

Article

When Interfaces “Act for You”: An Eye-Tracking Experiment on Delegation, Transparency Cues, and Trust in Agentic Shopping Assistants

Stefanos Balaskas ^{1,*}, Kyriakos Komis ², Ioanna Yfantidou ³ and Dimitra Skandali ⁴

¹ eGovernment & eCommerce Lab (Innovation & Entrepreneurship), Department of Business Administration, University of Patras, 26504 Patras, Greece

² Department of Electrical and Computer Engineering, School of Engineering, University of Patras, 26504 Patras, Greece; up1079081@upatras.gr

³ Department of Business and Management, Liverpool John Moores University (LJMU), Liverpool L3 5UL, UK; i.yfantidou@ljmu.ac.uk

⁴ Department of Economics, University of Peloponnese, 22100 Tripoli, Greece; dskandali@uop.gr

* Correspondence: s.balaskas@ac.upatras.gr

Abstract

Agentic shopping assistants increasingly move beyond recommending products to executing actions in users' workflows (e.g., adding items to cart, applying coupons, selecting shipping). This shift from advice to delegation raises questions about appropriate reliance, perceived control, and how interface cues support oversight when systems can act. We report a laboratory eye-tracking experiment using a chat-only e-commerce prototype in a mixed 2×2 design: action autonomy varied within participants (recommend-only vs. act-on-behalf, with undo/edit), and transparency cues varied between participants (minimal statements vs. preview + rationale describing what will happen and why). Three standardized shopping tasks were completed by 72 participants. Results included behavioral logs (task time, overrides), areas-of-interest (AOI)-based eye-tracking (chat attention and verification indicators), and post-task self-reports (trust, control, uneasiness, perceived transparency). Act-on-behalf autonomy reduced completion time, but it also increased unease, decreased trust and perceived control, and increased the likelihood of an override, suggesting a trade-off between efficiency and oversight. The autonomy-related penalties for trust and perceived control under act-on-behalf execution were lessened by preview + rationale transparency, which additionally enhanced perceived transparency, trust, and unease. This mechanism coincided with eye-tracking: transparency decreased verification latency during agent actions and redirected attention toward information supplied by assistants. Transparency did not reliably reduce overrides, suggesting that minimal effective transparency can streamline supervision and improve evaluations without eliminating corrective behavior.



Academic Editor: Wei Liu

Received: 3 February 2026

Revised: 21 February 2026

Accepted: 27 February 2026

Published: 1 March 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](https://creativecommons.org/licenses/by/4.0/).

Keywords: agentic interfaces; delegated autonomy; transparency cues; trust in automation; perceived control; oversight and verification; eye tracking; e-commerce assistants

1. Introduction

AI systems undertake more than simply offering recommendations; they are able to execute tasks in a user's workflow, which changes the nature of the interaction from advice to delegation and raises new questions about appropriate reliance [1–3]. Delegation-oriented accounts of supervisory control assert that autonomy is optimally conceptualized

as task-specific roles and authorizations, positing that effective delegation necessitates interfaces enabling users to assign, monitor, and amend actions within a collective task framework [4]. This development matters as users have to decide when to let the system act, when to step in, and how to weigh convenience against responsibility for the results. This has led to research that connects autonomy to trust, perceived control, and monitoring behavior [2,4,5].

The trust-in-automation literature indicates that trust is multifaceted and can be assessed through various self-report, behavioral, and physiological measures. However, even when behavioral evidence points to different dynamics, questionnaires are often preferred in empirical research [6]. Behavioral paradigms emphasize the distinction between expressing and exhibiting trust. Additionally, they demonstrate how vital it is to monitor choices that demonstrate supervision or dependence once users become familiar with how the system functions [4,7]. Real-world examples illustrate how “understanding” automation can become a crucial aspect of trust, surpassing competence and dependability. This suggests that measures confined to laboratory settings may overlook significant factors influencing both intervention and non-intervention [3]. These points collectively suggest that a credible assessment of agentic interfaces necessitates multi-layered measurement rather than relying solely on self-reporting.

However, it is still not clear how these dynamics change when automation moves from recommending to executing within interactive interfaces, and how informational cues can support calibration during that shift [1,2,8]. In driving automation, the same exposure can lead to different levels of trust, depending on what users are told up front (for example, predictability vs. faith) [2]. Failures can temporarily lower trust, but it can be rebuilt through reliable experiences that follow, which suggests that “transparency” interacts with learning over time instead of being a static add-on [2,7]. In multi-subsystem settings, a rise in the frequency and severity of failures consistently increases workload while diminishing performance and trust. However, certain effects reach a plateau, indicating that oversight requirements and trust responses are not linear and cannot be deduced from reliability alone [1]. Field observations demonstrate that operators can exhibit significantly different intervention styles under analogous conditions, suggesting that overarching attitudes may obscure substantial variations in the supervision of automation in practice [3]. This collectively indicates a deficiency in interface-level evidence regarding the mechanisms that facilitate appropriate oversight at the precise moments when systems can act.

Objective process measures are well suited to address this gap; however, a substantial portion of eye-tracking and XAI research has focused on interpretability validation or state inference, rather than evaluating specific autonomy and transparency indicators integrated into user workflows [9–11]. For example, eye tracking and explainable attention maps may serve to compare model saliency to human gaze to find out which visual features make people feel safe. This strengthens interpretability claims while leaving delegation and override behavior primarily up to the user [12,13]. Explainable eye-movement frameworks can also use fixation and saccade features to predict emotional states in VR or recognize levels of situational awareness in monitoring tasks. This demonstrates methodological promise without directly evaluating when users should verify, intervene, or override automated actions [5,8]. The unresolved issue is not the measurability of gaze, but rather the translation of attention and verification signals into actionable guidance for achieving “minimal effective transparency” in interfaces that encompass autonomy in execution.

To fulfill this demand, we conduct a laboratory-based eye-tracking experiment utilizing a chat-only e-commerce prototype that distinctly differentiates autonomy from transparency in a mixed 2×2 design across three standardized shopping tasks. Our contributions are threefold: first, we provide a causal test of how execution autonomy

(recommend-only vs. act-on-behalf) and transparency cues (minimal vs. preview + rationale) shape state trust, perceived control, and delegation-related unease; second, we complement self-reports with behavioral oversight outcomes (task time and override behavior) that directly operationalize supervision and correction; third, we add process evidence via AOI-based attention allocation and verification switching, linking established recommendations to measure trust beyond attitudes to a realistic agentic commerce workflow [6,7]. The study aims to find practical implications for the design of interfaces that facilitate appropriate reliance in systems capable of action, as delineated in the subsequent research questions and hypotheses.

