Computers in Human Behavior Reports 15 (2024) 100449

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Comparing and modeling the use of online recommender systems

Emma Engström [a,b,*], Irina Vartanova [a], Jennifer Viberg Johansson [c], Minna Persson [a], Pontus Strimling [a,d]

[a] Institute for Futures Studies, Stockholm, Sweden
[b] Department of Urban Planning and Environment, KTH Royal Institute of Technology, Stockholm, Sweden
[c] Centre for Research Ethics & Bioethics, Uppsala University, Uppsala, Sweden
[d] Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

ARTICLE INFO

Keywords:
Diffusion of innovations theory (DOI)
The unified theory of acceptance and use of technology 2 (UTAUT2)
Recommender systems
Adoption
Predictive modeling

ABSTRACT

This study explores a new way to model the adoption of AI, specifically online recommender systems. It aims to find factors that can explain the variation in usage in terms of differences between individuals and differences over technologies. We analyzed survey data from users of online platforms in the U.S. using a two-level structural equation model (SEM) ($N = 1007$). In this model, the dependent variable was the usage rate, which was defined as the share of time a person used a particular recommender system (e.g., "People You May Know") when they use the platform (e.g., Facebook). The individual responses (within-systems level) were clustered in the 26 recommender systems (between-systems level). We hypothesized that three technology-specific factors, adapted from the Diffusion of Innovations (DOI) theory and the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2), could explain the variations in usage at both levels: perceived performance expectancy (PE), perceived effort expectancy (EE), and perceived hedonic motivation (HM). Our estimated model showed that usage was associated with PE and HM at the within-system level and only with PE at the between-system level. A considerable part of the variation in usage across the 26 systems could be explained by PE only ($R^2 = 0.30$). The most important contribution to practitioners is that this study provides evidence for the idea that there are inherent, measurable differences across recommender technologies that affect their usage rates, and specifically it finds usefulness to be a key factor. This is potentially valuable for app developers and marketeers who look to promote the adoption of novel recommender systems. The main contribution to the literature is that it presents a proof-of-concept of a two-level model for AI adoption, conceptualizing it as an effect of both variations over users and variations over applications. This finding is potentially valuable for policymakers, as better predictive models might enable improved assessments of AI's social implications. In future studies, the two-level approach presented here could be applied to other forms of AI, such as voice assistants, chatbots, or Internet of Things (IoT).

1. Introduction

While recommender systems have become integral to the Internet experience, the extent to which the recommendations are actually used varies surprisingly widely across platforms. For example, suggestions influenced 80% of hours streamed on Netflix, according to Gomez-Uribe and Hunt (2016), while personalized playlists at Spotify accounted for only 17% of monthly content hours in 2017 (Spotify Technology, 2018). The reason for this variation has not yet been clarified, and this is explored in the current study.

1.1. Online recommenders

A recommender system is here defined broadly as: "any system that produces individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options" (Burke (2002); p. 331). It offers suggestions for products, location-based services, personalized fashion, as well as news, book, video, and music tracks, for example. It is also considered to be a central technology in the Internet of Things (IoT), "an emerging paradigm that envisions a networked infrastructure enabling different types of devices to be interconnected" (Felfernig et al., 2019, p.

* Corresponding author. Institute for Futures Studies, Stockholm, Sweden.
E-mail address: emma.engstrom@iffs.se (E. Engström).

https://doi.org/10.1016/j.chbr.2024.100449
Received 5 April 2024; Received in revised form 31 May 2024; Accepted 28 June 2024
Available online 6 July 2024
2451-9588/

285).

Recommender systems that offer personalized content make up one of the first types of Artificial Intelligence (AI) to reach the public. (AI here refers to software with abilities that are generally associated with intelligent beings, such as learning, problem solving, and perception (Copeland, 2022).) Already in 2019, AI in the form of deep learning technology was used for recommendations or ranking of content in the most popular platforms (Engström & Strimling, 2020). Today, AI is an inherent part of most recommender technologies and it is expected to enhance them even more in the future. Zhang et al. (2021) outlined the potential deployment of different types of AI to improve predictions in such systems, for instance by reinforcement learning, natural language processing, and computer vision. AI suits recommender systems because this technology tends to function well for predictions in contexts where the error tolerance is high and where there is an abundance of data.

In early research, Burke (2002) described three categories of recommenders: collaborative filters, which identify commonalities across users and generate suggestions by comparing their different ratings; content-based recommendations, in which different alternatives are linked by their associated systems; and utility-based recommenders, which make suggestions based on an assessment of the match between the user's needs and the options available. While the first of these categories highlights that recommender systems are *social* technologies that spread awareness of what content other users are consuming, the latter suggests that they are technologies that enhance *usefulness*. However, it is not fully understood what drives usage of recommender systems. Intuitively, it seems that users benefit from recommender systems by improved communication, collaboration, and efficiency, as well as fun or pleasure. As formulated by Zanker et al. (2019), online personalization systems are technologies that "provide the most relevant user experience" (p. 160). Then again, it is not obvious what "relevant" means for different people.

1.2. Aim and objectives

Rogers (2003) asked "What characteristics of innovations affect the rate at which they are adopted?" (p. 219). He argued that researchers had often seen innovations as *equivalent units*, which he considered to be an oversimplification, and he called for comparative studies in this field. To shed light on this aspect of innovation diffusion, the current study aims to develop a model of AI adoption that accounts for variations over *technologies* as well as differences over *individuals* (the conventional approach). Essentially, the research gap that it contributes to fill is the lack of comparative studies of AI adoption across applications. We concentrate on a subcategory of AI (recommender systems) as a proof-of-concept.

Specifically, we report on the findings from an online survey that asked platforms users (e.g., Facebook) to express how often they used its recommender system (e.g., "People You May Know") ($N = 1{,}007$, and 26 recommender systems). We then evaluated the average usage rate per recommender system, which we set as the dependent variable in a regression model. Subsequently, we examined various predictors that could potentially explain the variation in usage. The proposed model accounted for the average perception per technology in a group of potential users, and for individual variations from these averages, as depicted in Fig. 1. We focus on two research objectives. The first is to find factors that contribute to explain the variation in usage of recommender systems among individuals (subsection 2.2), and the second is to find factors that drive the variation in usage *across* technologies (subsection 2.3).

1.3. Contributions

The main contribution of this study is that it explores a new model for the usage of AI applications that could be used for predictions. It conceptualizes variations over AI adoption as an effect of variations over individuals' perceptions and variations over the average perception per application (Fig. 1). The investigation supports better understanding of AI diffusion in two main ways. First, by comparing usage across different technologies, we add an important perspective to the literature since earlier studies have mainly addressed differences in acceptance between individuals (Ayanwale & Ndlovu, 2024; Bouteraa et al., 2024). We have

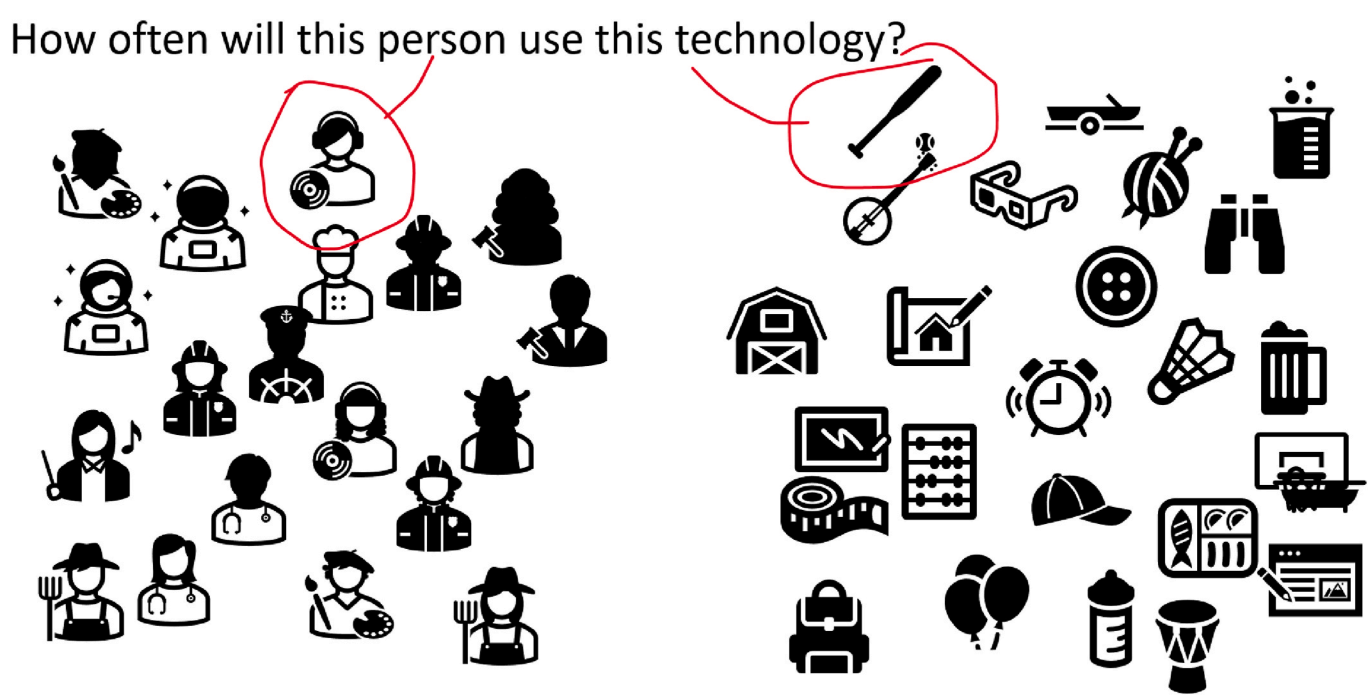

Fig. 1. How often will a particular person use a particular technology? The model presented postulates that this depends on two things: 1) How the technology is perceived by the population of potential users *in general*; for example, whether it is seen as relatively useful; and 2) How the technology is perceived by the *individual* in terms of the deviation from the average; for example, whether the person finds it more fun to use than people in general do.

not found any earlier study that has compared and modeled the difference in usage over a variety of AI applications in this way. Second, by focusing on real and specific instances of AI (recommender systems), we make the evaluations concrete for the participants in the survey. Previous research has often regarded broader conceptualizations of AI (Cabrera-Sánchez et al., 2021; Gansser & Reich, 2021). It is valuable to evaluate specific applications of AI (Schepman & Rodway, 2020), because generalizations may be biased due to cognitive heuristics or media coverage (Sloman & Lagnado, 2005).

Because we seek to find the features of recommender systems that correlate with usage, the study is relevant for platform developers and marketeers who develop and promote new systems. By contributing to developing models of AI diffusion more generally, it is also potentially useful for people who look to predict which applications of AI are likely to spread faster than others and why. Better predictions of the diffusion of different forms of AI are useful for policymakers and researchers who are interested in AI's future social impacts.

1.4. Scope and delimitations

Since the study focused on the acceptance of recommender systems, the dependent variable was usage of such systems *given* usage of the platform; therefore, all the respondents in the survey used the platform with the recommender. Further, because the study explored a novel methodological approach that depended on the analysis of many different technologies, the number of explanatory factors examined needed to be limited to ensure statistical robustness. Hence, we put emphasis on a few factors. All of these were centered on the technology's characteristics (e.g., usefulness), rather than the individuals' (e.g., gender). This is an important delimitation of the study. We concentrated on factors that were relevant for both levels of analysis to enable comparisons of the results at the two levels. To include as many systems as possible, we considered a wide range of recommender systems in terms of the type of content suggested.

2. Literature review and hypotheses development

This section first presents the theoretical foundation and then moves on to address the two research objectives.

