"I don't know anything about laptops!" – User Perception of Digital Product Advisors Adapting to Their Knowledge Levels

Kevin Schott
kevin.schott@gesis.org
GESIS – Leibniz Institute for the Social Sciences
Cologne, Germany

Andrea Papenmeier
a.papenmeier@utwente.nl
University of Twente
Enschede, Netherlands

Daniel Hienert
daniel.hienert@gesis.org
GESIS – Leibniz Institute for the Social Sciences
Cologne, Germany

Dagmar Kern
dagmar.kern@gesis.org
GESIS – Leibniz Institute for the Social Sciences
Cologne, Germany

Abstract

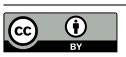
Conversational commerce uses digital assistants to support the search process and decision-making in e-commerce. Effective communication in these interactions can be facilitated by assistants adapting their communication style to users and supporting shared understanding. An open challenge in this context is adapting the presentation of complex product information to users with varying levels of domain knowledge. To investigate strategies for such knowledge-level adaptation, we set up a chatbot-assisted laptop search scenario. In a between-subjects experiment (n = 251), we examined novice and expert perceptions of product attribute recommendations presented as technical information only (T), or augmented with performance categories (TC), attribute explanations (TE), or both (TCE). For novices, approaches with explanations (TE, TCE) were perceived as more helpful and led to higher perceived learning than those without. Novices also rated the combined approach (TCE) more appropriate than the baseline (T) and TC in terms of information quantity, indicating that explanations are crucial to understand and benefit from performance categories. Critically, experts showed no significant differences across conditions, suggesting that providing supplementary information beneficial to novices did not detract from their experience. We distill these findings into four concrete design guidelines for inclusive text-based product advisors in technical domains: use TCE by default; keep a single inclusive interface; avoid standalone categories; and support user agency and personalize to the stated use case.

CCS Concepts

• **Information systems → Users and interactive retrieval**; • **Human-centered computing → HCI design and evaluation methods**.

Keywords

Conversational commerce, product search, digital assistant, informed decision-making, personalization, domain knowledge



1 Introduction

Recent advances in conversational technology have given rise to *conversational commerce*, a term introduced in 2015 [34, 60] and defined by Balakrishnan et al. as the "buying activity by a customer through a digital assistant" [7]. In this context, users' voice or text dialogue with a conversational agent (CA) is intended to mimic a natural exchange with a human shop assistant, gathering needs and preferences to recommend suitable products [59]. By acting as an interactive, socially present decision aid, digital product advisors can foster trust in e-commerce platforms, a crucial factor for purchase intentions [19, 62].

A subtle but important mechanism behind effective communication and advice is the *communication accommodation theory* (CAT): interlocutors adjust their language or non-verbal cues to reduce social distance [20]. Communication accommodation not only shapes human dialogue but has also been shown to improve user experiences in human-computer interaction [26]. This links CAT directly to research on intelligent user interfaces and personalization, i.e., the adaptation of "a service or a product in such a way that it fits to the preferences, cognition, requirements, or capabilities of specific persons within the confines of a specific setting" [27].

Prior research has demonstrated that communication accommodation strategies, such as lexical (e.g., [52, 54, 66]) and personality (e.g., [3, 50]) alignment, can effectively improve users' experience when interacting with CAs (see Section 2.3). Researchers have also presented technical frameworks for persona alignment (e.g., [31, 65]). By contrast, knowledge-level communication adaptation–a key aspect of *recipient design* (tailoring utterances to what the addressee is presumed to know and believe) [8, 46]–remains comparatively underexplored, despite knowledge being a core facet of a holistic user model [5]. Existing research has shown that LLMs can adapt the complexity of generated summaries based on user familiarity [57] and has presented conversational prototypes that infer user knowledge from dialogue signals to adapt responses [4]. However, user evaluations, especially in commercial scenarios, are scarce.

An open challenge, highlighted by prior calls for search systems in e-commerce to accommodate both low- and high-knowledge consumers (novices and experts) [45], is therefore to convey complex product information in an adaptive manner. Consumers with low domain knowledge tend to struggle with evaluating functional

product information [2], frequently relying on third-party opinions such as customer reviews [29, 35]. This discrepancy not only reduces novices' ability to make informed decisions but may also impact their trust and willingness to use product advisors for complex purchases in the first place. Conversely, experts may be negatively affected by redundant information [51]. While prototypes for general learning [14], learning-oriented search [64], and complex purchases [35] have highlighted the educational potential of conversational systems, a solid understanding of how complex product information should be proactively presented to users with varying knowledge levels has not yet been established. It is, therefore, currently unclear how text-based product advisors can best adapt the *type* and *amount* of provided information within conversational commerce.

To address this research gap, we report on a survey-based user study integrating both quantitative and qualitative data. We examined novices' and experts' perceptions of four product attribute recommendation types in a text-based laptop advisor, varying in the type and amount of presented product information. In our experiment with 251 participants, we found that the effectiveness of supplementary information varied by information type, especially for novices. Without attribute explanations, novices reported having learned less during the interaction. They also rated the combination of *performance categories* (e.g., "mid-range configuration") and *attribute explanations* (e.g., "The CPU [...] is a key factor in overall laptop speed and program performance.")–condition TCE–as more appropriate in information quantity than the baseline (T) and TC. In contrast, the amount and type of supplementary information did not significantly impact how experts perceived the conversation, indicating that accommodating novices' need for additional information is not detrimental to the experience of experts.

Our findings translate into concrete design guidelines for product advisors in technical domains that can serve diverse knowledge levels. We show that novices can be effectively supported without alienating experts by supplementing technical information with attribute explanations and performance categories. Accordingly, we advocate a single inclusive interface (TCE) as the default, avoiding standalone performance categories, supporting user agency, and personalizing to the stated use case. These guidelines aim to provide a practical, empirically backed method for implementing knowledge-level recipient design.

2 Related Work

This section presents the theoretical frameworks that inspired our research, examines search behaviors across expertise levels, and reviews related user studies on the personalization of chatbots.

2.1 Theoretical Frameworks

In human-human communication, interlocutors adapt to each other on several levels throughout the interaction. The *communication accommodation theory* [20] describes how individuals adjust their communication styles to either align (convergence) with or distinguish (divergence) themselves from their conversation partners through verbal and nonverbal aspects. Previous work has established that communication accommodation is not only relevant in human-human interaction (HHI) but can also enhance the perceived interaction experience in human-computer interaction (HCI) [26]. Complementing the communication accommodation theory, the principle of *recipient design* emphasizes that communicators actively tailor their utterances based on their hypotheses about the conversation partner's knowledge and beliefs [8, 46].

The *common ground theory* states that interlocutors continuously work on establishing and maintaining a common understanding to ensure effective communication. In the HCI context, digital systems employ various grounding techniques, such as providing feedback, asking clarifying questions, and building a shared knowledge base, to mitigate errors and misunderstandings [12, 58]. Communicating understanding through active grounding also helps product advisors to be perceived as more competent [42].

Next to these insights about communication accommodation, recipient design, and establishing common ground between interlocutors, the *cognitive load theory* [56] served as another inspiration for our research. The theory highlights the limited capacity of human working memory. It differentiates between intrinsic load (the inherent complexity of information), extraneous load (unnecessary distractions or details), and germane load (the cognitive effort to integrate new information) [55]. In educational contexts, Hollender et al. [24] suggested that software interfaces should adapt intrinsic load to users' domain knowledge levels and reduce extraneous load to optimize usability. This approach aligns with the *expertise reversal effect*, wherein novices benefit from guided instructions, whereas extra information can burden experts, who must reconcile it with their existing knowledge structures [51].

These theoretical frameworks can inform the implementation of *adaptive selling* [53] in e-commerce, which involves adjusting communication to the individual needs of customers to enhance satisfaction and encourage future interactions [44]. Together, the concepts presented in this section directly inform the design of our study conditions (see Section 3.2) and hypotheses (see Section 3.5).

