Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

What Moderates Generative AI Acceptance and Use? Disentangling Subscription Tier from Prior Usage Experience

Antun Biloš * and Bruno Budimir

Faculty of Economics and Business in Osijek, Josip Juraj Strossmayer University of Osijek, Trg Ljudevita Gaja 7, 31000 Osijek, Croatia; bruno.budimir@efos.hr
* Correspondence: antun.bilos@efos.hr

Abstract

Research on generative AI acceptance increasingly identifies new moderators, notably subscription tier, alongside the classic moderators of gender, age, and prior usage experience. Prior studies rarely ask whether these factors condition acceptance once their mutual confounding is addressed, and comparisons of paid and free users routinely contrast groups that also differ in prior usage experience. This study examines which factors moderate users' acceptance and use of generative AI. Drawing on an extended UTAUT2 model with personal innovativeness and trust, and a use construct capturing reported duration, frequency, and intensity of use, survey data from 765 Croatian users were analyzed using PLS-SEM. Subscription tier, gender, age, and prior usage experience were examined as candidate moderators, with subscription tier and prior usage experience entered jointly as interaction terms on the full sample. Habit showed the strongest association with both behavioral intention and use behavior. The subscription tier initially appeared to moderate the effect of behavioral intention on use behavior, but that difference did not persist once prior usage experience was modeled alongside it. Gender and age produced no robust differences. Prior usage experience was the only candidate whose moderation survived joint estimation, control variables, and adjustment for multiple testing: trust was strongly associated with behavioral intention among newer users and much less so among experienced ones. Trustworthiness signals therefore serve newer users, while sustained use among experienced users rests on habit.

Keywords: generative AI; UTAUT2; prior usage experience; subscription tier; moderation; habit; trust; PLS-SEM; personal innovativeness

Check for updates

Academic Editor: Mirjana Pejić Bach

Received: 1 July 2026
Revised: 24 August 2026
Accepted: 27 August 2026
Published: 4 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Generative AI tools such as ChatGPT have moved from novelty to routine within a remarkably short period, reshaping how people search for information, write, and perform everyday tasks. Understanding why individuals accept and continue to use these tools has become a central question for information systems research and for the firms that distribute them in consumer markets. Much of this work relies on the unified theory of acceptance and use of technology in its consumer form, UTAUT2, which has proven adaptable across a wide range of digital products [1,2] and has been applied extensively to generative AI [3–5].

A defining feature of this approach is its attention to moderators, the individual characteristics that condition how strongly the core determinants translate into behavioral intention and use behavior. UTAUT2 originally specified gender, age, and prior usage experience as moderators [2], and these have been carried into much of the generative

AI literature largely by convention. More recently, researchers have begun to introduce moderators tied to the commercial structure of AI products, most prominently subscription tier, the distinction between users on free and paid plans. The motivation is intuitive, since paying for a service can deepen commitment and alter behavior [6], and willingness to pay for tools such as ChatGPT Plus has emerged as an interesting subject of study [7]. These candidates do not carry equal weight in what follows. Subscription tier and prior usage experience form the core pair, because they are entangled with one another and because separating them is the contribution this study sets out to make. Gender and age are retained as supplementary candidates, and the question they address is whether the moderating effects that UTAUT2 attributes to them still operate in a technology that has widely diffused among consumers and in a sample composed entirely of its users.

The introduction of new moderators has exceeded the level of scrutiny regarding whether they structure anything once their mutual entanglement is genuinely considered. Subscription tier is the clearest case. Paid users are rarely a random segment of the user base, and they tend to be more experienced, more committed, and more frequent users from the outset [8]. Studies that compare paid and free users are therefore comparing groups that also differ in prior usage experience, and a difference attributed to payment cannot be separated from a difference in experience unless the two are modeled together. The standard demographic moderators raise a related concern. Gender and age shape adoption in many settings, but their influence often weakens in consumer technologies that have already diffused widely [9], and in samples composed entirely of users the relevant variation may be modest. Taken together, these observations point to a gap that is methodological as much as substantive. The field needs a way to distinguish moderators that structure acceptance from those that only appear once the confounding among them is addressed.

This study addresses that gap directly. We model the acceptance and use of generative AI with an extended UTAUT2 specification that adds personal innovativeness and trust, and we measure use through reported duration, frequency, and intensity rather than as a self-rated attitude. Using survey data from 765 generative AI users in Croatia, we treat subscription tier, gender, age, and prior usage experience as 4 candidate moderators. The primary evidence comes from models in which subscription tier and prior usage experience are entered jointly, together with their interactions with the paths at issue, so that each is estimated in the presence of the other and on the full sample. Group comparisons across the 4 candidates, preceded by measurement invariance testing, are reported alongside these models as descriptive support, as are two restricted subsamples in which one variable is held approximately constant while the other varies. Because the candidates are correlated, none of these approaches settles the question on its own, and the joint models are given precedence for the reasons set out in Section 3.5.

Three research questions guide the analysis. They are framed as questions rather than as directional hypotheses because part of the contribution is to test whether each candidate structures acceptance at all rather than to defend the claim that it does. The single exception concerns the relationship between trust and prior usage experience, for which a directional expectation is derived and stated in Section 2.5.

RQ1. *Does subscription tier moderate the acceptance and use of generative AI, and does any apparent moderation persist once differences in prior usage experience are accounted for?*

RQ2. *Do the standard demographic moderators of gender and age structure the acceptance and use of generative AI in a sample of established users?*

RQ3. *Does prior usage experience moderate the structural relationships in the model, and in particular, does the role of trust in forming behavioral intention differ across levels of prior usage experience?*

https://doi.org/10.3390/jtaer21090308

The analysis yields a narrower picture than the group comparisons alone would support. Habit shows the strongest association with both behavioral intention and use behavior, although it coexists with other significant determinants of intention. Of the 4 candidates, only prior usage experience conditions the model in a way that survives joint estimation, the addition of control variables, and adjustment for multiple testing, and it does so on the path from trust to behavioral intention. The apparent moderation by subscription tier of the path from intention to use does not persist once prior usage experience is modeled alongside it, while a weaker indication involving facilitating conditions remains unresolved. Gender and age structure little among the users sampled. The contribution is twofold. Methodologically, the study shows why entangled candidate moderators cannot be adjudicated by group comparison. Substantively, it locates the effect of prior usage experience on the deliberative rather than the habitual side of the model.

The remainder of the paper is organized as follows. Section 2 develops the theoretical background and situates the candidate moderators within the literature. Section 3 describes the sample, the instrument, and the analytical strategy. Section 4 reports the measurement and structural models, predictive relevance, joint interaction models, multigroup comparisons, restricted subsamples, and a set of robustness checks. Section 5 discusses the findings, their theoretical and practical implications, and the limitations of the study. Section 6 concludes the manuscript.

2. Theoretical Background and Literature Review

2.1. The UTAUT2 Model and Generative AI Acceptance

Research on the adoption of consumer technologies has converged on the Unified Theory of Acceptance and Use of Technology in its consumer-oriented form, UTAUT2 [2]. The model synthesizes earlier acceptance theories into a compact set of determinants of behavioral intention and use: performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, price value, and habit. Its consumer orientation, and in particular the addition of hedonic motivation, price value, and habit to the original UTAUT, makes it suitable for discretionary technologies that people adopt outside any organizational setting [1,9].

Generative AI is a paradigmatic example of such a technology. Tools such as ChatGPT diffused into everyday personal use with unusual speed, and a growing body of work has applied UTAUT2 to explain why individuals accept and continue to use them [3–5]. Recent applications continue to place habit at the center of the picture. In a study of 611 higher education students, habit made the largest contribution to behavioral intention and was the principal predictor of use behavior [10], and a study of graduate students identified habit and behavioral intention as the most important predictors of actual use [11]. Within the e-commerce and digital services literature, UTAUT and UTAUT2 have likewise been used to study chatbots, conversational agents, and AI-enabled services [12–14], which establishes the model as an appropriate baseline for the present study.

Alternatives to UTAUT2 exist, and the most developed of them for the present domain is the Artificially Intelligent Device Use Acceptance Model, which suggests that the acceptance of intelligent agents involves dimensions beyond those captured by general acceptance models [15,16]. AIDUA describes acceptance as a staged appraisal in which social norms and the hedonic value of an interaction inform a subsequent appraisal of effort and performance, which in turn produces an emotional response that governs willingness to accept the agent. Its distinctive contribution is the explicit place it gives to affect and to the perception of machine agency. Several characteristics of generative AI are consistent with that emphasis. Conversational systems invite anthropomorphic responses, their outputs are produced by processes users cannot inspect and are therefore encountered under

uncertainty, they are increasingly used for creative rather than routine work, and users form impressions of the intelligence of the system that have no clear counterpart in the acceptance of conventional software.

UTAUT2 is nevertheless retained here, for three reasons. First, the question motivating this study concerns the moderation of well-established acceptance relationships rather than the discovery of new AI-specific constructs, and UTAUT2 provides the canonical structure within which moderation has traditionally been examined, including the very moderators whose status is at issue. AIDUA is a model of appraisal sequence rather than of moderation, and it carries no comparable apparatus of boundary conditions. Second, retaining an established model keeps the analysis comparable with the large body of generative AI acceptance research built on UTAUT2 [9,17]. Third, the two extensions adopted here address the aspect of machine agency most directly relevant to the research questions, since trust captures the relational dimension of relying on a system whose behavior cannot be fully predicted. AI-specific affective constructs such as aversion and anxiety, and evaluative constructs such as perceptions of algorithmic bias or of effects on employment and society, are deliberately left outside the model rather than partially incorporated. Their exclusion is a boundary of this study and is revisited in Section 5.7.

2.2. Extending the Model with Personal Innovativeness and Trust

Two extensions are added to the standard UTAUT2 predictors of behavioral intention. The first is personal innovativeness, the degree to which an individual is willing to try out new information technologies [18]. Innovative individuals adopt new tools earlier and engage with them more readily, and personal innovativeness has been shown to predict the intention to use generative AI in particular [3,17]. Its inclusion is especially relevant for a technology as novel as generative AI, where early adoption is driven in part by a tendency toward experimentation. The second extension is trust. Generative AI tools produce outputs whose accuracy users often cannot verify directly, which makes reliance on them an act conducted under uncertainty. Research on trust in automation and in AI shows that trust affects whether people are willing to depend on a system whose behavior they cannot fully predict [19,20], and recent work extends this reasoning to conversational AI [21,22]. In the present model, trust is conceived as a determinant of behavioral intention, since users form an intention to continue using generative AI in part because they perceive its outputs as reliable, consistent, and secure. The theoretical interest of trust lies less in its average effect than in how that effect may depend on prior usage experience, a point elaborated in Section 2.5.

2.3. Moderators in UTAUT2 and Their Reach Among Users

A defining feature of the UTAUT tradition is its treatment of moderators. The original specifications proposed that gender, age, and prior usage experience condition the strength of the core relationships, so that the same determinant matters more for some users than for others [1,2]. These three have been carried into much of the subsequent literature, often by convention rather than by deliberate test. They do not stand in the same relation to the present study. Prior usage experience belongs to the core pair, alongside subscription tier, because the two are entangled and because separating them is the object of the analysis. Gender and age are examined as supplementary candidates, and the question they address is how far the classic boundary conditions of UTAUT2 extend into a technology that has already diffused widely among consumers.

There are reasons to expect the influence of the demographic moderators to be weaker in the present setting than in the contexts where the model was developed. They were theorized and validated largely in organizational settings and at earlier stages of technology

diffusion, where access and exposure varied substantially across groups. In a widely diffused consumer technology, and in a sample composed entirely of people who already use the technology, much of the between-group variation that demographic moderators exploit is compressed [9]. Recent evidence on generative AI is mixed on this point. A UTAUT2 study of graduate students reports that age, gender, and level of study exercise no moderating influence [11], while a study of 1190 students using an extended UTAUT model finds that gender and year of study do condition acceptance [23]. The expectation that gender and age structure little among users is therefore a question to be tested rather than a hypothesis to be defended, and it motivates their explicit examination as candidate moderators rather than their assumed inclusion.

2.4. Subscription Tier as a Candidate Moderator

Alongside the standard moderators, the commercialization of generative AI has prompted interest in a moderator tied to the business model of the product: subscription tier, the distinction between users on free and paid plans. The subscription tier is defined here as a proxy for financial commitment and the level of access. It is conceptually distinct from prior usage experience, which reflects behavioral engagement rather than what a user pays for. The two are nevertheless empirically entangled, and the case for treating subscription tier as a moderator can be stated in three steps whose competing implications make it a candidate for testing rather than a hypothesis to be defended.

First, paying for a service is not behaviorally neutral. Evidence from digital markets indicates that payment can deepen engagement and alter subsequent behavior, which makes subscription a potential conditioning variable [6]. The freemium literature similarly treats the move from free to paid use as a meaningful behavioral threshold [24], and willingness to pay for tools such as ChatGPT Plus has become an interesting research topic [7].

Second, paid status is conceptually difficult to isolate. Users who choose to pay tend to be more committed and more experienced, so premium status is bound up with prior engagement rather than standing apart from it [8]. A difference observed between paid and free users may therefore reflect who pays rather than the act of paying. In the present sample, this entanglement is pronounced, since paid users differ from free users in experience, employment, education, and the number of tools they use.