2.1. Theoretical background

The most influential theories in this field are the Diffusion of Innovations theory (DOI) (Rogers, 1962), the Theory of Reasoned Action (Fishbein, 1980), the Technology Acceptance Model (TAM) (Davis, 1989), as well as the Innovation Diffusion Theory (Moore & Benbasat, 1991). Of these, the current study draws theoretically on DOI and TAM. DOI is one of the oldest theories within social science, and TAM is the most parsimonious and generic model within technology acceptance, one of the most studied topics in Information Systems research (Nikou, 2019). A key difference between DOI and TAM is that the former conceptualizes diffusion as a *group* phenomenon, while the latter considers technology acceptance as a *personal* behavior. Thus, while DOI is concentrated on broader dynamics of innovation adoption in a population (Ayanwale & Ndlovu, 2024), TAM focuses on the individual. In the current study, we consider both perspectives. This way, we build on earlier research by Nikou (2019), studying smart home technology, and Ayanwale and Ndlovu (2024), focusing on chatbot technology, both of which integrated DOI and TAM.

According to DOI, the central drivers of diffusion are (Rogers, 1983): the innovation itself, which could be a novel idea, process, or object, for example; the adopters, the individuals, households or organizations using the innovation; communication channels, which pertain to how adopters get to know about the innovations, such as mass media or word of mouth; time, as most innovations take time to spread; and the social system, which concerns how the different influences are linked to each other, such as the nature of social relationships and their links to various media and organizational or government structures. In this study, we focus on the first driver: the innovation itself and its characteristics. DOI and TAM and its extensions have important similarities when it comes to the technology-specific characteristics that have been identified as drivers of adoption.

The original version of TAM specifically relates to employees, and it accounts for two constructs that are theorized to drive the behavioral intention to accept a new technology, that is, the individual's disposition to use it (Davis, 1989; Davis et al., 1989); the first one is perceived usefulness, which is the degree to which a particular technology enhances job performance, and the other is perceived ease of use, which reflects how much effort is needed to use it. Based on a review of TAM in addition to seven other models, Venkatesh et al. (2003) developed UTAUT, which additionally accounts for the effect of social influence, the extent to which users perceive that important others believe that they should use a particular technology; and facilitating conditions, the users' perceptions of the resources and support that are available. Since TAM and UTAUT were developed for employee adoption, the latter model was later adapted for non-organizational settings, such as consumer markets, in UTAUT2 (Venkatesh et al., 2012). It adds three constructs to UTAUT: hedonic motivation, the fun or pleasure derived from using the technology; price value, the value perceived by the consumer in relation to the price paid for using the technology; and habit, the habitual and natural use of the technology.

TAM and its extensions have successfully been used to model the acceptance of technologies such as instant messaging (Glass & Li, 2010), e-shopping (Ha & Stoel, 2009), social networks (Hartzel et al., 2016), and mobile navigation systems (Yang et al., 2021). Simliar models have also been applied to AI applications in particular (Ayanwale & Ndlovu, 2024; Bouteraa et al., 2024; Cabrera-Sánchez et al., 2021; Gansser & Reich, 2021; Li et al., 2024; Nikou, 2019; Nordhoff et al., 2020; Yilmaz et al., 2023). For example, an extension of TAM was used as the conceptual foundation for an acceptance scale for generative AI in educational settings, showing high reliability and validity (Yilmaz et al., 2023). One recent study of focused on AI in the form of chatbots for mental health, and the authors proposed an extension of UTAUT with the factors stigma, privacy concerns, and AI hesitancy (Li et al., 2024). However, while it is relatively common in the literature to extend UTAUT or UTAUT2 with more constructs (Cheng et al., 2020; Li et al., 2024; Schomakers et al., 2022; Suhail et al., 2024), we have not seen any earlier study that uses any of these models to *compare adoption* over many different technologies. This motivates the current study, in which we conceptualize adoption as a matter of both the individual's perceptions and the features of the technology.

2.2. Variations in usage over individuals

In this subsection, we focus on the first objective to explain the variation in usage of recommender systems among individuals. We formulate three hypotheses, based on earlier research.

2.2.1. Performance expectancy

In early studies on TAM, Davis (1989) and Davis et al. (1989) found that technologies that are perceived to be more useful are more likely to be accepted. In the more recent literature, a wide range of studies have confirmed that perceived usefulness, and the related construct of performance expectancy (PE) in UTAUT (Venkatesh et al., 2003), correlates with behavioral intention towards adoption. For example, Glass and Li (2010) found that perceived usefulness impacted the adoption of instant messaging in an analysis comparing non-adopters and adopters. Also, perceived performance expectancy was found to be a significant driver of the behavioral intention to use AI in three segments, mobility, household and health (Gansser & Reich, 2021). Perceived performance expectancy has moreover been found to be significant for recommender systems in e-commerce (Cabrera-Sánchez et al., 2020), which is particularly relevant for the current study. Applying an extended version of UTAUT to study the adoption of AI chatbots for mental health among U.S. adults with depression and anxiety symptoms, Li et al. (2024) reported that performance expectancy was positively related to intentions to use AI and continued usage. Bouteraa et al. (2024) drew on UTAUT and social cognitive perspectives to study students' adoption of chatGPT and found that it was associated with performance expectancy. Consistently, a study of among university students in the United Arab Emirates (UAE) reported that performance expectancy was associated with the intention to adopt robots in higher education as they estimated an extended UTAUT2 model (Suhail et al., 2024). On the other hand, Ayanwale and Ndlovu (2024) did *not* find a direct association between usefulness and the behavioral intention to use chatbots for educational uses in a study of undergraduate students. These mixed results call for further research. We hypothesize, in consistency with most earlier studies:

H1a. Individual differences in perceived performance expectancy are positively (+) associated with usage of recommender systems.

2.2.2. *Effort expectancy*

The second factor in TAM that has been proposed to influence the behavioral intention to accept new IT applications is perceived ease of use (Davis, 1989; Davis et al., 1989). Glass and Li (2010) reported that perceived ease of use impacted the adoption of instant messaging. The related construct in UTAUT, effort expectancy (EE), has also been observed to be associated with the behavioral intention to adopt AI in the household, mobility and health segments (Gansser & Reich, 2021). Consistently, effort expectancy was found to be a significant predictor of recommender system adoption (Oechslein et al., 2014). On the other hand, Cheng et al. (2020) reported that effort expectancy was *insignificant* when it came to predicting continuous use intentions of mobile news apps in India. In agreement, Ayanwale and Ndlovu (2024) did *not* find a direct association between perceived ease of use and behavioral intention to use chatbots for educational uses. Again, we find that the results are mixed, which motivates further research. We hypothesize:

H1b. Individual differences in perceived effort expectancy are positively (+) associated with usage of recommender systems.

2.2.3. *Hedonic motivation*

UTAUT2 includes another construct, hedonic motivation (HM), that is, the fun or pleasure derived from using a technology (Venkatesh et al., 2012). It has been found to be a predictor of AI adoption in many earlier studies. For example, Suhail et al. (2024) found that hedonic motivation was associated with students' adoption intentions around robot, and Bellet and Banet (2023) reported that it was related to usage intentions for automated shuttles. Moreover, in a study of the acceptance of mobile health apps among German participants, hedonic motivation was the only significant predictor in both of the studied categories (lifestyle and therapy), and the authors concluded that the acceptance of such apps was determined by emotional factors, rather than utilitarian ones, such as usefulness (Schomakers et al., 2022). In agreement, this factor was the strongest predictor of the acceptance in a study of (conditionally) automated passenger cars (Nordhoff et al., 2020). This result was consistent with findings in a study of the acceptance of automated road transport systems (Madigan et al., 2017). Hedonic motivation has also been reported as a predictor for the adoption of recommender systems specifically (Cabrera-Sánchez et al., 2020; Oechslein et al., 2014). Therefore, we hypothesize:

H1c. Individual differences in perceived hedonic motivation are positively (+) associated with usage of recommender systems.

We did not consider the UTAUT2 constructs of social influence, facilitating conditions, price value and habit, as they are of limited importance for comparisons across recommender systems; for social influence, it is not obvious to another user when a person uses such a system, which means that the importance of peer pressure is unlikely to have considerable effect. Price value and facilitating conditions are likewise unlikely to be very influential, since recommender systems are free to use, and online platforms manage their functioning. The UTAUT2 construct of habit was not included since this factor conceptually overlaps with the dependent variable (usage rate).

2.3. *Differences over technologies*

In this subsection, we address the second objective, finding technology-specific factors that can explain differences in adoption over systems. This objective is associated with three hypotheses that are theoretically grounded in DOI.

2.3.1. *Relative advantage*

Primarily, we considered DOI's conception of perceived relative advantage of using an innovation as compared to the idea or technology it replaces (Rogers, 2003). This aspect may differ between recommender systems and it may be important to explain the variation in how often they are used. It has been confirmed as a key factor in earlier studies on AI; for example, relative advantage was found to be related to students' behavioral intention to use chatbots in Ayanwale and Ndlovu (2024). Moreover, relative advantage can be viewed as DOI's counterpart to the construct of perceived performance expectancy in UTAUT2, and many earlier studies have found that this factor is positively associated with adoption intentions for AI-driven technologies (Bellet & Banet, 2023; Gansser & Reich, 2021; Nordhoff et al., 2020). For example, Bellet and Banet (2023) found that perceived usefulness was associated with usage intentions of automated shuttles among French citizens. Nordhoff et al. (2020) reported that acceptance of conditionally automated passenger cars was associated with performance expectancy among European car drivers. However, most of these studies consider only *one* application at the time, and to our knowledge there are no studies that have evaluated whether PE can explain the variation in adoption *across* AI innovations. Therefore, we hypothesize:

H2a. Average perceived performance expectancy is positively (+) associated with average usage per recommender system.

2.3.2. *Complexity*

The second factor in DOI that has been theorized to influence innovation diffusion is its perceived complexity (Rogers, 2003), which can be viewed as a counterpart to the construct of perceived effort expectancy in UTAUT2. This is one of the main factors in the TAM literature, and its importance for adoption has been confirmed in many earlier studies; for example, Nikou (2019) investigated the relationship between perceived ease of use and the intention to use smart home technology, and found a significant direct effect. Moreover, Gansser and Reich (2021) observed that perceived effort expectancy was related to the behavioral intention to use AI. On the other hand, in their meta-analysis, Tamilmani et al. (2021) concluded that the most questioned path in the UTAUT2 literature was the one between effort expectancy and behavioral intentions. Nevertheless, this factor seems to be important for AI, and it might vary over applications, and so we hypothesize:

H2b. Average perceived effort expectancy is positively (+) associated with average usage per recommender system.

2.3.3. *Compatibility*

The third factor that has been theorized to influence diffusion is perceived compatibility with user values or past experiences, which include socio-cultural values, existing practices, ideas, or needs (Rogers, 2003). This factor can mean many different things in different contexts. Rogers (2003) exemplified it by referring to the slow adoption of improved rice varieties ("miracle rice") due to taste incompatibility among villagers in southern India. This factor has also been found to influence the adoption of AI; for example, Ayanwale and Ndlovu (2024) found that perceived compatibility with learning styles among students was associated their behavioral intention to use chatbots. In the case of a recommender system, this factor could relate to the enjoyment or pleasure derived from using it, which may vary across systems and could drive differences in adoption rates. In this context, the notion of compatibility with values in DOI thus relates to the concept of hedonic motivation in UTAUT2. A meta-analysis of UTAUT2 research found that 29 out of 33 (88%) studies that had examined the relationship between hedonic motivations and use intentions had reported significant values (Tamilmani et al., 2021). Therefore, we account for this factor in our last hypothesis:

H2c. Average perceived hedonic motivation is positively (+) associated with average usage per recommender system.