2.2 Differences in Search Behavior between Novices and Experts

Prior work demonstrates that novice and expert users apply different search behaviors and strategies. Rowley [45] describes that novices, characterized by lower domain knowledge, are more inclined to browse and explore information without a fixed goal, whereas experts engage in more directed searches, formulating precise queries and effectively filtering out irrelevant results. Alba and Hutchinson [2] provide additional insights into the underlying cognitive mechanisms that differentiate the two groups. They note that when experts process new product information, they draw on well-organized mental representations that enable them to retrieve relevant details, supporting informed decision-making. In contrast, novices tend to focus more on surface-level cues such as brand names and price. In terms of decision-making, experts typically interpret new data in the context of preexisting knowledge and analyze functional product attributes. Novices, in contrast, rely more on explicit product descriptions and external factors like reviews as evaluation criteria. This can render them more vulnerable to external persuasion and simple heuristic processing, such as price-quality inferences. In addition to these differences, novice

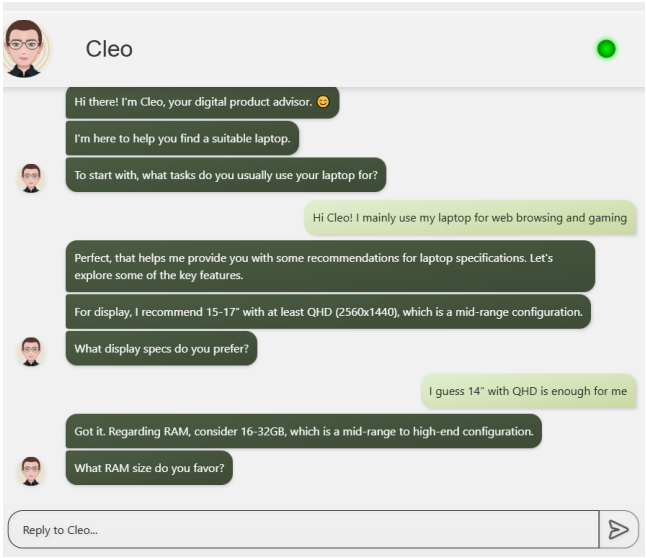


Figure 1: Interface for our product advisor "Cleo" (condition TC – technical information + performance categories).

consumers tend to be less confident in their purchase decisions compared to those with high domain knowledge [47].

Collectively, these findings underscore the importance of designing (product) search systems that accommodate varying levels of user knowledge and provide adaptive support to improve novices' search outcomes. They also highlight the importance of equipping novices with product knowledge in order to help them make informed, self-reliant decisions.

2.3 Chatbot Personalization

Personalizing interaction is a cornerstone of intelligent user interface (IUI) research. IUIs aim to represent, reason, and act on models of the user and their context [11, 32]. Ait Baha et al. [1] classify chatbot personalization into three dimensions: (i) linguistic style (e.g., vocabulary, sentence structure, tone), (ii) personal traits (e.g., personality, empathy), and (iii) persona attributes (e.g., identity, occupation, interests).

The adaptation of chatbot communication to the user, such as through lexical alignment (reusing users' wording) has proven beneficial in prior work. For instance, lexical alignment has been shown to enhance recall and understanding [54], and to reduce perceived effort, frustration, mental workload, and cognitive demand [26, 52]. Positive trends (though non-significant) have also been observed for perceived information and service quality, enjoyment, satisfaction, and trustworthiness [65]. Similarly, aligning chatbots' communication behavior with users' personality traits (e.g., introversion/extroversion) has been shown to improve engagement, purchasing behavior, trust, and usage intention [3, 50].

Although lexical and personality alignment are relatively well-studied, the adaptation of CAs to the user's domain knowledge–a crucial aspect of the user model [5]–remains comparatively underexplored. For instance, the findings of Thakkar et al. [57] show that LLMs can adjust the complexity of generated summaries based on user familiarity, and An et al. [4] aimed to implement recipient design in a conversational prototype by capturing different knowledge types identified through conversational cues. However, user evaluations of such systems are still emerging. For example, Higuchi and Inaba [22] demonstrated in a user study that dynamically tailoring candidate questions about news articles to users' comprehension levels can improve users' understanding compared to a baseline.

Together, these findings underscore the value of personalized chatbots. We extend this line of work by isolating knowledge-level adaptation and testing how variations in the type and amount of presented product information influence novices' and experts' perceptions in conversational commerce.

3 Study Design

Building on the identified need to understand how product advisors can effectively present complex technical information to users with varying levels of domain knowledge, our study addressed the following research question:

RQ: How should text-based product advisors adapt their presentation of technical product information to meet the needs of users with different levels of domain knowledge (novices and experts)?

To answer this research question, we conducted a controlled online user study employing a between-subjects experimental design. This approach allowed us to systematically manipulate the type and amount of supplementary information presented alongside technical product attribute recommendations, and to measure their distinct effects on novice and expert users. The study focused on the purchase of a new laptop, chosen as a representative use case of consumer electronics, an area characterized by complex technical specifications. Laptops were selected not only because of their technical complexity but also because electronic devices account for a substantial share of online purchases, representing approximately one-quarter of the global e-commerce market according to eCommerceDB (ECDB)[1].

3.1 Apparatus - Digital Product Advisor

For our experiment, we developed "Cleo", a text-based product advisor interface for laptop search (see Figure 1). To isolate the effects of the information presentation format and ensure consistent experiences across participants in each condition, Cleo was implemented as a rule-based system with a predefined interaction flow. Within this controlled structure, Cleo proactively asks users about their intended use of the laptop and their preferences for five different laptop attributes (display, RAM, storage, CPU, and GPU) in a predefined order designed to mimic the query behavior of human shop assistants [40]. User responses regarding the intended use were categorized as either "basic" (e.g., text editing or web browsing) or "advanced" use cases (e.g., video editing, programming, or gaming). This classification determines the specific predefined technical recommendation Cleo provides for each subsequent attribute query (e.g., recommending 8-16 GB of RAM for a user with "basic" use cases and 16-32 GB of RAM for a user with "advanced" use cases).

[1] https://ecdb.com/blog/leading-ecommerce-categories-in-top-markets/5134 last accessed: 28-09-2025

Table 1: Example attribute (RAM) recommendations for "basic" and "advanced" use cases under each condition. To isolate the effect of information presentation format, message structure and wording are held constant across use cases, and only the recommended attribute values (and corresponding performance-category labels where applicable) differ.

Condition	"Basic" use cases recommendation	"Advanced" use cases recommendation
Technical Only (**T**)	Regarding RAM, consider 8–16 GB.	Regarding RAM, consider 16–32 GB.
Technical + Performance Categories (**TC**)	Regarding RAM, consider 8–16 GB, which is an entry-level to mid-range configuration.	Regarding RAM, consider 16–32 GB, which is a mid-range to high-end configuration.
Technical + Attribute Explanations (**TE**)	Regarding RAM, consider 8–16 GB. RAM (working memory) helps the laptop handle multiple tasks at once and run demanding programs smoothly.	Regarding RAM, consider 16–32 GB. RAM (working memory) helps the laptop handle multiple tasks at once and run demanding programs smoothly.
Combined (**TCE**)	Regarding RAM, consider 8–16 GB, which is an entry-level to mid-range configuration. RAM (working memory) helps the laptop handle multiple tasks at once and run demanding programs smoothly.	Regarding RAM, consider 16–32 GB, which is a mid-range to high-end configuration. RAM (working memory) helps the laptop handle multiple tasks at once and run demanding programs smoothly.

The experimental manipulation occurs in the supplementary information presented alongside these technical recommendations across conditions, as detailed in the following section.

3.2 Information Presentation Formats (T, TC, TE, and TCE)

We tested four information presentation formats for Cleo's laptop attribute recommendations as our experimental conditions. Each provided users with a specific type of product information or a combination of types. The following subsections describe these conditions in detail, and Table 1 presents example messages for the attribute RAM.[2]

3.2.1 T – Technical Information Only (Baseline). In condition T, the product advisor offers only numerical specifications, accompanied by standard abbreviations (e.g., QHD, RAM, SSD, GPU) for all five laptop attributes. This format serves as the baseline for all subsequent conditions, reflecting how technical specifications are typically presented in e-commerce contexts.

