	11.39%
Frequency of hotel search on chatbot	Never	240	75.95%
	Rarely	51	16.14%
	Sometimes	17	5.38%
	Often	7	2.22%
	Always	1	0.32%

5. Results

Our hypothesized model was tested using partial least squares (PLS) path modeling (Wold, 1982), a variance-based structural equation modeling (SEM) approach, to investigate the relationship between an individual's perceived autonomy, perceived competence, cognitive effort, performance satisfaction, process satisfaction, and system satisfaction. PLS-SEM has been used frequently in IS research (Urbach & Ahlemann, 2010), and recent HCI studies (Busch et al., 2021; Canziani & MacSween, 2021; Wang et al., 2017). We chose PLS-SEM because it supports exploratory studies with smaller sample sizes and has higher statistical power when structural models are complex (Hair et al., 2014, pp. 19–20). PLS is also appropriate for our study as we studied new phenomenon with newly developed measurement model (Urbach & Ahlemann, 2010). The data were analyzed using SmartPLS 3.0 (Ringle et al., 2005), statistical software for PLS-SEM with graphical user interface. For the purpose of this study, we split our data into two sub-sets to test the effect of chatbot interface and menu-based interface (Website). In PLS, the evaluation of a model consists of two sequential stages: the validation of the measurement model, and the estimation of the structural model. These are presented below.

5.1. Measurement model

As the first step in PLS-SEM analysis, measurement model validation helps to verify the reliability and validity of constructs. The purpose of this step is to examine how well the items load on the constructs. All six constructs in our study are reflective constructs and were assessed using a variety of metrics, including composite reliability, Cronbach's alpha, convergent validity, and construct validity.

Convergent and discriminant validity of the items were assessed through the analysis of the outer loadings and the average variance extracted (AVE). In Tables 2 and 3, the AVE values range from 0.750 to 0.903 (Chatbot dataset), and 0.717 to 0.905 (Menu-based interface dataset), well above the recommended 0.50 threshold, indicating that each construct explains more than half of the variance of its indicators (Hair et al., 2014, p. 107), so convergent validity was established at the construct level. Tables 4 and 5 display PLS loadings and cross-loadings of the Chatbot and Menu-based interface datasets. In both datasets, the outer loadings of all items are greater than the accepted threshold of 0.70.

Discriminant validity refers to the extent to which a construct is different from other constructs by empirical standards (Hair et al., 2014). It can be assessed by examining the cross-loadings of indicators, where factor loadings for items associated with the construct should be greater than loadings of the same items on other constructs. The second criterion is the Fornell-Larker criterion that compares the square root of the AVE values with the correlation of the latent variable construct. In both datasets, the square roots of the AVEs were greater than the inter-construct correlations, so discriminant validity is established.

Reliability was assessed using composite reliability (CR) and Cronbach's alpha. The composite reliabilities were above the recommended 0.708 level, ranging from 0.942 to 0.974 in Chatbot dataset, and 0.938 to 0.974 in Menu-based interface dataset. Cronbach's alphas were also above the 0.70 level, ranging from 0.919 to 0.964 in Chatbot dataset, and 0.920 to 0.965 in Menu-based interface dataset. Hence, the

[2] NASA TLX page https://humansystems.arc.nasa.gov/groups/tlx/tlxapp.php.

Table 2
Construct correlations, consistency, and reliability of the Chatbot dataset (N = 316).

Construct	CA	CR	AVE	Construct				
				AU	CP	CE	PS	SS
Perceived Autonomy	0.933	0.947	0.750	**0.866**				
Perceived Competence	0.964	0.974	0.903	0.637	**0.950**			
Cognitive Effort	0.919	0.942	0.804	−0.271	−0.305	**0.897**		
Performance Satisfaction	0.945	0.960	0.858	0.643	0.811	−0.261	**0.926**	
System Satisfaction	0.963	0.973	0.900	0.726	0.564	−0.336	0.684	**0.949**

Notes.
- CA: Cronbach's Alpha, CR: Composite Reliability, AVE: Average Variance Extracted.
- Bold numbers on the diagonal are the square root of the AVE, off-diagonal elements are correlations among constructs.
- AU: Perceived Autonomy, CP: Perceived Competence, CE: Cognitive Effort, PS: Performance Satisfaction, SS: System Satisfaction.

Table 3
Construct correlations, consistency, and reliability of the Menu-based interface dataset (N = 316).

Construct	CA	CR	AVE	Construct				
				AU	CP	CE	PS	SS
Perceived Autonomy	0.920	0.938	0.717	**0.847**				
Perceived Competence	0.965	0.974	0.905	0.641	**0.951**			
Cognitive Effort	0.923	0.944	0.809	−0.225	−0.331	**0.900**		
Performance Satisfaction	0.937	0.955	0.841	0.694	0.839	−0.320	**0.917**	
System Satisfaction	0.959	0.970	0.891	0.708	0.627	−0.213	0.730	**0.944**

Notes.
- CA: Cronbach's Alpha, CR: Composite Reliability, AVE: Average Variance Extracted.
- Bold numbers on the diagonal are the square root of the AVE, off-diagonal elements are correlations among constructs.
- AU: Perceived Autonomy, CP: Perceived Competence, CE: Cognitive Effort, PS: Performance Satisfaction, SS: System Satisfaction.

Table 4
PLS loadings and cross-loadings of the Chatbot dataset (N = 316).

Construct	Item	AU	CP	CE	PS	SS
Perceived Autonomy	AUB1	**0.808**	0.569	−0.272	0.566	0.622
	AUB2	**0.840**	0.531	−0.197	0.531	0.575
	AUB3	**0.846**	0.494	−0.174	0.504	0.591
	AUB4	**0.908**	0.542	−0.234	0.579	0.671
	AUB5	**0.884**	0.546	−0.243	0.560	0.638
	AUB6	**0.906**	0.617	−0.279	0.596	0.668
Perceived Competence	CPB1	0.618	**0.928**	−0.328	0.772	0.587
	CPB2	0.601	**0.964**	−0.298	0.763	0.529
	CPB3	0.589	**0.948**	−0.237	0.752	0.504
	CPB4	0.610	**0.961**	−0.296	0.792	0.521
Cognitive Effort	CEB1	−0.229	−0.280	**0.900**	−0.269	−0.301
	CEB2	−0.243	−0.232	**0.870**	−0.200	−0.305
	CEB3	−0.267	−0.296	**0.904**	−0.222	−0.323
	CEB4	−0.233	−0.284	**0.912**	−0.241	−0.278
Performance Satisfaction	PSB1	0.626	0.724	−0.244	**0.913**	0.692
	PSB2	0.602	0.777	−0.239	**0.939**	0.637
	PSB3	0.570	0.733	−0.213	**0.913**	0.588
	PSB4	0.587	0.768	−0.269	**0.940**	0.619
System Satisfaction	SSB1	0.684	0.533	−0.324	0.635	**0.951**
	SSB2	0.692	0.516	−0.328	0.647	**0.948**
	SSB3	0.669	0.549	−0.315	0.656	**0.933**
	SSB4	0.710	0.543	−0.308	0.660	**0.963**

Table 5
PLS loadings and cross-loadings of the Menu-based interface dataset (N = 316).