In brief, the results show a consistent efficiency–oversight trade-off when the interface shifts from recommending to acting: act-on-behalf autonomy reduced state trust and perceived control and increased delegation-related unease, while also shortening task completion time and increasing the likelihood of overrides. The preview + rationale transparency cue set successfully increased perceived transparency and improved key evaluations (higher trust, lower unease), and—most importantly—attenuated the trust and control penalties associated with act-on-behalf execution (a reliable Autonomy × Transparency moderation for both outcomes). Process evidence from eye tracking reinforced this account: preview + rationale transparency increased visual attention to the assistant/chat region (greater gaze share and dwell time), particularly under act-on-behalf autonomy, and reduced A1-only verification latency (faster gaze shift to outcome-relevant controls). However, transparency did not increase verification switching and did not reliably reduce override likelihood, suggesting that minimal effective transparency may primarily improve understanding and streamline supervision rather than eliminating corrective intervention.

2. Literature Review and Hypothesis Development

2.1. Action Autonomy in Agentic Interfaces: Delegation, Trust, Control, and Behavioral Trade-Offs

Autonomy in agentic systems is frequently regarded as a singular “more-or-less” dial; however, the literature indicates it comprises a collection of capacities—perception, decision, and execution—that redefine responsibility and control dynamics [14,15]. Theoretical research on artificial agents elucidates that “acting” is not merely “recommending more strongly”, but represents a unique form of delegated agency with distinct ramifications for oversight and accountability [16]. Simultaneously, syntheses of Levels-of-Automation taxonomies indicate that research and design practices are impeded by inconsistent terminology and conceptual ambiguity between automation and autonomy, complicating the comparison and translation of empirical findings into interface requirements [17]. This indicates that autonomy must be defined in specific, user-oriented terms—specifically, what the system accomplishes—prior to theorizing its effects.

Autonomy is often portrayed as enhancing efficiency in various applied fields; however, the evidence suggests that these benefits are conditional and exaggerated, particularly in idealized representations of the technology’s capabilities. Studies of the implementation of autonomous vehicles for the logistics of e-commerce have noted the discrepancy between the results of simulation and the feasibility of the technology, particularly in critical situations such as health crises, showing that the technology is not as “autonomous” as might be assumed [18–20]. In workplace AI, case-based evidence contends that dependable autonomy necessitates “keeping the organization in the loop”, wherein decision workflows and continuous customization practices determine whether AI enhances or detracts from socio-technical functioning [15]. These viewpoints converge on a fundamental assertion: autonomy is not merely a technical attribute but also a governance challenge, emphasizing trust and supervision.

Research on trust additionally shows that autonomy can lead to miscalibration (under-trust/overtrust) along with dynamic shifts tied to cognitive state and uncertainty [21–23]. Multimodal approaches integrate attention, stress, and performance metrics to assess trust over time, offering the potential for adaptive teaming while also eliciting concerns regarding their feasibility and generalizability beyond controlled environments [12,14]. Related experimental research indicates that conveying uncertainty through various modalities alters perceived transparency and decision-making strategies, possibly leading to trade-offs between informativeness and user burden [21]. Oversight-oriented accounts assert that effective autonomy necessitates dependable error detection and introduce signal-detection logic to differentiate sensitivity from response bias—an essential rectification to trust-centric interpretations [19]. Nonetheless, numerous studies continue to inadequately assess the behavioral costs of autonomy in genuine interaction flows.

Our research addresses this gap by distinguishing between recommend-only and act-on-behalf autonomy within a commerce-like interface, linking it to trust, perceived control, unease, and observable corrective behavior (such as overrides and time), supplemented by attention-based verification indicators. Based on this positioning, we formalize the anticipated autonomy-related effects by analyzing how the transition from recommendation to execution alters users' evaluations and corrective actions within the shopping workflow, resulting in the following research question and hypotheses:

RQ1. Action autonomy effects. How does action autonomy (recommend-only vs. act-on-behalf execution) influence users' evaluations and behavioral oversight during shopping tasks?

H1a (Trust). *State trust will be lower when the assistant acts on the user's behalf than when it only recommends.*

H1b (Control). *Perceived control will be lower under act-on-behalf execution than under recommend-only.*

H1c (Unease). *Delegation unease will be higher under act-on-behalf execution than under recommend-only.*

H1d (Behavioral trade-off). *Act-on-behalf execution will reduce task completion time but increase corrective intervention (overrides) compared to recommend-only.*

2.2. Transparency Cues as a Design Intervention: Rationale and Action Preview Effects

Transparency cues are widely proposed as a practical remedy to the “black-box” problem, but research on human factors shows that transparency is more than just adding explanations [24–26]. It involves providing users a clear idea of what the system did or will do and why. In recommender contexts, explainability can boost trust when it helps users relate outputs to understandable reasons. However, as Shin et al. [27] demonstrate, users also assess whether explanations are causable—that is, whether they are genuinely comprehensible and foster emotional confidence—as opposed to merely being superficially instructive. This distinction is important because transparency may be viewed as “window dressing” if it does not fit users' decision-making criteria and cognitive frameworks. Evaluations should therefore differentiate between explanation quality, user comprehension, and appropriate reliance [24,28,29].

Empirical research indicates that transparency can enhance outcomes; however, the effects differ across trust-as-attitude versus trust-as-behavior, user capabilities, and explanation formats [27,30,31]. In competitive decision-making tasks, XAI enhanced decision accuracy and behavioral trust, even when self-reported trust remained unchanged. Furthermore, explanations may induce placebo-like effects in low-capacity users, suggesting that

“explanations” can serve as heuristic cues rather than merely aids for comprehension [25]. Feature-based explanations might not help people do better and could make them rely too much on AI when they don’t understand how it works. However, by anchoring intuition about results, features, and AI limitations, example-based explanations can more successfully support comprehensive human-AI performance [24]. These trends collectively point to a conflict between efficiency and accountability: transparency must control the risk that decision-speeding assistance could impair performance when AI guidance is flawed [29].