3.2.2 TC – Technical + Performance Categories. Condition TC extends condition T by adding performance categories ("entry-level," "mid-range," "high-end") alongside numerical specifications. Alba and Hutchinson [2] note that categorization is part of consumers' cognitive structures, with novices often attending to basic-level categories rather than technical specifications due to limited domain knowledge. Accordingly, we add plain-language performance category labels to support novice comprehension.

3.2.3 TE – Technical + Attribute Explanations. In condition TE, the product advisor omits performance categories and adds a plain-language sentence explaining each attribute's function and impact. It also replaces technical abbreviations (e.g., "RAM") with less specialized terms (e.g., "working memory"). This format targets two novice needs identified by Alba and Hutchinson [2]: (i) interpreting technical specifications (via explanations and simplified terminology) and (ii) assessing relative importance (by clarifying each attribute's practical impact).

3.2.4 TCE – Technical + Categories + Explanations. Finally, condition TCE combines both types of supplementary information: the performance categories from TC and the attribute explanations from TE. Technical specifications and performance categories are paired with an explanatory sentence clarifying abbreviations and describing each attribute's purpose. This format aims to maximize clarity and support informed decision-making, particularly for novice users.

3.3 Procedure and Scenario

We conducted the study online using SoSci Survey[3], requiring participants to access the survey via a desktop or laptop device. Participants first provided informed consent and answered closed demographic questions. They then rated their subjective domain knowledge of laptops using the item "How would you classify your knowledge of laptops?" on a 7-point Likert scale (1 = "No knowledge/non-expert" to 7 = "High knowledge/expert", adapted from [13]). The use of this subjective measure is supported by findings from prior work, which demonstrated statistically significant correlations between subjective and objective knowledge across various complex domains, including technical products [13, 43] and financial investments [21]. Following the demographic and knowledge assessment, participants were presented with the scenario and task description (adapted from [41]):

> "*Imagine that your laptop stopped working, and you are now searching for a new one. Your task will be to answer Cleo's questions so that they can understand your needs and preferences.*"

Participants were then randomly assigned and routed to one of the four interface variants (T, TC, TE, or TCE) and began interacting with Cleo, the product advisor (see Figure 1). Cleo sequentially

[2]The script containing the categories and explanations for all laptop attributes across all conditions is available at https://osf.io/be9jt/?view_only=c7ecc97a48454e578e860bfd180ea578.

[3]https://www.soscisurvey.de/en/index

asked about the participant's intended use of the laptop and preferences for the five laptop attributes (display, RAM size, storage, CPU, and GPU). Participants responded via free-text input, and Cleo returned attribute-level recommendations; no products were shown and no ranked product list was presented. After completing the interaction, participants were redirected to our questionnaire and answered several closed and one open question about their experience with Cleo. The study was approved by our institute's ethics committee.

3.4 Measurements

This section outlines the dependent variables examined in the study and the instruments used to measure them.

3.4.1 Measures for Appropriate Quantity, Relevance, Learning of Information Provided, and Trust. These measures directly assessed key user experience aspects relevant to our research question and the theoretical framework described in Section 2.1. While our framing draws on cognitive load theory, we did not employ its standard measurement instruments [6] (e.g., Paas' mental-effort rating scale [39]), as they are validated for extended instructional tasks and were deemed ill-suited to our study's brief chatbot interactions. Instead, we used lightweight single-item constructs on 5-point Likert scales (1 = "strongly disagree" to 5 = "strongly agree") that are closely aligned with the aim of accommodating users' domain knowledge and are immediately evaluable post-interaction:

Perceived appropriateness of information quantity: Whether the amount of information met user needs, using the statement "The chatbot gave me the appropriate amount of information." (inspired by [9])

Perceived learning: Whether the interaction subjectively increased the participant's knowledge, particularly valuable for novices, using the statement "I learned new information about laptop aspects through the chatbot." (inspired by [18, 23, 48])

Perceived relevance: The extent to which the information was applicable to the user's search, using the statement "The information provided by the chatbot was relevant." (inspired by [9])

Trust: Confidence in the chatbot, a critical factor for adoption and reliance, using the statement "I can trust this chatbot." (inspired by [19, 62])

3.4.2 Measures for Perceived Helpfulness. We assessed the perceived helpfulness of the supplementary information to determine the specific utility of each information component (performance categories, attribute explanations, or both in combination), relating to the practical value of the provided adaptations. In conditions TC, TE, and TCE, participants were shown screenshots of two example attribute recommendations for their assigned condition. They rated the statement "How helpful did you find the [categorization/explanation/supplementary information] for understanding the recommendation?" on a 5-point Likert scale ranging from 1 = "Not at all helpful" to 5 = "Very helpful."

3.4.3 Attribute Recommendation Accuracy. To evaluate perceptions of the core technical recommendations, participants rated their accuracy on a 5-point Likert scale (1 = "Not at all accurate" to 5 = "Totally accurate") with the residual answer option "I don't know." This measure ensured that the perceived plausibility of Cleo's recommendations did not confound participants' evaluations of the supplementary information.

3.4.4 Qualitative Feedback. Finally, participants responded to an open-text question: "What additional information would have helped you to answer Cleo's questions about your preferences?" This qualitative feedback provided insight into how novices and experts perceived the supplementary information presented under the different conditions.

3.5 Hypotheses

This section presents our hypotheses regarding the perceptions of novices and experts, both across and within experimental conditions. Our hypotheses are broadly guided by the assumption that novices, who often struggle with unfamiliar technical product information, will benefit from supplementary information [2]. Conversely, experts may find such information redundant or less impactful, aligning with concepts like the *expertise reversal effect* [51]. We therefore expect divergent responses to variations in the type and amount of information provided, depending on user expertise.

For *perceived appropriateness*, *learning*, *relevance*, and *helpfulness*, we generally expect novices to prefer more comprehensive information, while experts' perceptions will be less influenced by supplementary details, or they may prefer less:

H1a: Novices deem the combination of performance categories and attribute explanations (TCE) as most appropriate, while each type of supplementary information alone (TC, TE) is rated as more appropriate than no supplementary information (T).
H1b: Experts perceive no supplementary information (T) as most appropriate.
H1c: Novices rate conditions with supplementary information (TC, TE, TCE) higher in appropriateness than experts.
H1d: Experts perceive the no-supplementary-information condition (T) as more appropriate than novices do.

H2a: Novices experience the highest perceived learning with TCE, while TC and TE lead to more perceived learning than T.
H2b: Experts show no significant differences in perceived learning between conditions.
H2c: Novices report higher perceived learning in conditions TC, TE, and TCE compared to experts.

H3a: Novices perceive TCE as most relevant, while TC and TE are deemed more relevant than T.
H3b: Experts perceive no supplementary information (T) as most relevant.
H3c: Novices perceive conditions TC, TE, and TCE as more relevant than experts do.
H3d: Experts find condition (T) more relevant than novices.

H5a: Novices perceive the performance categories combined with attribute explanations (TCE) as most helpful.
H5b: Experts show no significant differences in perceived helpfulness across conditions TC, TE, and TCE.

H5c: Novices rate conditions TC, TE, and TCE as more helpful than experts.

Finally, building on Kuhail et al.'s finding that chatbot personality alignment can enhance *trust* [3], we hypothesize that alignment with a user's domain knowledge will have a similar effect:

H4a: Novices show most trust with TCE, while TC and TE enhance novices' trust compared to T.
H4b: Experts show increased trust with no supplementary information (T).
H4c: Novices exhibit higher trust in conditions TC, TE, and TCE than experts.
H4d: Experts show higher trust than novices in condition T.