Third, these two observations can be reconciled only empirically. If subscription tier structures acceptance on its own, differences between paid and free users should persist when prior usage experience is entered into the same model. If they do not persist, then what the comparison detected is a property of the users who pay rather than an effect of paying, and subscription tier is better described as a marker of an engagement profile than as a moderator in its own right. Which paths subscription tier would condition, if it conditioned any, is a further question. A theoretical case can be made for the path from facilitating conditions to use behavior, since a paid tier alters the resources and capabilities actually available to a user, and facilitating conditions is the construct that captures the perception of those resources. No comparable case can be made in advance for the path from behavioral intention to use behavior, and any moderation of that path is therefore treated as an exploratory finding rather than as a theoretically anticipated one. The subscription tier is introduced as one of several candidate moderators to be tested, and the analytical strategy described in Section 3.5 is designed to answer the question.

2.5. Prior Usage Experience and the Changing Role of Trust

Among the original UTAUT2 moderators, prior usage experience has the strongest theoretical claim to structure acceptance, because it indexes a process rather than a fixed

attribute. As users accumulate experience with a technology, the basis of their behavior changes. Early use is deliberate and is shaped by evaluative judgments about whether the tool is useful and whether it can be relied upon. With repetition, behavior becomes increasingly automatic and is carried by habit, the learned tendency to use a technology with little conscious deliberation [25–28]. Habit research describes this development as gradual rather than abrupt, at rates that vary considerably across individuals and behaviors [29], and work on technology habits notes that digital tools occupy an unusually wide range of cues and contexts through which such automation can develop [30].

This development has a specific implication for the role of trust. For new users, reliance on a generative AI tool is uncertain, and trust in the reliability of the tool is a meaningful determinant of the intention to continue using it. For experienced users, the deliberative weighing of trustworthiness becomes less decisive, for reasons on which two studies converge. Research on trust in automated systems distinguishes the trust a user brings to a system from the trust learned through accumulated interaction, and suggests that the learned component stabilizes as evidence about performance accumulates [31], which is consistent with the calibration view of trust in automation [19,20]. Continuance research arrives at the same place by another route, suggesting that continued use is governed by the confirmation of expectations and the satisfaction that follows from it rather than by beliefs formed before use [32], that trust acts on continuance largely through experienced performance and satisfaction rather than directly [33], and that the beliefs underlying usage decisions are re-weighted as usage accumulates [34]. Recent work on researchers using generative AI reports a similar pattern, with the relationship between trust and usage intention weakening as familiarity increases [35]. The contribution of trust to behavioral intention is therefore expected to weaken as prior usage experience grows, and this is the one-directional expectation stated in the present study.

A parallel implication concerns the link between behavioral intention and use behavior. If behavior comes to follow established routine rather than current deliberation, the translation of intention into use should also weaken as prior usage experience accumulates [25]. The implication is stated here because it follows from the same reasoning, and it is tested alongside the expectation concerning trust in Section 4.5. The same reasoning applies to the habit paths themselves, which are examined in the robustness checks reported in Section 4.8. Accordingly, habit enters the present model in a single capacity, as a direct determinant of both behavioral intention and use behavior, consistent with its position in UTAUT2 and in habit research. The maturation process is indexed by prior usage experience, which is the moderator, and habit is one of the determinants whose relative weight that process is expected to alter. Framing prior usage experience in this way places it not as one demographic moderator among others but as the variable most likely to restructure the acceptance process, and it is the focus of the third research question.

3. Materials and Methods

3.1. Research Design and Data Collection

The study draws on a cross-sectional survey of generative AI users in Croatia. Data were collected through a highly structured questionnaire administered online using the computer-assisted web interviewing (CAWI) method on the Alchemer platform (Alchemer LLC, Louisville, CO, USA). The questionnaire was developed in Croatian and English, and the Croatian version is the basis for the present analysis. Before the main fielding, the instrument was piloted with 23 respondents drawn from the target population to check the clarity of items and the framing of the constructs, and several Alchemer data-quality features were enabled, including randomized presentation order of response options and enforced response formatting.

Data were collected in March 2025. Respondents in Croatia were recruited by the research team. The primary technique was snowball or chain-referral sampling, in which the survey was distributed by direct email and private messages to close contacts who were asked to forward it within their networks, supplemented by trained student fieldworkers who distributed the questionnaire to the target segment, and by the authors' own personal and professional contacts. The recruitment therefore combined snowball, convenience, and to some degree purposive sampling [17]. A screening question retained only respondents who had used generative AI tools within the preceding two years.

A total of 1005 respondents accessed the questionnaire. Of these, 767 completed it, 129 were disqualified because they reported never having used generative AI tools, and 109 submitted partial responses, including 43 that the platform quarantined for indicators of poor response quality such as extremely short completion times or inconsistent answering. The 767 completed responses constituted the valid sample. Two further cases were removed before analysis: one with the entire habit block missing and one with the subscription variable missing. The final analytic sample is 765 respondents.

The study utilizes a non-probability sample and is skewed toward young, female, and student respondents. Women make up 56.7% of respondents, 44.6% are aged 18 to 24, and the median age is about 27 years. Roughly half are students or combine study with work, and about half live in a large city. A majority, 61%, report more than 12 months of experience with generative AI tools. This composition reflects the recruitment strategy and is the principal limitation of the study. The sample supports the analysis of relationships among acceptance constructs within a population of users, but it does not support claims about the prevalence of generative AI use in the wider population, and the findings should be read with that limitation in mind.

3.2. Measures

The questionnaire consisted of 46 items, with 8 capturing demographic and usage background, and 38 measuring determinants of acceptance and use. All reflective constructs were measured on 7-point Likert scales anchored from complete disagreement to complete agreement, following the original UTAUT2 instrument for comparability [2]. Performance expectancy, effort expectancy, facilitating conditions, habit, and trust were each measured with four items, and social influence, hedonic motivation, price value, personal innovativeness, and behavioral intention with three. Item wordings were adapted to the generative AI context from established scales. Personal innovativeness was adapted from Agarwal and Prasad [18] using the formulation by Gansser and Reich [17], and trust was operationalized with items reflecting reliance on the tool when its content is perceived as reliable, accurate, expectation-consistent, and secure, adapted from Mohd Rahim et al. [36]. The full instrument, with item wordings, response scales, and sources, is reported in Appendix A, Table A1.

Use was measured through reported usage behavior rather than as a self-rated attitude. Following Venkatesh et al. [37], use was captured through three indicators: duration, measured as the average hours of use per week; frequency, measured as how often the tools are used; and intensity, measured as the respondent's assessment of the extent of current use. Each indicator was recorded on a seven-point scale with context-adapted anchors. This behavioral operationalization is adopted because earlier research efforts found that an agreement-based use measure was empirically difficult to separate from behavioral intention [38].

3.3. Grouping Variables and the Subscription Confound

Four grouping variables define the candidate moderators. Subscription tier distinguishes paid users (n = 182, coded 1) from free users (n = 583, coded 0). Gender distinguishes men (n = 331) from women (n = 434). Age is split at the sample median into a younger group (27 and under, n = 390) and an older group (over 27, n = 375). Experience distinguishes users with more than 12 months of use (n = 469) from those with 12 months or less (n = 296).

The 4 candidate moderators do not carry equal theoretical weight. Subscription tier and prior usage experience form the core pair whose disentanglement motivates the study, since the two are empirically confounded and the first is frequently introduced into generative AI research without controlling for the second. Gender and age are retained because UTAUT2 specifies them, and they serve to test whether the standard demographic moderators still operate in a widely diffused consumer technology examined within a population of established users. The analysis therefore treats the first pair as the substantive question and the second pair as a boundary condition of the original model.

Two of the grouping variables are dichotomized, and the cut points require justification. Age is split at the sample median, which maximizes the size of both groups but discards information about the underlying continuum. Experience is split at 12 months, and this cut is not one choice among many. The 5 response categories are heavily unbalanced: the 4 categories below 12 months hold 134 respondents in total against 469 in the highest category, so no lower cut yields groups of comparable size. Twelve months is also the only boundary at which the sample supports a comparison of adjacent categories, contrasting 6 to 12 months (n = 162) with more than 12 months (n = 469). Both dichotomizations were examined for artifacts. Age was re-specified as a continuous moderator in an interaction model, and the experience boundary was tested through the adjacent-category comparison; both checks are reported in Section 4.8. The loss of information that dichotomization entails is acknowledged in Section 5.7.

The subscription variable deserves particular attention because paid status is not randomly distributed across the sample. Paid users are disproportionately male, more experienced, more often employed, more highly educated, and they report using roughly twice as many distinct tools as free users do on average, 4.66 versus 2.25. This systematic profile means that any difference observed between paid and free users is entangled with differences in experience and engagement, which motivates the disentangling design described in Section 3.5.

Because the disentangling design compares subgroups that were not randomly assigned, both subsamples were also examined for equivalence on observed characteristics. Within the experienced group, paid and free users differ significantly on every characteristic examined. Within the free group, experienced and inexperienced users do not differ in gender or settlement size but do differ in the remaining characteristics, most notably in age. The full comparison, with test statistics, exact p-values, and effect sizes, is reported in Table S9, and its implications for interpretation are discussed in Sections 4.7 and 5.7.

3.4. Model Specification

The model is an extended UTAUT2 specification estimated with all constructs treated as reflective. Behavioral intention is modeled based on nine predictors: performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, price value, habit, personal innovativeness, and trust. Use is modeled based on three predictors: facilitating conditions, habit, and behavioral intention. The two extensions to the standard UTAUT2 model, personal innovativeness and trust, were retained as predictors of behavioral intention on theoretical grounds developed in Section 2 (Figure 1).

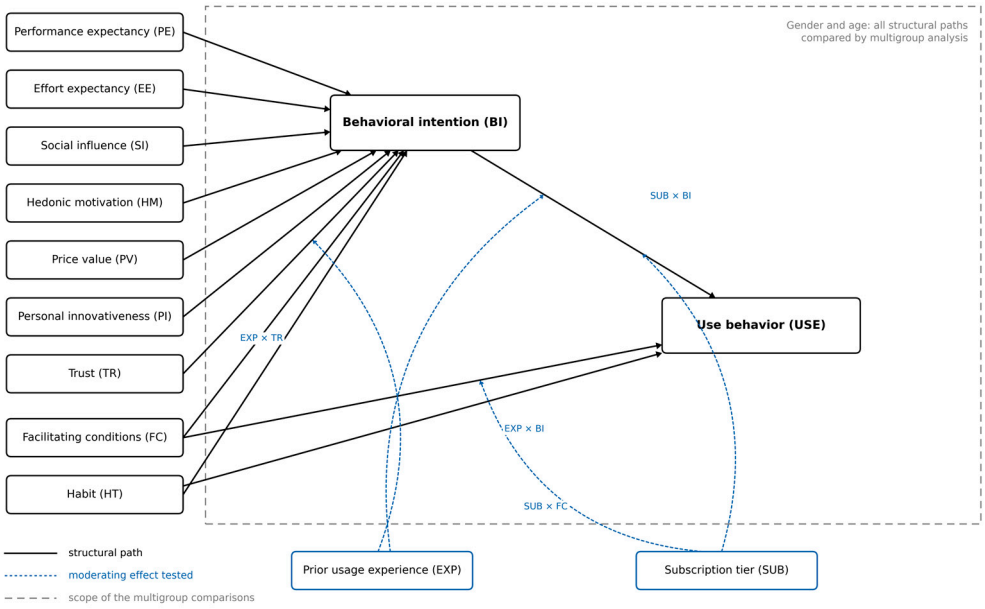


Figure 1. Extended UTAUT2 research model. *Note.* Solid arrows denote structural paths. Dotted blue arrows denote the moderating effects examined for prior usage experience (EXP) and subscription tier (SUB), each drawn onto the specific structural path it moderates. The dashed gray frame marks the scope of the multigroup comparisons, in which gender and age are examined across all structural paths. Full construct names are used throughout, with abbreviations given in parentheses for reference to the tables.

One measurement decision is carried over from the related research for consistency. The fourth facilitating-conditions item (FC4) was retained despite a comparatively low loading, in line with the original UTAUT2 instrument and with the related study on the same construct. A robustness check reported in Section 4.4 confirms that retaining or removing this item leaves the structural estimates unchanged. The same model was specified identically across the full sample and the two analytical subsamples.

Alongside this specification, the same model was estimated as a set of joint interaction models on the full sample. This second specification addresses three limitations of group comparison: splitting the sample reduces statistical power, comparison across groups assumes measurement invariance, and a moderator can enter only as a dichotomy. In the joint models, prior usage experience enters as a quasi-continuous variable coded from 1 to 5 following the 5 response categories; subscription tier enters as a dummy variable coded 0 for free and 1 for paid users; and both are included simultaneously, along with their interactions, in the structural paths. Interaction terms were formed using the two-stage approach [39]. Three models are reported in the main text. M0 contains the main effects of both moderators without interactions. M1 adds 4 interaction terms: prior usage experience with trust on behavioral intention, prior usage experience with behavioral intention on use behavior, subscription tier with behavioral intention on use behavior, and subscription tier with facilitating conditions on use behavior. M3 adds 5 control variables, namely age, education, employment status, gender, and the number of reported AI tools used, each specified as a predictor of both endogenous constructs. Two further specifications are estimated as robustness checks: a variant of M1 with a binary experience indicator (M1b), used in Section 4.5 to express the interaction in conditional form, and an extension with two habit interactions (M2), reported in Section 4.8.