Construct	Item	AU	CP	CE	PS	SS
Perceived Autonomy	AUW1	**0.763**	0.591	−0.275	0.622	0.601
	AUW2	**0.770**	0.519	−0.165	0.546	0.568
	AUW3	**0.888**	0.516	−0.157	0.541	0.603
	AUW4	**0.893**	0.532	−0.212	0.585	0.583
	AUW5	**0.853**	0.490	−0.177	0.546	0.562
	AUW6	**0.902**	0.592	−0.152	0.662	0.660
Perceived Competence	CPW1	0.592	**0.933**	−0.266	0.798	0.609
	CPW2	0.633	**0.955**	−0.332	0.803	0.584
	CPW3	0.609	**0.956**	−0.364	0.793	0.605
	CPW4	0.607	**0.962**	−0.296	0.800	0.587
Cognitive Effort	CEW1	−0.172	−0.255	**0.901**	−0.232	−0.128
	CEW2	−0.148	−0.210	**0.832**	−0.189	−0.128
	CEW3	−0.234	−0.345	**0.927**	−0.342	−0.234
	CEW4	−0.228	−0.338	**0.935**	−0.337	−0.234
Performance Satisfaction	PSW1	0.701	0.768	−0.273	**0.888**	0.727
	PSW2	0.618	0.733	−0.308	**0.928**	0.611
	PSW3	0.596	0.780	−0.295	**0.922**	0.641
	PSW4	0.627	0.796	−0.299	**0.930**	0.694
System Satisfaction	SSW1	0.685	0.579	−0.191	0.684	**0.945**
	SSW2	0.677	0.570	−0.194	0.671	**0.944**
	SSW3	0.655	0.623	−0.201	0.717	**0.934**
	SSW4	0.654	0.594	−0.218	0.684	**0.953**

measurement scale's reliability is established. In conclusion, the results indicated that the measurement for six reflective constructs was psychometrically adequate for this study.

5.2. Structural model

Hypotheses 1 and 2 were tested by running the paired *t*-procedure. The data in this study is paired data since the Chatbot dataset and Menu-based interface dataset come from the same observational unit: each individual used both Chatbot and Menu-based systems. Hence, there is a natural link between Chatbot-related observations and Menu-based-related observations. This results in dependent individual perceptions of Chatbot and Menu-based systems. The within-subject *t*-test (paired samples *t*-test) is a parametric test that compares two means in paired data.

We used SPSS to calculate the mean score of two constructs, Perceived Autonomy, and Cognitive Effort, then ran the within-subject *t*-test analysis. Table 6 presents the results of the test. Hypothesis 1 was supported (p-value < 0.001), indicating that there is a significant difference in the level of perceived autonomy between Chatbots with NLP interface and Menu-based interface. The negative difference (−0.434) between perceived autonomy of Chatbot and menu-based interface suggests that Chatbots with NLP interface lead to a lower level of perceived autonomy, as we expected.

For Hypothesis 2, the mean of the difference between Chabot and the menu-based interface was positive (0.312) and this difference was

Table 6
Results of within-subject *t*-test on H1 & H2.

No.	Hypothesis Tested	Chatbot (n = 316)		Menu (n = 316)		Mean of Difference between Chatbot and Menu-based Interfaces	*t*-value	*p*-value
		Mean	Std. Deviation	Mean	Std. Deviation			
H1	Perceived Autonomy (Chatbot vs Menu-based)	3.533	1.026	3.966	0.834	−0.434	−7.28	0.000
H2	Cognitive Effort (Chatbot vs Menu-based)	2.540	1.133	2.229	1.093	0.312	4.39	0.000

significant (p-value < 0.001). This result contradicts our hypothesis, as we expected chatbots with NLP interface to be associated with a lower level of cognitive effort, compared to websites with the menu-based interface. Our findings show that chatbots with NLP interface lead to a higher level of cognitive effort than websites.

Next, we conducted a between-subjects *t*-test to examine differences between the two systems. We divided the dataset into two datasets based on what system the participants used first. The first dataset had responses from participants who used the Chatbot interface first, and the second dataset had responses from participants who used Menu-based interface first. We compared the perceived autonomy (Chatbot) and the cognitive effort (Chatbot) in the Chatbot dataset with perceived autonomy (Menu-based) and cognitive effort (Menu-based) in the Menu-based interface, respectively. We ran the two-sample *t*-test in SPSS. The results indicate there is a significant difference in the level of perceived autonomy between the two datasets (p-value < 0.001). The negative difference (−0.349) confirmed that the Chatbot interface leads to a lower level of perceived autonomy. Similar to within-subject *t*-test results, the between-subject *t*-test also concludes that the Chatbot interface leads to a higher level of cognitive effort (p-value = 0.045). Hence, the within-subject *t*-test and between-subject *t*-test produced similar findings. Table 7 presents the results of the between-subject *t*-test.

Other hypotheses (3 to 10) were tested by running nonparametric bootstrapping procedure to test the significance of various PLS-SEM results, such as path coefficients estimates (β) and coefficients of determination (R^2). The path coefficient indicates the strength of the relationship between dependent and independent variables in a structural model, while the R^2 value refers to the amount of explained variance in the dependent variable that is due to the independent variables in the structural model. The path significance levels (*t*-values) are calculated using the bootstrapping method. Figs. 2 and 3 show the SmartPLS results for structural model testing in Chatbot dataset and Menu-based dataset, respectively[3].

In both datasets, hypotheses 3, 4, 5, 6, 7, and 8 were supported. Hypothesis 9 was not supported, indicating there is no significant relationship between Cognitive Effort and Performance Satisfaction. Hypothesis 10 was supported in the Chatbot dataset (p-value = 0.004), but not supported in the Menu-based interface dataset (p-value = 0.833). This result indicates that the effect of Cognitive Effort on System Satisfaction depends on the type of system. The results of path coefficients in the Chatbot and Menu-based interface datasets are presented in Table 8.

According to the R^2 results, the model explains 57.41% variability of system satisfaction in Chatbot dataset, and 55.73% variability in Menu-based interface dataset. Table 9 displays R-square results.