Consequently, we focused on three factors in DOI (relative advantage, complexity, and compatibility). Since these are conceptually related to three factors in UTAUT2 (performance expectancy, effort expectancy, and hedonic motivation), we used the UTAUT2 constructs to operationalize the DOI factors. We disregarded the two remaining factors in DOI, perceived trialability, the ability to test and experiment

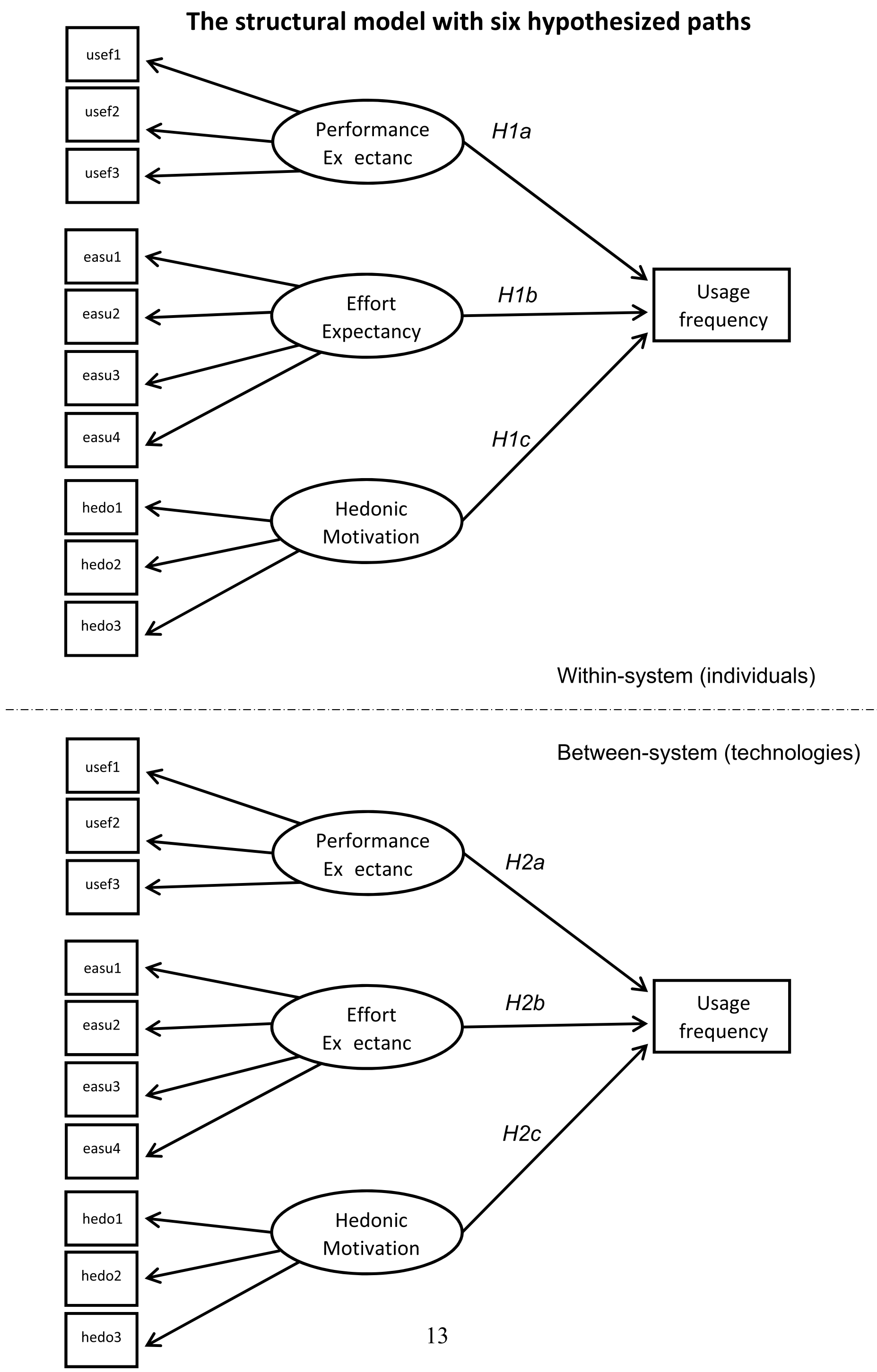

Fig. 2. The two-level SEM with the hypothesized paths (*H1a*-*H1c* and *H2a*-*H2c*), showing the within-system structure (individuals) and the between-system structure (technologies). Oval shapes stand for latent variables, and box shapes stand for observed variables, while arrows are paths and loadings.

with it, and perceived observability, the visibility of the innovation to other prospective adopters (Rogers, 2003). As the study focused on recommender system usage *among platform users,* all the systems were free and accessible to use, which means that trialability can be supposed to be high for all of them. Furthermore, the extent to which users rely on recommender systems is normally not visible to others, which means that observability is unlikely to vary considerably across systems. Hence, these aspects were not expected to have a large effect on the variation in usage.

3. Methodology

The empirical foundation for this study was an online survey among platform users, and their responses were analyzed using structural equation modeling (SEM).

3.1. Structural model

The dependent variable the regression was usage of recommender systems, an observed variable, and the independent variables were the perceived properties of such systems (PE, EE and HM), which were latent variables. To analyze variations over individuals and systems, respectively, we estimated a two-level model with hierarchically structured data. The dependent variable (usage) was thus regressed on predictors at two levels: within-system differences (individuals) and between-systems differences (recommender technologies). Thus, the individual responses at the within-system level were clustered in the 26 recommender systems at the between-system level. This allowed the residuals to not be independent across the different systems. Fig. 2 shows the structural model with the hypotheses.

3.2. Survey items

The questions in the survey were based on the constructs for perceived performance expectancy, perceived effort expectancy, and perceived hedonic motivation in UTAUT2 (Venkatesh et al., 2012). We adapted these items to fit recommender systems, and we only included those that were relevant for this technology (Table 2). The formulations were adapted to ensure that the respondents evaluated the specific recommender system in question, and not the platform. The responses were of Likert type, ranging from "*Strongly disagree*" (1) to "*Strongly agree*" (7). To elicit characteristics of the systems rather than the individuals, the respondents were asked whether they thought people in general would agree with the statements. As we focused on recommender systems, we surveyed *actual* (stated) usage instead of behavioral intentions, which is the dominant approach in TAM literature. We did this because the study was focused on existing platform users, and so we had good reasons to assume that all the respondents had access to the system in focus. Thus, usage was measured by a novel one-item construct, according to: "When I use [the platform, such as Netflix], I use [the recommender system, such as 'Top Picks for You'] … of the time", with a slider ranging from 0% to 100%. The full survey is presented in Supplementary Material S1.

3.3. Sampling and data collection

We designed the survey with the Qualtrics Survey Software (2023), and we used Prolific (https://www.prolific.com) to recruit the participants, distribute the survey, and collect the responses. The survey was based on a convenience sample of U.S. residents aged 18 years and above. It was conducted in June 2023. The respondents were recruited under informed consent, and they were notified that they could withdraw from the survey at any point without any implications. The survey covered basic demographic data, such as gender, age, and level of education, as well as a set of questions that considered platform usage as well as perceived characteristics and usage of recommender systems. To target platform users, the respondents were explicitly asked at the beginning of the survey which platforms they had experience of using, and if they were aware of the corresponding recommender system. From this subset, one platform was then randomly selected as the focus of further questions (e.g., the "Top Picks for You" system on Netflix). People who had no experience of any of the platforms were excluded from further questioning. Each respondent only evaluated *one* recommender system to ensure independence across the samples.

3.4. Data quality

We ensured the quality of the data by a range of measures. Especially, we focused on reducing the risk of common method variance (CMV), which happens when the variance in the data is an effect of the measurement method rather than the hypothesized factors (Podsakoff et al., 2003). We used clear instructions, avoided complex questions, and kept the survey brief. We pre-tested it among our colleagues to ensure that the questions were unambiguous and easy to understand. A pilot study was also run on the same platform as the main study; in this test, as well as in the final survey, we encouraged the respondents to comment on whether anything was hard to understand, and we asked if they had any other observations on the survey design. The feedback that we received confirmed that the survey was understandable.

Additionally, the survey did not include questions with factual answers, in agreement with recommendations (Goodman et al., 2013). To ensure respondent naïveté, the participants who had taken part in the pilot were not invited to take part in the actual survey. The survey completion time was short, generally below 6 min, which meant that there was a low risk of respondent fatigue, a known limitation of online samples (Goodman et al., 2013). Also addressing this issue, the survey included an item to ensure the respondents' full attention: we asked them to select "*strongly disagree*" to show that they had read the statement. This attention check was randomized among the statements that were presented in a matrix question, and all the participants who failed it were excluded from the analyzed data. To further reduce the risk of CMV, we used different scales for the dependent variable (usage), as compared to the independent variables (Podsakoff et al., 2012); specifically, we used a gliding scale for usage, and a Likert-type scale for perceived system characteristics. These items also appeared on different pages of the survey, thereby separating measurements in time and space. Moreover, to reduce potential acquiescence bias and left-side selection bias we used a descending scale ("*disagree*" – "*agree*") (Maeda, 2015). In addition, we randomized the order of the items for each respondent to mitigate any sequence effects (Podsakoff et al., 2003). At the beginning of the survey, we presented its purpose, and the respondents were asked to commit to answering the questions thoughtfully, stressing that we valued the quality of the data (see also Geisen, 2022). Providing respondents with a contextual awareness in the form of comprehensive instructions has been shown to increase motivation and reduce bias generally (Podsakoff et al., 2012).

3.5. Statistical analyses

We developed the two-level SEM in two stages: first, we confirmed the measurement model that related the latent constructs to the observed items, and then we analyzed the estimated paths in the structural model. To validate the measurement model, we evaluated the measurement scales for the three constructs to make sure that they satisfied measurement invariance requirements across the 26 systems, using the approach suggested by Jak et al. (2013). Measurement invariance implies that the scores on the measurement scales can be compared across recommender systems, and this ensures that the associations estimated at the between-system level are not biased. We used a two-level confirmatory factor analysis (CFA) with equal factor loadings across the within-system and between-system levels, and zero residual variance at the between-system level. We considered established

Table 1
The recommender systems examined, sorted by the name of the corresponding platform.

Platform	Service	Recommender system name	Example recommendation	Downloads on Google Play (2023)
Airbnb	Short-term housing	“Explore”	Accommodation	100
Amazon Kindle	E-book service	–	Books	100
Amazon Prime Video	Streaming service for films and TV-series	–	Films	500
Amazon Shopping	Online shopping from different retailers	“You might also like”	Products	500
Apple Books	E-book service	“For You”	Books	NA
Apple Music	Music streaming service	“For You”	Music tracks	100
Audible	Audio book service	“Discover”	Audo books	100
Disney+	Streaming service for films and TV-series	–	Films	100
HBO Max	Streaming service for films and TV-series	“For You"	Films	100
Instagram	Photo and video-sharing social networking	“Explore”	Photos	1000
MyFitnessPal	Workout app	“News Feed”	Workout routines	100
Netflix	Streaming service for films and TV-series	“Top Picks for You"	Films	1000
Pandora – Music & Podcasts	Internet radio and music streaming	“For You”	Music tracks	NA
PicsArt	Image sharing and social media service	–	Images	1000
Pinterest	Visual discovery social networking	“Browse"	Photos	500
Reddit	Social news and forum social network service	“Discover"	News	100
Snapchat	Ephemeral image and video messaging	“Stories"	Videos	1000
SoundCloud	Music streaming	“The Upload"	Music tracks	100
Spotify	Music streaming and media provider	“Made For You"	Music tracks	1000
Tripadvisor	Online travel web service	“Explore"	Travel destinations	100
Uber Eats	Fast food delivery service	“Browse"	Restaurants	100
Wattpad	User created fictional writing	“Other Stories"	Books	100
Wikipedia	User created encyclopedia	“Explore"	Encyclopedia entries	50
Wish	Marketplace for discount products	“Home"	Products	500
Yelp	Crowd-sourced reviews of businesses	“Great spots near you"	Restaurants	50
YouTube	Video streaming	“Home"	Videos	10 000

Table 2
The three constructs and the corresponding items, which were adapted from UTAUT2 to elicit perceptions of the recommender systems. Abbreviations: usef: Performance Expectancy; easu: Effort Expectancy; hedo: Hedonic Motivation.