We remain uncertain about the implications of transparency dynamics when systems transition from advisory roles to action execution, particularly in contexts where oversight becomes evident (e.g., through overrides) and trust may be reflected in diminished monitoring rather than affirmative attitudes [32–34]. Philosophical perspectives elucidate this notion by contending that explainability cultivates justified trust only to the extent that it alters established beliefs regarding trustworthiness and the necessity of oversight, occasionally transitioning the “trusted entity” towards an AI–user dyad rather than solely the AI [34]. High-stakes decision support reviews consistently emphasize the necessity for reliance calibration mechanisms that maintain agency, frequently concluding that simpler, action-relevant presentations are more effective in mitigating overreliance than intricate explanation interfaces [30,35–37].

Our research investigates the topic in question by evaluating a minimal, interface-native transparency cue—concise rationale coupled with action preview—in a commerce-oriented assistant that either suggests or performs actions, linking perceived transparency with trust, control, discomfort, and behavioral oversight (including overrides and time management) as well as attention-based verification. We therefore test whether preview + rationale transparency (vs a minimal cue) produces measurable improvements in perceived transparency and downstream evaluations and behaviors during shopping tasks. This leads to the following research question and hypotheses:

RQ2. Transparency cue effects. How does Transparency Cue Set (B0 minimal vs. B1 preview + rationale) influence perceived transparency, evaluations, and behavior?

H2a (Manipulation check). *Perceived transparency will be higher in B1 than in B0.*

H2b (Trust/control/unease). *Compared to B0, B1 will increase state trust and perceived control and decrease delegation anxiety/unease.*

H2c (Behavior). *Compared to B0, B1 will reduce override behavior.*

2.3. When Transparency Matters Most: Moderation by Autonomy and the Attention/Verification Mechanism

Transparency does not function uniformly in human–AI interactions; its efficacy is contingent upon the timing of autonomy transfer of responsibility from the human to the system, and whether transparency indicators stimulate verification rather than complacency [38–40]. Studies on automation monitoring indicate that delegation may diminish verification practices, despite redundancy being designed to enhance reliability. In safety-critical instances, having several individuals to monitor without individual accountability leads to fewer cross-checks and missed failures [38,40,41]. This phenomenon is considered to be due to social loafing rather than a lack of information [39]. This implies a key boundary condition: transparency alone is inadequate if autonomy diminishes users’ perceived obligation to monitor, necessitating increased focus on behavioral verification rather than self-reported trust.

Eye-tracking studies provide converging evidence that transparency cues affect attention allocation, albeit not always as intended [9,13,42,43]. In policy communication, visual

attention is a better predictor of recall and behavioral impact than exposure alone. This is the case given that slogans and emotionally charged elements tend to get more attention early on [43]. In clinical decision support, unsafe AI recommendations consistently garner attention, whereas various forms of explainable AI (XAI) did not succeed in enhancing focus on explanations or in “rescuing” decisions, indicating a discrepancy between perceived utility and actual attentional engagement [9]. In medical image inspection, analogous patterns arise: AI assistance shifts attention to interface components and enhances transitions, yet it may prolong task duration by disrupting optimized visual strategies [41]. Overall, transparency can reconfigure attention without necessarily improving verification quality, and it may impose attentional costs that must be treated as a design trade-off.

Work on appropriate trust further complicates the picture by questioning whether visual attention reliably indexes healthy distrust [9,10]. Peters et al. [44] show that misclassifications render it more challenging to pay attention, yet they rarely make judgments better, which implies that more attention doesn’t always mean better oversight. In contrast, process-control studies establish a direct connection between inadequate information sampling and complacency, leading to commission errors. They also illustrate that specific interventions, such as exposure to failures, can reinstate verification behavior [40]. These results suggest that verification behavior, not explanations on their own, is the mechanism through which transparency should be evaluated.

Design experiments manipulating transparency level reinforce this mechanism-based view [4,26,45]. In certain scenarios, highly transparent agent reasoning can make people less complacent. However, excessive or poorly structured information can confuse users and impair decision-making, leading to superficial trust [10]. Conversational search interface studies also show that high-visibility source presentations keep consumers more visually engaged, nevertheless they frequently fail to lead to behavioral use. They can likewise render learning harder at first unless they are actively integrated into the task flow [46]. Recent research on explanation scope substantiates that joint explanations enhance comprehension via mediated visual attention, contingent upon alignment with cognitive load and modality [1,25,47].

The question of when transparency matters most, or more specifically, whether succinct, action-linked transparency works best when autonomy increases execution power, remains unanswered. By testing preview + rationale cues under recommend-only versus act-on-behalf autonomy and tracking their effects through attention allocation and observable verification (overrides, gaze switching), our study narrows this gap and unifies autonomy, transparency, and oversight into a single behavioral mechanism:

RQ3. Autonomy × transparency interaction. Does a preview + rationale cue mitigate the evaluation and oversight costs of act-on-behalf execution?

H3 (Autonomy × Transparency). *Preview + rationale transparency (vs minimal) will have stronger beneficial effects when the assistant executes actions than when it only recommends—specifically, it will (a) increase trust, (b) increase perceived control, and (c) reduce delegation unease and corrective intervention (overrides) under act-on-behalf execution.*

Building on the main-effect predictions, we test a mechanism-oriented account: transparency should matter most at the moment responsibility shifts to the system (act-on-behalf autonomy), and its influence should be observable not only in self-reports and overrides but also in gaze-based attention and verification dynamics. Thus:

RQ4. Attention allocation and verification behavior. Do transparency cues shift visual attention toward assistant-provided information and verification behavior, consistent with the intended manipulation?

H4a (Attention allocation). Compared to B0, B1 will increase attention to assistant information (greater dwell/gaze share in the assistant/chat region).

H4b (Verification switching). Compared to B0, B1 will increase verification-related switching between assistant information and decision controls.

H4c (A1-only verification latency). In act-on-behalf tasks (A1), B1 will reduce action verification latency (faster gaze shift to updated cart/controls after an assistant action).