One reporting convention follows from this specification. Because subscription tier is a binary indicator, it is not standardized during estimation, and coefficients attached to it are partially standardized, which makes them incomparable with the remaining

coefficients. Such coefficients are reported in fully standardized form throughout, obtained by multiplying by the standard deviation of the indicator (0.426); the partially standardized values are given in the note to the corresponding table.

3.5. Analytical Strategy

The model was estimated using PLS-SEM in SmartPLS 4 (version 4.1.1.8) [40]. This approach was selected because the study is oriented toward prediction and the explanation of target constructs in a complex model with many predictors, a setting for which the method is appropriate and which does not require distributional assumptions [39]. Significance of path coefficients was evaluated using bootstrapping with 10,000 subsamples and bias-corrected and accelerated confidence intervals.

Internal consistency reliability was evaluated with Cronbach's alpha, composite reliability ρC, and ρA; convergent validity with average variance extracted and factor loadings; and discriminant validity with the heterotrait–monotrait ratio of correlations with bootstrap confidence intervals [41]. Collinearity among predictors was assessed using variance inflation factors (VIFs) against the conservative threshold of 3.3 [42], which also serves as a full-collinearity test for common method bias. As a complementary check, Harman's single-factor test was conducted by extracting a single unrotated component from all 38 measurement items. The structural model was assessed through the coefficient of determination, effect sizes, and out-of-sample predictive relevance using PLSpredict with 10 folds and 10 repetitions, benchmarked against a linear-model baseline following the procedure of Shmueli et al. [43].

The moderation analysis proceeded in three stages. First, each of the 4 candidate moderators was tested through measurement invariance assessment followed by multigroup analysis. Measurement invariance was established using the measurement invariance of composite models (MICOM) procedure [44], and group differences in path coefficients were tested using permutation-based multigroup analysis with 5000 permutations, a two-tailed test, and a fixed random seed. This stage identifies which candidate moderators are associated with differences in the structural relationships.

Second, and central to the design, two disentangling subsamples were constructed to separate the effect of subscription from that of experience, since the two are confounded in the full sample. The first subsample retains only experienced users and contrasts paid and free users within that group (n = 469, of whom 330 are free and 139 are paid). If a difference attributed to subscription in the full sample reflects experience rather than payment, it should diminish once experience is held roughly constant. The second subsample retains only free users and contrasts experienced and inexperienced users within that group (n = 583, of whom 253 are inexperienced and 330 are experienced). If a difference attributed to experience is genuine rather than an artifact of subscription, it should persist among free users, where payment is held constant. Examining an apparent moderator under the control of the variable most likely to confound it provides a check on whether an apparent moderator survives that control, although restriction constrains the confounding variable only approximately and does not equate the groups on other characteristics (Figure 2).

Third, the joint interaction models described in Section 3.4 were estimated on the full sample. Because these models retain all 765 respondents, require no assumption of measurement invariance across groups, treat prior usage experience as a quasi-continuous variable, and test a single coefficient rather than a set of 12, they carry the primary evidential weight in the moderation analysis. The multigroup comparisons and the disentangling subsamples are reported as descriptive support and as internal sensitivity checks.

Measurement invariance was assessed with the MICOM procedure for the 4 candidate moderators and, in addition, for the two disentangling subsamples, so that the comparisons

within those subsamples rest on the same evidentiary basis as those in the full sample. Output for all three MICOM steps is reported in Table S8.

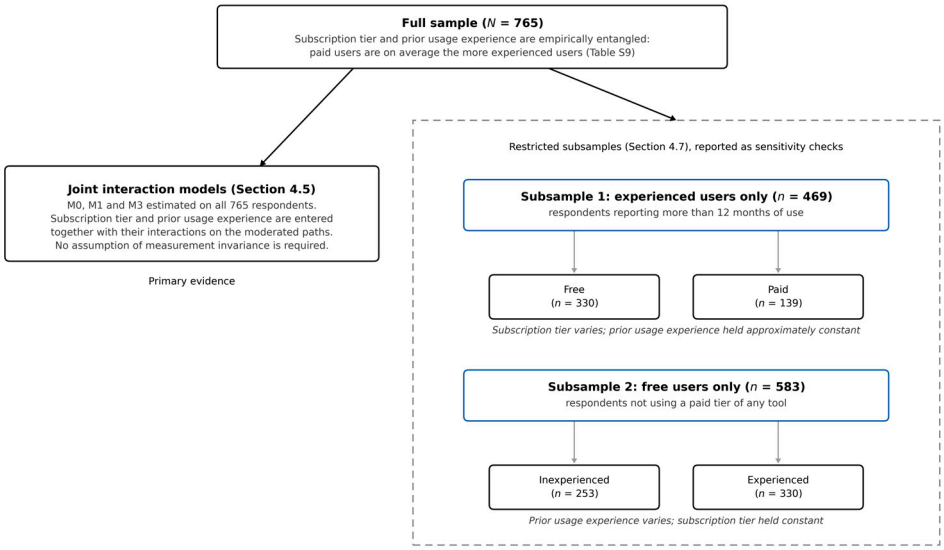


Figure 2. Design of the moderation analysis. *Note*. Restriction does not produce groups that are equivalent on other observed characteristics. The composition of both subsamples is compared in Table S9.

Because the multigroup stage compares 12 structural paths within each analysis, the number of tests is substantial and the risk of inflated Type I error is real. Two measures address it. First, all comparisons are reported rather than only those reaching significance, with the complete output in Table S7. Second, *p*-values are adjusted using the Benjamini–Hochberg procedure, which controls the false discovery rate. The adjustment is applied symmetrically to every moderator, including subscription tier, and not only to gender and age. In the primary reporting, each analysis constitutes a family of 12 tests, corresponding to the 12 structural paths compared within it. A stricter alternative, in which all six analyses form a single family of 72 tests, is reported in Table S7 as a sensitivity check. The adjacent-category comparison of experience described in Section 3.3 was conducted after these analyses and is adjusted within its own family of 12 paths rather than being folded into the 72.

The adjustment is embedded in a two-level reporting framework. The relationship between trust and prior usage experience was specified in advance as RQ3, and as a single pre-specified expectation it is also reported without adjustment, with the joint interaction model rather than the group comparison serving as its primary confirmation. All remaining comparisons are treated as exploratory and are reported with adjusted values throughout. The permutation procedure also yields tests of between-group differences in effect size; these are not adjusted and are not used to support any conclusion in this study.

Generative AI tools were used to support the preparation of this manuscript. Their role was limited to language and drafting assistance, as detailed in the Acknowledgments. The authors conducted data collection, PLS-SEM analysis, and interpretation of the results.

4. Results

4.1. Measurement Model

The reflective measurement model was assessed for reliability, convergent validity, and discriminant validity on the full sample (N = 765). Factor loadings exceeded the 0.70 threshold for all items with a single exception, the fourth facilitating-conditions item (FC4), which loaded at 0.503 and was retained for the reasons given in Section 3.4. The three

behavioral use indicators loaded well (duration 0.790, frequency 0.883, intensity 0.906), and the lowest loading within the habit construct was 0.776. Internal consistency reliability and convergent validity were satisfactory across all constructs, with Cronbach's alpha from 0.782 upward, ρ_A from 0.843 upward, ρ_C from 0.853 upward, and average variance extracted from 0.600 upward (Table 1). Factor loadings for all 38 items are reported in Table S1 of the Supplementary Materials.

Table 1. Construct reliability and convergent validity (N = 765).

Construct	Items	Cronbach's α	ρ_A	ρ_C	AVE
Performance expectancy (PE)	4	0.851	0.854	0.899	0.691
Effort expectancy (EE)	4	0.890	0.890	0.924	0.752
Social influence (SI)	3	0.919	0.920	0.949	0.861
Facilitating conditions (FC)	4	0.782	0.843	0.853	0.600
Hedonic motivation (HM)	3	0.900	0.901	0.938	0.834
Price value (PV)	3	0.882	0.888	0.927	0.809
Habit (HT)	4	0.864	0.892	0.906	0.707
Personal innovativeness (PI)	3	0.882	0.883	0.927	0.809
Trust (TR)	4	0.908	0.909	0.935	0.783
Behavioral intention (BI)	3	0.852	0.858	0.911	0.773
Use Behavior (USE)	3	0.827	0.854	0.896	0.742

Note. α = Cronbach's alpha; ρ_A = reliability coefficient; ρ_C = composite reliability, reported as CR in earlier PLS-SEM literature; AVE = average variance extracted. All constructs meet the recommended thresholds (α, ρ_A, and $\rho_C \geq 0.70$; AVE ≥ 0.50).

Discriminant validity was assessed using the heterotrait–monotrait ratio of correlations (HTMT) with bootstrap confidence intervals, and cross-checked against the Fornell-Larcker criterion. All point estimates fell below the conservative 0.85 threshold. The highest ratio was between facilitating conditions and effort expectancy (0.802), followed by use behavior and habit (0.766, 95% CI [0.716, 0.810]). For the facilitating conditions and effort expectancy pair, the upper bound of the confidence interval reached 0.855, marginally above 0.85 but well below the 0.90 threshold applied to conceptually proximate constructs [41]. The two constructs are therefore retained as distinct, and the proximity is reported here rather than left to the Supplementary Materials. The Fornell–Larcker criterion was satisfied for all 11 constructs, and no indicator loaded more strongly on a construct other than the one to which it was assigned. The ratio between behavioral intention and use behavior was 0.701, which indicates that intention and behavior are empirically distinct constructs in these data. The full HTMT matrix with confidence intervals, the Fornell-Larcker matrix, the cross-loadings, and the complete outer and inner VIFs are reported in Tables S2–S5. The threshold of 3.3 applied above is the full-collinearity criterion for the structural model; the VIFs for the reflective indicators are assessed against the conventional threshold of 5, and the highest observed value was 3.543.

One measurement decision warrants explicit comment. Use behavior is specified reflectively, which presumes that duration, frequency, and intensity share a common latent source rather than jointly composing the construct. The available evidence supports this specification. The three indicators are substantially intercorrelated (0.533 to 0.673), load at 0.790, 0.883, and 0.906, and yield an average variance extracted of 0.742. Confirmatory tetrad analysis, which would provide a direct empirical test of the specification, cannot be applied here because it requires at least 4 indicators per construct. The decision was therefore examined by re-estimating the model with use behavior specified formatively in Mode B, a check reported in Section 4.8. The indicators remain self-reported rather than objective behavioral records, a limitation taken up in Section 5.7.

4.2. Structural Model

The structural model was estimated using bootstrapping with 10,000 subsamples and bias-corrected and accelerated confidence intervals. Inner VIFs ranged from 1.22 to 2.20, well below the conservative threshold of 3.3, which indicates that collinearity is not a concern and, taken as a full-collinearity test, gives no indication of common method bias in the estimates [42]. Harman's single-factor test agrees: a single unrotated component extracted from all 38 measurement items accounts for 35.9% of the total variance, well below the 50% threshold, and 8 components have eigenvalues above one. Neither procedure is decisive on its own, and a single-source design cannot entirely rule out common method bias, as noted in Section 5.7. The model explained a substantial share of variance in both endogenous constructs, with R^2 of 0.641 for behavioral intention and 0.493 for use behavior (adjusted R^2 of 0.636 and 0.491, respectively). Path coefficients, bias-corrected and accelerated confidence intervals, and effect sizes are reported in Table 2.

Table 2. Structural model results (N = 765).

Path	β	95% BCa CI	f^2	p
Behavioral intention (R^2 = 0.641, adjusted R^2 = 0.636)				
HT → BI	0.373	[0.318, 0.427]	0.236	<0.001
HM → BI	0.193	[0.120, 0.270]	0.052	<0.001
TR → BI	0.182	[0.108, 0.260]	0.058	<0.001
PE → BI	0.153	[0.091, 0.220]	0.035	<0.001
PI → BI	0.111	[0.053, 0.168]	0.023	<0.001
SI → BI	0.068	[0.009, 0.129]	–	0.027
EE → BI	−0.050	[−0.124, 0.022]	–	n.s.
FC → BI	0.030	[−0.039, 0.103]	–	n.s.
PV → BI	0.023	[−0.030, 0.080]	–	n.s.
Use behavior (R^2 = 0.493, adjusted R^2 = 0.491)				
HT → USE	0.483	[0.412, 0.549]	0.255	<0.001
BI → USE	0.272	[0.205, 0.339]	0.075	<0.001
FC → USE	0.019	[−0.037, 0.071]	–	0.485

Note. β = standardized path coefficient; BCa CI = bias-corrected and accelerated bootstrap confidence interval from 10,000 subsamples; f^2 = effect size; p = significance from the same bootstrap procedure. A dash in the f^2 column indicates an effect size below the 0.02 relevance threshold; n.s. = not significant at $p < 0.05$. Confidence intervals that include zero correspond to the paths marked n.s. Construct abbreviations are defined in Table 1 and in the Abbreviations section.

Habit was the strongest determinant in the model. It was the leading predictor of behavioral intention (β = 0.373) and the leading predictor of use behavior (β = 0.483), with the largest effect size in each equation. Hedonic motivation, trust, performance expectancy, and personal innovativeness were further significant predictors of behavioral intention, and social influence reached significance with a small coefficient. Effort expectancy, facilitating conditions, and price value were not significant predictors of behavioral intention. In the use equation, behavioral intention was a significant predictor (β = 0.272) alongside habit, whereas facilitating conditions had no significant effect on use behavior (β = 0.019, p = 0.485).