Construct	Abbreviation	Item
Performance Expectancy	usef1	The [system] makes [platform] more useful in daily life.
	usef2	Using the [system] helps to accomplish things more quickly when using [platform].
	usef3	When using [platform], the [system] increases productivity.
Effort Expectancy	easu1	Learning how to use [system] is easy.
	easu2	The [system] is easy to use.
	easu3	The interaction with the [system] is clear and understandable.
	easu4	It is easy to become skillful at using the [system].
Hedonic Motivation	hedo1	[Platform] is more fun when using the [system].
	hedo2	The [system] makes using [platform] more entertaining.
	hedo3	Using the [system] makes using [platform] more enjoyable.

thresholds for fit indices to evaluate the model, specifically a Comparative Fit Index (CFI) above 0.95, a root mean square error of approximation (RMSEA) below 0.06, and Standardized Root Mean Squared Residual (SRMR) below 0.08 (Hu & Bentler, 1999). We evaluated SRMR for both the within-system and between-system levels, as CFI and RMSEA are not sensitive to between-system level model misspecifications.

To analyze the paths in the structural model, we estimated the effects of the latent variables on usage at the two levels; thus, we examined separately whether the perceived characteristics predicted usage at the within-system level (individuals) and the between-system level (technologies). To confirm that there was a difference in usage across the investigated technologies, we also used a Kruskal Wallis-test with 26 factor levels, one for each recommender system. The statistical analyses were conducted in R (R Core Team, 2023), using the lavaan package (Rosseel, 2012).

3.6. Recommender systems

Since we adopted a wide definition of recommender systems, the systems that we considered did not only regard consumer products, but any form of personalized content or services, such as albums, posts, movies, venues, creators, people, and user-generated content. Alternatively, we could have adopted a narrower focus and selected platforms of similar content, but because of our interest in between-system comparisons, it was necessary to include a large set of recommender systems. The breadth of content reflected the fact that the purpose of many recommender systems today is to enhance the user experience in many ways to increase the overall engagement. Our choice of recommender systems was based on platform usage statistics, and most of the selected platforms had at least 50 million downloads on Google Play (June 2023). In total, the survey included 26 recommender systems (Table 1). The included systems supply services within social media (Pinterest, Instagram, Snapchat, Reddit), e-commerce (Wish, Amazon Shopping), video streaming (Netflix, YouTube, Amazon Prime Video, Disney+, HBO Max), music streaming (Spotify, Pandora, Apple Music, Soundcloud), as well as e-book/audio book services (Apple Books, Amazon Kindle, Audible), food/entertainment (Uber Eats, Yelp), travel (TripAdvisor, Airbnb), workout (MyFitnessPal), creative (Wattpad, PicsArt), and encyclopedic services (Wikipedia).[1]

4. Results

This section presents the descriptive statistics, the measurement model evaluation and the results of the path analyses in the SEM.

4.1. Descriptive statistics

The sample included 1007 participants, who supplied survey responses for 26 recommender systems. Among these, the number of responses varied from 19 for the recommendations in PicsArt to 46 for “Great spots near you” in Yelp. In the sample, 59% were female, 39%

[1] The original survey also included Trivago and Deezer, however we excluded them because of the small number of responses, 6 and 13, respectively.

were male, and the rest selected "Other/prefer not to say". The mean (*M*) age was 41.6 years, and the standard deviation (*SD*) was 14.2. Further, 30% had graduated from college or university, 12% had an advanced degree, 31% were students, 26% were high school graduates, while the rest had no formal education.

4.2. Validating the measurement model

The measurement model accounted for factor analyses that specified the latent variables. In our estimated two-level model with three latent factors at each level, the CFI was 0.982, the RMSEA was 0.038, and the within-system level SRMR was 0.032, indicating a good fit (Table 3). While the between-level SRMR of 0.134 was relatively high, this was likely due to the small sample size at this level (26). The larger SRMR might thus be a reflection of the sample size rather than model misfit (Asparouhov & Muthén, 2018). We applied a χ^2 test to compare the model that had a three-factor structure at the between-system level and unrestricted variance covariance at the within-system level with the fully saturated model (unrestricted variance covariance at both levels). It showed no difference between the target and the null model (χ^2 difference = 21.5, degrees of freedom (*df*) = 42, p = 0.99), and we concluded that the model had an acceptable fit.

For each item, the standardized factor loadings were above 0.7, showing good convergent validity, and item reliability was also very high (Table 3). The composite reliabilities for the latent factors at the between-system and within-system levels were estimated using the approach in Geldhof et al. (2014). Cronbach's α was adequate (>0.7) (Nunnally, 1975) for all the constructs, both within-systems (PE = 0.865, EE = 0.884, HM = 0.901) and between-systems (PE = 0.961, EE = 0.914, HM = 0.985). However, despite the high factor loadings and the good fit, the two-level model with three constructs at each level did not converge. The likely explanation was that two of the items for effort expectancy had small variance at the between-system level (less than 1%). Also, there was a high correlation between PE and EE at the between-system level (0.981) (Table 4). This meant that regressing usage on these two factors combined would yield unreliable estimates. Therefore, in the final model, which is presented in the next subsection, we excluded EE from the between-level analysis. To estimate the individual-level effect of EE, we centered all EE items at the recommender system (cluster) means. For all combinations of constructs (except for PE and EE at the between-system level), the discriminant validity was acceptable, both at the within-system and the between-system level and the Fornell-Larcker (1981) criterion was met (Table 5).

4.3. Path relationships

This section presents the estimated paths in the final structural model, which had a good fit ($\chi^2 = 172.8$, $df = 61$, $p < 0.001$, CFI = 0.984, RMSEA = 0.043, an SRMR within = 0.030, and an SRMR between = 0.092), and adequate explanatory power for explaining the variation in usage at the within-system level ($R^2 = 0.33$) and the between-system level ($R^2 = 0.41$). At the within-system level (individuals), two of the three hypothesized factors, performance expectancy and hedonic motivation, were significant predictors (Table 6). The standardized parameter estimates were $\beta = 0.37$ ($p < 0.001$) (PE) and $\beta = 0.23$ ($p < 0.001$) (HM), highlighting usefulness as the factor with the highest effect.

Moreover, the results showed a large variance in terms of the average usage rate per system (Fig. 3). A Kruskal Wallis-test with 26 factor levels, one for each recommender system, confirmed that there was a significant difference in usage across the systems ($p < 0.001$). It varied from 24.3% (*SD* 29.9) for the "Upload" system in Soundcloud, to 68.2% (*SD* 27.1) for the "Browse" system in Pinterest, which suggests that people use the recommender feature about two-thirds of the time they use Pinterest, which is remarkably high (Table 7). Likewise, the perceived characteristics varied significantly, particularly HM, which ranged from 4.33 (*SD* 1.26) for the News Feed in MyFitnessPal to 5.61 (*SD* 0.75) for "Other Stories" in Wattpad.

At the between-system level (technologies), the path coefficient for performance expectancy was significant, with a standardized parameter estimate $\beta = 0.50$ ($p = 0.033$), while the path coefficient for hedonic motivation was insignificant (Table 8). These results suggested that, while a certain recommender system may bring about distinct levels of enjoyment for different users, the average level of perceived enjoyment per system does not vary that much, at least not to the extent that this factor can explain the variation in average usage across systems. The explanatory power of a simple linear regression model, associating the average performance expectancy with the average usage rate, was substantial ($R^2 = 0.30$) (Fig. 4). Hence, the differences in usage could be predicted relatively well with only this one construct.

We further performed two-level analyses on different subgroups in the sample, based on gender and age, the latter being categorized in two groups by the median (Appendix A: Table A1). In most of the subsamples, the patterns that we saw were qualitatively similar to the general results; however, there were some differences that were potentially meaningful, and which should be investigated in future research. For example, at the within-system level, there were divergences between the general group and older people for HM and EE (Table A1). For comparison, note that UTAUT2 (Venkatesh et al., 2012) accounts for gender, age and experience as moderating factors for HM, but not for EE. Thus, our subgroup analyses suggested that gender and age could be

Table 3
Evaluation of the measurement model: a two-level CFA with three factors (PE, EE and HM), indicating a good fit. Abbreviations: Est.: parameter estimates; *SE*: standard errors, *Std.*: standardized parameter estimates, and ICC: intraclass correlation coefficients; CR: composite reliability; AVE: average variance extracted; PE: Performance Expectancy; EE: Effort Expectancy; HM: Hedonic Motivation.

Parameter	Within-system (individuals)				CR	AVE	Between-system (technologies)						ICC
	Est.	*SE*	*p*	*Std.*			Est.	*SE*	*p*	*Std.*	CR	AVE	
Factor loadings													
PE					0.868	0.685					0.926	0.861	
usef1	1.151	0.038	<0.001	0.825			1.151	0.038	<0.001	1.000			0.034
usef2	1.291	0.040	<0.001	0.851			1.291	0.040	<0.001	1.000			0.038
usef3	1.216	0.041	<0.001	0.799			1.216	0.041	<0.001	1.000			0.049
EE					0.890	0.661					0.680	0.541	
easu1	0.814	0.032	<0.001	0.727			0.814	0.032	<0.001	1.000			0.007
easu2	0.862	0.027	<0.001	0.856			0.862	0.027	<0.001	1.000			0.018
easu3	0.831	0.028	<0.001	0.801			0.831	0.028	<0.001	1.000			0.037
easu4	0.852	0.026	<0.001	0.869			0.852	0.026	<0.001	1.000			0.009
HM					0.902	0.755					0.969	0.943	
hedo1	1.169	0.035	<0.001	0.860			1.169	0.035	<0.001	1.000			0.040
hedo2	1.068	0.033	<0.001	0.846			1.068	0.033	<0.001	1.000			0.050
hedo3	1.153	0.033	<0.001	0.895			1.153	0.033	<0.001	1.000			0.050

Table 4
Residual variance, factor variance and factor covariance. Abbreviations: Est.: parameter estimates; *SE*: standard errors, *Std.*: standardized parameter estimates; PE: Performance Expectancy; EE: Effort Expectancy; HM: Hedonic Motivation. Note that we excluded between-level EE in the final model due to the high correlation between PE and EE at the between-system level.

	Within-system (individuals)				Between-system (technologies)			
Parameter	Est.	*SE*	*p*	*Std.*	Est.	*SE*	*p*	*Std.*
		Residual variance						
PE								
usef1	0.622	0.040	<0.001	0.319	0.000	–	–	0.000
usef2	0.637	0.044	<0.001	0.276	0.000	–	–	0.000
usef3	0.841	0.049	<0.001	0.362	0.000	–	–	0.000
EE								
easu1	0.589	0.030	<0.001	0.471	0.000	–	–	0.000
easu2	0.272	0.018	<0.001	0.268	0.000	–	–	0.000
easu3	0.387	0.022	<0.001	0.359	0.000	–	–	0.000
easu4	0.235	0.016	<0.001	0.245	0.000	–	–	0.000
HM								
hedo1	0.479	0.030	<0.001	0.260	0.000	–	–	0.000
hedo2	0.453	0.027	<0.001	0.284	0.000	–	–	0.000
hedo3	0.329	0.024	<0.001	0.198	0.000	–	–	0.000
		Factor variance						
PE	1.000	–	–	1.000	0.054	0.024	0.024	1.000
EE	1.000	–	–	1.000	0.015	0.013	0.275	1.000
HM	1.000	–	–	1.000	0.060	0.025	0.017	1.000
		Factor covariance						
PE with EE	0.438	0.030	<0.001	0.438	0.027	0.015	0.062	0.981
PE with HM	0.812	0.016	<0.001	0.812	0.024	0.020	0.216	0.426
EE with HM	0.501	0.028	<0.001	0.501	0.002	0.013	0.858	0.079

Table 5
Assessment of the discriminant validity. The diagonal elements in italics show the square root of the AVE, and the off-diagonal elements account for the correlation coefficients between the different constructs. Abbreviations: PE: Performance Expectancy; EE: Effort Expectancy; HM: Hedonic Motivation. *EE is grayed out as it was excluded at the between-system level in the final model due to its strong correlation with PE.