Next, we compared the structural models of Chatbot dataset and Menu-based interface dataset. For comparing path coefficients, we calculated the differences between path coefficients of Chatbot dataset and Menu-based interface dataset, then ran the paired *t*-test (Ferguson, 1971; Sarstedt & Wilczynski, 2009) using the bootstrapping results. Since hypothesis 9 is not supported, and hypothesis 10 is partially supported, we did not compare the path coefficients associated with these hypotheses. Table 10 presents the results of paired *t*-tests on path coefficients, and effect size. The results indicate that there are significant differences in path coefficients between Chatbot and Menu-based interface. Table 11 presents summary of the results.

The results also suggest that the effect of control variables, including demographic characteristics, the frequency of Chatbot and Menu-based interface travel websites use, and task type (Business vs Personal) on the dependent variables vary across the systems. Age, gender, frequency use, and task type have no effect on performance satisfaction and system satisfaction in both Chatbot and website interfaces. Education has no effect on system satisfaction in both cases and has no effect on performance satisfaction in Chatbot interface. However, education has significant effect on performance satisfaction (p-value = 0.006) in Menu-based interface case.

6. Discussion

The findings of this study contribute to our understanding of how user experience differs between Chatbots with NLP interface and menu-based interface. First, as we proposed in Hypothesis 1, chatbots with NLP interface were associated with a lower level of perceived autonomy than the menu-based interface. This finding suggests that in human-chatbot interaction, individuals perceive less control over the system since they only interact with the system through natural language and have no other control over the system. In the websites with the menu-based interface, users control the system by navigating around the websites, typing texts, clicking on different parts of the websites, such as icons, hyperlinks, menu bars, or checkboxes. Thus, in the menu-based interface scenarios, users would have more control over the system and their perception of autonomy is higher.

Contrary to our expectation, chatbots with NLP-interface were associated with a higher level of cognitive effort than with menu-based interface. There could be several possible explanations of this finding. The first explanation is related to user experience with similar systems. Over 75% of participants of the study had not used chatbots to search for travel information prior to the study. By comparison, only 12.34% of participants had no prior experience making flight reservations using a menu-based website, and only 16.46% had no experience making hotel reservations using a menu-based website. Hence, most of our participants were first time users of chatbots but they were experienced users of menu-based websites. First-time users normally do not provide enough information for chatbots to execute search tasks, so the chatbots must request additional questions. Alternatively, the results can be attributed to the fact that interacting with chatbots requires users to provide commands, understand messages from chatbots, and reply to chatbots' questions until they receive the desired results. This process requires users to be more aware of the conversation, and hence, increases their cognitive load. Forlizzi and Ford (2000) proposed an initial framework of experiences related to interaction design, the framework includes four components: sub-consciousness, cognition, narrative, and storytelling. According to the authors, sub-consciousness experiences refer to automatic or fluent experiences, while cognitive experiences refer to experiences where users have to put more thought into what they are doing (Forlizzi & Ford, 2000). There may also be a

[3] We have repeated the analysis using Hayes PROCESS V4.0. All mediation relationships depicted in Figs. 2 and 3 were confirmed at a 95% confidence level with 5000 bootstrap samples.

Table 7
Results of between-subjects *t*-test on H1 & H2.

No.	Hypothesis Tested	Chatbot (n = 152)		Menu (n = 164)		Mean of Difference between Chatbot and Menu-based Interfaces	*t*-value	*p*-value
		Mean	Std. Deviation	Mean	Std. Deviation			
H1	Perceived Autonomy (Chatbot vs Menu-based)	3.708	0.894	4.057	0.765	−0.349	−3.71	0.000
H2	Cognitive Effort (Chatbot vs Menu-based)	2.344	0.959	2.120	1.020	0.225	2.02	0.045

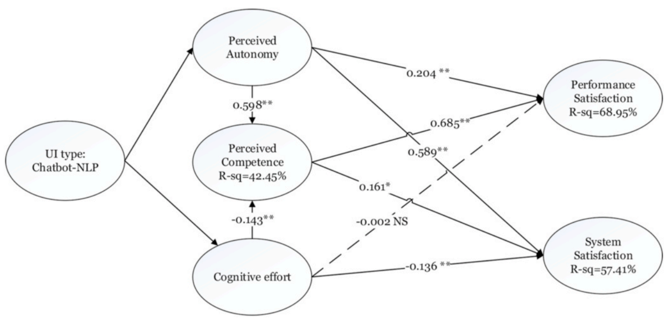


Fig. 2. PLS Model – Chatbot dataset.
Note: NS: not significant; * *p*-value ≤ 0.1; ** *p*-value ≤ 0.01.

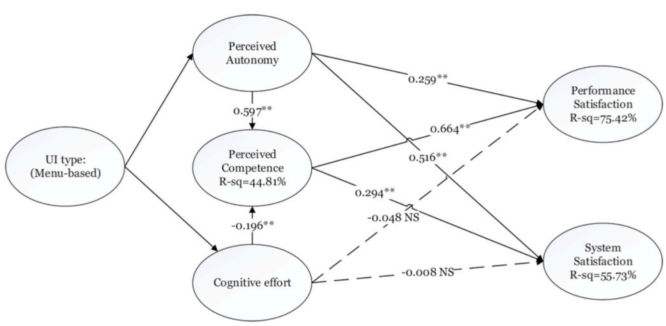


Fig. 3. PLS Model – Menu-based interface dataset.
Note:NS: not significant; * *p*-value ≤ 0.1; ** *p*-value ≤ 0.01.

compounding effect of the two explanations. When a user uses a product frequently and is familiar with the product, he or she forms a sub-consciousness experience that makes it possible to use the product without much thinking (for instance, an experienced iPhone user makes a call on iPhone). In other words, using the product has become the user's habit, a routine behavior he or she performs on a regular basis. When a certain task becomes a habit, the human brain spends less effort on the task and switches to automatic mode (Duhigg, 2012). We have cognitive experiences when we interact with new or confusing products, or when completing complex tasks that require attention (for instance, an experienced iPhone user uses an Android phone for the first time). In our study, all participants who used the Hipmunk website and Hipmunk chatbot for the first time engaged in a cognitive experience. However, they were more familiar with browsing travel websites than using travel chatbots, so they already had sub-consciousness experience with travel websites that they could build on. Hence, it took less cognitive effort for them to shift from a cognitive to a sub-conscious experience when they first used the Hipmunk website, compared to when they first used the Hipmunk chatbot. Another explanation is a menu-based website

Table 9
Results of R-Square in the Chatbot and Menu-based interface dataset.