4.3. Predictive Relevance

Out-of-sample predictive relevance was assessed using PLSpredict with 10 folds and 10 repetitions, benchmarked against a linear-model baseline following the procedure of Shmueli et al. [43]. The Q^2predict values were positive for both endogenous constructs (0.628 for behavioral intention and 0.470 for use behavior) and for all 6 indicators, ranging from 0.224 for the use duration indicator to 0.540 for one of the behavioral intention indicators. The model therefore predicts more accurately than the mean of the training sample.

The comparison against the linear-model benchmark is less favorable, and it is reported here in full. On root mean squared error, the PLS-SEM model produced a lower value than the benchmark for 1 of the 6 indicators, use duration, was effectively tied on a behavioral intention indicator, where the difference was 0.002, and was outperformed on the remaining 4. On mean absolute error, the model produced a lower value for use intensity alone and was outperformed on the other 5. The differences are small in absolute terms, ranging from 0.002 to 0.054, but the criterion applied here classifies a model that fails to outperform the linear benchmark on the majority of indicators as having low predictive power, and that is the classification these results warrant. The complete output, with RMSE, MAE, and Q^2predict for the PLS-SEM model, the linear-model benchmark, and the indicator-average benchmark, is reported in Table S6.

The model accounts for a substantial share of the variance in behavioral intention and use behavior within the sample, and the structural relationships reported below are estimated with that explanatory purpose in view, but its out-of-sample predictive performance does not exceed that of a simple linear benchmark. The moderation analyses that follow concern differences in structural relationships across groups and across levels of a moderator, and they do not depend on predictive accuracy. The limited predictive power is nonetheless relevant to any forecasting use of the model, and it is addressed again in Section 5.7.

4.4. Robustness of the Facilitating-Conditions Item

Because the fourth facilitating-conditions item (FC4) loaded below the conventional threshold, the model was re-estimated with that item removed to confirm that its retention did not affect the substantive results. The two specifications were almost identical. The R^2 values were unchanged for behavioral intention (0.641) and changed negligibly for use behavior (0.493 with the item, 0.494 without). Every structural coefficient was stable to the thousandth, with facilitating conditions remaining non-significant for both behavioral intention and use in each version. Removing the item raised the construct's average variance extracted from 0.600 to 0.722 and its Cronbach's alpha from 0.782 to 0.808, closing the gap between alpha and ρ_A. The item also carries a significant outer weight in the base model (0.101, $p = 0.005$, 95% CI [0.024, 0.165]), so its contribution to the construct is small but not null. The item was retained for consistency with the original UTAUT2 instrument and with the related study on the same construct, since its retention has no bearing on the structural estimates.

4.5. Moderation: Joint Interaction Models

The moderation analysis begins with the joint interaction models, which carry the primary evidential weight for the reasons set out in Section 3.5. Three models are reported in Table 3. M0 estimates the main effects of subscription tier and prior usage experience without interactions, M1 adds the 4 interaction terms, and M3 adds the 5 control variables. Each was estimated on the full sample of 765 respondents using bootstrapping with 10,000 subsamples.

Two features of the table require comment before the estimates are read. First, coefficients involving subscription tier are reported in fully standardized form for the reason given in Section 3.4. Second, and more consequentially, several coefficients in M1 and M3 are conditional rather than average effects. In a model containing an interaction term, the coefficient of a moderated path is its effect at the value zero of the moderator. Because subscription tier is coded zero for free users, the facilitating conditions to use and intention to use coefficients in M1 are the effects among free users rather than across the whole sample, which is why the facilitating conditions coefficient changes sign relative to Table 2. The change is a property of the specification, not an instability in the model, and the conditional effects recovered in this way agree closely with the corresponding group estimates reported in Section 4.6.

Table 3. Joint interaction models estimated on the full sample (N = 765).

Effect	M0	M1	M3
Interaction terms			
EXP × TR → BI	–	−0.069 (0.001)	−0.072 (0.001)
EXP × BI → USE	–	0.024 (0.234)	0.007 (0.709)
SUB × BI → USE	–	−0.013 (0.669)	0.000 (0.991)
SUB × FC → USE	–	0.069 (0.015)	0.052 (0.039)
Main effects of the moderators			
EXP → BI	0.052 (0.045)	0.032 (0.210)	0.039 (0.164)
EXP → USE	0.115 (<0.001)	0.125 (<0.001)	0.077 (0.008)
SUB → BI	−0.002 (0.913)	0.003 (0.897)	−0.028 (0.232)
SUB → USE	0.201 (<0.001)	0.196 (<0.001)	0.092 (0.001)
Selected structural paths			
HT → BI	0.363 (<0.001)	0.367 (<0.001)	0.384 (<0.001)
TR → BI	0.182 (<0.001)	0.171 (<0.001)	0.168 (<0.001)
HT → USE	0.428 (<0.001)	0.431 (<0.001)	0.437 (<0.001)
BI → USE	0.229 (<0.001)	0.239 (<0.001)	0.185 (<0.001)
FC → USE	0.011 (0.677)	−0.025 (0.401)	−0.021 (0.473)
Control variables			
AGE → USE	–	–	−0.063 (0.027)
EDU → USE	–	–	0.115 (0.001)
TOOLS → USE	–	–	0.202 (<0.001)
GEN → BI	–	–	0.099 (0.027)
WORK → USE	–	–	0.045 (0.487)
Explained variance			
R^2 (BI)	0.643	0.650	0.657
Adjusted R^2 (BI)	0.638	0.644	0.649
R^2 (USE)	0.544	0.549	0.592
Adjusted R^2 (USE)	0.541	0.544	0.585

Note. Cells report standardized path coefficients with bootstrap p-values in parentheses, from 10,000 subsamples. M0 = main effects of both moderators; M1 = M0 with 4 interaction terms; M3 = M1 with 5 control variables. EXP = prior usage experience, entered as a quasi-continuous variable; SUB = subscription tier, entered as a dummy variable; AGE = age; EDU = education; TOOLS = number of tools used; GEN = gender; WORK = employment status. Coefficients involving SUB are fully standardized; the corresponding partially standardized values in M1 and M3 are 0.460 and 0.216 for SUB → USE, 0.006 and −0.065 for SUB → BI, −0.030 and 0.001 for SUB × BI → USE, and 0.161 and 0.121 for SUB × FC → USE. Coefficients of moderated paths in M1 and M3 are conditional effects at the zero value of the moderator, as explained in the text. A dash indicates that the effect is not part of the model. Control variables were specified as predictors of both endogenous constructs; the paths not shown here are reported in Table S10.

The interaction between prior usage experience and trust on behavioral intention is negative and significant in both models (β = −0.069, p = 0.001 in M1; β = −0.072, p = 0.001 in M3), with effect sizes of 0.019 and 0.021. The contribution of trust to behavioral intention therefore weakens as prior usage experience increases. The result survives the Benjamini–Hochberg adjustment within the family of 4 interaction terms in both models (adjusted p = 0.004) and is unaffected by the addition of age, education, employment status, gender, and the number of tools used. It is the only interaction in the model that meets all of these conditions. A variant with a binary experience indicator (M1b) expresses the same result in a more directly interpretable form: the conditional effect of trust on behavioral intention is 0.284, 95% CI [0.191, 0.377], among inexperienced users and 0.103, 95% CI [0.021, 0.187], among experienced users, intervals that do not overlap.

The interaction between subscription tier and facilitating conditions on use behavior is positive and significant in M1 (β = 0.069, p = 0.015), but 4 qualifications attach to it. It

weakens by roughly a quarter once the control variables enter (β = 0.052, p = 0.039), it survives adjustment in M1 but not in M3 (adjusted p = 0.030 and 0.078), its effect size is small in both models (f^2 = 0.008 and 0.005), and it moderates a path that is not significant on average, since facilitating conditions do not predict use behavior in the base model, in M0, in M1, or in M3. The result is therefore reported as an indication that facilitating conditions may operate differently for paying users, and it is not treated as established.

The two remaining interactions are not significant in any model. Subscription tier does not moderate the intention-to-use path once prior usage experience is modeled alongside it (p = 0.669 in M1, p = 0.991 in M3), which is the direct test of the difference reported in Section 4.6. Prior usage experience does not moderate that path either (p = 0.234 in M1, p = 0.709 in M3), and its coefficient is positive in both models. This last result stands against the group comparisons, in which the intention-to-use path is weaker among more experienced users. The disagreement is examined in Sections 4.6 and 4.7 and is resolved in favor of the joint models, with the consequence that no claim is made in this study that the intention-to-use link weakens with experience.

The main effects tell a second story. In M0, subscription tier predicts use behavior with a coefficient of 0.201 and an effect size of 0.082, the largest in that equation after habit. Once the control variables enter in M3, the coefficient falls to 0.092 and the effect size to 0.014, below the threshold of practical relevance, while the number of tools used reaches 0.202 with an effect size of 0.069, and education and age also contribute. Prior usage experience follows the same pattern in weaker form, falling from 0.115 to 0.077. Habit is indifferent to the specification: its coefficient on use behavior is 0.428 in M0, 0.431 in M1, and 0.437 in M3. Finally, the models describe a subscription variable that serves as a marker of an engagement profile rather than an independent determinant of use, and a habit effect that does not depend on how the model is specified (Figure 3).

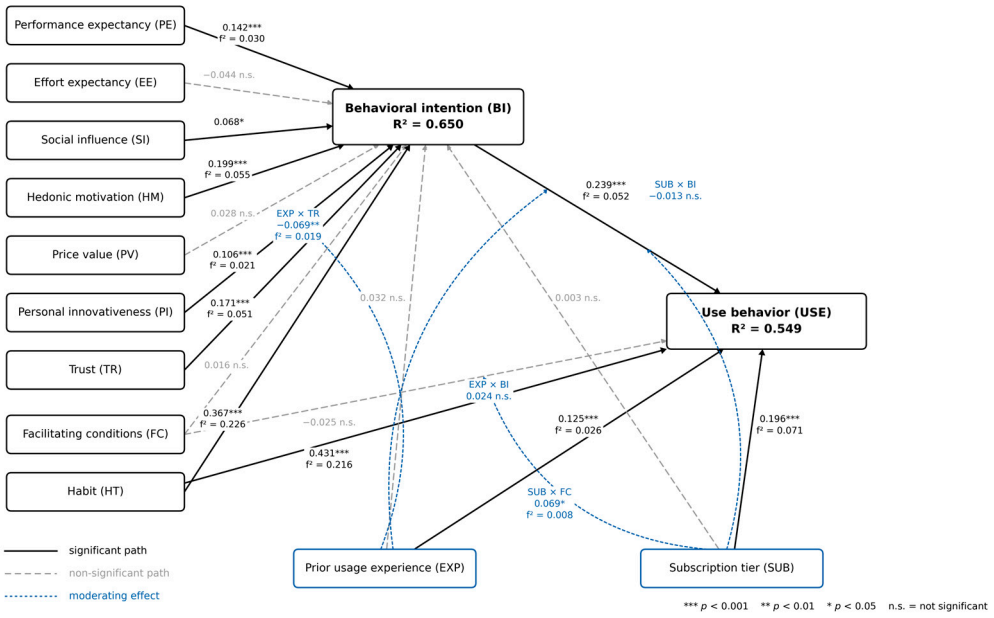


Figure 3. Joint interaction model (M1) estimated on the full sample (N = 765). *Note.* Solid arrows denote significant paths, dashed gray arrows non-significant paths, and dotted blue arrows the moderating effects, each drawn onto the specific path it moderates. Path values are standardized coefficients with significance levels, and effect sizes are reported for the 4 interaction terms and for the remaining paths where f^2 reaches the 0.02 relevance threshold. Reporting conventions for coefficients involving subscription tier are given in Section 3.4 and in the note to Table 3, and Q^2predict values for the base model in Section 4.3.

4.6. Multigroup Comparisons

The multigroup comparisons are reported next, as descriptive support for the joint models rather than as the primary evidence. Each of the 4 candidate moderators was examined through measurement invariance assessment using the MICOM procedure followed by permutation-based multigroup analysis with 5000 permutations and a two-tailed test. Compositional invariance, assessed in MICOM Step 2, was fully established for the gender and experience groupings. For the subscription grouping, it held for 9 out of 11 constructs, the exceptions being behavioral intention (c = 0.999, permutation 5% quantile 0.999, $p = 0.011$) and use behavior (c = 0.996, quantile 0.999, $p < 0.001$), and for the age grouping, it failed for price value, social influence, and use behavior. All comparisons in all analyses are reported in Table S7; Table 4 summarizes those that reach significance before adjustment.

Table 4. Permutation-based multigroup analysis of the 4 candidate moderators.

Moderator (Groups)	MICOM Step 2	Path	Group A	Group B	*p* (Diff)
Subscription (free 583/paid 182)	Partial	BI → USE	0.293	0.109	0.025
		FC → USE	−0.012	0.172	0.005
Gender (women 434/men 331)	Full	PE → BI	0.214	0.076	0.042
Age (younger 390/older 375)	Partial	HM → BI	0.108	0.274	0.030
Experience (inexp. 296/exp. 469)	Full	TR → BI	0.289	0.097	0.016

Note. Group A and Group B columns report the path coefficient in each group, in the order listed under "groups." Paths reaching significance before adjustment for multiple testing are shown; all 12 structural paths in each analysis are reported in Table S7, together with raw and adjusted *p*-values under both family definitions. No difference in this table survives adjustment. Construct abbreviations are defined in Table 1 and in the Abbreviations section.