	Within-system (individuals)			Between-system (technologies)		
	PE	EE	HM	PE	EE*	HM
PE	*0.827*			*0.928*		
EE	0.438	*0.813*		0.981	*0.736*	
HM	0.812	0.501	*0.869*	0.426	0.079	*0.971*

moderating factors in a between-level analysis, in correspondence with UTAUT2, highlighting the need for more research. Given the limited sample size, we did not have sufficient statistical power to fully examine these variations further in the current study.

5. Discussion

This study examined drivers behind differences in usage of recommender technologies, hypothesizing that three factors adapted from DOI and UTAUT2 could explain the variation in usage at two levels of analysis.

5.1. Predictors of variations over individuals

Our first objective was to find factors that could explain the variation in usage among individuals. The estimated model generated two significant paths, one between perceived performance expectancy and usage ($\beta = 0.37$, $p < 0.001$), and one between perceived hedonic motivations and usage ($\beta = 0.23$, $p < 0.001$) (Fig. 5). The path between effort expectancy and usage was insignificant. As a result, *H1a* and *H1c* were supported, while *H1b* was dismissed (Table 9). The strong link between performance expectancy and adoption substantiated earlier findings on AI in the mobility, household and health segments by Gansser and Reich (2021). It was also in agreement with earlier results around the drivers behind the continued use of AI chatbots (Li et al., 2024), as well as research on chatGPT specifically (Bouteraa et al., 2024) and robots (Suhail et al., 2024).

Our results around the importance of hedonic motivation was consistent with earlier studies on the adoption of different types of technologies that use AI, such as mobile health apps (Schomakers et al., 2022), robots (Suhail et al., 2024), automated shuttles (Bellet & Banet, 2023), automated road transport systems (Madigan et al., 2017) and conditionally automated passenger cars (Nordhoff et al., 2020). Notably, our findings were also consistent with Cabrera-Sánchez et al. (2020), who reported that performance expectancy and hedonic motivation were the most important factors to explain the intention to use recommendation systems in e-commerce. This was expected, since we also focused on recommender systems, however in a broader sense.

Table 6
Within-system level summary of the final structural model coefficients, showing that PE and HM could explain the variance between individuals. Abbreviations: Est.: parameter estimates; *SE*: standard errors, *Std.*: standardized parameter estimates (β); PE: Performance Expectancy; EE: Effort Expectancy; HM: Hedonic Motivation. N.S. not significant. *Significant at the 5% level. **Significant at the 1% level. *** Significant at the 0.1% level.

Within-system path relationships (individuals)							
Hypothesis	Path	Est.	*SE*	*z* value	*P* (>|z|)	*Std.*	Comment
H1a	PE → Usage	10.42***	1.66	6.29	<0.001	0.37***	Supported
H1b	EE → Usage	0.15	0.92	0.16	0.873	0.005 (N.S.)	Not supported
H1c	HM → Usage	6.42***	1.70	3.79	<0.001	0.23***	Supported

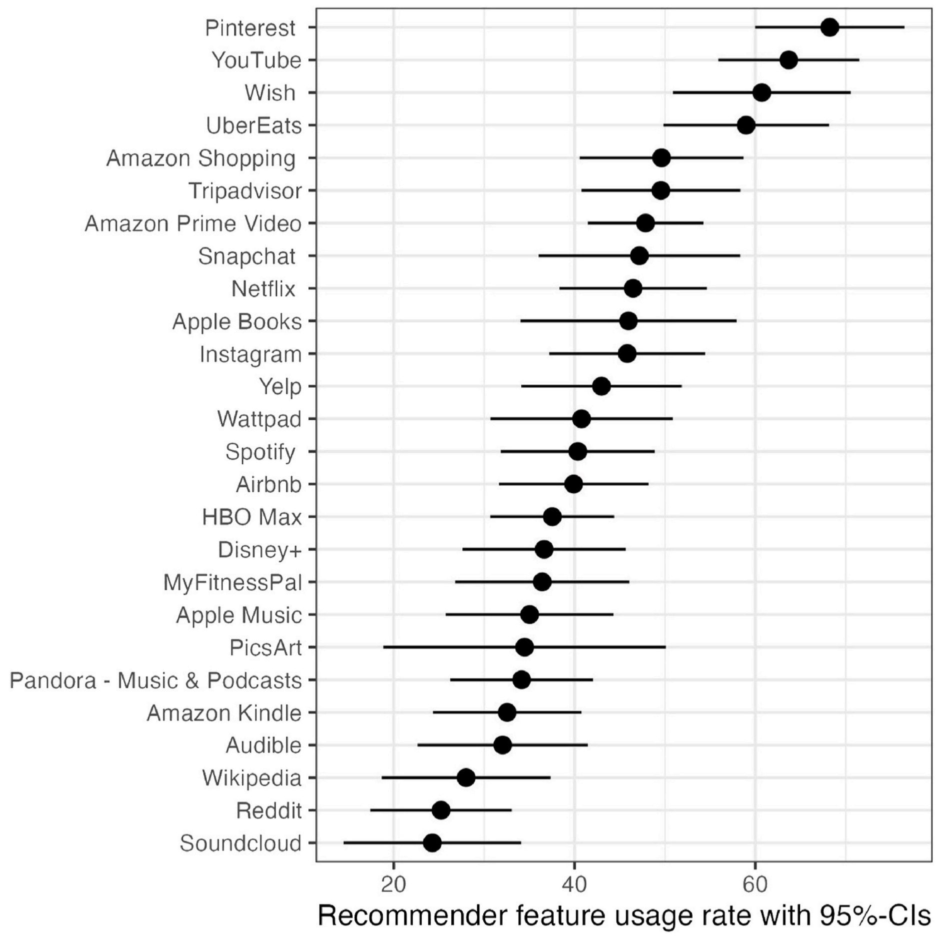

Fig. 3. Mean usage per recommender system expressed as the percentage of the time spent on the platform, with 95% confidence intervals, showing a significant variation across the systems.

The finding that the path between effort expectancy and usage was unimportant was in discord with early results on TAM (Davis, 1989); nevertheless, it confirmed several other studies. For example, Cheng et al. (2020) reported that effort expectancy did *not* relate to use intentions pertaining to mobile news apps in India. Our results also confirmed findings in Ayanwale and Ndlovu (2024) on the use of chatbots for educational uses. Consistently, in their meta-analysis, Tamilmani et al. (2021) concluded that the most questioned path in UTAUT2 was the one between effort expectancy and use intentions.

5.2. Drivers behind differences over technologies

The second objective was to find factors that can explain the variation in usage *across* recommender systems. Given a variety of recommender systems, this perspective addresses which one of them will be used more often and why, for example. There was only one significant path in our data, the one between performance expectancy and usage ($\beta = 0.50$ and $p = 0.033$). Our findings thus supported *H2a*, while *H2c* was refuted (Table 9). (As noted above, it was not possible to investigate *H2b*, due to the strong correlation between EE and PE at this level.) This result contributes to the development of predictive models for AI diffusion, because it suggests that usage can partly be predicted by perceived usefulness ($R^2 = 0.30$).

The stark link between usefulness and adoption that we found confirms many earlier studies at the individual level; in particular, a meta-analysis of UTAUT2 reported that the most adopted path in this literature – with the most significant values – was the one between performance expectancy and behavioral intention (Tamilmani et al., 2021); of the 50 examined studies that investigated this path, only 10% found a non-significant path relationship. Recent research on AI specifically also supports this finding (Bellet & Banet, 2023; Bouteraa et al., 2024; Nordhoff et al., 2020; Suhail et al., 2024). However, while it has been shown before that PE is a central factor for the adoption of *specific* AI technologies, the current study is unusual in the sense that it highlights the importance of this factor when it comes to modeling *differences* in adoption across AI technologies.

5.3. Theoretical implications

The main theoretical contribution of this study was that it presents and evaluates a new model of AI acceptance that bridges two strands of research: the DOI and the TAM frameworks. The study provides support for the proposition that there are inherent differences across technologies that affect their relative adoption rates. Even though it may seem clear that some systems spread faster than others, as they have different functions in our lives, it is not obvious that there are *commonalities* among the technologies that spread faster. While we recognize that there are many aspects that may underlie the variation in usage over recommender systems, such as the type of content suggested, our results suggest that such aspects, at least partially, can be represented by the concept of usefulness (Fig. 5).

The results showed that *distinct* factors could explain the variations in usage over the two levels (Fig. 5); in particular, they suggested that performance expectancy was a good predictor for usage of recommender systems at *both* levels of analysis, while hedonic motivation was only significant at one level. This difference across the two levels of analysis

Table 7
Usage rate and perceived characteristics per recommender system (*M* and *SD*), sorted by the name of the corresponding platform. Abbreviations: PE: Performance Expectancy; EE: Effort Expectancy; HM: Hedonic Motivation. Usage varied between 24.3% (*SD* 29.9) for the "Upload" system in Soundcloud, and 68.2% (*SD* 27.1) for the "Browse" system in Pinterest.

Platform with the system	Usage (%)	PE (1–7)	EE (1–7)	HM (1–7)
Airbnb	39.9 (26.2)	4.62 (1.25)	5.99 (0.65)	5.26 (1.13)
Amazon Kindle	32.6 (27.2)	4.27 (1.38)	5.82 (0.90)	4.38 (1.44)
Amazon Prime Video	47.9 (20.5)	4.57 (1.25)	5.79 (0.95)	4.49 (1.21)
Amazon Shopping	49.6 (27.6)	4.85 (1.08)	6.03 (0.69)	4.83 (1.13)
Apple Books	46.0 (32.2)	5.01 (1.39)	6.19 (0.63)	5.15 (1.31)
Apple Music	35.0 (28.6)	4.60 (1.14)	5.56 (1.00)	4.79 (1.15)
Audible	32.1 (30.5)	3.97 (1.42)	5.85 (0.74)	4.59 (1.44)
Disney+	36.6 (28.9)	4.33 (1.37)	5.80 (0.91)	4.34 (1.28)
HBO Max	37.5 (23.0)	4.73 (0.99)	5.89 (0.70)	4.66 (0.88)
Instagram	45.8 (27.6)	4.23 (1.59)	5.60 (1.03)	5.34 (1.08)
MyFitnessPal	36.4 (28. 9)	3.80 (1.26)	5.51 (1.00)	4.33 (1.26)
Netflix	46.5 (26.4)	4.87 (1.06)	6.02 (0.74)	4.72 (0.87)
Pandora	34.1 (25.6)	4.36 (1.13)	5.84 (0.80)	4.79 (0.95)
PicsArt	34.5 (34.1)	4.46 (1.87)	5.30 (1.62)	4.93 (1.81)
Pinterest	68.2 (27.1)	4.76 (1.43)	5.91 (1.16)	5.41 (1.23)
Reddit	25.2 (24.4)	3.88 (1.31)	5.52 (0.90)	4.64 (1.36)
Snapchat	47.2 (35.7)	4.07 (1.79)	5.77 (1.23)	5.07 (1.61)
Soundcloud	24.3 (29.9)	4.47 (1.41)	5.55 (1.00)	4.81 (1.29)
Spotify	40.4 (27.6)	4.69 (0.99)	5.84 (0.88)	5.11 (0.92)
Tripadvisor	49.6 (28.5)	4.97 (1.09)	5.82 (0.72)	5.23 (0.88)
UberEats	59.0 (30.7)	4.79 (1.43)	5.98 (0.83)	4.99 (1.22)
Wattpad	40.8 (23.6)	5.09 (0.87)	5.91 (0.65)	5.61 (0.75)
Wikipedia	28.0 (28.4)	4.43 (1.43)	5.71 (0.74)	5.33 (1.05)
Wish	60.7 (28.7)	5.08 (1.25)	6.17 (0.64)	5.27 (1.21)
Yelp	43.0 (30.1)	4.91 (1.34)	5.91 (0.71)	4.99 (1.20)
YouTube	63.7 (25.3)	4.79 (1.26)	5.80 (1.04)	5.13 (1.00)

suggests that new knowledge was obtained in the between-technologies analysis. This implies that models that only account for differences between individuals cannot be used to fully capture the drivers behind differences across AI technologies. It highlights the need for more research on the between-systems level. Accounting for differences over technologies in a macroscopic approach is still unusual as compared to most earlier studies, which tend to focus on one innovation at a time. In summary, our findings contribute to the development of a new theory of AI diffusion, and they suggest that the concept of usefulness should be a central feature in such a theory. The theoretical approach that we have presented is possibly relevant for other categories of AI such as voice assistants or IoT.