	R-Square (Chatbot)	R-Square (Menu-based)
Competence	42.45%	44.81%
Performance Satisfaction	68.95%	75.42%
System Satisfaction	57.41%	55.73%

Table 10
Results of paired *t*-test on path coefficients and Effect size.

No.	Hypothesis	*β*-diff (Chatbot vs Menu-based Interface)	*t*-Value (Chatbot vs Menu-based Interface)	*p*-Value (Chatbot vs Menu-based Interface)	*Effect size*
H3	Autonomy - > Competence	0.003	2.037	0.042	≤0.001
H4	Cognitive Effort - > Competence	0.055	35.414	≤0.001	1.10
H5	Autonomy - > Performance Satisfaction	−0.050	−34.576	≤0.001	1.16
H6	Autonomy - > System Satisfaction	0.084	40.777	≤0.001	1.13
H7	Competence - > Performance Satisfaction	0.023	14.920	≤0.001	0.46
H8	Competence - > System Satisfaction	−0.154	−72.790	≤0.001	1.98

Table 8
Results of path coefficients in the Chatbot and Menu-based interface datasets.

No.	Hypothesis	Chatbot			Menu-based Interface		
		β (Chatbot)	t-value (Chatbot)	p-Value (Chatbot)	β (Menu-based)	t-value (Menu-based)	p-Value (Menu-based)
H3	Autonomy - > Competence	0.598	12.970	≤0.001	0.597	11.421	≤0.001
H4	Cognitive Effort - > Competence	−0.143	3.035	0.002	−0.196	3.949	≤0.001
H5	Autonomy - > Performance Satisfaction	0.204	3.990	≤0.001	0.259	5.614	≤0.001
H6	Autonomy - > System Satisfaction	0.589	9.500	≤0.001	0.516	7.209	≤0.001
H7	Competence - > Performance Satisfaction	0.685	12.739	≤0.001	0.664	13.875	≤0.001
H8	Competence - > System Satisfaction	0.161	2.480	0.013	0.294	4.073	≤0.001
H9	Cognitive Effort - > Performance Satisfaction	−0.002	0.048	**0.962**	−0.048	1.594	**0.111**
H10	Cognitive Effort - > System Satisfaction	−0.136	2.868	**0.004**	−0.008	0.211	**0.833**

Table 11
Summary of the results.

No.	Hypothesized Path	Results
H1	Chatbots lead to a lower level of perceived autonomy than menu-based interfaces.	Supported
H2	Chatbots lead to a lower level of cognitive effort than menu-based interfaces.	Contradict proposed hypothesis
H3	Perceived autonomy has a positive effect on perceived competence.	Supported
H4	Cognitive effort has a negative effect on perceived competence.	Supported
H5	Perceived autonomy has a positive effect on performance satisfaction.	Supported
H6	Perceived autonomy has a positive effect on system satisfaction.	Supported
H7	Perceived competence has a positive effect on performance satisfaction.	Supported
H8	Perceived competence has a positive effect on system satisfaction.	Supported
H9	Cognitive effort has a negative effect on performance satisfaction.	Not Supported
H10	Cognitive effort has a negative effect on system satisfaction.	Partial Supported

provides visual cues, including menu options and labels of text input controls that focus and structure the attention of the user, thus helping them to learn how to complete the task using the system. A typical chatbot interface lacks such visual cues, resulting in a potentially less focused and less structured approach to the task and, consequently, in more cognitive effort on the part of the user.

The results suggest that higher perceived autonomy leads to higher perceived competence. When users feel they are in control over the systems and their tasks, they are more likely to think that they have the necessary abilities, knowledge, and skills to execute the task effectively. Interestingly, we found no significant relationship between cognitive effort and competence, indicating that individuals' mental activities have no influence on how they perceive their abilities to complete the job effectively. The results also show a significant positive effect of autonomy on both performance satisfaction and system satisfaction. When individuals feel in control of the systems and the task, they are more pleased with their own performance as well as with the system they used to complete the task. On the other hand, when users perceive that their autonomy is low when using the system, they may find their performance less satisfactory and feel frustrated with the system. These findings are in line with the fundamental tenets of the SDT, which identifies the need for autonomy as one of the key drivers of intrinsic motivation (Ryan & Deci, 2000b).

Our findings also show that perceived competence has a positive impact on performance satisfaction and system satisfaction. When individuals believe in their ability to complete a task successfully, they are more likely to be satisfied with their own performance and with the system they use. When individuals feel they lack such ability, they feel uncertain about their performance, which negatively affects their performance satisfaction. As a result, lower level of perceived competence leads to less satisfaction in performance and system. This is also consistent with the SDT, which identifies the need for competency as other key drivers of intrinsic motivation (Ryan & Deci, 2000b).

Surprisingly, our hypothesis regarding the negative effect of cognitive effort on performance satisfaction was not supported. A potential explanation is that the hypothesized negative relationship between cognitive effort and performance satisfaction may be canceled out by the psychological need to justify higher cognitive effort by inflating the perception of performance. Specifically, individuals who expend significant cognitive effort may experience cognitive dissonance if they believe that their performance was sub-optimal. Because they cannot retroactively change the amount of cognitive effort they have committed to the task, they can alter their subjective assessment of their own performance to more satisfactory, thus avoiding cognitive dissonance (Festinger, 1957).

Furthermore, our findings suggest that the type of interface (menu-based vs. chatbot) moderates the relationship between cognitive effort and system satisfaction. Specifically, cognitive effort has a significant negative effect on system satisfaction for individuals using a chatbot interface. However, the relationship between cognitive effort and system satisfaction was found insignificant for the menu-based interface. Because menu-based interface users reported lower cognitive effort compared to chatbot users on average, it is possible that there was not enough variability in cognitive effort within the menu-based group, which made it impossible to detect a significant relationship between cognitive effort and system satisfaction. A more theoretical explanation would suggest the individuals make different attributions for the need to expend additional cognitive effort depending on the user interface type. Specifically, they attribute additional cognitive effort to the system in the case of the chatbot, and to other factors, perhaps task difficulty, in the case of a more familiar menu-based interface. This difference in attributions can be further explained either by the higher autonomy of the chatbot interfaces, or by its novelty and, therefore, salience. The viability of these explanations needs to be explored in future research.

Three demographic characteristics (age, gender, and education) were found to have no effect on performance satisfaction and system satisfaction. The type of task (business or personal task) has no effect on performance satisfaction. However, the effect of the task type on system satisfaction was moderated by the interface type. While task type had no effect on system satisfaction in the chatbot interface, it was found to be a significant predictor of system satisfaction in the menu-based interface. When using a menu-based interface, completing a business task was associated with higher system satisfaction than a personal task. This finding may be explained by the level of order and complexity of the menu-based interface. Research suggest that web interfaces characterized by low complexity and high order are associated with more positive affective responses when users are in a goal-oriented mode (telic state) (Deng & Poole, 2010). In our experiment, the business task was had a more well-defined and explicit goal and thus was expected to elicit a more telic state in the study participants compared to the personal task. Our study participants presented with the business task were more satisfied with the menu-based interface, which they perceived as less complex. Hence, our finding of the interaction effect between the task type and interface type is consistent with the prior research on user interfaces.