4 Participants

We recruited participants on Prolific[4] with eligibility criteria requiring residence in the UK or USA, English as their primary language, and age ≥ 18 years. To ensure data quality, we required a Prolific approval rate of at least 95% and a minimum of 20 previous submissions, and we included an attention-check question [38]. One participant was excluded for being under 18, yielding a final sample of 251 participants (124 male, 126 female, one preferred not to disclose; age 18–77, $M = 41.2$, $SD = 13.6$).

Following prior work that classifies consumers by subjective domain knowledge (e.g., [30, 36, 61]), we defined novices as ratings 1–4 and experts as ratings 5–7 on the 7-point subjective domain-knowledge scale before data collection. The threshold was specified before data collection based on an archival dataset collected with the same item in a comparable population (participants recruited via Prolific, located in the UK, and speaking English as their primary language)[5]. In that dataset ($N = 135$), counts by scale point (1–7) were 3, 17, 14, 32, 37, 22, and 10, yielding grouped totals of 1–4 = 66 (48.9%) and 5–7 = 69 (51.1%), which informed our recruitment quotas. Using stratified random assignment, participants from each stratum (novice/expert) were assigned with equal probability to one of four conditions (T, TC, TE, TCE). Recruitment continued until the stopping rule was met: at least 30 novices and 30 experts in each condition. The resulting groups showed distinct means on the 1–7 scale: novices ($M = 3.0$, $SD = 1.0$) and experts ($M = 5.6$, $SD = 0.8$). Final per-condition counts were: novices (T = 30, TC = 30, TE = 30, TCE = 35) and experts (T = 31, TC = 30, TE = 35, TCE = 30).

Participants took part in the study for an average of 6:40 minutes (SD = 4:13 minutes) and received 1.20 GBP in compensation, equivalent to an hourly rate of approximately 10 GBP.

5 Results

This section presents our findings for the user study. First, we describe the results of our statistical analyses and evaluate our hypotheses. We then present the qualitative insights we gathered from the participants' responses to our open question.

[4]https://www.prolific.com/
[5]The anonymized dataset is available at https://osf.io/be9jt/?view_only=c7ecc97a48454e578e860bfd180ea578.

5.1 Statistical Analysis

Because our data was not normally distributed, we used Kruskal-Wallis tests with post-hoc Dunn's tests to test for significant differences between our four conditions among novices and among experts. Because this analysis involves multiple pairwise post-hoc comparisons per dependent variable, we controlled for multiple testing using the Benjamini-Hochberg procedure with a false discovery rate of $q = .05$. Unlike Bonferroni-style family-wise error control, this approach limits the expected proportion of false positives among the set of significant findings while retaining more statistical power. We applied BH correction within each family of post-hoc comparisons per dependent variable and expertise group. Accordingly, we report BH-adjusted post-hoc p-values, denoted p_{adj}. For novice-expert comparisons within a condition, we used Mann-Whitney U tests. To quantify effect sizes for non-parametric tests, we report Cliff's Delta (δ) [63]. Means and standard deviations for each dependent variable are presented in Table 2. Meanwhile, Figure 2 visualizes our results and indicates significant differences.

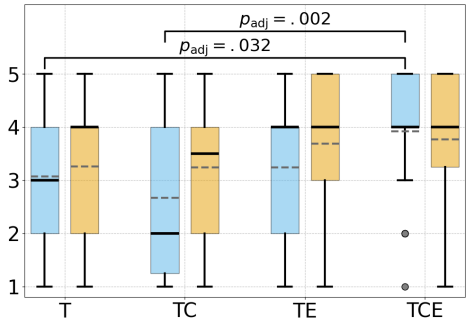
5.1.1 Perceived Appropriateness of Information Quantity. A Kruskal-Wallis test revealed a significant difference among the four conditions for novices ($H = 13.736$, $p = .003$). Dunn's post hoc tests indicated that novices perceived TCE as significantly more appropriate than TC ($p_{\text{adj}} = .002$, $\delta = .478$, large effect), and more appropriate than T ($p_{\text{adj}} = .032$, $\delta = .372$, medium effect). The differences between TC and T, TE and T, TE and TC, and TCE and TE were not significant (all $p_{\text{adj}} > .05$), thereby **only partly supporting H1a**. For experts, no significant differences emerged among the four conditions ($H = 4.926$, $p = .177$), thus **rejecting H1b**. Mann–Whitney U tests comparing novices and experts within each condition revealed no significant differences (all $p > .05$), thus also **rejecting H1c** and **H1d**.

5.1.2 Perceived Learning. Among novices, the Kruskal-Wallis test showed a significant effect of condition on perceived learning ($H = 18.525$, $p < .001$). Dunn's post hoc tests revealed that both TE and TCE were rated significantly higher than T (TE vs. T: $p_{\text{adj}} = .013$, $\delta = .371$; TCE vs. T: $p_{\text{adj}} = .003$, $\delta = .459$) and higher than TC (TE vs. TC: $p_{\text{adj}} = .013$, $\delta = .381$; TCE vs. TC: $p_{\text{adj}} = .003$, $\delta = .471$); all reflecting medium effect sizes. TE and TCE did not differ ($p_{\text{adj}} = .693$), thereby **partly supporting H2a**. For experts, no significant differences were observed among the four conditions ($H = 1.335$, $p = .721$), **supporting for H2b**. When comparing novices and experts within each condition using Mann–Whitney U tests, novices rated TCE significantly higher ($U = 752.5$, $p = .002$, $\delta = .433$, medium effect), while no other comparisons reached significance (all $p > .05$), thereby **partly supporting H2c**.

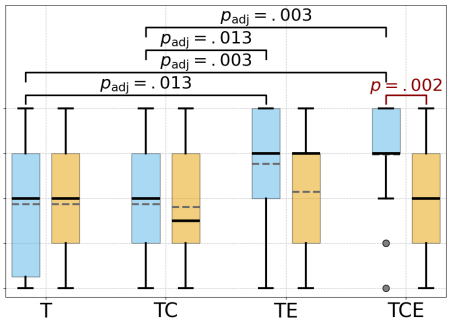
5.1.3 Perceived Relevance. Kruskal-Wallis tests for both novices ($H = 5.916$, $p = .116$) and experts ($H = 4.769$, $p = .190$) revealed no significant differences among the four conditions. Mann–Whitney U tests comparing novices and experts within each condition also showed no significant differences, although TE exhibited a potential trend toward significance ($U = 392.5$, $p = .062$, $\delta = .252$, small effect). No other comparisons approached significance, **rejecting H3a**, **H3b**, **H3c**, and **H3d**.

Table 2: Means (standard deviations) for each dependent variable (scales ranging from 1 = "strongly disagree" to 5 = "strongly agree") by condition, separated for novices and experts.

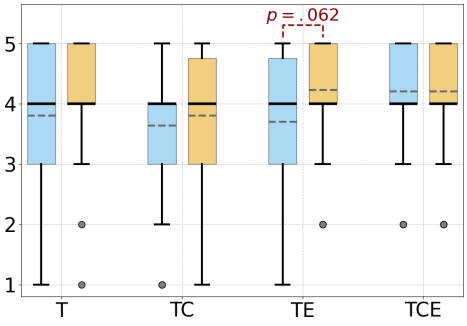
	T		TC		TE		TCE	
	Novices	Experts	Novices	Experts	Novices	Experts	Novices	Experts
Appropriateness	3.10 (1.35)	3.26 (1.15)	2.67 (1.45)	3.25 (1.43)	3.31 (1.35)	3.69 (1.13)	3.91 (1.21)	3.96 (1.14)
Learning	2.86 (1.41)	2.87 (1.34)	2.87 (1.36)	2.86 (1.21)	3.73 (1.28)	3.14 (1.44)	3.97 (0.95)	3.20 (1.22)
Relevance	3.79 (1.15)	4.00 (0.97)	3.63 (1.13)	3.86 (1.04)	3.88 (1.03)	4.23 (0.88)	4.20 (0.90)	4.36 (0.86)
Trust	3.38 (0.90)	3.29 (1.07)	2.83 (1.05)	3.18 (1.25)	3.46 (1.39)	3.69 (0.96)	3.80 (1.05)	3.76 (1.01)
Helpfulness	–	–	2.97 (1.35)	4.04 (1.00)	3.88 (1.21)	4.14 (0.91)	4.20 (1.11)	3.92 (1.15)