5.4. *Practical implications*

The emerging role of AI in society has been discussed intensely in recent years, for example in terms of its role in farming and forestry management (Galaz et al., 2021), its ethical implications in finance (Svetlova, 2022), its position to promote sustainable societies and a circular economy (Kar et al., 2022), its function in smart cities and future transportation systems (Herath & Mittal, 2022), as well as its contribution to make the health care sector more effective (Davenport & Kalakota, 2019). Many argue for faster AI adoption, for instance by highlighting that the technology has a lot of potential to address global sustainability goals (Vinuesa et al., 2020). However, for AI to have political, social, and economic influence in these sectors, widespread adoption is a prerequisite. And, to promote acceptance of beneficial AI, it is useful to better understand the mechanisms that drive its adoption. Recommender systems make up a useful starting point here, because they make up applications of AI that have been available to consumers for a relatively long time, which implies that usage is not limited by awareness or accessibility, but rather behavioral intentions among potential users. By investigating this category of AI, our results are thus potentially useful for researchers and policymakers who are interested in AI's diffusion in the near future. They are also relevant for organizations that aim to develop and market the characteristics that drive the adoption of AI features, such as companies within e-commerce, gaming, and online streaming enterprises, as well as other platforms and developers of online applications or desktop software.

5.5. *Limitations*

Since we explored a novel approach in this study, it was impossible to entirely follow established methods in the technology acceptance literature. Thus, when the results in this study are interpreted, it is important to consider that it has limitations, as detailed below.

5.5.1. *Statistical analyses*

While the two-level SEM allowed us to find predictors of usage at two levels of analysis, it meant that we could only assess a few independent variables per system since we wanted to compare adoption across many systems, which was a limitation. We also needed to consider a relatively wide range of recommender systems in terms of the type content recommended, because we wanted to include as many systems as possible to increase the statistical power of the regression. Therefore, another limitation is regarding the wide range of recommender systems included. Arguably, systems with such varying content, ranging from restaurants to music tracks, do not lend themselves to comparisons, and

Table 8
Between-system level summary of the final structural model coefficients. Abbreviations: Est.: parameter estimates; *SE*: standard errors, *Std.*: standardized parameter estimates (β); PE: Performance Expectancy; EE: Effort Expectancy; HM: Hedonic Motivation. N.S.: not significant. *Significant at the 5% level. **Significant at the 1% level. *** Significant at the 0.1% level. This shows that PE was significant at both levels, while HM was only significant at the within-system level.

Between-system path analyses (technologies)							
Hypothesis	Path	Est.	*SE*	*z* value	*P* (>|z|)	*Std.*	Comment
H2a	PE → Usage	21.39*	10.01	2.14	0.033	0.50*	Supported
H2b	Not possible to examine due to the correlation between PE and EE						
H2c	HM → Usage	9.30	8.95	1.04	0.299	0.23 (N.S.)	Not supported

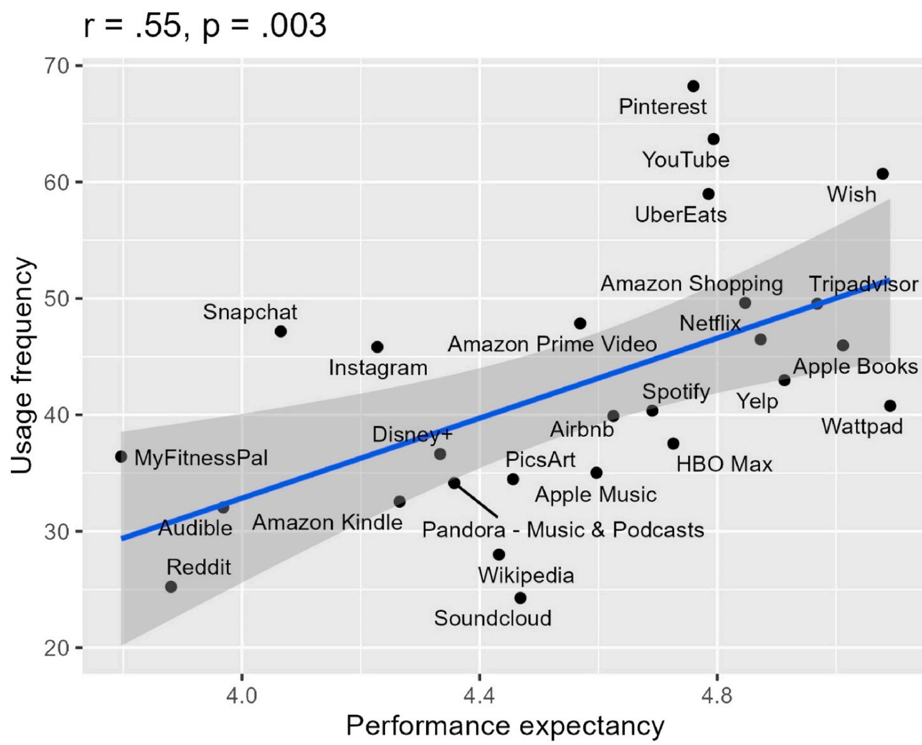

Fig. 4. Scatter plot of the average usage rate per system vs. the average perceived performance expectancy, denoted by the corresponding platform. This shows that this factor had considerable explanatory power ($R^2 = 0.30$).

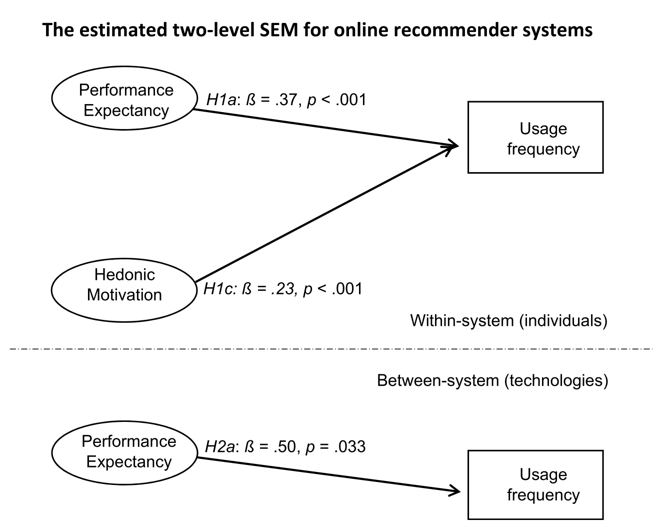

Fig. 5. Path analyses in the two-level model, showing the significant paths with standardized parameter estimates (β) and p values.

an alternative approach with a narrower scope could have been to focus on a subcategory of recommender systems; however, such as approach would be methodologically limited by the lower number of platforms to include, which would reduce the statistical power of the analysis. The broad scope that we adopted here seemed relevant, given that the platforms did not cluster along the type of content recommended (music, news, etc.) (Fig. 3).

Further, despite that UTAUT2 is a well-established theory, we found that there was a significant correlation between performance expectancy and effort expectancy at the system level, which meant that we could only include one of them at the system level in the final model, which was another limitation. Nonetheless, this finding may have highlighted an important aspect of recommender systems since such technologies might be useful precisely *because* they are easy to use. This means that they provide quick and convenient access to content that the user may otherwise have found by more cumbersome methods. The correlation between PE and EE is supported by earlier research, which shows that usefulness and ease of use are related for AI technologies. For example, Gansser and Reich (2021) found that convenience comfort was a strong predictor for performance expectancy for AI within the mobility and household sectors. They also reported that performance expectancy was a strong predictor of usage intentions, concluding that a key motivator for AI acceptance was that the technology could make life easier and more comfortable (Gansser & Reich, 2021), highlighting the association between effort expectancy and performance expectancy for AI. Furthermore, Nordhoff et al. (2020) examined the interrelationships between performance and effort expectancy within the context of automated cars. They reported on considerable support for a positive relationship between perceived ease of use and perceived usefulness when it comes to acceptance, that is, vehicles that are seen as easier to use are also generally found to be more useful.

5.5.2. *Data quality*

Regarding data collection, we followed established practices in the UTAUT2 literature, implying that the dependent variable, reflecting acceptance intentions, was measured at the same time as the independent variables. However, this approach involved the risk for common method bias. To mitigate this concern, we took several measures relating to survey design and implementation. Nevertheless, we cannot fully disregard the possibility that some of the variance can be attributed to CMV; however, note that the potential noise introduced by CMV was unlikely to have had a significant impact on the main results, because it would not influence the between-system variance, which was the focus of the study.

Another limitation regarded the empirical data, which was based on a convenience sample recruited over Prolific. Thus, when interpreting these results, it is important to keep in mind that the sample that was not representative of the general U.S. population. Nevertheless, in some

Table 9
Summary of findings in the tests at the two levels of analysis. Abbreviations: PE: Performance Expectancy; EE: Effort Expectancy; HM: Hedonic Motivation.

Hypothesis	Path	Level of analysis	Description	Finding
H1a	PE → Usage	Within-system (individuals)	Individual differences in perceived performance expectancy are positively (+) associated with usage of recommender systems.	Supported
H1b	EE → Usage		Individual differences in perceived effort expectancy are positively (+) associated with usage of recommender systems.	Not supported
H1c	HM → Usage		Individual differences in perceived hedonic motivation are positively (+) associated with usage of recommender systems.	Supported
H2a	PE → Usage	Between-system (technologies)	Average perceived performance expectancy is positively (+) associated with average usage per recommender system.	Supported
H2b	EE → Usage		Average perceived effort expectancy is positively (+) associated with average usage per recommender system.	Not possible to examine
H2c	HM → Usage		Average perceived hedonic motivation is positively (+) associated with average usage per recommender system.	Not supported

regards, our sample was similar to the U.S. population at large. In the sample, 30% had graduated from college or university while the share of U.S. adults (25 years+) with a bachelor's degree or higher is 34.3% (2018–2022) (United States Census Bureau, 2024). Further, the average age in our sample was 41.6 years, which is near the U.S. average of 38.9 years (United States Census Bureau, 2023). On the other hand, 59% in the sample were female, while the corresponding number in the U.S. is 50% (United States Census Bureau, 2024).

When comparing an MTurk convenience sample with a population-based online panel, Weinberg et al. (2014) found that participants in the former group tended to be younger, better educated and more often female. This partly reflected our sample, and it indicates that online convenience samples are likely to be more computer literate than national samples; on the other hand, it also suggests that the variation in recommender system usage that we found may be larger in the general U.S. population. Previous research has shown that data from self-selected internet-based surveys can be as valid as those obtained via traditional methods, and many studies based on online convenience samples have replicated previous experimental and observational results (Berinsky et al., 2012; Clifford et al., 2015; Mullinix et al., 2015; Rice & Winter 2020; Weinberg et al., 2014). On the other hand, a recent decline in MTurk's data quality has been observed (Tang et al., 2022), and comparisons between different online data collection services have found that Prolific generally provides more reliable data than MTurk and undergraduate student samples (Douglas et al., 2023; Peer et al., 2022). Therefore, we used Prolific here.

5.6. Future research

While general conclusions drawn from this study were limited by the sample and the specific cultural context in focus (platform users in the U. S.), we find that the methodological approach holds a lot of promise. We recommend that the two-level method be extended to account for a larger set of explanatory factors, more categories of AI innovations, and other cultural settings, as detailed below.