None of the differences in Table 4 survives adjustment for multiple testing. Under the primary criterion, in which each analysis forms a family of 12 tests, the smallest adjusted value is 0.060, for the facilitating-conditions-to-use difference by subscription tier, followed by 0.150 for the intention-to-use difference in the same analysis and 0.192 for the trust-to-intention difference by experience. Under the stricter alternative, in which all six analyses form a single family, every adjusted value is 0.360 or above. Both sets of values are reported in Table S7. This outcome is the reason the joint interaction models rather than the group comparisons carry the argument of this study: the interaction test of the trust relationship survives adjustment within its own family, whereas no group comparison does.

The failure of compositional invariance in the subscription grouping deserves particular attention, because it affects use behavior, the construct to which both differences attributed to subscription tier lead, and behavioral intention, the source of one of them. Where compositional invariance is not established, a difference in path coefficients cannot be attributed unambiguously to a difference in the underlying structural relationship, since the construct may not be formed identically in the two groups, and part of the observed difference may reflect how the composite is built rather than the strength of the relationship. The MICOM Step 3 results separately indicate that paid and free users differ in the mean value of use behavior, which is consistent with a substantive difference in engagement. However, Step 3 cannot repair a Step 2 failure, and the two steps are kept distinct in interpretation. The subscription comparisons in the use equation are therefore reported without being used to support a conclusion, and the corresponding test in the joint interaction model, which requires no invariance assumption, is treated as decisive.

The individual comparisons can be summarized briefly. The subscription grouping produced two differences, both in the use equation: the intention-to-use path was stronger among free users than among paid users (0.293 versus 0.109, $p = 0.025$), and the facilitating-

conditions-to-use path differed as well (−0.012 versus 0.172, $p = 0.005$). The MICOM Step 3 comparison showed that paid and free users differ systematically at the construct level, with paid users scoring higher on use behavior, personal innovativeness, habit, performance expectancy, behavioral intention, and price value, consistent with a more engaged user profile. Both path differences are examined directly in Section 4.5, where the corresponding interaction terms are estimated on the full sample, and in Section 4.7, where experience is held constant by design.

The two demographic moderators produced one difference each. Among women, the performance expectancy-to-intention path was stronger than among men (0.214 versus 0.076, $p = 0.042$), and among older users, the hedonic motivation-to-intention path was stronger than among younger users (0.274 versus 0.108, $p = 0.030$). Neither difference reached the use equation, nor was it anticipated on theoretical grounds; the age analysis rests on partial invariance, and neither survives adjustment. Gender and age are therefore treated as not structuring the model in these data. Whether the composition of the sample suppressed genuine variation is a separate question, taken up in Section 5.3.

The experience grouping produced the difference that motivates the analysis. The trust-to-intention path was substantially stronger among inexperienced users than among experienced users (0.289 versus 0.097, $p = 0.016$). Taken on its own, the comparison does not survive adjustment (adjusted $p = 0.192$), and it is one of 12 tests within that analysis. Its evidential status rests instead on the interaction test reported in Section 4.5, which uses the full sample, requires no invariance assumption, and does survive adjustment. The intention-to-use path was also directionally weaker among experienced users (0.337 versus 0.223), but the difference was not significant ($p = 0.107$), and the corresponding interaction in the joint models is non-significant and opposite in sign, so no conclusion is drawn from it.

4.7. Disentangling Subsamples

The two disentangling subsamples are reported as internal sensitivity checks rather than as independent replications, since the respondents in both are part of the full sample analyzed above. Their purpose is to show whether the pattern identified in the joint models also appears when the confounded variable is held constant by design. The results are reported in Table 5.

Table 5. Path differences within the two restricted subsamples.

Subsample	Path	Group A	Group B	p (Diff)	Outcome
Cut 1: experienced only (free 330/paid 139)	BI → USE	0.218	0.121	0.334	difference disappears
	FC → USE	−0.014	0.111	0.127	difference disappears
Cut 2: free only (inexp. 253/exp. 330)	TR → BI	0.297	0.100	0.034	difference confirmed
	BI → USE	0.388	0.218	0.039	difference confirmed

Note. Construct abbreviations are defined in Table 1 and in the Abbreviations section.

Measurement invariance was assessed for both subsamples. In the first, compositional invariance failed for use behavior ($p = 0.003$) while the remaining 10 constructs passed. In the second, all 11 constructs passed, including use behavior ($p = 0.217$) and trust ($p = 0.170$). The asymmetry is informative: the invariance problem in the use construct is tied to grouping by subscription tier rather than to grouping by experience, which means the comparisons in the second subsample rest on firmer measurement ground than those in the first. Full MICOM output for both subsamples is reported in Table S8.

Neither subsample is balanced on observed characteristics, and this must be stated plainly. Within the experienced group, paid and free users differ significantly on every char-

acteristic examined, including gender, age, education, employment status, settlement size, the number of tools used, and all three use indicators, with effect sizes reaching r = 0.383 for the number of tools and r = 0.360 for education. Within the free group, experienced and inexperienced users do not differ in gender or settlement size, but the inexperienced group is on average about 8 years older (35.5 versus 27.2 years, r = 0.362) and more often employed. Holding one variable constant by restriction therefore does not produce comparable groups, and the first subsample in particular remains open to the objection that a difference it fails to detect may be absent for reasons other than the variable held constant. The full comparison is reported in Table S9. The joint interaction models are not exposed to this objection in the same way, since M3 includes age, education, employment status, gender, and the number of tools used as controls, and the trust interaction is unaffected by their inclusion.

The first subsample retains only experienced users and contrasts paid and free users within that group (n = 469). Both subscription differences observed in the full sample become non-significant. The intention-to-use path no longer differs between paid and free users (0.218 versus 0.121, p = 0.334), and neither does the facilitating conditions-to-use path (−0.014 versus 0.111, p = 0.127). Two cautions apply. The subsample is smaller than the full sample, so a loss of significance may reflect reduced statistical power rather than the absence of a difference, and compositional invariance for use behavior is not established here either. The joint interaction models address both concerns, since they retain all 765 respondents and require no invariance assumption. They show the intention-to-use interaction to be absent and the facilitating conditions interaction to persist, so the subsample result agrees with the first of these findings and not with the second, which is a further reason to treat the facilitating conditions indication as unresolved. One incidental difference emerged in the social influence-to-intention path (p = 0.036); it was unanticipated, does not survive adjustment, and is not built upon.

The second subsample retains only free users and contrasts experienced and inexperienced users within that group (n = 583). The trust-to-intention difference persists at almost exactly the magnitude observed in the full sample (0.297 versus 0.100, p = 0.034), which shows that payment cannot account for it. The intention-to-use path is also weaker among experienced users here (0.388 versus 0.218, p = 0.039), but this difference is not supported by the joint models, in which the corresponding interaction is non-significant and positive in sign. Where the two sources disagree, the joint models take precedence for the reasons given in Section 3.5, and the intention-to-use pattern is reported as an unreplicated subsample difference rather than as a finding. Neither difference survives adjustment within this analysis (adjusted p = 0.234 for both). Taken together, the two subsamples are consistent with the joint models on the central point, that subscription tier does not structure the model once prior usage experience is accounted for, whereas the role of trust differs by experience in a way that payment cannot explain. They are supporting rather than independent evidence for it.

4.8. Robustness Checks

Four additional specifications were estimated to test whether the results depend on particular modeling choices. The first extends M1 with two habit interactions, addressing the possibility that experience strengthens the habit paths rather than weakening the deliberative ones. In this model (M2), the interaction of prior usage experience with habit on use behavior is positive but nearly misses reaching significance (β = 0.063, p = 0.054, 95% CI [−0.004, 0.126]), with a negligible effect size (f^2 = 0.006) and an adjusted p-value of 0.108 within the family of six interaction terms. The interaction on behavioral intention is absent (β = −0.009, p = 0.704). A binary variant (M2b), for which bootstrap confidence intervals

were computed rather than p-values, generates conditional habit effects on use behavior of 0.376, 95% CI [0.266, 0.480] for inexperienced users and 0.495, 95% CI [0.395, 0.587] for experienced users. These intervals overlap, while the corresponding intervals for trust do not. The direction is therefore consistent with a strengthening of habit across levels of prior usage experience, but the evidence does not support asserting it. The trust interaction is unaffected in M2 ($\beta = -0.067$, $p = 0.003$).

The second specifies use behavior as a formative construct estimated in Mode B. The outer weight of the duration indicator is not significant (0.068, $p = 0.160$), while those of frequency (0.373) and intensity (0.657) are both significant ($p < 0.001$), so this specification would in effect remove duration from the construct. The structural estimates are essentially unchanged (habit on use behavior 0.484 versus 0.483; behavioral intention on use behavior 0.280 versus 0.272), and R^2 rises from 0.493 to 0.515, a change expected under Mode B because the weights are estimated so as to maximize explained variance and therefore not an argument for the formative specification. Since the reflective specification is supported by the indicator intercorrelations and loadings reported in Section 4.1, produces the same structural conclusions, and keeps all three indicators, it is retained.

The third addresses the concern that dividing age at the median created an artifact. With age entered as a continuous moderator, neither interaction is significant: age with hedonic motivation on behavioral intention generates $\beta = 0.022$, $p = 0.435$, $f^2 = 0.001$; and age with effort expectancy on behavioral intention generates $\beta = -0.050$, $p = 0.065$, $f^2 = 0.006$. The direct effects are also non-significant (behavioral intention $p = 0.088$, use behavior $p = 0.611$). The difference detected in the median-split comparison is therefore not reproduced when age is treated as continuous, and the conclusion that age does not structure the model does not depend on dichotomization.

The fourth addresses the 12-month cut point for experience. Comparing the adjacent categories of 6 to 12 months ($n = 162$) and more than 12 months ($n = 469$) reproduces the trust-to-intention difference (0.313 versus 0.097, $p = 0.030$; adjusted $p = 0.360$ within its own family of 12 paths). Since no difference distinguishes the categories below 12 months, the change appears to occur after the first year rather than gradually across it, which also explains why the binary specification of experience yields a larger interaction coefficient than the quasi-continuous one. One caution applies: compositional invariance fails for behavioral intention in this comparison ($p = 0.000$), so it is reported as an indication of where the change is located rather than as an independent test.

5. Discussion

5.1. Habit as the Strongest Predictor of Intention and Use

Habit was the strongest predictor in both equations of the base model, with a coefficient of 0.373 and an effect size of 0.236 on behavioral intention and a coefficient of 0.483 and an effect size of 0.255 on use behavior. The result is a statement of relative magnitude rather than of exclusivity. Hedonic motivation, trust, performance expectancy, and personal innovativeness were all significant predictors of behavioral intention, with coefficients ranging from 0.111 to 0.193, whereas social influence was significant with a smaller coefficient. In the use equation, behavioral intention remained a significant predictor alongside habit. Habit is the largest single contributor to the model, and it is not the only one.

The magnitude attributed to habit also depends on what else the model contains. When subscription tier and prior usage experience are included as predictors in M0, the habit coefficient on use behavior falls from 0.483 to 0.428, while the explained variance in use rises from 0.493 to 0.544. Part of what the base model assigns to habit therefore belongs to an engagement profile that the base model does not measure. Beyond that point the coefficient is stable, at 0.428 in M0, 0.431 in M1, and 0.437 in M3, so the addition of 5 control

variables leaves it essentially unchanged. The sequence describes a habit effect that is robust to specification once the relevant covariates are present, and that was somewhat overstated when they were absent.

This pattern is consistent with recent evidence from other generative AI settings. In a study of 611 higher education students, habit was the largest contributor to behavioral intention at 0.489, ahead of performance expectancy, social influence, and facilitating conditions, and it predicted use behavior at 0.325, although behavioral intention did not contribute in that model [10]. A study of graduate students likewise identified habit and behavioral intention as the most important predictors of actual use [11]. The convergence places generative AI, despite its novelty, alongside the established technologies for which habit has long been shown to drive continued use [25,28].

The habit at issue deserves comment, since generative AI use tends to be more creative, exploratory, and varied compared to traditional cases of technology-driven habits. In the dominant account, habits are learned responses that are controlled by recurring situational cues and operate with limited reference to current goals [27,45]. Applied here, the cue is the situation in which something has to be drafted, summarized, or checked, rather than the content of any particular task, so the variability of the output is not in itself an obstacle to habituation. Research on technology habits suggests that digital tools provide an unusually wide range of cues, contexts, and rewards, which may amplify rather than weaken habitual mechanisms [30]. At the same time, behavioral intention remained a significant predictor of use across all specifications estimated here, indicating that goal direction is not displaced by habit in this setting [37]. Whether habit around generative AI differs in kind from habit around more routine systems is a question these measures cannot settle, and it is a natural target for a study designed around it.

The two equations are also informative in contrast. Several determinants that shape intention, including hedonic motivation, trust, performance expectancy, and personal innovativeness, do not appear in the use equation, where habit and intention carry the explanatory weight. Behavioral intention and use behavior are empirically distinct in these data, as the discriminant validity results confirm, and they are predicted by different configurations of determinants. This supports measuring use through reported usage behavior rather than treating intention as a proxy for it.