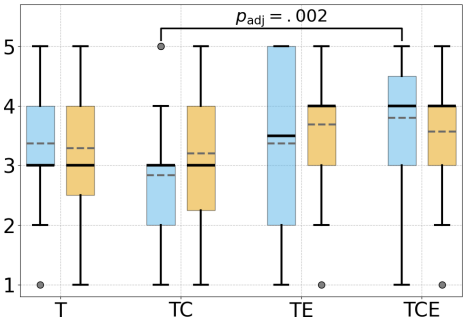
(a) Perceived appropriateness of information quantity



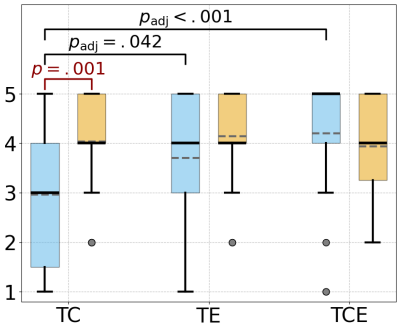
(b) Perceived learning



(c) Perceived relevance



(d) Trust



(e) Perceived helpfulness

Figure 2: Box plots comparing novices' (blue) and experts' (orange) ratings across our dependent variables. Black horizontal lines represent medians, and dashed grey lines indicate means. Black and red brackets indicate significant differences determined by Dunn's post-hoc ($p_{adj} < .05$) and Mann–Whitney U tests ($p < .05$), respectively.

5.1.4 Trust. For novices, the Kruskal-Wallis test indicated a significant difference across conditions ($H = 12.713$, $p = .005$). Dunn's post hoc tests showed that TCE was rated significantly higher than TC ($p_{adj} = .002$, $\delta = .500$, large effect), while no other pairwise comparisons were significant, thereby **partly supporting H4a**. In contrast, experts showed no significant differences among the four conditions ($H = 4.308$, $p = .230$), **rejecting H4b**. Mann–Whitney U tests revealed no significant novice–expert differences within any condition (all $p > .05$), **rejecting H4c** and **H4d**.

5.1.5 Perceived Helpfulness (only for TC, TE, and TCE). A Kruskal-Wallis test among novices ($H = 15.989$, $p < .001$) showed a significant difference in perceived helpfulness across TC, TE, and TCE. Dunn's post hoc tests indicated that novices found both TE and TCE significantly more helpful than TC (TE vs. TC: $p_{adj} = .042$, $\delta = .319$, small effect; TCE vs. TC: $p_{adj} < .001$, $\delta = .553$, large effect), while TE and TCE did not differ significantly ($p_{adj} = .086$, $\delta = .240$, small effect), thereby **partly supporting for H5a**. Experts showed no significant differences among TC, TE, and TCE ($H = .484$, $p = .785$), **supporting H5b**. Mann–Whitney U tests further revealed that experts rated TC as significantly more helpful than novices ($U = 240.5$, $p = .001$, $\delta = .465$, medium effect), while no other novice–expert comparisons were significant (all $p > .05$), thus **rejecting H5c**.

5.1.6 Perceived Accuracy of Attribute Recommendations. Both novices ($M = 3.83$, $SD = 0.86$, $n = 72$, 53 "I don't know") and experts ($M = 3.94$, $SD = 0.85$, $n = 113$, 13 "I don't know") perceived the attribute recommendations provided by the product advisor as fairly accurate.

5.2 Qualitative Feedback

To analyze participants' responses to our open-ended question, "What additional information would have helped you to answer Cleo's questions about your preferences?", we performed a two-stage thematic analysis [10]. First, an independent researcher uninvolved in the study design and data collection conducted an initial inductive coding (i.e., generating initial codes) on all participants' responses in close discussion with the first author. This stage collaboratively developed the coding scheme and produced a final codebook.[6] Subsequently, the independent researcher and the first author independently recoded all responses using the finalized scheme. Responses could receive multiple codes when they encompassed several distinct themes. The coders then reconciled disagreements through discussion until consensus, following negotiated-agreement best practices [33].

The qualitative feedback revealed clear divergences between novices ($n = 125$) and experts ($n = 126$). The most significant difference was in *Attribute Confusion*, where participants struggled to understand technical terms, what each attribute does or is relevant for, and asked for further explanations. For example, P17 (novice in condition TC), expressed, *"If [Cleo] could have given me more information about the different computer terms that [they were] using."*. This was the most mentioned theme by novices (49% overall), most pronounced in conditions T (60%) and TC (60%), with lower frequencies in TE (43%) and TCE (34%). For experts, this was a minor issue (15% overall), with the highest rate appearing in condition TC (23%), followed by T (16%), TE (11%), and TCE (10%). Conversely, experts' primary request was for *Missing Specs & Attributes*–indicating a desire to specify or receive information on additional attributes like price, brand, or weight–(36% overall), with frequencies highest in conditions TE and TCE (both 43%), compared to T (32%) and TC (23%). This theme was far less frequent among novices (12% overall), where it was highest in condition TE (17%), followed by TCE (14%) and TC (13%), and was rarest in T (3%). *P235* (expert in condition TC), for example, stated, *"Maybe give me some brand names of those GPUs or CPUs"*. Other differences were smaller: only experts mentioned needing *Improved Recommendations* (suggestions that the product advisor's attribute recommendations should be refined, 4%), while only novices called the chatbot a *Non-preferred Medium* (and preferring other mediums such as going to a physical store or reading online reviews, 2%). Finally, more than twice as many experts as novices reported that they received enough information and would not have needed additional information compared to novices (13% vs. 6%). For example, *P71* (expert in condition TC) noted: *"Nothing needed as I am very comfortable in laptop technology"*.

Despite these differences, a key commonality was the desire for a *Provision of Options*–where novices noted they would have liked to be provided more than one attribute value recommendation to choose from or a comparison of different value options. *P77* (novice in condition TC) articulated, *"Giving more options for a given category (RAM, GPU, Storage, etc.) and not just what Cleo thinks is the best option"*. Overall, this theme was mentioned by 14% of novices and 17% of experts. For novices, this desire was highest in conditions TC (20%) and TE (17%), and lower in TCE (11%) and T (10%). For experts, rates were highest in TE and TCE (20% each) and lower in T and TC (13% each). Similarly, both groups expressed a desire for more *Personalization* (i.e., tailoring the advisor's questions and explanations to a user's specific use cases, 13% for experts vs. 8% for novices) and better *Interaction Quality* (such as wanting to ask questions, have a more natural conversation, or have the system confirm its understanding, 14% for novices, 11% for experts).

[6]The full codebook is available at https://osf.io/be9jt/?view_only=c7ecc97a48454e578e860bfd180ea578.

6 Discussion & Implications

Our findings reveal a consistent picture of the novice experience, where the type of supplementary information provided is critical. Novices rated conditions with attribute explanations (TE and TCE) as significantly more helpful than performance categories alone (TC), and they found the combination of both (TCE) significantly more appropriate than the baseline (T) and TC. Qualitative data provides insight into this preference: the theme of *Attribute Confusion* was the major concern for novices, voiced by 60% in the explanation-free baseline (T) and TC conditions. This confusion was mentioned only around half as frequently in the TCE condition, suggesting that novices must first understand an attribute's function and impact before they can benefit from the more abstract performance categories–an interpretation also supported by their significantly higher reported learning in TCE versus TC. Critically, our findings suggest that designing for novices must not come at the expense of the expert experience: experts showed no significant perceptual differences across conditions. Their qualitative feedback indicates that they had different priorities, with their primary request being for *Missing Specs & Attributes* like brand and price, which suggests the supplementary information was not as relevant to their evaluation.