6.1. Theoretical contributions

This empirical study makes several contributions to IS research. First, the study builds on the SDT (Ryan & Deci, 2000b) and develops a research model for explaining and predicting the effects of incorporating AI-based chatbot capabilities into a user interface on user outcomes, including performance and system satisfaction. The model posits that the effect of AI-based chatbot capabilities, such as NLP, on user outcomes need to be viewed through the lens of the user's need to satisfy their need for autonomy and competence. An NLP-based interface that would negatively influence user perceptions of their own autonomy and competence is likely to result in less favorable user outcomes. The results confirm that perceived autonomy and perceived competence are two essential predictors of system and performance satisfaction. When the needs for autonomy and competence are met, users work effectively and are more satisfied with their performance and the system they use. When users feel they have more autonomy over the way they perform a task, they tend to experience the task as more interesting and work more effectively, which results in performance satisfaction and system satisfaction. While both basic psychological needs (perceived autonomy and perceived competence) are important, we found that perceived competence was a stronger predictor of performance satisfaction, and perceived autonomy was a stronger predictor of system satisfaction. An explanation of this finding may be that competence is more associated with users' ability to execute the task, thus it plays a more important role in evaluating their performance. Meanwhile, autonomy is more related to system design, so it has a stronger effect on system satisfaction.

Second, this study contributes to SDT literature by illustrating the effect of perceived autonomy on perceived competence, the two constructs that are treated as independent in SDT literature. Our analysis shows that perceived autonomy positively influences perceived competence, at least in the context of relatively straightforward, IT-supported tasks such as those used in this study. When users believe they have control over the system and can complete the task the way they want, they gain higher confidence in their ability to do the task effectively. We also find that the effect of perceived autonomy on perceived competence is significantly higher in the chatbot interface compared to the menu-based interface. While our research model and hypotheses were tested in the chatbot and menu-based interfaces, we believe that the results can be generalized to other IS applications that include information searching. However, the boundary conditions for such finding need to be carefully explored. For example, it could be that in highly unstructured and difficult tasks, a higher level of autonomy may lead to lower perceived confidence. Third, the role of cognitive effort was identified as important toward perceived competence. The cognitive effort has a negative impact on perceived competence, meaning users would feel less confident in finishing the tasks successfully when the mental effort required is high. Thus, this research adds value to both SDT literature and cognitive load literature.

Finally, the findings of this study raise the important question of what AI-enabled systems should be compared to. The traditional approach is to compare the performance of AI-enabled systems against human-level performance. However, in situations where AI-enabled systems are considered as alternatives to traditional rule-based systems, a three-way comparison may be in order. Just because an AI-enabled interface may achieve near human-level performance on speech recognition or question answering tasks, it may not always be the preferred way of interacting with a business, if compared to a well-designed menu-based interface.

6.2. Practical implications

Our findings have important implications for the development and deployment of NLP-based interfaces such as chatbots, and also other interfaces in general. Perceived autonomy and perceived competence are critical factors that influence user satisfaction with their performance and the system. To ensure that user needs are satisfied, the interface design must allow users to feel they have control over the system and they are capable of using the system. In the chatbot interface, a user shifts part of the task to the chatbot. Thus, the user perceives less control over the task is performed, which leads to lower perceived autonomy, and lower level of satisfaction with their performance and system, compared with using a menu-based interface. Chatbot developers should try to design dialog sequences that make users feel they are still in control of the interaction. One option is to have a menu element into a chatbot interface, combining the menu-driven GUI design with the conversational user interface (CUI), to a hybrid design (Pathak 2017, p. 55). In the hybrid GUI + CUI design, instead of designing a chatbot that only responds to user typed commands, designers can add a button-based menu as multiple-choice options that the user can choose. For example, when a user wants to book a flight, instead of asking "what airlines do you prefer?" he or she can display airlines options as buttons, and the user only needs to click or tap on buttons. This approach can help improve the decision-making process as it is easy to navigate and the user can make response quicker. However, if there are too many buttons, users may think that they do not need to type requests, which contradicts the idea of a chatbot. Therefore, chatbot design needs to balance between directly asking the chatbot a question and clicking a button-based menu. Moreover, the results indicate that perceived autonomy positively influences perceived competence, which also leads to a positive impact on performance satisfaction and system satisfaction. Hence, it is important to design interfaces that help increase perceived autonomy.

The findings also show that NLP-base interfaces require more cognitive effort, compared to menu-based interfaces, which has a negative on perceived competence, performance, and system performance satisfaction. The role of cognitive effort is well acknowledged in HCI research but is less studied in the context of AI interfaces. It is assumed that AI-based interfaces help reduce cognitive effort by making interaction with a system more similar to interaction with another human. But are interactions with other humans less cognitively demanding as compared to interactions with other systems? Extensive research in communications and social cognition suggest that cognitive processes involved in human interaction with other animate objects are more complex than those involved in interaction with inanimate objects (Fiske & Taylor, 2013). Therefore, AI-enabled interfaces may make HCI more effort-intensive, and this needs to be taken into account when making decisions about developing and deploying AI-based interfaces. On the other hand, the high cognitive effort associated with chatbot that was observed in this study may simply be due to the lower familiarity of the users with the chatbot. Therefore, one of the possible solutions to overcoming the challenges of cognitive effort might be to develop training modules that new users can follow to learn how to use the chatbots. For instance, after the first welcome message, a chatbot can quickly explain to new users what it can do and give examples of commands. It would help set guidelines and expectations, so users can use the chatbot easily and have pleasant experiences.

6.3. Limitations and future research

The results of the study need to be interpreted in light of the following limitations. First, the study does not examine the effect of relatedness (the third basic psychological need according to SDT (Ryan & Deci, 2000b). Some companies design chatbots that have attitude, tone, and style that reflects personality, which can lead to a natural conversation between users and chatbots. Creating a chatbot with personality would enable an organization to drive customer engagement and reinforce its brand image (Nash & Castellanos, 2018). Unlike the menu-based interface, a chatbot may form a close relationship with its users and would satisfy the users' need to interact with other people. Thus, future research should look into the role of relatedness in the chatbot context.