First, the approach presented should be investigated with larger samples to enable thorough investigations of the differences between women/men, younger/older individuals, as well as varying levels of experience. This is particularly relevant given that age, gender and experience are moderating factor in UTAUT2 (Venkatesh et al., 2012). Moreover, other independent variables than those considered here are conceivable; for example, we linked compatibility in DOI with hedonic motivation in UTAUT2, although we did not explicitly investigate the effect of compatibility with existing *values* on adoption. There are many aspects that may influence adoption in a negative way, since algorithmic decision support raise important social and ethical concerns (Christiano, 2022; Barocas & Selbst, 2016; Engström, 2022; Milano et al., 2020; O'Neil, 2016); for example, Milano et al. (2020) identified six issues related to the ethical impacts of recommender systems, arguing that a variety of stakeholders needs to be considered. Christiano (2022) reasoned that personalization algorithms can challenge democracy by polarizing the electorate, spreading fake news, and making vulnerable users susceptible to manipulation, thus posing a threat to political equality. Engström (2022) discussed various possible conflicts of interests between providers and users as related to privacy, autonomy, and transparency, with a focus on the spreading mechanisms of recommender systems. Both collaborative filters and content-based recommendations use detailed information about users to improve precision in suggestions (Burke, 2002), highlighting privacy preservation as a central challenge in their use. This is underlined by the fact that there are strong economic incentives involved in the deployment of these systems (Zanker et al., 2019). Outlining a research agenda for AI tools, Venkatesh (2022) discussed some particularities of AI that may influence their usage, including their blackboxed nature, which entails little or no understanding of their underlying algorithms; their tendency to make mistakes and to learn over time; in addition to model bias, which could cause unequal treatment and discrimination. Venkatesh (2022) further reasoned that environmental variables, in addition to factors in UTAUT, might have an effect on employee AI adoption intentions.

Related, Cabrera-Sánchez et al. (2020) reported that perceived trust and technology fear influenced usage intentions for recommender systems in e-commerce. These factors could also vary across apps, given that they have different providers. Studies on news recommendations have suggested that concerns about privacy and personal autonomy matter for users when it comes to the adoption of systems for algorithmic selection that is guided by past behavior (Monzer et al., 2020; Thurman et al., 2019). Thurman et al. (2019) found that the appreciation of algorithmic news curation varied with the data that were used to drive the systems, and argued that adoption depends on the *context*. Hence, it may be that privacy concerns could explain a part of the variance in usage of recommender systems for *news*, but not for other types of content. Detailing these patterns is a suggested area for future research.

Second, while our results suggested that the two-level approach was suitable for recommender systems, such technologies have some unusual features; for example, they cannot be used individually, but only as part of some other application, and they are generally free of charge. This means that the variation in usage across other forms of consumer AI may be explained by *other* factors than those considered here (e.g., price value). Given the proliferation of AI applications in recent years, it would be interesting to use the method explored in this study to assess other sets of AI innovations, for instance home automation technologies or IoT.

Third, we recommend that future research accounts for other cultural contexts and consider cross-cultural comparisons with regards to AI adoption. Online recommender systems are global, and there might be important cultural differences among users' beliefs that could affect usage in different countries. Previous studies have found that online privacy concerns vary over the world (Bellman et al., 2004; Engström et al., 2023; Milberg et al., 2000; Trepte et al., 2017), which indicates that there are cultural differences in the relationship between values related to privacy and AI adoption, which should be considered in future research. In this study, we have focused on the U.S. market, which may in fact be an *outlier* in international comparisons, since many of the big technology developers have their headquarters in Silicon Valley, U.S. Thus, American consumers may be forerunners of AI adoption, and variations in usage may be higher in other countries. The findings in Engström et al. (2023) suggested that privacy concerns were lower in countries with higher Internet penetration, which is a sign that such considerations are less important in the U.S. population than in other countries. In particular, privacy and user autonomy may be important for European consumers, considering the enactment of the EU General Data Protection Regulation (GDPR) on May 25, 2018, as well as the Digital Services Act package (DSA), and the Digital Markets Act (DMA). These regulations include enhanced protection of user data as well as measures to ensure user autonomy and limit online profiling, that is, the use of information of past user behavior to predict and direct future actions. It is possible that the implementation of these policies will increase the adoption of recommender systems in the EU if consumers perceive that they reduce privacy risks. Article 29 in DSA specifically accounts for obligations for very large online platforms, asserting that such platforms should clearly state the main parameters used in their recommender systems. They should also allow users to change those parameters, and there should be an option which is not based on profiling. Such privacy settings could spur adoption of recommender systems, given that usefulness is kept.

6. Conclusion

This study aimed to develop and test a model of recommender systems usage that considered variations over technologies as well as variations over individual users. We estimated the model using data from 1007 users and 26 recommender systems. To our knowledge, no earlier study has systematically compared usage across different recommender systems online and tried to explain the variation in this way. The objectives were to find factors that might explain differences between individuals' usage and differences over technologies.

We learned that differences in usage among individuals could be explained relatively well by two constructs adapted from UTAUT2: perceived performance expectancy (usefulness), as well as perceived hedonic motivation (enjoyment). These findings confirmed earlier research about the variation in adoption of AI-driven technologies over individuals. However, our results found no support for the association between effort expectancy and usage, in contrast with many earlier studies. We also learned that usage varied significantly over the 26 systems and that the variation in usage could be explained relatively effectively by only one factor: the perceived average performance expectancy per system. We did not find support for the hypothesis that it could be explained by hedonic motivations at this level of analysis. These results suggested that it is possible to predict the relative adoption of one recommender technology over another one with this one factor (usefulness). The finding that different factors were significant at the two levels of analysis underlined the relevance of a two-level model for AI adoption, and we suggest that future research considers whether the identified factors in this model apply to other types of AI. Such studies could, for example, examine less prevalent forms of AI, such as home automation. This would contribute to better informed forecasts of the diffusion of various categories of AI.

While this study focused on three technology-specific constructs derived from DOI and UTAUT2, it did not thoroughly investigate the effect of a technology's compatibility with values, and this is a proposed path for future research. Especially, concerns about privacy and data integrity may be important. Our analysis was based on a sample in the U.S., and we suggest that future research be conducted in other cultural contexts, where, for example, privacy concerns may play a larger role on AI adoption. A particularly interesting case would be to study consumer attitudes towards the adoption of recommender systems in the EU after the enactment of the DSA, which specifically regards recommender systems in Article 29.

CRediT authorship contribution statement

Emma Engström: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Irina Vartanova:** Software, Formal analysis. **Jennifer Viberg Johansson:** Writing – review & editing. **Minna Persson:** Writing – review & editing, Investigation, Data curation. **Pontus Strimling:** Writing – review & editing, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Data availability

Part of the data will be shared upon reasonable request in view of privacy/ethical restrictions.

Acknowledgment

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbr.2024.100449.

Appendix A. Subgroup analyses

In this section, we show two-level SEM results for various subgroups in the sample (Table A1). While the results for each subgroup were generally consistent with the results for the general population, there were some deviations. At the within-system level, there were divergences between the general group and older people for hedonic motivation, although the p value was relatively low ($p = 0.124$), and between the general group and older people for effort expectancy. Moreover, there were divergences at the between-system level between the general group and men as well as younger people for performance expectancy. These results suggested that the average perceived performance expectancy was not associated with usage for men and younger people.

Table A.1
Within-level summary of the structural model coefficients. Younger and older refer to people aged below and above the median in the sample. Abbreviations: *Std.*: standardized parameter estimates; PE: Performance Expectancy; EE: Effort Expectancy; HM: Hedonic Motivation; N.S.: not significant; *Significant at the 5% level. **Significant at the 1% level. *** Significant at the 0.1% level.

Level	Path	All	Women	Men	Younger	Older
		Std. (*p* value)	*Std.* (*p* value)	*Std.* (*p* value)	*Std.* (*p* value)	*Std.* (*p* value)
Within-system	PE → Usage	0.37*** (<0.001)	0.41*** (<0.001)	0.29** (0.002)	0.30*** (<0.001)	0.43*** (<0.001)
	EE → Usage	0.005 (0.873)	0.03 (0.481)	−0.04 (0.502)	−0.07 (0.183)	0.09* (0.033)
	HM → Usage	0.23***(<0.001)	0.19* (0.015)	0.33** (0.001)	0.29*** (<0.001)	0.15 (0.124)
Between-system	PE → Usage	0.50* (0.029)	0.67* (0.016)	0.50 (0.491)	0.30 (0.218)	0.78* (0.029)
	HM → Usage	0.23 (0.299)	0.13 (0.608)	−0.05 (0.889)	0.43 (0.094)	0.18 (0.592)

References

Asparouhov, T., & Muthén, B. (2018). SRMR in Mplus. https://www.statmodel.com/download/SRMR2.pdf.

Ayanwale, M. A., & Ndlovu, M. (2024). Investigating factors of students' behavioral intentions to adopt chatbot technologies in higher education: Perspective from expanded diffusion theory of innovation. *Comp Human Behavior Reports, 14*, Article 100396. https://doi.org/10.1016/j.chbr.2024.100396

Barocas, S., & Selbst, A. (2016). Big Data's Disparate impact. *California Law Review, 104* (3). https://www.jstor.org/stable/24758720.

Bellet, T., & Banet, A. (2023). UTAUT4-AV: An extension of the UTAUT model to study intention to use automated shuttles and the societal acceptance of different types of automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour, 99*, 239–261. https://doi.org/10.1016/j.trf.2023.10.007

Bellman, S., Johnson, E. J., Kobrin, S. J., & Lohse, G. L. (2004). International differences in information privacy concerns: A global survey of consumers. *The Information Society, 20*(5), 313–324. https://doi.org/10.1080/01972240490507956

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online Labor markets for experimental research: Amazon.com's mechanical Turk. *Political Analysis, 20*(3), 351–368. https://doi.org/10.1093/pan/mpr057

Bouteraa, M., Bin-Nashwan, S. A., Al-Daihani, M., Dirie, K. A., Benlahcene, A., Sadallah, M., Zaki, H. O., Lada, S., Ansar, R., Fook, L. M., & Chekima, B. (2024). Understanding the diffusion of AI-generative (ChatGPT) in higher education: Does students' integrity matter? *Comp Human Behavior Reports, 14*, Article 100402. https://doi.org/10.1016/j.chbr.2024.100402

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction, 12*(4), 331–370. https://doi.org/10.1023/A:1021240730564

Cabrera-Sánchez, J.-P., Ramos-de-Luna, I., Carvajal-Trujillo, E., & Villarejo-Ramos, Á. F. (2020). Online recommendation systems: Factors influencing use in E-commerce. *Sustainability, 12*(21). Article 21.

Cabrera-Sánchez, J.-P., Villarejo-Ramos, Á. F., Liébana-Cabanillas, F., & Shaikh, A. A. (2021). Identifying relevant segments of AI applications adopters–Expanding the UTAUT2's variables. *Telematics and Informatics, 58*, Article 101529.

Cheng, Y., Sharma, S., Sharma, P., & Kulathunga, K. (2020). Role of personalization in continuous use intention of mobile news apps in India: Extending the UTAUT2 model. *Information, 11*(1), 33. https://doi.org/10.3390/info11010033

Christiano, T. (2022). Algorithms, manipulation, and democracy. *Canadian Journal of Philosophy, 52*(1), 109–124. https://doi.org/10.1017/can.2021.29

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from Mechanical Turk valid for research on political ideology? *Research & Politics, 2*(4), Article 205316801562207. https://doi.org/10.1177/2053168015622072

Copeland, B. J. (2022). Artificial intelligence. In *Encyclopedia britannica*. https://www.britannica.com/technology/artificial-intelligence.

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare J, 6*(2), 94–98. https://doi.org/10.7861/futurehosp.6-2-94

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8). Article 8.