3. Research Methodology

3.1. Experimental Design

A lab-based, randomized experiment utilizing a mixed 2×2 design was conducted for this purpose [48,49]. The within-subject factor is Action Autonomy, which has two levels: recommend-only (the assistant gives suggestions and the participant performs all the actions) and act-on-behalf (the assistant does a set of shopping actions in the interface, and the participant can undo or edit them) [50–52]. The between-subject factor is Transparency Cue Set. There are two levels: minimal (short action statements without explanation or preview) and preview + rationale (short “what/why” transparency, including a short reason and an action preview). Participants perform three standardized shopping tasks within a controlled e-commerce prototype, which includes product listing/detail pages, cart, coupon and shipping modules, and a final “checkout review” screen without actual payment. They interact with a chat-only assistant integrated into the interface during this process. The autonomy order is counterbalanced (recommend-only → act-on-behalf vs. act-on-behalf → recommend-only), and the assignment of tasks to autonomy is rotated so that each task appears under both autonomy modes in the sample, educing task-difficulty confounding.

In the recommend-only condition, the assistant suggests a product (and, if necessary, coupon/shipping choices) and provides brief instructions. The participant performs the corresponding clicks. In the act-on-behalf condition, the assistant executes key actions (such as adding items to the cart, applying a coupon if applicable, choosing shipping, and moving on to the review) that are clearly displayed in the interface. Participants are able to modify or undo their choices. Transparency is implemented via standardized message templates. In the minimal condition, the assistant gives recommendations and actions without any explanation. In the preview + rationale condition, the assistant additionally provides a short explanation of the action chain and an in-depth description of the constraint (e.g., price/rating/delivery). Table 1 demonstrates the 2×2 manipulations, and Figures 1 and 2 illustrate the difference in transparency (B0 vs. B1) when the recommend-only operation (A0) is applied, where the cart remains empty until the participant executes the selection.

Table 1. Experimental conditions in the 2×2 Autonomy \times Transparency design (operationalization of manipulations).

Autonomy Mode (Within-Subject)	B0 Minimal (No Preview/Rationale)	B1 Preview + Rationale (Brief “What/Why” + Action Preview)
A0 Recommend-only (assistant suggests; user executes)	Chat content: Product recommendation (+optional coupon/shipping suggestion). User action: Participant clicks Add-to-cart/coupon/shipping. Interface cue: Cart/order module remains unchanged until user action.	Chat content: Recommendation + brief constraint-based rationale + preview of intended action sequence. User action: Participant clicks Add-to-cart/coupon/shipping. Interface cue: Cart/order module remains unchanged until user action.

Table 1. Cont.

Autonomy Mode (Within-Subject)	B0 Minimal (No Preview/Rationale)	B1 Preview + Rationale (Brief "What/Why" + Action Preview)
A1 Act-on-behalf (assistant executes; user can undo/edit)	<p>Chat content: Assistant executes predefined actions (add-to-cart; apply coupon if applicable; select shipping; proceed to review) and confirms completion.</p> <p>Interface cue: Cart/order module updates automatically; user can undo/edit.</p>	<p>Chat content: Same executed actions + brief rationale + action preview (constraint checks + planned action chain).</p> <p>Interface cue: Cart/order module updates automatically; user can undo/edit.</p>

Note. Transparency is manipulated between participants (Prototype A = B0; Prototype B = B1). Autonomy varies within participants (A0 vs. A1), counterbalanced in order and rotated across tasks.

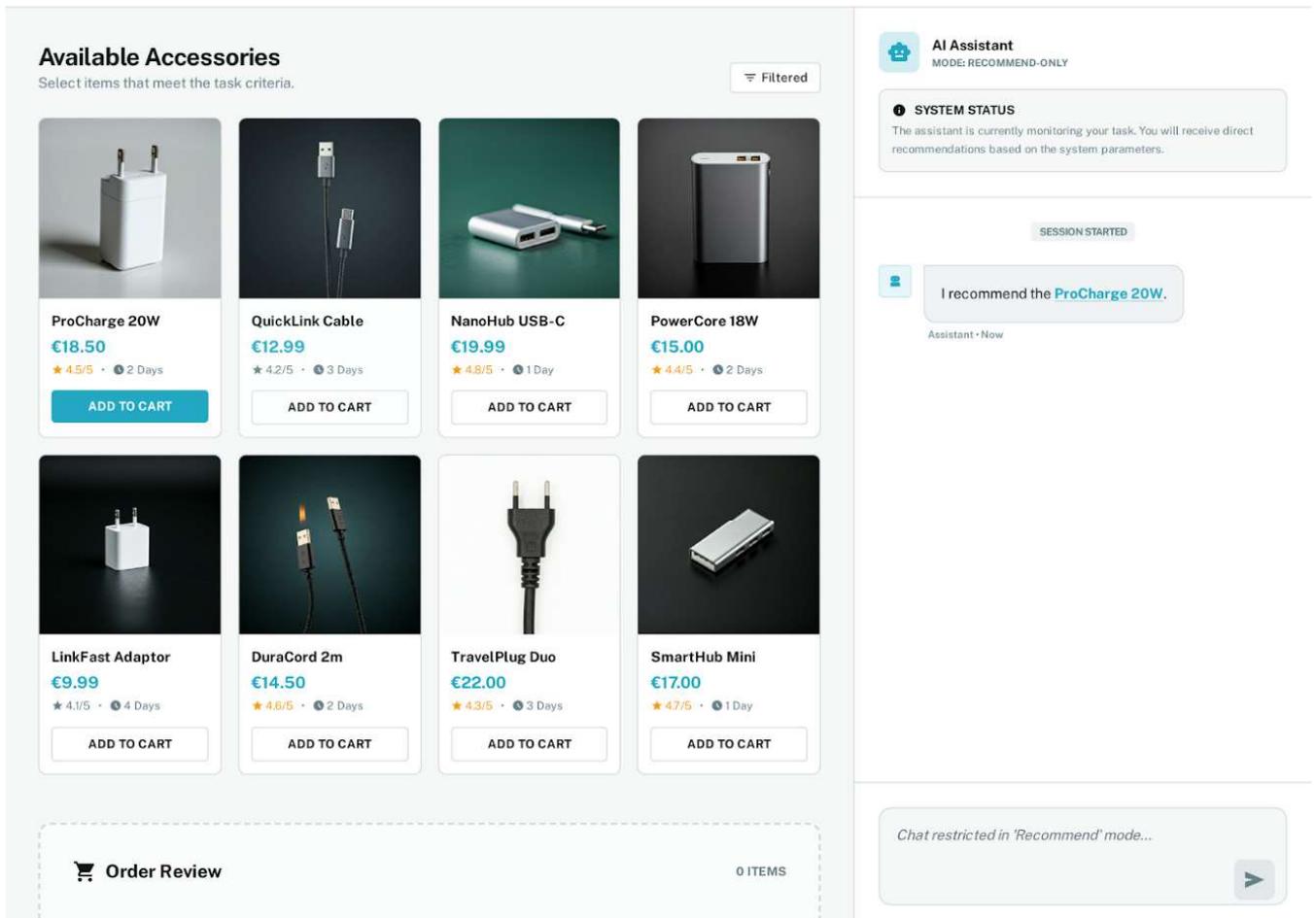


Figure 1. Prototype A (B0 Minimal) recommendation state task 1.