Another limitation is related to the fact that the study used only one GUI and one chatbot interface, and thus the results of the study may not be fully representative of all GUI and chatbot interfaces respectively. Our choice to go with only one GUI and chatbot interface was dictated to maintain experimental control and to control the complexity of the experimental design. However, future research needs to focus on how different features of each of the interface methods influence user experiences and outcomes. For example, a more granular examination of interface characteristics such as depth, breadth, and congruence of menu-based systems, as well as the emotional character and NLP method of the chatbot interface may yield additional insights.

Additionally, differences in prior experience of the participants with the different interface types have likely influenced the results of the experiment. Most of our participants never used chatbots to search for travel information before, while the majority of them were familiar with travel websites and with GUI interface in general. Our *post hoc* analysis suggests that experience with the chatbot may influences user perceptions and experiences, but future research should conduct longitudinal studies to examine the effect of user experience on user perceptions, experiences, and satisfaction with chatbots over a longer period.

Another limitation is related to the demographics of the study participants. This study recruited university students as respondents, with 49.37% of our sample about 18–21 years of age. This might have influenced our results since different generations would have different travel habits. For example, the millennial generation (between the ages of 16 and 34 at the time of this writing) is more likely to plan for international trips compared to the Boomers generation (Gelfeld, 2017). Young travelers are also more familiar with travel booking through online travel agencies, such as Expedia or Priceline (Barton et al., 2013). Future studies could expand to other demographic groups and increase the sample size by utilizing Amazon's Mechanical Turk. Furthermore, we suggest conducting the experiment on different tasks, such as insurance purchasing or financial advising, to investigate the differences across industries and task types.

From the methodological perspective, the reliance on self-reported measures as dependent variables is a significant limitation. The need to rely on self-reported measures was dictated by our decision to use commercially available chatbot and UI applications in order to maintain the element of contextual realism in our experiment. However, future researchers could employ system-captured measures of user behavior and outcomes such as the task completion time, the task completion rate, and leverage system log files to gain additional insight in user interactions with chatbots vis-a-vis menu-based systems.

7. Conclusions

The goal of this study was to examine the effect of chatbots on individual motivation factors in the context of task performance. Rooted in the SDT theory and HCI literature, examined the differential effect of chatbot (vs. menu-based) interactions on perceived autonomy, perceived competence, and cognitive effort, and on the effect these factors have on each other and on user satisfaction with the task performance and the system itself. The results point to a direct negative effect of chatbot interaction on perceived autonomy and cognitive effort, leading to lower perceived confidence and lower satisfaction with the systems. Interestingly, in cases of chatbot interactions, we found a positive relationship between cognitive effort and performance satisfaction. The findings paint a complex picture of how AI-enabled systems fits within human work processes, and human motivation. The findings also suggest that new advanced technology is not always the right solution to organizational problems, and may indeed result in unintended and even negative consequences, especially if user issues are not adequately addressed. Ultimately, it is AI will likely permeate all organizational processes, but it needs to be designed in such a way as to augment human activity, and to ensure that the basic motivational needs of human actors continue to be satisfied.

Credit author statement

Quynh N. Nguyen: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft. Anna Sidorova: Conceptualization, Methodology, Writing – review & editing. Russell Torres: Writing – review & editing, Formal analysis.

Declarations of competing interest

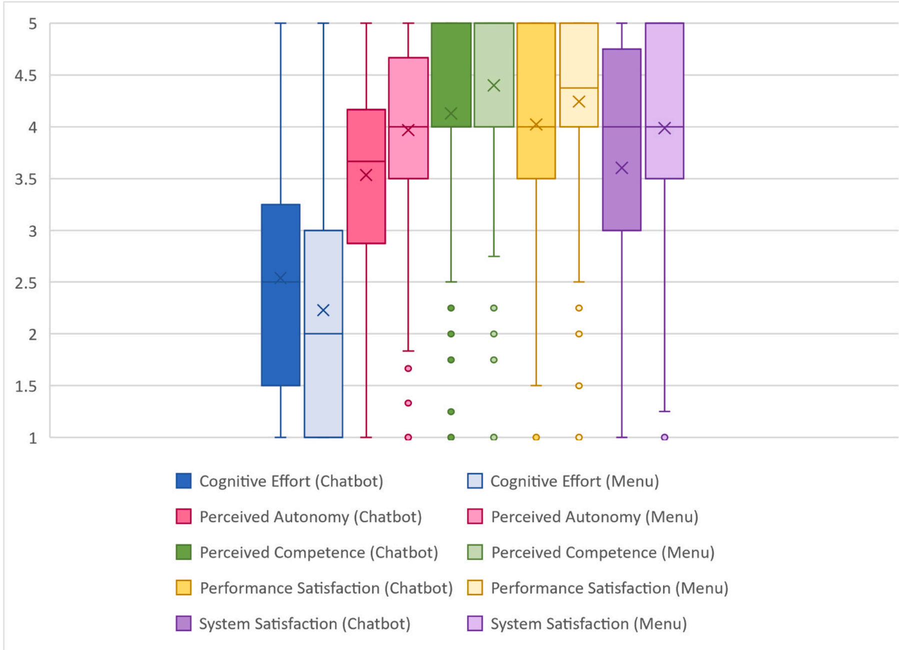
None.

Appendix

Appendix 1
Survey questions

Construct	Item Code	Items	Sources
Perceived autonomy	AU1	I felt like I could decide how to complete search task.	Adapted from Deci and Ryan (2000) and Gagné (2003)
	AU2	I felt like I could pretty much be myself when completing search task.	
	AU3	When doing search task, I had an opportunity to decide for myself how to go about my work.	
	AU4	I felt like I had flexibility to decide how to complete search task.	
	AU5	I could control how search task was done.	
	AU6	I felt like I could manage my own work while completing search task.	
Perceived competence	CP1	I feel confident in my ability to complete search task.	Adapted from Deci and Ryan (2000) and Gagné (2003)
	CP2	I am capable of completing search task.	
	CP3	I am able to complete search task.	
	CP4	I feel able to meet the challenge of completing search task.	
Cognitive effort	CE1	I needed a lot of thinking when completing the search task using the website/chatbot	Adapted from Hu et al., 2017
	CE2	I often contemplated when completing the search task using the website/chatbot.	
	CE3	Generally speaking, completing search task using the website/chatbot was cognitively demanding.	
	CE4	When completing the search task using the website/chatbot, I invested high mental effort.	
Performance satisfaction	PS1	I am satisfied with my performance in completing search task.	Developed by authors
	PS2	I think I did well in completing search task.	
	PS3	I completed search task correctly.	
	PS4	I believe that I performed well in completing search task.	
System satisfaction	SS1	All things considered, I am very satisfied with the Hipmunk website/chatbot system.	Developed by authors
	SS2	Overall, my interaction with the Hipmunk website system/chatbot is very satisfying.	
	SS3	The Hipmunk website/chatbot system performed in a satisfactory manner.	
	SS4	I am satisfied with the Hipmunk website/chatbot system.	