5.2. Subscription Tier as an Apparent Moderator

Subscription tier appeared, on first inspection, to moderate acceptance. In the full sample, paid and free users differed on two paths in the use equation: the intention-to-use path was stronger among free users (0.293 versus 0.109, $p = 0.025$), and the facilitating conditions-to-use path differed in the opposite direction (−0.012 versus 0.172, $p = 0.005$). Had the analysis stopped there, subscription tier could have been reported as a moderator of generative AI use and given a behavioral reading in terms of the commitment that paying induces. Two features of the evidence argue against that reading before any further modeling is considered. Neither difference survives adjustment for the number of tests within its own analysis, and compositional invariance is not established in this grouping for use behavior, the construct to which both paths lead, or for behavioral intention, from which one of them departs.

Modeling both candidate moderators jointly separates the two paths. The intention-to-use difference does not persist. The corresponding interaction is non-significant in M1 ($p = 0.669$) and in M3 ($p = 0.991$), and the difference is also absent when experience is held constant by restriction (0.218 versus 0.121, $p = 0.334$). What the group comparison detected does not appear once prior usage experience is included in the same model, and no independent moderating role for subscription tier on that path is supported. The facilitating

conditions interaction behaves differently. It remains significant in M1 (0.069, $p = 0.015$) and weakens by roughly a quarter when the control variables enter (0.052, $p = 0.039$). It does not survive adjustment in M3 (adjusted $p = 0.078$), its effect size is small in both models (0.008 and 0.005), and it moderates a path that is not significant on average in any specification estimated here. It is reported as an indication that facilitating conditions may operate differently for paying users, and it is not treated as established.

The main effect of subscription tier follows a clearer path. In M0 it predicts use behavior with a coefficient of 0.201 and an effect size of 0.082, the largest in that equation after habit. With the control variables in M3, the coefficient falls to 0.092 and the effect size to 0.014, below the threshold of practical relevance, while the number of tools used reaches 0.202 with an effect size of 0.069 and education and age also contribute. The subsample comparison points the same way. Within the experienced group, paid and free users differ significantly on every characteristic examined, with the largest differences on the number of tools used and on education. Subscription tier in these data is best understood as a marker of an engagement profile that includes prior usage experience, breadth of tool use, education, and age, rather than as an independent determinant of use.

Two qualifications keep this conclusion within its evidence. The first concerns causal language. The association between subscription tier and the use paths weakens or disappears when experience and the engagement covariates are included, which is a statement about what these variables share in this sample and not a claim that paid status reflects experience in any causal sense. The second concerns the scope of the null result. Nothing here shows that payment is inert. In a design built to identify it, using matched comparisons and a difference-in-differences estimator applied to a long behavioral panel, conversion to a premium tier substantially increased consumption and community engagement [6]. A cross-sectional survey can neither detect an effect of that kind nor rule it out. What it can show is that in a sample where paying users are also the more experienced and more engaged users, an unadjusted comparison cannot separate the two. Researchers introducing subscription tier as a moderator without that separation risk mistaking a feature of who pays for an effect of paying [8].

5.3. The Limited Reach of the Standard Demographic Moderators

Gender and age, the demographic moderators carried forward from the original model, showed little effect in these data. Each produced a single difference on an intention path, a stronger performance expectancy effect among women (0.214 versus 0.076, $p = 0.042$) and a stronger hedonic motivation effect among older users (0.274 versus 0.108, $p = 0.030$). Neither reached the use equation, neither was anticipated on theoretical grounds, and neither survives adjustment. The age comparison additionally rests on partial invariance.

The null result for age is not an artifact of the median split. With age entered as a continuous moderator, neither interaction reaches significance, at 0.022 ($p = 0.435$) for hedonic motivation on intention and -0.050 ($p = 0.065$) for effort expectancy on intention, and the direct effects are not significant either. The difference detected by the median-split comparison is not reproduced when the variable is treated as continuous, so the conclusion does not depend on the dichotomization.

This outcome should not be read as evidence that gender and age are irrelevant to acceptance of generative AI. It is consistent with the expectation that demographic factors lessen in influence within broadly adopted consumer technology when the analysis is confined to users, where the between-group variation these moderators rely on is compressed [9], and with a recent UTAUT2 study of graduate students in which age, gender, and level of study showed no moderating influence [11]. The evidence is not uniform, however. In a study of 1190 students using an extended UTAUT model, gender and year of

study did moderate acceptance [23]. The samples differ in a way that matters. The present sample skews young, female, and student-heavy, and differences reported in larger and more heterogeneous samples may simply be compressed out of it. The safer reading is that among the users sampled here, gender and age do little to shape how the determinants translate into intention and use, and that the null result is bounded by the composition of the sample rather than established for the population. The result also says nothing about who becomes a user in the first place, a question this design cannot address.

5.4. Prior Usage Experience and the Receding Role of Trust

Prior usage experience was the only candidate that moderated the model robustly, and it did so on one path. The interaction between experience and trust on behavioral intention is negative and significant with the quasi-continuous specification (-0.069, $p = 0.001$) and remains so when age, education, employment status, gender, and the number of tools used are added (-0.072, $p = 0.001$). It survives adjustment within its family of interaction terms in both models (adjusted $p = 0.004$), and it is the only interaction estimated here that meets all of these conditions. The binary variant states the same result in conditional form: the effect of trust on behavioral intention is 0.284, 95% CI [0.191, 0.377], among inexperienced users and 0.103, 95% CI [0.021, 0.187], among experienced users, intervals that do not overlap. The group comparisons agree, at 0.289 versus 0.097 in the full sample and 0.297 versus 0.100 among free users alone, where payment cannot account for the difference.

The change appears to occur after the first year of use rather than be spread evenly across it. The comparison of adjacent categories reproduces the difference at the 12-month boundary (0.313 versus 0.097, $p = 0.030$), while no difference distinguishes the categories below it, which also explains why the binary specification generates a larger interaction coefficient than the quasi-continuous one. The evidence on this point is weaker than for the interaction itself. Compositional invariance fails for behavioral intention in that comparison, and the highest response category is open at the upper end, so users with 13 months and users with several years of use are grouped together. Habit research offers a reason not to expect a sharp threshold in any case, since automaticity develops gradually and its rate varies substantially across individuals and behaviors [29]. A 12-month boundary offers a rough overview of a gradual process, and it is best read as the point at which a difference becomes detectable in these data rather than as the moment at which something changes.

Why the contribution of trust should recede has more than one explanation, and these explanations are complementary rather than competing. Research on trust in automated systems distinguishes the trust a user brings to a system from the trust that is learned through accumulated interaction and holds that the learned component stabilizes as evidence about the performance of the system accumulates [31]. A stabilized assessment is less available to serve as an active input to a decision that is made repeatedly, which is consistent with the calibration view of trust in automation more generally [19,20]. The continuance literature arrives at a similar conclusion by another route. Continued use is governed by the confirmation of expectations and by the satisfaction that follows from it rather than by beliefs formed before use [32], and trust in a technology acts on continuance largely through experienced performance, disconfirmation, and satisfaction rather than directly [33]. Longitudinal evidence shows that the beliefs and attitudes underlying usage decisions, and the weight each of them carries, change as usage accumulates [34]. In both cases, the diminished role of trust reflects a broader re-weighting process, not something specific to this technology.

Independent evidence points in the same direction. Two studies of researchers using generative AI, one with a three-wave longitudinal design and one with a vignette experiment, report a negative interaction in which the relationship between trust and

usage intention weakens as familiarity and experience increase [35]. That research supplies through its designs what the present study cannot, since temporal separation and experimental manipulation address an inference that a single cross-section leaves open. The convergence is worth stating with caution rather than as confirmation. The moderating role of experience on this path is not a settled question in the literature, and the two designs differ from the present one in population and in measurement. What can be claimed reasonably is that independent lines of evidence agree on the direction.

Whether the receding role of trust is matched by a strengthening of habit is a separate question, and the answer here is qualified. Extending the model with habit interactions leaves the interaction of experience with habit on use behavior positive but short of significance (0.063, $p = 0.054$), with a negligible effect size (0.006) and an adjusted p-value of 0.108, and the conditional habit effects for inexperienced and experienced users, 0.376, 95% CI [0.266, 0.480], and 0.495, 95% CI [0.395, 0.587], have overlapping intervals where the corresponding intervals for trust do not. The direction is consistent with a consolidation of habit across levels of prior usage experience, and the evidence does not support asserting it. A further expectation from habit theory is not met at all. Accounts in which behavior comes to follow established routine rather than current intention predict that the intention-to-use link should loosen with experience [25]. The group comparisons are directionally consistent with that prediction, but the corresponding interaction is positive and non-significant in every model estimated here, and the joint models take precedence for the reasons given in Section 3.5. No claim is made in this study that the intention-to-use link weakens with prior usage experience.

Two limits govern how far any of these findings can be taken. The first is the design. The comparisons are between more and less experienced respondents observed at a single point in time, not observations of the same individuals as they accumulate experience, so what the data establish is a difference across experience levels that is consistent with a process of maturation rather than a demonstration of one. The account offered above is an interpretation of that difference and is presented as such. The second limit is measurement, and it belongs here rather than among the limitations, because it bears directly on which pieces of evidence carry weight. Compositional invariance failed for use behavior in the subscription and age groupings, for behavioral intention in the subscription grouping and in the adjacent-category comparison, and for price value and social influence in the age grouping, so several of the group differences rest on less secure measurement ground than their p-values suggest. The joint interaction models require no invariance assumption, retain all 765 respondents, and test a single coefficient rather than 12, which is why the central claim of this section rests on them while the group comparisons are reported as support.

5.5. Theoretical Contributions

Three contributions follow from the analysis. The first is methodological in substance and theoretical in consequence. When candidate moderators are correlated with one another, an unadjusted group comparison cannot establish that any one of them conditions the model, because the split on one variable is simultaneously a partial split on the others. The present study makes that point with a worked case. Subscription tier produced two significant path differences in the group comparisons, and neither survived as an established moderating effect once prior usage experience and the engagement covariates were entered into the same model. The intention-to-use interaction was absent in every joint model, and the facilitating-conditions interaction weakened once the control variables entered, failed adjustment for multiple testing in that model, and retained a negligible effect size throughout. Its main effect on use behavior remained significant but lost more

than half of its magnitude and fell below the threshold of practical relevance. The trust interaction with experience behaved differently: it held its magnitude and its significance under adjustment in every specification in which it was estimated. That contrast is not a matter of degree, and it follows from the two designs themselves. Multigroup analysis requires measurement invariance that was not established for the construct at issue, divides the sample and loses power, and tests 12 coefficients per analysis, whereas a joint interaction model requires no invariance assumption, retains the full sample, and tests one. Where candidate moderators are entangled, the interaction model should carry the evidential weight and the group comparison should be read as description.

The same case supports a second contribution, which concerns how subscription tier should be theorized in this literature. UTAUT2 accommodates the monetary aspect of consumer technology through price value, which is a perception rather than a structural position, and it offers no account of paid status itself [2]. The commercial turn in generative AI invites researchers to fill that gap by treating subscription tier as a moderator, and the present results argue against doing so directly. In these data, paid status carries information about breadth of tool use, education, age, and prior usage experience. Once those are modeled, it loses more than half of its association with use behavior, and its effect size falls below the threshold of practical relevance. Read as a construct, subscription tier here indexes a position within an engagement profile rather than a distinct behavioral condition created by payment. That is a claim about this sample and this design, and it is compatible with the finding that payment changes behavior when a design is able to isolate it [6]. What follows for theory is that subscription tier requires either a design that separates paying from prior engagement or an explicit treatment as a proxy, and that the freemium literature on why users upgrade is the more appropriate point of departure than the moderator apparatus of UTAUT2 [8,24].

The third contribution concerns the other member of the core pair. It locates the effect of prior usage experience more precisely than the framework does. UTAUT2 specifies experience as a moderator of numerous paths, including those from social influence, facilitating conditions, and habit [2]. In these data, it moderated one path, and that path belongs to the deliberative side of the model rather than to the habitual one. The trust contribution to intention weakened with experience while the habit paths did not measurably strengthen, which refines the framework rather than confirming it. The refinement also connects the UTAUT2 moderator debate to a body of literature it rarely engages with. Continuance research has established that the beliefs governing continued use are reweighted as usage accumulates, and that trust operates on continuance largely through experienced performance and satisfaction rather than directly [32–34]. Experience as a moderator in UTAUT2 has no mechanism attached to it, and the continuance account supplies one. Treating the two studies together offers a more specific prediction than either does alone, namely that experience should reduce the influence of belief-based factors on intention, while automatic factors remain unaffected.

5.6. Implications for Practice

The results carry different implications for acquiring users and for retaining them, and the distinction that matters is not who users are or what they pay but where they stand on the continuum from deliberation to routine. Providers who treat their user base as a single population, or who segment it by subscription plan, are working with a division that these results do not support.

For users in their first year, trust is a live determinant of the intention to continue, with a conditional effect of 0.284 in the binary specification. The practical levers that follow are specific rather than general. They include visible source attribution for factual

claims, explicit signaling of the confidence or uncertainty attached to an output, plain statements of the tasks a system is not suited to, and verification paths that let a user check a claim without leaving the tool. Research involving generative AI has led to compatible recommendations, suggesting conservative default settings for less experienced users, issuing warnings when responses lack verifiable sources, and encouraging cross-checking of key outputs [35]. Interventions of this kind are ordinarily justified by concerns about overreliance, and the present results add a second justification: for the users who most need them, trustworthiness signals are also the determinant most closely tied to continued intention.