These results can be viewed through several theoretical lenses. From a *communication accommodation theory* [20] and *recipient design* [8, 46] perspective, our findings can be interpreted as an instance of effective knowledge-level accommodation. This form of persona-based personalization for technical products appears to create more suitable interactions for novices, in line with prior work showing positive effects of communication accommodation in chatbots [3, 50, 52, 54]. Our findings may also relate to *common ground theory* [12]: providing explanations (TE and TCE) seems to have facilitated the establishment of a knowledge base in novices, as indicated by their significantly higher perceived learning compared to conditions T and TC. While Cleo's rule-based nature limited dynamic grounding, this initial knowledge-building is a key step toward a shared user-system understanding [58]. Additionally, while we did not measure cognitive load directly, our findings indicate a nuanced view of the *expertise reversal effect* [3] and *cognitive load theory* [56] as inspirations for our hypotheses. While novices seem to have benefited from the supplementary information in TCE as expected, it seems this information did not concern experts, conflicting with our assumptions. The concise, sequential information may have allowed experts to efficiently filter or skim content not immediately relevant to them, served as quick confirmations rather than distractions, or still offered some utility to them, preventing the adverse extraneous cognitive load the expertise reversal effect would predict.

Beyond our primary hypotheses, the qualitative analysis revealed shared desires that our quantitative metrics did not capture. Both groups requested more personalized recommendations and the ability to compare different attribute values rather than receiving a single recommendation. Notably, both novices and experts mentioned personalization most frequently in condition TE (13% novices vs. 20% experts), suggesting that while participants rated the chatbot's supplementary information as broadly relevant across all conditions (with no significant differences found), attribute explanations provided objective information that users would have liked to be catered to their stated use cases. Meanwhile, participants' desire for greater control and agency in their decision-making process reflects established HCI principles (e.g., [37, 49]).

Based on these insights, we propose the following design guidelines for text-based product advisors in technical domains:

1. **Default to TCE as information presentation format:** technical specs + performance categories + attribute explanations.
2. **Use one inclusive interface:** don't hide supplementary information for experts.
3. **Do not present performance categories in isolation:** pair performance categories with an attribute explanation.
4. **Support user agency & personalization:** offer options, let users add missing specs, and tailor explanations to the stated use case.

7 Limitations

Several limitations should be considered when interpreting our findings. First, our study focused on the domain of laptop search. Different accommodation and explanation strategies may be more suitable in other domains, suggesting a need for further domain-specific investigations of knowledge adaptation approaches.

Second, we operationalized expertise using an a priori threshold and used stratified random assignment. Although novices and experts differed substantially in mean knowledge scores and condition-specific responses–suggesting that the categorization was fit for purpose–the approach relies on self-reports, which can be biased [28], and binary thresholds on continuous measures entail information loss and reduced statistical efficiency [17]. We therefore treat "novice" and "expert" as coarse groupings for design and exposition. Future work should supplement self-reports with objective, standardized knowledge tests (e.g., laptop-component quizzes) and analyze domain knowledge as a continuous moderator to obtain more granular inference. The limitation of self-reporting also applies to our measurement of learning.

Finally, the experimental design employed a rule-based advisor and scripted prompts to isolate presentation effects. This control precluded free-form, mixed-initiative dialogue and limited opportunities for dynamic grounding [15], thereby reducing ecological validity. Future studies should replicate the investigation in more open conversational settings.

8 Future Work

Our research opens several promising directions. First, future work could explore alternative communication accommodation strategies, such as varying formality or message length, and examine how our findings translate to voice-based and more interactive conversational agents. The latter would also enable analysis of the dynamic establishment of common ground, e.g., via conversation analysis (e.g., [16]) or by measuring grounding costs (e.g., [25]). Furthermore, developing and evaluating methods for real-time user assessment could enable dynamic adaptation of explanatory depth within a single inclusive interface–scaling detail in response to explicit requests, confusion detected through behavioral or physiological signals, or knowledge inferred from conversational cues (building on [4]) while retaining the minimal explanation that accompanies any category. Our presentation-format findings could also inform prompt engineering or instruction tuning for LLM-based product advisors, where domain-knowledge signals guide adaptive explanation depth.

Second, several design alternatives merit investigation. One promising approach is progressive elaboration: present any performance category with a concise one-sentence explanation, and make additional details accessible via tooltips or expandable message bubbles. This may be especially valuable when users' expertise varies across attributes. Another avenue is to integrate multimedia (e.g., images, videos) to support supplementary information and facilitate learning for users.

Third, future studies could compare user preferences for system guidance versus user autonomy. Analyzing how novices and experts differ in their preference for a question-driven guided interaction versus a direct-query model, as well as their query behavior, could inform the design of more flexible, mixed-initiative systems.

9 Conclusion

Novices in e-commerce often lack sufficient domain knowledge to evaluate complex products by attributes, relying instead on third-party opinions. Conversational commerce offers the opportunity to adapt to users' knowledge levels, provide educational value, and thus support informed decision-making. Our research, therefore, examined how different combinations of technical information, performance categories, and attribute explanations influence user perception in a laptop search scenario.

Our findings demonstrate that approaches including attribute explanations (TE and TCE) were perceived as significantly more helpful and enhanced perceived learning for novices, who perceived the combined approach (TCE) specifically as more appropriate than the baseline (T) and TC in terms of information quantity. This indicates the need for the right kinds of supplementary information to help bridge novices' knowledge gaps. Crucially, we did not detect a significant negative impact on expert users, who showed no significant preference across conditions.

We distill these findings into four design guidelines for text-based product advisors in technical domains: use TCE by default; stick to one inclusive interface; do not present performance categories without an explanation; and preserve user agency (by offering options and letting users add missing specs) and personalize to the stated use case. Together, these guidelines aim to operationalize knowledge-level recipient design for novices without alienating experts. Future research should continue exploring this strategy across different product domains and with more interactive conversational systems.

Acknowledgments

This work was supported by the German Research Foundation (DFG) as part of the "VACOS 2" project (no. 388815326). We thank our study participants and our anonymous reviewers for their helpful feedback. We would also like to thank Jan Lattenkamp for his assistance with the technical implementation.

References

[1] Tarek Ait Baha, Mohamed El Hajji, Youssef Es-Saady, and Hammou Fadili. 2023. The Power of Personalization: A Systematic Review of Personality-Adaptive Chatbots. *SN Computer Science* 4, 5 (30 Aug 2023), 661. doi:10.1007/s42979-023-02092-6

[2] Joseph W. Alba and J. Wesley Hutchinson. 1987. Dimensions of Consumer Expertise. *Journal of Consumer Research* 13, 4 (1987), 411–454. http://www.jstor.org/stable/2489367

[3] Mohammad Amin Kuhail, Mohamed Bahja, Ons Al-Shamaileh, Justin Thomas, Amina Alkazemi, and Joao Negreiros. 2024. Assessing the Impact of Chatbot-Human Personality Congruence on User Behavior: A Chatbot-Based Advising System Case. *IEEE Access* 12 (2024), 71761–71782. doi:10.1109/ACCESS.2024.3402977

[4] Sungeun An, Robert Moore, Eric Young Liu, and Guang-Jie Ren. 2021. Recipient Design for Conversational Agents: Tailoring Agent's Utterance to User's Knowledge. In *Proceedings of the 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 30, 5 pages. doi:10.1145/3469595.3469625

[5] Farshid Anvari and Hien Minh Thi Tran. 2013. Persona ontology for user centred design professionals. In *Proceedings of the 4th International Conference on Information Systems Management and Evaluation*, John Blooma, Mathews Nkhoma, and Nelson Leung (Eds.). Academic Conferences and Publishing International, United Kingdom, 35–44. International Conference on Information Systems Management and Evaluation (4th : 2013) ; Conference date: 13-05-2013 Through 14-05-2013.