ACTIVE OBJECTIVE

Task 1: Buy an accessory under €20, rating ≥ 4.3 , delivery ≤ 3 days. (No coupon)

The interface displays a grid of eight recommended accessories:

Product	Price	Rating	Delivery
ProCharge 20W	€18.99	4.6★	2 days
UltraLink USB-C	€12.50	4.4★	3 days
MagSafe Adapter	€19.00	4.3★	1 day
Silicon Case	€15.00	4.1★	5 days
QuickCharge Hub	€24.99	4.8★	2 days
Braided Cable	€9.99	4.2★	4 days
Screen Protector	€7.50	4.5★	2 days
Earbud Tips	€5.00	4.0★	1 day

Task 1 Active (ASSISTANT MODE-B1)

Shopping Assistant

I recommend the **ProCharge 20W** because it perfectly meets your constraints (price, rating, and delivery speed) and represents the best overall match in the current inventory.

LOGIC VERIFICATION

- Price Check: €18.99 < €20.00
- Rating Check: 4.6 > 4.3
- Delivery Window: 2 Days < 3 Days

PROPOSED ACTION PIPELINE

Add to cart → Select shipping → Proceed

Execute Recommendation ⚡

Order review module is empty

Your cart is currently empty. Select a product to begin the checkout sequence.

[Refresh Grid](#)

Need Help? Contact Lab Supervisor

RESEARCH PROTOTYPE – NO REAL PAYMENT INFORMATION REQUESTED

Figure 2. Prototype B (B1 Preview + rationale) recommendation state task 1.

3.2. Participants

For a controlled lab session, we recruited $N = 80$ adults. Social media and university online calls and postings were used to recruit participants from the general adult population, mainly those who were community members or affiliated with the university. English was used to administer the study. Age ≥ 18 , fluency in English, and prior online shopping experience were prerequisites for eligibility (assessed by a quick pre-study questionnaire). Following the application of predetermined eye-tracking quality and manipulation-check criteria, $N = 72$ usable participants (roughly 35 from each transparency group) made up the final analytical sample. The age range of the recruited sample ($N = 80$) was 18–45 years old ($M = 29.38$, $SD = 7.13$). Gender was reported by one participant as non-binary ($n = 2$), female ($n = 48$), and male ($n = 29$). High school ($n = 19$), bachelor's ($n = 36$), master's ($n = 17$), doctorate ($n = 6$), and other ($n = 2$) were the educational levels. Basic demographic information and past interactions with AI assistants were collected to outline the sample. The frequency of regular online shopping was reported by the participants ($M = 3.14$, median = 3). Additionally, prior experience with AI tools was prevalent ($M = 3.13$, median = 3). In particular, 23 participants (29.1%) stated that they had “Never” used AI for shopping, while the remaining participants (Occasionally = 19; Once = 13; Monthly = 8; Weekly = 11; Daily = 5) said they had at least some prior exposure.

3.3. Experimental Materials, Apparatus and Procedure

We utilized a Gazepoint GP3 remote eye tracker (Gazepoint Research Inc., Vancouver (British Columbia), Canada) to record eye movements [53]. This research-grade USB video eye tracker works at 60 Hz. It maintains an accuracy ranging from 0.5–1.0° of visual angle, with head motion tolerance of 25 cm (horizontal) × 11 cm (vertical) and ±15 cm in depth [54,55]. The experiment was conducted on a desktop computer with a 24-inch screen. Using the Gazepoint software (Version 7.2.0), we were able to collect eye-tracking data and analyze it based on areas of interest (AOI). At the start of the session, a 9-point calibration was performed and repeated once if calibration quality was unsatisfactory [54,55]. Participants were comfortably seated, allowing for unrestricted head movement within the device's tracking range. An I-VT approach (velocity threshold 30°/s, minimum fixation duration 80 ms) is used to retrieve fixations [56]. AOIs are consistently defined across tasks (see Appendix A, Table A1).

The stimuli consisted of interactive interface screens in a controlled e-commerce prototype that were displayed in a normal browser environment. Each participant performed three shopping tasks, each lasting 2–4 min. The tasks included product browsing/selection, reviewing an order (cart), selecting the shipping option (Tasks 1–2), and applying a coupon if applicable (Task 2 only). The last screen was a “checkout review” screen without real payment. The timing of each task was based on events. Each task started with a standard task_start event and ended with a task_complete event on the review screen. Participants encountered a short, blank transition screen (1–2 s) between tasks to standardize re-orientation before the subsequent task.

After submitting informed consent, participants filled out a short baseline questionnaire that inquired about their demographics, online shopping frequency, prior AI-assistant use, and baseline scales. Followed by calibration and a short training step. The next step was to perform the three tasks while engaging to the chat-only assistant that was built into the interface. In the recommend-only autonomy mode, the assistant provided product recommendations (and, if applicable, coupon and shipping suggestions) and short instructions. Participants completed all UI actions through clicking. In the act-on-behalf mode, the assistant carried out a set of key actions that were already planned out (such as adding to the cart, applying a coupon when appropriate, choosing shipping, and going to the review) that were clearly shown in the interface. Participants could undo or change their choices, which assisted with override measurement. At the event level, all interactions were recorded, including timestamps and action types. After each task, participants filled out short surveys about their trust, perceived control, how anxious or uneasy they felt about delegating, and how transparent they thought the process was. The session concluded with a post-study questionnaire (comprising manipulation checks and willingness-to-delegate items) and a brief debriefing. Each participant's session lasted about 10 to 12 min.