For experienced users, the same levers are unlikely to do much work. The conditional effect of trust falls to 0.103 in this group, with a confidence interval that approaches zero, while habit remains the strongest determinant of use across all specifications. Retention among these users is a question of sustaining a cue-response structure rather than of further trust-building. In cue-based accounts of habit, what maintains a behavior is reliable encounter with the situation that triggers it [27,45], and digital tools are distinctive in the range of cues and contexts they can occupy [30]. The operational translation is placement rather than persuasion. A tool encountered at the point where work is already happening, through integration into the applications and workflows a user occupies daily, presents its cue without requiring a deliberate decision to visit it. Stable and predictable entry points serve the same function, and interface changes that relocate or reshape those entry points carry a cost that a purely usability-based assessment will not register. One qualification applies throughout. Behavioral intention remained a significant predictor of use in every model estimated here, so routine does not displace deliberation in this population, and a retention strategy resting on habit alone would be building on part of the mechanism.

For analysts inside provider organizations, the subscription result has a direct operational reading, though not a causal one. Paid and free users differ substantially in observed behavior and, in this sample, in every characteristic examined, including tenure, breadth of tool use, education, and age. A comparison of the two groups that does not adjust for these characteristics will attribute to the plan what belongs to the users who select into it, and will overstate what a change in plan can be expected to produce. The number of tools used was the strongest single control in the model, at 0.202 with an effect size of 0.069, which suggests it as a practical segmentation variable alongside tenure. Segmenting by breadth of tool use and by time since adoption is likely to be more informative than segmenting by plan, and it is available from data that providers already hold.

The timing evidence carries a further implication. No difference was observed between the experience categories below 12 months, and the difference in the role of trust appeared only in the comparison across that boundary. If this pattern holds, the window in which trust-oriented design does useful work is longer than a typical onboarding sequence, and cohort measurement that stops at three or six months will miss the transition entirely. A final caution governs all of the above. The predictive assessment reported in Section 4.3 placed the model below a linear benchmark on most indicators, which is a statement of low predictive power. These implications concern which levers merit attention and which segmentations are defensible, and they are not a basis for forecasting the behavior of individual users.

5.7. Limitations and Future Research

The sample is a non-probability sample recruited through snowball, convenience, and purposive methods within a single country, and it skews toward the young, female, and student segments. It supports the analysis of relationships among users and does not warrant claims about the prevalence of generative AI use in the wider population.

The composition also bounds the moderator results in a specific way. Between-group variation in age and in the circumstances associated with gender is compressed in a sample of this kind, so the demographic null results reported in Section 5.3 are weaker evidence of absence than the same results would be in a heterogeneous sample, and they should not be generalized beyond it.

The measurement of prior usage experience is the clearest limitation of the design, and it is acknowledged rather than defended. The item offers 5 ordered categories with the highest one open at the upper end, so respondents with 13 months of use and respondents with several years of use cannot be distinguished. Treating experience as quasi-continuous therefore adds resolution only below the 12-month boundary, and it is precisely above that boundary that the interesting variation is likely to sit. The dichotomization at 12 months loses further information, and while the 12-month point is where the difference becomes detectable in these data, no claim is made that it is where anything changes. The parallel concern for age was addressed by estimating age as a continuous moderator, which reproduced the null result, but the median split remains a simplification in the group comparisons that use it.

The strategy of holding one variable constant by restriction has a limitation that the results make visible. Restriction produced subsamples that are not comparable on observed characteristics. Within the experienced group, paid and free users differ significantly on every characteristic examined, and within the free group the inexperienced are on average about 8 years older and more often employed. A difference that a restricted subsample fails to detect may therefore be absent for reasons other than the variable held constant, and a difference it does detect may reflect the remaining imbalance. The joint interaction models with controls address this by modeling the covariates rather than restricting on them, and the trust interaction is unaffected by their inclusion, but the controls are limited to characteristics that were observed. Income, professional dependence on generative AI, and the intensity of task reliance were not measured and could not be adjusted for.

The use construct rests on three self-reported ordinal indicators of duration, frequency, and perceived intensity, treated as continuous, rather than on objective behavioral records. Respecifying it in Mode B left the structural conclusions unchanged and did not support removing any indicator on substantive grounds, and confirmatory tetrad analysis, which would test the reflective specification directly, cannot be applied to a construct with three indicators. The reflective form is therefore supported by the indicator intercorrelations, the loadings, and a robustness check rather than by a direct test. All measures come from a single instrument administered at one point in time. Neither the full-collinearity test nor Harman's single-factor test provides evidence that common method bias distorts the estimates, and a single-source design cannot exclude it.

Predictive performance is a limitation in its own right. Benchmarked against a linear model, the PLS-SEM estimates produced a lower error on 1 of 6 indicators by the root mean squared error and on 1 of 6 by the mean absolute error, indicating low predictive power. The differences are small, and the criterion is directional rather than a measure of magnitude, but the conclusion stands that the model accounts for variance better than it predicts new observations. The findings should be read as explanatory rather than as a basis for prediction.

The treatment of multiple testing leaves one result unresolved and one framework decision open to disagreement. No difference in any group comparison survived Benjamini–Hochberg adjustment under either family definition, which is why the conclusions of this study rest on the interaction models rather than on the comparisons. The interaction between subscription tier and facilitating conditions is the remaining unresolved case. It is significant without controls, weakens with them, fails adjustment in the controlled model,

has a negligible effect size, and moderates a path that is not significant on average. It is reported in full and requires replication before it is interpreted. The two-tier criterion applied here, in which the a priori question about trust and experience is reported without adjustment alongside the interaction test and everything else is treated as exploratory, is defensible but not the only defensible choice, and all adjusted values are reported so that readers can apply a stricter one.

The model does not include constructs specific to generative AI that other frameworks would include, such as perceptions of AI ethics and bias, creative empowerment, the experience of human-AI collaboration, and concerns about effects on employment and society. The instrument was designed for comparability with the UTAUT2 corpus and does not measure them, so their exclusion is a boundary of this study rather than a judgment about their relevance. Two further characteristics that plausibly distinguish paying from non-paying users were also unmeasured, namely tolerance for risk and expectations about the quality of the technology. Personal innovativeness is included in the model and provides partial evidence for the first of these, since the invariance results indicate that paid users score higher on it, but this falls well short of a test. Each of these constructs is a candidate for the design that would replace the present one.

The direction of future work follows from these limits. A longitudinal design that follows users as they accumulate experience would test directly what the present comparison across experience levels can only suggest, and would resolve whether the receding role of trust and the consolidation of habit occur within individuals. The habit interaction reported here fell just short of significance, so a study powered for that specific question is warranted. Separating payment from prior engagement calls for a different design again, whether a matched panel with behavioral records or a setting in which access is assigned rather than chosen [6]. Replication of the trust interaction in more representative samples and in other cultural settings would establish how far the pattern travels, and a measure of experience with an open upper range would allow the shape of the relationship above 12 months to be examined rather than assumed.

6. Conclusions

This study asked which of several candidate moderators structures the acceptance and use of generative AI once the confounding among them is taken seriously. Treating subscription tier, gender, age, and prior usage experience as candidates to be tested rather than as assumed conditions, and estimating them jointly rather than through separate group comparisons, produced a narrower answer than the group comparisons alone would have supported. Habit was the strongest determinant of both behavioral intention and use behavior, though it coexists with other significant determinants of intention, and its magnitude depends on which covariates the model contains. Subscription tier appeared to condition two paths in the use equation, and that appearance did not persist for the intention-to-use path once prior usage experience was modeled alongside it. A weaker indication involving facilitating conditions remains unresolved and is reported as such. Gender and age structured little among the users sampled. Prior usage experience was the one candidate whose effect survived joint estimation, the addition of controls, and adjustment for multiple testing, and it did so on a single path, the contribution of trust to behavioral intention, which was substantially weaker among experienced respondents.

The contribution is twofold. Methodologically, the study shows why entangled candidate moderators cannot be adjudicated by group comparison. A split on one such variable is simultaneously a partial split on the others, and the comparison inherits an ambiguity that no amount of care in reporting can remove. Modeling the candidates jointly resolves it, retains the whole sample, and requires no assumption of measurement

invariance across groups. Substantively, the study locates the effect of experience on the deliberative rather than the habitual side of the model, and connects the moderating role that UTAUT2 assigns to experience with the mechanism that the continuance literature supplies for it, namely the reweighting of belief-based determinants as usage accumulates.

These conclusions carry boundaries that should accompany them. The evidence is cross-sectional, so the difference between more and less experienced respondents is consistent with a maturation process without demonstrating one. The model also accounts for variance better than it predicts new observations, so the contribution is explanatory and associational rather than predictive. The sample is drawn from a single country and skews young and student-heavy, which compresses the variation that demographic moderators depend on and limits what the null results for gender and age can establish. The effect of experience is also unlikely to be uniform across settings. The complexity and consequences of tasks users assign to generative AI vary. Consequently, the importance of trust in low-stakes drafting tasks might differ from its role when errors have significant costs. Therefore, the reduced impact observed here may be more prominent in certain task settings than others and could also depend on the user's level of AI literacy. Longitudinal designs, designs able to separate payment from prior engagement, and replication in more representative samples and other cultural settings are the routes by which these boundaries would be established rather than assumed.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jtaer21090308/s1, Table S1: Factor loadings of all 38 measurement items in the base model; Table S2: Cross-loadings of all 38 items on all 11 constructs; Table S3: Heterotrait-monotrait ratios of correlations with bias-corrected and accelerated 95% confidence intervals; Table S4: Fornell-Larcker criterion; Table S5: Collinearity statistics; Table S6: Predictive assessment against the linear-model and indicator-average benchmarks; Table S7: Permutation-based multigroup comparisons: all twelve structural paths in each of the seven analyses; Table S8: Measurement invariance of composite models (MICOM) for all four candidate moderators and both restricted subsamples; Table S9: Equivalence of the two restricted subsamples on observed characteristics; Table S10: Control variable paths and explained variance across estimated models.

Author Contributions: Conceptualization, A.B.; Methodology, B.B. and A.B.; Software, B.B.; Validation, A.B. and B.B.; Formal Analysis, B.B.; Investigation, B.B.; Resources, A.B.; Data Curation, B.B.; Writing—Original Draft Preparation, A.B.; Writing—Review and Editing, A.B. and B.B.; Visualization, A.B.; Supervision, A.B.; Project Administration, A.B.; Funding Acquisition, A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research is the result of the project DIMPACT (Transformation of Marketing Practices in the Era of Disruptions: Integrative Effects of Digital Innovations, Changes in Consumer Behaviour and Contemporary Market Challenges, 581-UNIOS-11), funded by the European Union—NextGenerationEU. The views and opinions expressed are solely those of the author(s) and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Institutional Review Board Statement: Ethical review and approval were waived for this study, as ethical approval is not required for anonymous, survey-based research in which participants are healthy adults who give informed consent and the study does not collect sensitive or identifiable personal data. Under Croatian legislation, mandatory prior approval by an ethics committee is tied to the health-care and biomedical domain (the Health Care Act, Official Gazette No. 100/2018, with subsequent amendments, and the national framework governing clinical trials); it does not extend to anonymous, non-interventional social-science surveys of the present kind. Research conduct in higher education and science is governed by the Act on Higher Education and Scientific Activity (Official Gazette No. 119/2022) and the codes of ethics adopted thereunder. The study was

conducted in accordance with the principles of the Declaration of Helsinki and applicable codes of good research practice.

Informed Consent Statement: Informed consent was obtained from all participants. Before starting the anonymous online questionnaire, participants were informed of the purpose of the study and of the anonymous and voluntary nature of participation, and provided consent by choosing to proceed.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: During the preparation of this manuscript, the authors used Claude Opus 4.8 and Claude Opus 5 (Anthropic, San Francisco, CA, USA) and ChatGPT 5.5 (OpenAI) for language refinement and assistance with drafting and structuring the text, and Grammarly for grammar, spelling, and punctuation checking. These tools were not used to generate or analyze data or to interpret the results. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
AVE	Average variance extracted
BI	Behavioral intention
EE	Effort expectancy
FC	Facilitating condition
Gen AI	Generative artificial intelligence
HM	Hedonic motivation
HT	Habit
HTMT	Heterotrait–monotrait ratio of correlations
MGA	Multigroup analysis
MICOM	Measurement invariance of composite models
PE	Performance expectancy
PI	Personal innovativeness
PLS-SEM	Partial least squares structural equation modeling
PV	Price value
SI	Social influence
TR	Trust
USE	Use behavior
UTAUT2	Unified theory of acceptance and use of technology 2
VIF	Variance inflation factor

Appendix A

The full measurement instrument is reported in Table A1. The items are reported in their English version. The sample analyzed here completed the Croatian version, translated from the same sources; the two versions differ only in the response scheme of the education item, as described in Section 3.1. Wordings are reproduced verbatim from that version, adapted to the generative AI context from the sources listed in the final column. The questionnaire additionally collected 8 background items covering gender, year of birth, the generative AI tools respondents had used, length of use, use of a paid tier, employment or student status, settlement size, and highest completed level of education. These are described in Sections 3.1 and 3.3 and are not repeated here.