[6] Ebrahim Babaei, Tilman Dingler, Benjamin Tag, and Eduardo Velloso. 2025. Should we use the NASA-TLX in HCI? A review of theoretical and methodological issues around Mental Workload Measurement. *International Journal of Human-Computer Studies* 201 (2025), 103515. doi:10.1016/j.ijhcs.2025.103515

[7] Janarthanan Balakrishnan and Yogesh K. Dwivedi. 2021. Conversational commerce: entering the next stage of AI-powered digital assistants. *Annals of Operations Research* (12 Apr 2021). doi:10.1007/s10479-021-04049-5

[8] Mark Blokpoel, Marlieke van Kesteren, Arjen Stolk, Pim Haselager, Ivan Toni, and Iris van Rooij. 2012. Recipient design in human communication: simple heuristics or perspective taking? *Front Hum Neurosci* 6 (Sept. 2012), 253. doi:10.3389/fnhum.2012.00253

[9] Simone Borsci, Martin Schmettow, Alessio Malizia, Alan Chamberlain, and Frank van der Velde. 2023. A confirmatory factorial analysis of the Chatbot Usability Scale: a multilanguage validation. *Personal and Ubiquitous Computing* 27, 2 (01 Apr 2023), 317–330. doi:10.1007/s00779-022-01690-0

[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 2 (Jan. 2006), 77–101.

[11] Saša Brdnik, Tjaša Heričko, and Boštjan Šumak. 2022. Intelligent User Interfaces and Their Evaluation: A Systematic Mapping Study. *Sensors* 22, 15 (2022). doi:10.3390/s22155830

[12] Susan E Brennan. 1998. The grounding problem in conversations with and through computers. In *Social and cognitive approaches to interpersonal communication* (1 ed.). Lawrence Erlbaum, 201–225.

[13] Merrie Brucks. 1985. The Effects of Product Class Knowledge on Information Search Behavior*. *Journal of Consumer Research* 12, 1 (06 1985), 1–16. arXiv:https://academic.oup.com/jcr/article-pdf/12/1/1/5067463/12-1-1.pdf doi:10.1086/209031

[14] Pengshan Cai, Hui Wan, Fei Liu, Mo Yu, Hong Yu, and Sachindra Joshi. 2022. Learning as Conversation: Dialogue Systems Reinforced for Information Acquisition. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (Eds.). Association for Computational Linguistics, Seattle, United States, 4781–4796. doi:10.18653/v1/2022.naacl-main.352

[15] Herbert H. Clark and Susan E. Brennan. 1991. *Grounding in communication.* American Psychological Association, Washington, DC, US, 127–149. doi:10.1037/10096-006

[16] Gregorio Convertino, Helena M. Mentis, Mary Beth Rosson, John M. Carroll, Aleksandra Slavkovic, and Craig H. Ganoe. 2008. Articulating common ground in cooperative work: content and process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1637–1646. doi:10.1145/1357054.1357310

[17] Jamie DeCoster, Marcello Gallucci, and Anne-Marie R. Iselin. 2011. Best Practices for Using Median Splits, Artificial Categorization, and their Continuous Alternatives. *Journal of Experimental Psychopathology* 2, 2 (2011), 197–209. arXiv:https://doi.org/10.5127/jep.008310 doi:10.5127/jep.008310

[18] Santosh D'Mello and Arthur Graesser. 2012. Malleability of students' perceptions of an affect-sensitive tutor and its influence on learning. *Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference, FLAIRS-25* (01 2012), 432–437.

[19] David Gefen, Elena Karahanna, and Detmar W. Straub. 2003. Trust and TAM in Online Shopping: An Integrated Model. *MIS Quarterly* 27, 1 (2003), 51–90. http://www.jstor.org/stable/30036519

[20] Howard Giles and Tania Ogay. 2007. *Communication Accommodation Theory.* Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 293–310.

[21] Elizabeth Goldsmith and Ronald E. Goldsmith. 1997. Gender Differences in Perceived and Real Knowledge of Financial Investments. *Psychological Reports* 80, 1 (1997), 236–238. arXiv:https://doi.org/10.2466/pr0.1997.80.1.236 doi:10.2466/pr0.1997.80.1.236

[22] Tomoya Higuchi and Michimasa Inaba. 2024. Interactive Dialogue Interface for Personalized News Article Comprehension. In *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Tatsuya Kawahara, Vera Demberg, Stefan Ultes, Koji Inoue, Shikib Mehri, David Howcroft, and Kazunori Komatani (Eds.). Association for Computational Linguistics, Kyoto, Japan, 329–332. doi:10.18653/v1/2024.sigdial-1.29

[23] Starr Roxanne Hiltz. 1994. *The virtual classroom: learning without limits via computer networks.* Ablex Publishing Corp., USA.

[24] Nina Hollender, Cristian Hofmann, Michael Deneke, and Bernhard Schmitz. 2010. Integrating cognitive load theory and concepts of human–computer interaction. *Computers in Human Behavior* 26, 6 (2010), 1278–1288. doi:10.1016/j.chb.2010.05.031 Online Interactivity: Role of Technology in Behavior Change.

[25] Leila Homaeian, James R. Wallace, and Stacey D. Scott. 2021. Joint Action Storyboards: A Framework for Visualizing Communication Grounding Costs. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 28 (April 2021), 27 pages. doi:10.1145/3449102

[26] Shen Huiyang and Wang Min. 2022. Improving Interaction Experience through Lexical Convergence: The Prosocial Effect of Lexical Alignment in Human-Human and Human-Computer Interactions. *International Journal of Human–Computer Interaction* 38, 1 (2022), 28–41. arXiv:https://doi.org/10.1080/10447318.2021.1921367 doi:10.1080/10447318.2021.1921367

[27] Mourad Jbene, Abdellah Chehri, Rachid Saadane, Smail Tigani, and Gwanggil Jeon. 2025. Intent detection for task-oriented conversational agents: A comparative study of recurrent neural networks and transformer models. *Expert Systems* 42, 2 (2025), e13712. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/exsy.13712 doi:10.1111/exsy.13712

[28] Samuel C. Karpen. 2018. The Social Psychology of Biased Self-Assessment. *American Journal of Pharmaceutical Education* 82, 5 (2018), 6299. doi:10.5688/ajpe6299

[29] Paul E. Ketelaar, Lotte M. Willemsen, Laura Sleven, and Peter Kerkhof. 2015. The Good, the Bad, and the Expert: How Consumer Expertise Affects Review Valence Effects on Purchase Intentions in Online Product Reviews. *Journal of Computer-Mediated Communication* 20, 6 (11 2015), 649–666. arXiv:https://academic.oup.com/jcmc/article-pdf/20/6/649/19492556/jjcmcom0649.pdf doi:10.1111/jcc4.12139

[30] Eun-Jung Lee. 2016. How perceived cognitive needs fulfillment affect consumer attitudes toward the customized product: The moderating role of consumer knowledge. *Computers in Human Behavior* 64 (2016), 152–162. doi:10.1016/j.chb.2016.06.017

[31] Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and Bill Dolan. 2016. A Persona-Based Neural Conversation Model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Katrin Erk and Noah A. Smith (Eds.). Association for Computational Linguistics, Berlin, Germany, 994–1003. doi:10.18653/v1/P16-1094

[32] Mark T. Maybury and Wolf Wahlster. 1998. *Intelligent user interfaces: an introduction.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1–13.

[33] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (Nov. 2019), 23 pages. doi:10.1145/3359174

[34] Chris Messina. 2024. Conversational commerce. When the point of sale comes to the… | by Chris Messina | Chris Messina | Medium. https://medium.com/chris-messina/conversational-commerce-92e0bccfc3ff. Accessed: 2024-02-15.