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, prolific, CloudResearch, Qualtrics, and SONA. *PLoS One, 18*(3), Article e0279720. https://doi.org/10.1371/journal.pone.0279720

Engström, E. (2022). AI's fast and furtive spread by infusion into technologies that are Already in use—a critical assessment. In A. Hanemaayer (Ed.), *Artificial intelligence and its discontents* (pp. 63–82). Springer International Publishing. https://doi.org/10.1007/978-3-030-88615-8_4.

Engström, E., Eriksson, K., Björnstjerna, M., & Strimling, P. (2023). Global variations in online privacy concerns across 57 countries. *Comp Human Behavior Reports, 9*, Article 100268. https://doi.org/10.1016/j.chbr.2023.100268

Engström, E., & Strimling, P. (2020). Deep learning diffusion by infusion into preexisting technologies – implications for users and society at large. *Technology in Society, 63*, Article 101396. https://doi.org/10.1016/j.techsoc.2020.101396

Felfernig, A., Polat-Erdeniz, S., Uran, C., Reiterer, S., Atas, M., Tran, T. N. T., Azzoni, P., Kiraly, C., & Dolui, K. (2019). An overview of recommender systems in the internet of things. *Journal of Intelligent Information Systems, 52*(2), 285–309. https://doi.org/10.1007/s10844-018-0530-7

Fishbein, M. (1980). A theory of reasoned action: Some applications and implications. *Nebraska Symposium on Motivation, 27*, 65–116.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1). Article 1.

Galaz, V., Centeno, M. A., Callahan, P. W., Causevic, A., Patterson, T., Brass, I., Baum, S., Farber, D., Fischer, J., Garcia, D., McPhearson, T., Jimenez, D., King, B., Larcey, P., & Levy, K. (2021). Artificial intelligence, systemic risks, and sustainability. *Technology in Society, 67*, Article 101741. https://doi.org/10.1016/j.techsoc.2021.101741

Gansser, O. A., & Reich, C. S. (2021). A new acceptance model for artificial intelligence with extensions to UTAUT2: An empirical study in three segments of application. *Technology in Society, 65*, Article 101535.

Geisen, E. (2022). Improve data quality by using a commitment request instead of attention checks. *Qualtrics blog*. https://www.qualtrics.com/blog/attention-checks-and-data-quality/. (Accessed 1 February 2024).

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*(1), 72–91. https://doi.org/10.1037/a0032138

Glass, R., & Li, S. (2010). Social influence and instant messaging adoption. *Journal of Computer Information Systems, 51*, 24–30.

Gomez-Uribe, C. A., & Hunt, N. (2016). The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems, 6*(4), 1–19. https://doi.org/10.1145/2843948

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of mechanical Turk samples: Data collection in a flat world. *Journal of Behavioral Decision Making, 26*(3), 213–224. https://doi.org/10.1002/bdm.1753

Google Play. (2023). Android apps on Google play. https://play.google.com/store/apps?hl=en&gl=US.

Ha, S., & Stoel, L. (2009). Consumer e-shopping acceptance: Antecedents in a technology acceptance model. *Journal of Business Research, 62*(5), 565–571. https://doi.org/10.1016/j.jbusres.2008.06.016

Hartzel, K. S., Marley, K. A., & Spangler, W. E. (2016). Online social network adoption: A cross-cultural study. *Journal of Computer Information Systems, 56*(2), 87–96. https://doi.org/10.1080/08874417.2016.1117367

Herath, H. M. K. K. M. B., & Mittal, M. (2022). Adoption of artificial intelligence in smart cities: A comprehensive review. *International Journal of Information Management Data Insights, 2*(1), Article 100076. https://doi.org/10.1016/j.jjimei.2022.100076

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jak, S., Oort, F. J., & Dolan, C. V. (2013). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(2), 265–282. https://doi.org/10.1080/10705511.2013.769392

Kar, A. K., Choudhary, S. K., & Singh, V. K. (2022). How can artificial intelligence impact sustainability: A systematic literature review. *Journal of Cleaner Production, 376*, Article 134120. https://doi.org/10.1016/j.jclepro.2022.134120

Li, L., Peng, W., & Rheu, M. M. J. (2024). Factors predicting intentions of adoption and continued use of artificial intelligence chatbots for mental health: Examining the role of UTAUT model, stigma, privacy concerns, and artificial intelligence hesitancy. *Telemedicine and E-Health, 30*(3), 722–730. https://doi.org/10.1089/tmj.2023.0313

Madigan, R., Louw, T., Wilbrink, M., Schieben, A., & Merat, N. (2017). What influences the decision to use automated public transport? Using UTAUT to understand public acceptance of automated road transport systems. *Transportation Research Part F: Traffic Psychology and Behaviour, 50*, 55–64. https://doi.org/10.1016/j.trf.2017.07.007

Maeda, H. (2015). Response option configuration of online administered Likert scales. *International Journal of Social Research Methodology, 18*(1), 15–26. https://doi.org/10.1080/13645579.2014.885159

Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *AI & Society, 35*(4), 957–967. https://doi.org/10.1007/s00146-020-00950-y

Milberg, S. J., Smith, H. J., & Burke, S. J. (2000). Information privacy: Corporate management and national regulation. *Organization Science, 11*(1), 35–57. https://doi.org/10.1287/orsc.11.1.35.12567

Monzer, C., Moeller, J., Helberger, N., & Eskens, S. (2020). User perspectives on the news personalisation process: Agency, trust and utility as building blocks. *Digital Journalism, 8*(9), 1142–1162. https://doi.org/10.1080/21670811.2020.1773291

Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research, 2*(3), 192–222. https://doi.org/10.1287/isre.2.3.192

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science, 2*(2), 109–138. https://doi.org/10.1017/XPS.2015.19

Nikou, S. (2019). Factors driving the adoption of smart home technology: An empirical assessment. *Telematics and Informatics, 45*, Article 101283. https://doi.org/10.1016/j.tele.2019.101283

Nordhoff, S., Louw, T., Innamaa, S., Lehtonen, E., Beuster, A., Torrao, G., Bjorvatn, A., Kessel, T., Malin, F., Happee, R., & Merat, N. (2020). Using the UTAUT2 model to explain public acceptance of conditionally automated (L3) cars: A questionnaire study among 9,118 car drivers from eight European countries. *Transportation Research Part F: Traffic Psychology and Behaviour, 74*, 280–297. https://doi.org/10.1016/j.trf.2020.07.015

Nunnally, J. C. (1975). Psychometric theory—25 Years ago and now. *Educational Researcher, 4*(10), 7–21. https://doi.org/10.3102/0013189X004010007

Oechslein, O., Fleischmann, M., & Hess, T. (2014). An application of UTAUT2 on social recommender systems: Incorporating social information for performance expectancy. In *2014 47th Hawaii international conference on system sciences* (pp. 3297–3306). https://doi.org/10.1109/HICSS.2014.409

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy* (1st ed.). Crown.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods, 54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*(1), 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Qualtrics Survey Software. (2023). Qualtrics first release: 2005 copyright year 2023 location: Provo, Utah, USA. *Qualtrics* (February 2023) [Computer software] https://www.qualtrics.com.

R Core Team. (2023). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing [Computer software] https://www.R-project.org/.

Rice, S., & Winter, S. (2020). A practical guide for using electronic surveys in aviation research: Best practices explained. *International Journal of Aviation, Aeronautics, and Aerospace*. https://doi.org/10.15394/ijaaa.2020.1457

Rogers, E. (1962). *Diffusion of innovations* (1st ed.). Free Press of Glencoe.

Rogers, E. M. (1983). *Diffusion of innovations* (3rd ed.). Free Press ; Collier Macmillan.

Rogers, E. (2003). *Diffusion of innovations* (5th ed.). The Free Press.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards artificial intelligence scale. *Comp Human Behavior Reports, 1*, Article 100014.

Schomakers, E.-M., Lidynia, C., Vervier, L. S., Calero Valdez, A., & Ziefle, M. (2022). Applying an extended UTAUT2 model to explain user acceptance of lifestyle and therapy mobile health apps: Survey study. *JMIR mHealth and uHealth, 10*(1), Article e27095. https://doi.org/10.2196/27095

Sloman, S. A., & Lagnado, D. (2005). The problem of induction. *The cambridge handbook of thinking and reasoning* (pp. 95–116).

Spotify Technology, S. A. (2018). Form F-1 registration statement under the securities Act of 1933. *As filed with the Securities and Exchange Commission*. February 28, 2018. https://www.sec.gov/Archives/edgar/data/1639920/000119312518063434/d494294df1.htm#rom494294_4

Suhail, F., Adel, M., Al-Emran, M., & AlQudah, A. A. (2024). Are students ready for robots in higher education? Examining the adoption of robots by integrating UTAUT2 and TTF using a hybrid SEM-ANN approach. *Technology in Society, 77*, Article 102524. https://doi.org/10.1016/j.techsoc.2024.102524

Svetlova, E. (2022). AI ethics and systemic risks in finance. *AI and Ethics, 2*(4), 713–725. https://doi.org/10.1007/s43681-021-00129-1

Tamilmani, K., Rana, N. P., & Dwivedi, Y. K. (2021). Consumer acceptance and use of information technology: A meta-analytic evaluation of UTAUT2. *Information Systems Frontiers, 23*(4), 987–1005. https://doi.org/10.1007/s10796-020-10007-6

Tang, J., Birrell, E., & Lerner, A. (2022). *Replication: How well do my results generalize now? The external validity of online privacy and security surveys*. https://www.usenix.org/conference/soups2022/presentation/tang.

Thurman, N., Moeller, J., Helberger, N., & Trilling, D. (2019). My friends, editors, algorithms, and I: Examining audience attitudes to news selection. *Digital Journalism, 7*(4), 447–469. https://doi.org/10.1080/21670811.2018.1493936

Trepte, S., Reinecke, L., Ellison, N. B., Quiring, O., Yao, M. Z., & Ziegele, M. (2017). A cross-cultural perspective on the privacy calculus. *Social Media + Society, 3*(1), Article 2056305116688035. https://doi.org/10.1177/2056305116688035

United States Census Bureau. (2023). America is getting older. *Press Release Number CB23-106*. June 22, 2023 https://www.census.gov/newsroom/press-releases/2023/population-estimates-characteristics.html.

United States Census Bureau. (2024). QuickFacts United States. https://www.census.gov/quickfacts/fact/table/US/SEX255222.

Venkatesh, V. (2022). Adoption and use of AI tools: A research agenda grounded in UTAUT. *Annals of Operations Research, 308*(1–2), 641–652. https://doi.org/10.1007/s10479-020-03918-9

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 425–478.

Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly, 36*(1), 157–178.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications, 11*(1), 233. https://doi.org/10.1038/s41467-019-14108-y

Weinberg, J., Freese, J., & McElhattan, D. (2014). Comparing data characteristics and results of an online factorial survey between a population-based and a crowdsource-recruited sample. *Sociological Science, 1*, 292–310. https://doi.org/10.15195/v1.a19

Yang, L., Bian, Y., Zhao, X., Liu, X., & Yao, X. (2021). Drivers' acceptance of mobile navigation applications: An extended technology acceptance model considering drivers' sense of direction, navigation application affinity and distraction perception. *International Journal of Human-Computer Studies, 145*, Article 102507.

Yilmaz, F. G. K., Yilmaz, R., & Ceylan, M. (2023). Generative artificial intelligence acceptance scale: A validity and reliability study. *International Journal of Human-Computer Interaction*, 1–13. https://doi.org/10.1080/10447318.2023.2288730

Zanker, M., Rook, L., & Jannach, D. (2019). Measuring the impact of online personalisation: Past, present and future. *International Journal of Human-Computer Studies, 131*, 160–168. https://doi.org/10.1016/j.ijhcs.2019.06.006

Zhang, Q., Lu, J., & Jin, Y. (2021). Artificial intelligence in recommender systems. *Complex & Intelligent Systems, 7*(1), 439–457. https://doi.org/10.1007/s40747-020-00212-w