3.4. Measurement Scales and Areas of Interest

Three types of measures were employed: baseline self-report, objective process data (eye tracking and behavioral logs), and self-report after the task or experiment. Self-report measures were obtained at baseline, subsequent to each task, and at the conclusion of the session utilizing 5-point Likert scales (1 = strongly disagree, 5 = strongly agree). The baseline measures assessed trust propensity, need for control, privacy/data concern, and shopping self-efficacy (see Appendix A, Table A1). After the task, we evaluated state trust (ST), perceived control (PC), delegation anxiety/unease (DA), and perceived transparency (PT). At the same time, eye-tracking recorded how people allocated their attention and verified their behavior across predefined AOIs (Chat, Product information, Decision controls, Navigation/filters). The prototype also recorded the duration of a task

along with how often people used the override feature. End-of-study measures recorded overall trust/acceptance and willingness to delegate, in addition to manipulation checks. Data quality screening adhered to established standards for eye-tracking validity and compliance with manipulation checks. To support replicability, Appendix A reports the exact item set per construct; internal consistency reliability (Cronbach's α) was evaluated for each multi-item scale, and all reported scales showed acceptable reliability for research use ($\alpha \geq 0.70$), with values reported in the Results section.

The e-commerce prototype retained track of event-level interaction logs that had timestamps and action types. The primary behavioral outcomes were task completion time (time_sec, from task_start to task_complete) and override behavior, assessed through reversals of choices made by the assistant. Override events included taking away or changing the application of a coupon, undoing adding to the cart, changing shipping after making a choice, and removing or replacing the product that the assistant suggested. These were aggregated per task as override_count and any_override (0/1). Compliance was optionally coded as whether the participant's final decisions matched the assistant's recommendation for all relevant parts (always for the product and when appropriate for the coupon/shipping). All behavioral events were recorded at the event level with timestamps and then aggregated to the task level to align with the mixed-effects modeling structure (up to three observations per participant).

AOIs were defined a priori to capture four functional regions of the interface: Chat, Product information, Decision controls (add-to-cart/coupon/shipping/proceed), and Navigation/filters. Across participants and tasks, AOIs remained the same. The borders included a small pixel margin (5–10 px) to cut down on misassignment due to small gaze errors. We employed an I-VT approach with a speed limit of $30^\circ/s$ and a minimum fixation time of 80 ms to find fixations [57,58]. The main results of the eye-tracking study were time-to-first-fixation (TTFF), dwell time, fixation count, and AOI transitions (for example, Chat \leftrightarrow Controls). There were also derived indices reflecting reliance on the assistant (for example, chat dwell share) and control-checking.

We consider Chat \leftrightarrow Controls transitions and the A1-only action verification latency as verification-oriented attention because they operationalize targeted information sampling between (i) assistant-provided guidance (recommendation/execution messages) and (ii) outcome-relevant decision state and controls (cart contents, coupon/shipping selection, proceed/undo). In traditional monitoring and visual inspection frameworks, such systematic transitions between an information source and the locus of action/outcome are typically regarded as checking or sampling behavior rather than undirected exploration, especially when Areas of Interest (AOIs) are predetermined by functional role and evaluated in relation to task structure and action events [39,40]. In this manner, transitions measure how often participants check "assistant information" and "decision controls" back and forth, and action verification latency measures how quickly they visually confirm the new state after an assistant action in act-on-behalf tasks.

The percentage of valid gaze samples (valid_gaze_pct) was used to check the validity of eye-tracking for each task. Tasks with less than 70% valid gaze were marked. Participants were included for analysis if a minimum of 2 out of 3 tasks satisfied the validity criterion and successfully passed manipulation checks. For self-report quality, item nonresponse was negligible; scale scores were computed when at least 75% of items per scale were present (otherwise set to missing), and the mixed-effects framework handled missing outcomes at the task level via maximum likelihood estimation.

4. Data Analysis and Results

The data analysis was conducted in Google Colab, using the standard Colab runtime (Python 3.12.12; R 4.5.2) [59]. Analyses were implemented primarily in Python (pandas, numpy, scipy, statsmodels), with plotting in matplotlib. For mixed-effects models and post hoc/simple-effects comparisons, we additionally used R packages commonly employed for these estimators (lme4/lmerTest, emmeans; and for count/GLMM sensitivity models where needed, glmmTMB).

We applied mixed-effects models at the task level (with up to three observations per participant) to test all of the hypotheses [60–62]. To account for repeated measures, we used a random intercept for each participant. The fixed effects comprised autonomy (recommend-only versus act-on-behalf), transparency (minimal versus preview plus rationale), their interaction, and controls for task and counterbalanced autonomy order [61,62]. We utilized linear mixed models to look at self-reported outcomes such as state trust, perceived control, delegation unease, and perceived transparency [62,63]. We additionally employed a linear mixed model to look at completion time, which we log-transformed if it was skewed [58,64,65]. The analysis of override behavior was conducted chiefly using a mixed-effects logistic model (any override vs. none), supplemented by a count model (Poisson) when override counts provided additional insights [57,66–68]. The structure of the eye-tracking outcomes was the same: gaze share to the assistant/chat region was used to test attention allocation (with chat dwell as a secondary measure); verification switching (chat ↔ controls transitions) was tested using a mixed-effects count model; and optional verification latency was tested only within act-on-behalf tasks using a linear mixed model [67,69,70]. The primary confirmatory emphasis was the Autonomy × Transparency interaction; when significant, it was analyzed using estimated marginal means/simple effects (transparency effects within each autonomy mode) [60,61]. Minimal diagnostics included skewness and residual checks for time and dwell/latency, overdispersion for counts, and clear reporting of exclusions based on predefined eye-tracking validity (valid gaze ≥70% per task; inclusion required ≥2 valid tasks) and manipulation-check compliance, with effect estimates presented alongside 95% confidence intervals and *p*-values [57,58,64].

4.1. Sample, Exclusions, and Descriptives

The final analytic sample consisted of 72 participants after applying the preregistered inclusion criteria, which mandated eye-tracking validity of at least 70% per task, a minimum of two-thirds of tasks being valid for each participant, and adherence to manipulation-check protocols. In total, eight participants were excluded: four due to the eye-tracking inclusion criterion and four due to failure in the manipulation check.