[35] Lidiya Murakhovs'ka, Philippe Laban, Tian Xie, Caiming Xiong, and Chien-Sheng Wu. 2023. Salespeople vs SalesBot: Exploring the Role of Educational Value in Conversational Recommender Systems. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 9823–9838.

doi:10.18653/v1/2023.findings-emnlp.657

[36] Myungwoo Nam, Jing Wang, and Angela Y. Lee. 2012. The Difference between Differences: How Expertise Affects Diagnosticity of Attribute Alignability. *Journal of Consumer Research* 39, 4 (03 2012), 736–750. arXiv:https://academic.oup.com/jcr/article-pdf/39/4/736/17931392/39-4-736.pdf doi:10.1086/664987

[37] Jakob Nielsen. 1994. Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA) *(CHI '94)*. Association for Computing Machinery, New York, NY, USA, 152–158. doi:10.1145/191666.191729

[38] Daniel M. Oppenheimer, Tom Meyvis, and Nicolas Davidenko. 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology* 45, 4 (2009), 867–872. doi:10.1016/j.jesp.2009.03.009

[39] Fred G. W. C. Paas. 1992. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology* 84, 4 (1992), 429–434. doi:10.1037/0022-0663.84.4.429

[40] Andrea Papenmeier, Alexander Frummet, and Dagmar Kern. 2022. "Mhm..." – Conversational Strategies For Product Search Assistants. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) *(CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 36–46. doi:10.1145/3498366.3505809

[41] Andrea Papenmeier, Dagmar Kern, Daniel Hienert, Alfred Sliwa, Ahmet Aker, and Norbert Fuhr. 2021. Starting Conversations with Search Engines - Interfaces that Elicit Natural Language Queries. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra ACT, Australia) *(CHIIR '21)*. Association for Computing Machinery, New York, NY, USA, 261–265. doi:10.1145/3406522.3446035

[42] Andrea Papenmeier and Elin Anna Topp. 2023. Ah, Alright, Okay! Communicating Understanding in Conversational Product Search. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 49, 5 pages. doi:10.1145/3571884.3604318

[43] P.S. Raju, Subhash C. Lonial, and W. Glynn Mangold. 1995. Differential Effects of Subjective Knowledge, Objective Knowledge, and Usage Experience on Decision Making: An Exploratory Investigation. *Journal of Consumer Psychology* 4, 2 (1995), 153–180. doi:10.1207/s15327663jcp0402_04

[44] Sergio Román and Dawn Iacobucci. 2010. Antecedents and consequences of adaptive selling confidence and behavior: a dyadic analysis of salespeople and their customers. *Journal of the Academy of Marketing Science* 38, 3 (01 Jun 2010), 363–382. doi:10.1007/s11747-009-0166-9

[45] Jennifer Rowley. 2000. Product search in e-shopping: a review and research propositions. *Journal of Consumer Marketing* 17, 1 (01 Jan 2000), 20–35. doi:10.1108/07363760010309528

[46] Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 4 (1974), 696–735. doi:10.1353/lan.1974.0010 Volume 50, Number 4 (Part 1), December 1974.

[47] Kevin Schott, Andrea Papenmeier, Daniel Hienert, and Dagmar Kern. 2024. What Did I Say Again? Relating User Needs to Search Outcomes in Conversational Commerce. In *Proceedings of Mensch Und Computer 2024* (Karlsruhe, Germany) *(MuC '24)*. Association for Computing Machinery, New York, NY, USA, 129–139. doi:10.1145/3670653.3670680

[48] Ali Sher. 2009. Assessing the Relationship of Student-Instructor and Student-Student Interaction to Student Learning and Satisfaction in Web-based Online Learning Environment. *Journal of Interactive Online Learning* 8 (01 2009).

[49] Ben Shneiderman, Maxine Cohen, Steven Jacobs, Catherine Plaisant, Nicholas Diakopoulos, and Niklas Elmqvist. 2017. *Designing the User Interface Strategies for Effective Human-Computer Interaction, Global Edition.* Pearson Deutschland. 624 pages. https://elibrary.pearson.de/book/99.150005/9781292153926

[50] Michael Shumanov and Lester Johnson. 2021. Making conversations with chatbots more personalized. *Computers in Human Behavior* 117 (2021), 106627. doi:10.1016/j.chb.2020.106627

[51] Paul Chandler Slava Kalyuga, Paul Ayres and John Sweller. 2003. The Expertise Reversal Effect. *Educational Psychologist* 38, 1 (2003), 23–31. https://doi.org/10.1207/S15326985EP3801_4

[52] Laura Spillner and Nina Wenig. 2021. Talk to Me on My Level – Linguistic Alignment for Chatbots. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 45, 12 pages. doi:10.1145/3447526.3472050

[53] Rosann L. Spiro and Barton A. Weitz. 1990. Adaptive Selling: Conceptualization, Measurement, and Nomological Validity. *Journal of Marketing Research* 27, 1 (1990), 61–69. http://www.jstor.org/stable/3172551

[54] Sumit Srivastava, Mariët Theune, and Alejandro Catala. 2023. The Role of Lexical Alignment in Human Understanding of Explanations by Conversational Agents. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 423–435. doi:10.1145/3581641.3584086

[55] John Sweller. 2010. Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review* 22, 2 (01 Jun 2010), 123–138. doi:10.1007/s10648-010-9128-5

[56] John Sweller. 2011. CHAPTER TWO - Cognitive Load Theory. Psychology of Learning and Motivation, Vol. 55. Academic Press, 37–76. doi:10.1016/B978-0-12-387691-1.00002-8

[57] Jay Thakkar, Suresh Kolekar, Shilpa Gite, Biswajeet Pradhan, and Abdullah Alamri. 2024. Evaluating the Adaptability of Large Language Models for Knowledge-aware Question and Answering. *International Journal on Smart Sensing and Intelligent Systems* 17, 1 (2024). doi:10.2478/ijssis-2024-0021

[58] Antonia Tolzin and Andreas Janson. 2025. Uncovering the mechanisms of common ground in human–agent interaction: review and future directions for conversational agent research. *Internet Research* ahead-of-print, ahead-of-print (01 Jan 2025). doi:10.1108/INTR-06-2023-0514

[59] Manos Tsagkias, Tracy Holloway King, Surya Kallumadi, Vanessa Murdock, and Maarten de Rijke. 2021. Challenges and research opportunities in eCommerce search and recommendations. *SIGIR Forum* 54, 1, Article 2 (feb 2021), 23 pages. doi:10.1145/3451964.3451966

[60] Sven Tuzovic and Stefanie Paluch. 2018. *Conversational Commerce – A New Era for Service Business Development?* Springer Fachmedien Wiesbaden, Wiesbaden, 81–100. doi:10.1007/978-3-658-22426-4_4

[61] Debbie Vigar-Ellis, Leyland Pitt, and Pierre Berthon. 2015. Knowing what they know: A managerial perspective on consumer knowledge. *Business Horizons* 58, 6 (2015), 679–685. doi:10.1016/j.bushor.2015.07.005 SPECIAL ISSUE: THE MAGIC OF SECRETS.

[62] Preeti Virdi, Arti D. Kalro, and Dinesh Sharma. 2020. Online decision aids: the role of decision-making styles and decision-making stages. *International Journal of Retail & Distribution Management* 48, 6 (01 Jan 2020), 555–574. doi:10.1108/IJRDM-02-2019-0068

[63] Zhiyuan Wan, Xin Xia, David Lo, and Gail C. Murphy. 2021. How does Machine Learning Change Software Development Practices? *IEEE Transactions on Software Engineering* 47, 9 (2021), 1857–1871. doi:10.1109/TSE.2019.2937083

[64] Yuyu Yang, Kelsey Urgo, Jaime Arguello, and Robert Capra. 2025. Search+Chat: Integrating Search and GenAI to Support Users with Learning-oriented Search Tasks. In *Proceedings of the 2025 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '25)*. Association for Computing Machinery, New York, NY, USA, 57–70. doi:10.1145/3698204.3716446

[65] Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing Dialogue Agents: I have a dog, do you have pets too?. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Iryna Gurevych and Yusuke Miyao (Eds.). Association for Computational Linguistics, Melbourne, Australia, 2204–2213. doi:10.18653/v1/P18-1205

[66] Zhenqi Zhao, Mariët Theune, Sumit Srivastava, and Daniel Braun. 2024. Exploring Lexical Alignment in a Price Bargain Chatbot. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 40, 7 pages. doi:10.1145/3640794.3665576