Appendix 2. Boxplots of Measurement of Constructs



References

Al-Maskari, A., & Sanderson, M. (2010). A review of factors influencing user satisfaction in information retrieval. *Journal of the Association for Information Science and Technology, 61*(5), 859–868. https://doi.org/10.1002/asi.21300

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior, 85*, 183–189. https://doi.org/10.1016/J.CHB.2018.03.051

Ariely, D. (2000). Controlling the information flow: Effects on consumers' decision making and preferences. *Journal of Consumer Research, 27*(2), 233–248. https://doi.org/10.1086/314322

Au, N., T Ngai, E. W., E Cheng, T. C., Kong, H., Croteau, A.-M., & DeLone, W. (2008). Extending the understanding of end user information systems satisfaction formation: An equitable needs fulfillment model approach. *MIS Quarterly, 32*(1), 43–66. http://www.jstor.org/stable/25148828.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies, 4*(3), 114–123.

Barton, C., Haywood, J., Jhunjhunwala, P., & Bhatia, V. (2013). *Traveling with millennials*. http://image-src.bcg.com/Images/BCG Traveling with Millennials Mar 2013_tcm9-96752.pdf.

Burton-jones, A., & Grange, C. (2012). From use to effective use: A representation theory perspective. *Information Systems Research, 24*(3), 632–658. https://doi.org/10.1287/isre.1120.0444

Busch, P. A., Hausvik, G. I., Ropstad, O. K., & Pettersen, D. (2021). Smartphone usage among older adults. *Computers in Human Behavior, 121*, 106783. https://doi.org/10.1016/J.CHB.2021.106783

Canziani, B., & MacSween, S. (2021). Consumer acceptance of voice-activated smart home devices for product information seeking and online ordering. *Computers in Human Behavior, 119*, 106714. https://doi.org/10.1016/J.CHB.2021.106714

Chattaraman, V., Kwon, W. S., & Gilbert, J. E. (2012). Virtual agents in retail web sites: Benefits of simulated social interaction for older users. *Computers in Human Behavior, 28*(6), 2055–2066. https://doi.org/10.1016/j.chb.2012.06.009

Chen, K. C., & Jang, S. J. (2010). Motivation in online learning: Testing a model of self-determination theory. *Computers in Human Behavior, 26*(4), 741–752. https://doi.org/10.1016/j.chb.2010.01.011

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems, 92*, 539–548. https://doi.org/10.1016/j.future.2018.01.055

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Deng, L., & Poole, M. S. (2010). Affect in web interfaces: A study of the impacts of web page visual complexity and order. *MIS Quarterly*, 711–730. https://doi.org/10.2307/25750702

Duhigg, C. (2012). The power of habit: Why we do what we do in life and business. *Random House, 34*(Issue 10).

Fallows, D. (2005). *How women and men use the Internet*. Pew Research Center. http://www.pewinternet.org/2005/12/28/part-5-functions-of-the-internet-how-men-and-women-use-it-as-a-tool-to-communicate-transact-get-information-and-entertain-themselves/.

Ferguson, G. A. (1971). *Statistical analysis in psychology and education*. McGraw-Hill.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Fiske, S. T., & Taylor, S. E. (2013). *Social cognition : From brains to culture* (2nd ed.). SAGE Publications.

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *Interactions*. https://doi.org/10.1145/3085558

Forlizzi, J., & Ford, S. (2000). The building blocks of experience: An early framework for interaction designers. In *Proceedings of the 3rd conference on designing designing interactive systems: Processes, practices, methods, and techniques*. https://doi.org/10.1145/347642.347800

Gagné, M. (2003). The role of autonomy support and autonomy orientation in prosocial behavior engagement. *Motivation and Emotion, 27*(3), 199–223. https://doi.org/10.1023/A:1025007614869

Gelfeld, V. (2017). *Travel research: 2018 travel trends*. https://doi.org/10.26419/res.00179.001

Grand View Research. (2017). Chatbot market size and share analysis | industry report, 2014 - 2025. In *Grand View research*. https://www.grandviewresearch.com/industry-analysis/chatbot-market?utm_source=Paid_PR&utm_medium=Referral&utm_campaign=PRNewswire_19sep&utm_term=RD&utm_content=RD.

Grand View Research. (2021). *Chatbot market size*. Share & Growth Report. https://www.grandviewresearch.com/industry-analysis/chatbot-market.

Hair, J. F. J., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2014). A primer on partial least squares structural equation modeling (PLS-SEM). In *Long range planning* (Vol. 46). https://doi.org/10.1016/j.lrp.2013.01.002. Issues 1–2.

Harper, S., Michailidou, E., & Stevens, R. (2009). Toward a definition of visual complexity as an implicit measure of cognitive load. *ACM Trans . Appl . Percpt, 6*(10). https://doi.org/10.1145/1498700.1498704

Hill, J., Randolph Ford, W., & 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. https://doi.org/10.1016/j.chb.2015.02.026

Hong, W., Thong, J. Y. L., & Tam, K. Y. (2004). The effects of information format and shopping task on consumers' online shopping behavior: A cognitive fit perspective. *Journal of Management Information Systems, 21*(3), 149–184. https://doi.org/10.1080/07421222.2004.11045812

Hu, P. J.-H., Hu, H., & Fang, X. (2017). Examining the mediating roles of cognitive load and performance outcomes in user satisfaction with a website: A field quasi-experiment. *MIS Quarterly, 41*(3), 975–987. https://www.misq.org/skin/frontend/default/misq/pdf/appendices/2017/V41I3Appendices/12604_RN_HuHuFangAppendices.pdf.

Igbaria, M., & Iivari, J. (1995). The effects of self-efficacy on computer usage. *Omega*. https://doi.org/10.1016/0305-0483(95)00035-6

Insider Intelligence. (2021). *Chatbot market 2021: Stats, trends, size & ecosystem research*. https://www.businessinsider.com/chatbot-market-stats-trends.