Table 2 shows descriptive statistics for Transparency × Autonomy. Perceived transparency was descriptively higher in the preview + rationale condition (B1) than in the minimal transparency condition (B0) across both autonomy modes. Trust and perceived control were lowest in the act-on-behalf with minimal transparency (A1 + B0) condition and highest in the act-on-behalf with preview + rationale (A1 + B1) condition. In contrast, delegation unease was higher in the A1 condition than in the A0 condition, and lower in the B1 condition than in the B0 condition within each autonomy mode. Behaviorally, completion times were shorter under act-on-behalf autonomy (A1; ~150 s) than under recommend-only autonomy (A0; ~177–180 s). Overrides were more frequent under A1 than A0, with small descriptive reductions under B1 relative to B0. For the A1-only action verification latency metric, descriptives suggested shorter latency under B1 than B0 (Table 3).

Table 2. Descriptive statistics by Transparency × Autonomy.

Transparency	Autonomy	Trust	Control	Unease	Perceived Transparency	Time (s)	Override Count	Any Override % (n)	Chat Gaze Share	Dwell Chat (ms)	Chat–Controls Switches
B0 minimal	A0 recommend-only	3.97 (0.50)	4.59 (0.64)	3.72 (0.49)	4.18 (0.54)	176.83 (28.60)	0.41 (0.63)	33.3% (18)	0.278 (0.047)	15,294 (2994)	7.09 (2.77)
B0 minimal	A1 act-on-behalf	3.61 (0.47)	4.03 (0.58)	4.44 (0.59)	4.29 (0.72)	150.14 (32.21)	0.89 (1.02)	57.4% (31)	0.282 (0.057)	17,521 (4295)	13.11 (3.18)
B1 preview + rationale	A0 recommend-only	4.52 (0.46)	4.68 (0.60)	3.24 (0.56)	5.19 (0.58)	179.73 (29.09)	0.35 (0.56)	31.5% (17)	0.333 (0.050)	19,964 (2615)	6.33 (3.14)
B1 preview + rationale	A1 act-on-behalf	5.12 (0.51)	4.83 (0.52)	3.88 (0.63)	5.16 (0.51)	150.15 (29.60)	0.74 (0.76)	55.6% (30)	0.382 (0.056)	25,000 (3916)	9.24 (3.49)

Note. Values are M (SD) at the task level. “Any override” is the percentage of tasks with ≥1 override (task count in parentheses). “Chat gaze share” is the proportion of valid gaze time allocated to the Chat AOI. “Chat–Controls switches” indicates the number of transitions between the Chat and Controls AOIs per task.

Table 3. A1-only verification latency descriptive.

Transparency	n (Tasks)	n (Participants)	Action Verification Latency (ms)
B0 minimal	54	36	850.0 (169.8)
B1 preview + rationale	54	36	522.6 (157.8)

4.2. RQ1-Autonomy Effects (H1a–H1d)

Mixed-effects models with random intercepts for participant (Table 4), indicated that act-on-behalf autonomy (A1) was associated with lower trust ($\beta = -0.345$, 95% CI $[-0.505, -0.185]$, $p < 0.001$) and lower perceived control ($\beta = -0.614$, 95% CI $[-0.829, -0.400]$, $p < 0.001$), but higher delegation unease ($\beta = +0.753$, 95% CI $[+0.543, +0.963]$, $p < 0.001$). For behavioral outcomes, A1 tasks completed faster ($\beta = -25.832$ s, 95% CI $[-30.541, -21.123]$, $p < 0.001$) and showed greater override likelihood (OR = 3.034, 95% CI $[1.338, 6.875]$, $p = 0.0079$). Transparency (preview + rationale; B1) was associated with increased trust ($\beta = +0.552$, 95% CI $[+0.354, +0.751]$, $p < 0.001$) and decreased unease ($\beta = -0.476$, 95% CI $[-0.697, -0.256]$, $p < 0.001$). The Autonomy × Transparency interaction was statistically reliable for trust ($\beta = +0.960$, $p < 0.001$) and control ($\beta = +0.719$, $p < 0.001$). Planned simple-effects analyses showed that B1 increased trust in both autonomy modes, with a larger increase under act-on-behalf (A1: $\Delta = +1.513$, $p < 0.0001$; A0: $\Delta = +0.552$, $p < 0.0001$). For perceived control, B1 improved control only under act-on-behalf (A1: $\Delta = +0.805$, $p < 0.0001$), with no reliable effect under recommend-only (A0: $\Delta = +0.085$, $p = 0.454$) (Table 5).

The RQ1 models also include transparency and the Autonomy × Transparency term. We report transparency main effects next (RQ2) and then interpret interactions (RQ3) as preregistered.

Table 4. Mixed-effects model estimates for self-report and behavioral outcomes (Autonomy, Transparency, and Autonomy × Transparency effects).

DV	Term	β	95% CI	p-Value
Trust (ST)	Autonomy	-0.345	$[-0.505, -0.185]$	<0.001
Trust (ST)	Transparency	+0.552	$[+0.354, +0.751]$	<0.001
Trust (ST)	Autonomy × Transparency	+0.960	$[+0.738, +1.183]$	<0.001
Control (PC)	Autonomy	-0.614	$[-0.829, -0.400]$	<0.001
Control (PC)	Transparency	+0.085	$[-0.139, +0.310]$	0.454
Control (PC)	Autonomy × Transparency	+0.719	$[+0.423, +1.015]$	<0.001
Unease (DA)	Autonomy	+0.753	$[+0.543, +0.963]$	<0.001
Unease (DA)	Transparency	-0.476	$[-0.697, -0.256]$	<0.001
Unease (DA)	Autonomy × Transparency	-0.090	$[-0.381, +0.200]$	0.540

Table 4. Cont.

DV	Term	β	95% CI	<i>p</i> -Value
Time (s)	Autonomy	−25.832	[−30.541, −21.123]	<0.001
Time (s)	Transparency	+3.258	[−2.733, +9.250]	0.284
Time (s)	Autonomy × Transparency	−3.598	[−10.172, +2.977]	0.282

Logistic GLMM				
DV	Term	OR	95% CI	<i>p</i>
Any override	Autonomy	3.034	[1.338, 6.875]	0.0079
Any override	Transparency	0.918	[0.408, 2.066]	0.836
Any override	Autonomy × Transparency	1.010	[0.332, 3.073]	0.986

Table 5. Planned simple-effects tests: Transparency (B1 – B0) within each Autonomy mode.