Table A1. Measurement instrument. Item codes correspond to those used in Table 1 and in the Supplementary Material. Construct abbreviations follow the Abbreviations section.

Construct	Code	Item Wording	Source
Reflective constructs measured on seven-point Likert scales (1 = strongly disagree, 7 = strongly agree)			
Performance expectancy (PE)	PE1	I find Gen AI useful in my daily life	Venkatesh et al. [2]
	PE2	Using Gen AI helps me accomplish things more quickly	
	PE3	Using Gen AI increases my productivity	
	PE4	Using Gen AI increases my chances of achieving things that are important to me	
Effort expectancy (EE)	EE1	My interaction with Gen AI is clear and understandable	Venkatesh et al. [2]
	EE2	It is easy for me to become skillful at using Gen AI	
	EE3	I find Gen AI easy to use	
	EE4	Learning how to use Gen AI is easy for me	
Social influence (SI)	SI1	People whose opinions I value prefer that I use Gen AI	Venkatesh et al. [2]
	SI2	People who are important to me think that I should use Gen AI	
	SI3	People who influence my behavior think that I should use Gen AI	
Facilitating conditions (FC)	FC1	I have the knowledge necessary to use Gen AI	Venkatesh et al. [2]
	FC2	Gen AI tools are compatible with the technology I use	
	FC3	I have the resources necessary to use Gen AI	
	FC4	I can get help from others when I have difficulties using Gen AI	
Hedonic motivation (HM)	HM1	Using Gen AI is fun	Venkatesh et al. [2]
	HM2	Using Gen AI is very entertaining	
	HM3	Using Gen AI is enjoyable	
Price value (PV)	PV1	Gen AI is a good value for money	Venkatesh et al. [2]
	PV2	At the current prices, Gen AI tools provide a good value	
	PV3	Gen AI tools are reasonably priced	
Habit (HT)	HT1	I am addicted to using Gen AI	Venkatesh et al. [2]
	HT2	Using Gen AI has become natural to me	
	HT3	I must use Gen AI	
	HT4	The use of Gen AI has become a habit for me	
Personal innovativeness (PI)	PI1	If I hear about new technology, I would look for ways to experiment with it	Agarwal and Prasad [18]; Gansser and Reich [17]

Table A1. *Cont.*

Construct	Code	Item Wording	Source
	PI2	Among my peers, I am usually the first to try out new technologies	
	PI3	I like to experiment with new technologies	
Trust (TR)	TR1	I will use Gen AI if I feel that the content is trustworthy	Mohd Rahim et al. [36]
	TR2	I will use Gen AI if I feel that Gen AI is secure	
	TR3	I will use Gen AI if I feel that Gen AI meets my expectations	
	TR4	I will use Gen AI if I feel that Gen AI provides reliable information	
Behavioral intention (BI)	BI1	I intend to continue using Gen AI in the future	Venkatesh et al. [2]
	BI2	I plan to continue using Gen AI frequently	
	BI3	I will always try to use Gen AI in my daily life	
Use behavior, measured on seven-point ordinal scales with context-adapted anchors			
Use behavior (USE)	USE_dur	On average, how many hours per week do you use Gen AI tools? (1 = less than 1 h; 7 = more than 6 h)	Venkatesh et al. [37]
	USE_freq	How often do you use Gen AI tools? (1 = a few times a year or less; 7 = a few times a day or more)	
	USE_int	How do you consider the extent of your current use of Gen AI tools? (1 = extremely low use; 7 = heavy use)	

Note. $N = 765$. All 38 measurement items are listed. Facilitating conditions item FC4 was retained despite a low factor loading for the reason given in Section 4.4.

References

1. Venkatesh, V.; Morris, M.G.; Davis, G.B.; Davis, F.D. User acceptance of information technology: Toward a unified view. *MIS Q.* **2003**, *27*, 425–478. [CrossRef]
2. Venkatesh, V.; Thong, J.Y.L.; Xu, X. Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Q.* **2012**, *36*, 157–178. [CrossRef]
3. Strzelecki, A. To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interact. Learn. Environ.* **2024**, *32*, 5142–5155. [CrossRef]
4. Habibi, A.; Muhaimin, M.; Danibao, B.K.; Wibowo, Y.G.; Wahyuni, S.; Octavia, A. ChatGPT in higher education learning: Acceptance and use. *Comput. Educ. Artif. Intell.* **2023**, *5*, 100190. [CrossRef]
5. Camilleri, M.A. Factors affecting performance expectancy and intentions to use ChatGPT: Using SmartPLS to advance an information technology acceptance framework. *Technol. Forecast. Soc. Change* **2024**, *201*, 123247. [CrossRef]
6. Bapna, R.; Ramaprasad, J.; Umyarov, A. Monetizing freemium communities: Does paying for premium increase social engagement? *MIS Q.* **2018**, *42*, 719–735. [CrossRef]
7. Jo, H. Uncovering the reasons behind willingness to pay for ChatGPT-4 premium. *Int. J. Hum.–Comput. Interact.* **2025**, *41*, 994–1009. [CrossRef]
8. Mäntymäki, M.; Islam, A.K.M.N.; Benbasat, I. What drives subscribing to premium in freemium services? A consumer value-based view of differences between upgrading to and staying with premium. *Inf. Syst. J.* **2020**, *30*, 295–333. [CrossRef]
9. Tamilmani, K.; Rana, N.P.; Dwivedi, Y.K. Consumer acceptance and use of information technology: A meta-analytic evaluation of UTAUT2. *Inf. Syst. Front.* **2021**, *23*, 987–1005. [CrossRef]
10. Rashid, S. Habit predicting higher education EFL students' intention and use of AI: A nexus of UTAUT-2 model and metacognition theory. *Educ. Sci.* **2025**, *15*, 756. [CrossRef]
11. Yıldırım-Samuk, Ş.; Altay, İ.F. Exploring graduate students' acceptance and use of generative AI: An application of the UTAUT2 model. *Contemp. Educ. Technol.* **2026**, *18*, ep666. [CrossRef] [PubMed]

12. de Andrés-Sánchez, J.; Gené-Albesa, J. Explaining policyholders' chatbot acceptance with an unified technology acceptance and use of technology-based model. *J. Theor. Appl. Electron. Commer. Res.* **2023**, *18*, 1217–1237. [CrossRef]
13. Silva, F.A.; Shojaei, A.S.; Barbosa, B. Chatbot-based services: A study on customers' reuse intention. *J. Theor. Appl. Electron. Commer. Res.* **2023**, *18*, 457–474. [CrossRef]
14. Song, Y.; Tan, H. When generative AI meets abuse: What are you anxious about? *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 215. [CrossRef]
15. Gursoy, D.; Chi, O.H.; Lu, L.; Nunkoo, R. Consumers acceptance of artificially intelligent (AI) device use in service delivery. *Int. J. Inf. Manag.* **2019**, *49*, 157–169. [CrossRef]
16. Ma, X.; Huo, Y. Are users willing to embrace ChatGPT? Exploring the factors on the acceptance of chatbots from the perspective of AIDUA framework. *Technol. Soc.* **2023**, *75*, 102362. [CrossRef]
17. Gansser, O.A.; Reich, C.S. A new acceptance model for artificial intelligence with extensions to UTAUT2: An empirical study in three segments of application. *Technol. Soc.* **2021**, *65*, 101535. [CrossRef]
18. Agarwal, R.; Prasad, J. A conceptual and operational definition of personal innovativeness in the domain of information technology. *Inf. Syst. Res.* **1998**, *9*, 204–215. [CrossRef]
19. Lee, J.D.; See, K.A. Trust in automation: Designing for appropriate reliance. *Hum. Factors* **2004**, *46*, 50–80. [CrossRef] [PubMed]
20. Glikson, E.; Woolley, A.W. Human trust in artificial intelligence: Review of empirical research. *Acad. Manag. Ann.* **2020**, *14*, 627–660. [CrossRef]
21. Choung, H.; David, P.; Ross, A. Trust in AI and its role in the acceptance of AI technologies. *Int. J. Hum.–Comput. Interact.* **2023**, *39*, 1727–1739. [CrossRef]
22. Choudhury, A.; Shamszare, H. Investigating the impact of user trust on the adoption and use of ChatGPT: Survey analysis. *J. Med. Internet Res.* **2023**, *25*, e47184. [CrossRef] [PubMed]
23. Xu, J.; Li, Y.; Shadiev, R.; Li, C. College students' use behavior of generative AI and its influencing factors under the unified theory of acceptance and use of technology model. *Educ. Inf. Technol.* **2025**, *30*, 19961–19984. [CrossRef]
24. Hamari, J.; Hanner, N.; Koivisto, J. "Why pay premium in freemium services?" A study on perceived value, continued use and purchase intentions in free-to-play games. *Int. J. Inf. Manag.* **2020**, *51*, 102040. [CrossRef]
25. Limayem, M.; Hirt, S.G.; Cheung, C.M.K. How habit limits the predictive power of intention: The case of information systems continuance. *MIS Q.* **2007**, *31*, 705–737. [CrossRef]
26. Kim, S.S.; Malhotra, N.K.; Narasimhan, S. Research note—Two competing perspectives on automatic use: A theoretical and empirical comparison. *Inf. Syst. Res.* **2005**, *16*, 418–432. [CrossRef]
27. Wood, W.; Neal, D.T. A new look at habits and the habit-goal interface. *Psychol. Rev.* **2007**, *114*, 843–863. [CrossRef] [PubMed]
28. Polites, G.L.; Karahanna, E. Shackled to the status quo: The inhibiting effects of incumbent system habit, switching costs, and inertia on new system acceptance. *MIS Q.* **2012**, *36*, 21–42. [CrossRef]
29. Lally, P.; van Jaarsveld, C.H.M.; Potts, H.W.W.; Wardle, J. How are habits formed: Modelling habit formation in the real world. *Eur. J. Soc. Psychol.* **2010**, *40*, 998–1009. [CrossRef]
30. Bayer, J.B.; LaRose, R. Technology habits: Progress, problems, and prospects. In *The Psychology of Habit: Theory, Mechanisms, Change, and Contexts*; Verplanken, B., Ed.; Springer: Cham, Switzerland, 2018; pp. 111–130. [CrossRef]
31. Hoff, K.A.; Bashir, M. Trust in automation: Integrating empirical evidence on factors that influence trust. *Hum. Factors* **2015**, *57*, 407–434. [CrossRef] [PubMed]
32. Bhattacherjee, A. Understanding information systems continuance: An expectation-confirmation model. *MIS Q.* **2001**, *25*, 351–370. [CrossRef]
33. Lankton, N.; McKnight, D.H.; Thatcher, J.B. Incorporating trust-in-technology into expectation disconfirmation theory. *J. Strateg. Inf. Syst.* **2014**, *23*, 128–145. [CrossRef]
34. Bhattacherjee, A.; Premkumar, G. Understanding changes in belief and attitude toward information technology usage: A theoretical model and longitudinal test. *MIS Q.* **2004**, *28*, 229–254. [CrossRef]
35. Kleine, A.-K.; Schaffernak, I.; Lermer, E. When do researchers trust AI? Trust as a mediator and experience as a boundary condition in researchers' intentions to use generative AI. *Comput. Hum. Behav. Artif. Hum.* **2026**, *8*, 100310. [CrossRef]
36. Mohd Rahim, N.I.; Iahad, N.A.; Yusof, A.F.; Al-Sharafi, M.A. AI-based chatbots adoption model for higher-education institutions: A hybrid PLS-SEM-neural network modelling approach. *Sustainability* **2022**, *14*, 12726. [CrossRef]
37. Venkatesh, V.; Brown, S.A.; Maruping, L.M.; Bala, H. Predicting different conceptualizations of system use: The competing roles of behavioral intention, facilitating conditions, and behavioral expectation. *MIS Q.* **2008**, *32*, 483–502. [CrossRef]
38. Biloš, A.; Budimir, B. Understanding the adoption dynamics of ChatGPT among generation Z: Insights from a modified UTAUT2 model. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 863–879. [CrossRef]
39. Hair, J.F.; Hult, G.T.M.; Ringle, C.M.; Sarstedt, M. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 3rd ed.; Sage Publications: Thousand Oaks, CA, USA, 2022.

40. Ringle, C.M.; Wende, S.; Becker, J.-M. *SmartPLS 4*; SmartPLS GmbH: Bönningstedt, Germany, 2024. Available online: https://www.smartpls.com (accessed on 27 June 2026).
41. Henseler, J.; Ringle, C.M.; Sarstedt, M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J. Acad. Mark. Sci.* **2015**, *43*, 115–135. [CrossRef]
42. Kock, N.; Lynn, G.S. Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *J. Assoc. Inf. Syst.* **2012**, *13*, 546–580. [CrossRef]
43. Shmueli, G.; Sarstedt, M.; Hair, J.F.; Cheah, J.-H.; Ting, H.; Vaithilingam, S.; Ringle, C.M. Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *Eur. J. Mark.* **2019**, *53*, 2322–2347. [CrossRef]
44. Henseler, J.; Hubona, G.; Ray, P.A. Using PLS path modeling in new technology research: Updated guidelines. *Ind. Manag. Data Syst.* **2016**, *116*, 2–20. [CrossRef]
45. Wood, W.; Rünger, D. Psychology of habit. *Annu. Rev. Psychol.* **2016**, *67*, 289–314. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.