Jang, M., Jung, Y., & Kim, S. (2021). Investigating managers' understanding of chatbots in the Korean financial industry. *Computers in Human Behavior, 120*, 106747. https://doi.org/10.1016/j.chb.2021.106747

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Lee, Y., Lee, J., & Hwang, Y. (2015). Relating motivation to information and communication technology acceptance: Self-determination theory perspective. *Computers in Human Behavior, 51*(PA), 418–428. https://doi.org/10.1016/j.chb.2015.05.021

Marakas, G. M., Yi, M. Y., & Johnson, R. D. (1998). The multilevel and multifaceted character of computer self-efficacy: Toward clarification of the construct and an integrative framework for research. *Information Systems Research, 9*(2), 126–163. https://doi.org/10.1287/isre.9.2.126

McTear, M. (2010). *The role of spoken dialogue in user–environment interaction* (pp. 225–254). Human-Centric Interfaces for Ambient Intelligence.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine*. https://doi.org/10.1109/MRA.2012.2192811

Mou, Y., & Xu, K. (2017). *The media inequality: Comparing the initial human-human and human-AI social interactions*. https://doi.org/10.1016/j.chb.2017.02.067

Munro, M. C., Huff, S. L., Marcolin, B. L., & Compeau, D. R. (1997). *Understanding and measuring user competence*. Information and Management. https://doi.org/10.1016/S0378-7206(97)00035-9

Nash, K. S., & Castellanos, S. (2018). *Businesses get Into the 'flo' with chatbots - CIO journal. - WSJ. Wall Street Journal*. https://blogs.wsj.com/cio/2018/05/23/businesses-get-into-the-flo-with-chatbots/.

Nikou, S. A., & Economides, A. A. (2017). Mobile-based assessment: Integrating acceptance and motivational factors into a combined model of self-determination theory and technology acceptance. *Computers in Human Behavior, 68*, 83–95. https://doi.org/10.1016/j.chb.2016.11.020

Oliver, R. L., & Desarbo, W. S. (1988). Response determinants in satisfaction judgments. *Source Journal of Consumer Research, 14120178*(4), 495–507. http://www.jstor.org/stable/2489156.

Ringle, C. M., Wende, S., & Will, A. (2005). *SmartPLS 3.0*. Http://Www.Smartpls.De.

Roca, J. C., & Gagné, M. (2008). Understanding e-learning continuance intention in the workplace: A self-determination theory perspective. *Computers in Human Behavior, 24*(4), 1585–1604. https://doi.org/10.1016/j.chb.2007.06.001

Ryan, R. M., & Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54–67. https://doi.org/10.1006/ceps.1999.1020

Ryan, R. M., & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, R. M., & Deci, E. L. (2002). Overview of self-determination theory: An organismic-dialectical perspective. In *Handbook of self-determination research* (pp. 3–33).

Sarstedt, M., & Wilczynski, P. (2009). More for less? A comparison of single-item and multi-item measures. *Betriebswirtschaft, 69*(2), 211–227.

Sawers, P. (2016). *Hipmunk embraces bots with an A.I. travel assistant for Facebook Messenger and Slack | VentureBeat. Venturebeat*. https://venturebeat.com/2016/06/23/hipmunk-embraces-bots-with-an-a-i-travel-assistant-for-facebook-messenger-and-slack/.

Schrepp, M., Sandkühler, H., & Thomaschewski, J. (2021). How to create short forms of UEQ+ based questionnaires? In C. Wienrich, P. Wintersberger, & B. Weyers (Eds.), *Mensch und Computer 2021 - workshopband. Gesellschaft für Informatik e.V*. https://doi.org/10.18420/muc2021-mci-ws01-230

Semeraro, G., Andersen, V., Andersen, H. H. K., de Gemmis, M., & Lops, P. (2008). User profiling and virtual agents: A case study on e-commerce services. *Universal Access in the Information Society, 7*(3), 179–194. https://doi.org/10.1007/s10209-008-0116-1

Sheehan, B. T. (2018). *Customer service chatbots: Anthropomorphism, adoption and word of mouth*. Queensland University of Technology https://eprints.qut.edu.au/121188/1/Benjamin_Sheehan_Thesis.pdf.

Shroff, R. H., & Keyes, C. J. (2017). A proposed framework to understand the intrinsic motivation factors on university students' behavioral intention to use a mobile application for learning. *Journal of Information Technology Education: Research, 16*(16), 143–168. http://www.informingscience.org/Publications/3694.

Shumanov, M., & Johnson, L. (2021). Making conversations with chatbots more personalized. *Computers in Human Behavior, 117*, 106627. https://doi.org/10.1016/j.chb.2020.106627

Sidorova, A., Evangelopoulos, N., Torres, R. R., & Johnson, V. (2013). *A survey of core research in information systems*. Springer US.

Smith, A. M. (2008). Control, responsibility, and moral assessment. *Philosophical Studies, 138*(3), 367–392. https://doi.org/10.1007/s11098-006-9048-x

Sørebø, Ø., Halvari, H., Gulli, V. F., & Kristiansen, R. (2009). The role of self-determination theory in explaining teachers' motivation to continue to use e-learning technology. *Computers & Education, 53*(4), 1177–1187. https://doi.org/10.1016/j.compedu.2009.06.001

Teo, T., Lee, C. B., Chai, C. S., & Choy, D. (2009). Modelling pre-service teachers' perceived usefulness of an ICT-based student-centred learning (SCL) curriculum: A Singapore study. *Asia Pacific Education Review, 10*(4), 535–545. https://doi.org/10.1007/s12564-009-9051-y

Torres, R., & Sidorova, A. (2015). The effect of business process configurations on user motivation. *Business Process Management Journal, 21*(3), 541–563. https://doi.org/10.1108/BPMJ-09-2013-0131

Urbach, N., & Ahlemann, F. (2010). Structural equation modeling in information systems research using Partial Least Squares. *Journal of Information Technology Theory and Application, 11*(2), 5–40. https://www.researchgate.net/publication/228467554.

Wang, X., Wallace, M. P., & Wang, Q. (2017). Rewarded and unrewarded competition in a CSCL environment: A coopetition design with a social cognitive perspective using PLS-SEM analyses. *Computers in Human Behavior, 72*, 140–151. https://doi.org/10.1016/J.CHB.2017.02.045

Wixom, B. H., & Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research, 16*(1), 85–102. https://doi.org/10.1287/isre.1050.0042

Wold, H. (1982). Soft modelling: The basic design and some extensions. *Systems under Indirect Observation: Causality-Structure-Prediction, 1–54*.

Youssef, A. B., Chollet, M., Jones, H., Sabouret, N., Pelachaud, C., & Ochs, M. (2015). Towards a socially adaptive virtual agent. *International Conference on Intelligent Virtual Agents*, 3–16.

Zylka, J., Christoph, G., Kroehne, U., Hartig, J., & Goldhammer, F. (2015). Moving beyond cognitive elements of ICT literacy: First evidence on the structure of ICT engagement. *Computers in Human Behavior, 53*, 149–160. https://doi.org/10.1016/j.chb.2015.07.008