DV	Autonomy	Δ (B1 – B0)	SE	95% CI	<i>p</i> -Value
Trust (ST)	A0 recommend-only	0.552	0.100	[0.354, 0.751]	<0.0001
Trust (ST)	A1 act-on-behalf	1.513	0.100	[1.314, 1.711]	<0.0001
Control (PC)	A0 recommend-only	0.085	0.114	[−0.139, 0.310]	0.454
Control (PC)	A1 act-on-behalf	0.805	0.114	[0.580, 1.030]	<0.0001

Note. Simple effects estimated via emmeans, averaged over task_id and order_autonomy; 95% confidence intervals shown.

4.3. RQ2-Transparency Effects

4.3.1. Manipulation Check: Transparency (H2a)

Transparency was successfully manipulated. In a mixed-effects model controlling for autonomy, task, and autonomy order, preview + rationale transparency (B1) increased perceived transparency (PT) relative to minimal transparency (B0), $\beta = +1.019$, 95% CI [0.796, 1.241], $p < 0.001$. A sensitivity check using participant-level mean PT yielded the same conclusion (B0 M = 4.231 vs. B1 M = 5.176; Welch $t = 12.098$, $p < 0.001$).

4.3.2. Transparency Effects (H2b–H2c)

In the primary task-level models (Table 6), preview + rationale transparency (B1) was associated with higher trust (ST; $\beta = +0.552$, 95% CI [0.354, 0.751], $p < 0.001$) and lower unease (DA; $\beta = -0.476$, 95% CI [−0.697, −0.256], $p < 0.001$). Consistent with the presence of moderation by autonomy (reported below), the overall main effect of transparency on perceived control (PC) was not statistically reliable ($\beta = +0.085$, $p = 0.454$). In the logistic mixed model, transparency did not significantly predict the likelihood of override (OR = 0.918, $p = 0.836$). Additionally, there was no consistent transparency main effect in the optional secondary time model ($\beta = +3.258$ s, $p = 0.284$).

We next evaluate whether transparency moderated autonomy effects, which is the confirmatory focus of RQ3.

Table 6. RQ2 key tests (Transparency main effect).

DV	Key Term	β	95% CI	<i>p</i> -Value	Decision
Perceived transparency (PT)	Transparency	+1.019	[0.796, 1.241]	<0.001	Supported
Trust (ST)	Transparency	+0.552	[0.354, 0.751]	<0.001	Supported
Control (PC)	Transparency	+0.085	[−0.139, 0.310]	0.454	Not supported (main effect)

Table 6. Cont.

DV	Key Term	β	95% CI	<i>p</i> -Value	Decision
Unease (DA)	Transparency	-0.476	[-0.697, -0.256]	<0.001	Supported
Time (s) (secondary)	Transparency	+3.258	[-2.733, 9.250]	0.284	Not supported
Logistic GLMM					
DV	Key term	OR	95% CI	<i>p</i> -value	Decision
Any override	Transparency	0.918	[0.408, 2.066]	0.836	Not supported

Note: **Bold** indicates primary/confirmatory tests; all other entries are secondary/exploratory and formatting is for readability only (it does not change statistical interpretation).

4.4. RQ3: Autonomy × Transparency Interaction (H3)

The Autonomy × Transparency interaction was statistically reliable for trust and perceived control, indicating that the impact of autonomy depended on whether transparency was minimal (B0) or preview + rationale (B1) (Table 7). The interaction for trust (ST) was significant and positive ($\beta = +0.960$, 95% CI [0.738, 1.183], $p < 0.001$). Transparency increased trust in both autonomy modes, according to planned simple-effects analyses, but the increase was significantly greater under act-on-behalf autonomy (A1: $\Delta = +1.513$, 95% CI [1.314, 1.711], $p < 0.0001$) than under recommend-only (A0: $\Delta = +0.552$, 95% CI [0.354, 0.751], $p < 0.0001$). The interaction was also positive for perceived control (PC) ($\beta = +0.719$, 95% CI [0.423, 1.015], $p < 0.001$). Simple effects indicated no reliable transparency effect under recommend-only autonomy (A0: $\Delta = +0.085$, $p = 0.454$), but a strong improvement in control under act-on-behalf (A1: $\Delta = +0.805$, 95% CI [0.580, 1.030], $p < 0.0001$) (Table 8). On the other hand, neither the override likelihood ($p = .986$) nor the delegation unease (DA; $p = 0.540$) interaction were statistically reliable; the optional time model also revealed no interaction ($p = 0.282$).

Table 7. Interaction effects (Autonomy × Transparency) across outcomes.

DV	Model	Interaction Term (β /OR)	95% CI	<i>p</i>	H3 Supported?
Trust (ST)	LMM	$\beta = +0.960$	[+0.738, +1.183]	<0.001	Yes
Control (PC)	LMM	$\beta = +0.719$	[+0.423, +1.015]	<0.001	Yes
Unease (DA)	LMM	$\beta = -0.090$	[-0.381, +0.200]	0.540	No
Any override	Logistic GLMM	OR = 1.010	[0.332, 3.073]	0.986	No
Time (s) (secondary)	LMM	$\beta = -3.598$	[-10.172, +2.977]	0.282	No (secondary)

Table 8. Simple effects (B1 – B0) within each Autonomy mode (EMMs).

DV	Autonomy	Δ (B1 – B0)	SE	95% CI	<i>p</i> -Value
Trust (ST)	A0 recommend-only	0.552	0.100	[0.354, 0.751]	<0.0001
Trust (ST)	A1 act-on-behalf	1.513	0.100	[1.314, 1.711]	<0.0001
Control (PC)	A0 recommend-only	0.085	0.114	[-0.139, 0.310]	0.454
Control (PC)	A1 act-on-behalf	0.805	0.114	[0.580, 1.030]	<0.0001

The interaction plots demonstrate how transparency significantly impacts the influence of autonomy on users’ assessments. Under minimal transparency (B0), shifting from recommend-only (A0) to act-on-behalf (A1) is associated with lower trust and lower perceived control. These autonomy-related penalties are lessened under preview + rationale transparency (B1); the pattern is particularly noticeable for trust, suggesting that increased transparency can offset (and descriptively reverse) the trust cost of increased autonomy.

The annotated interaction p -values confirm that these moderation patterns are statistically reliable for both outcomes in Figure 3.

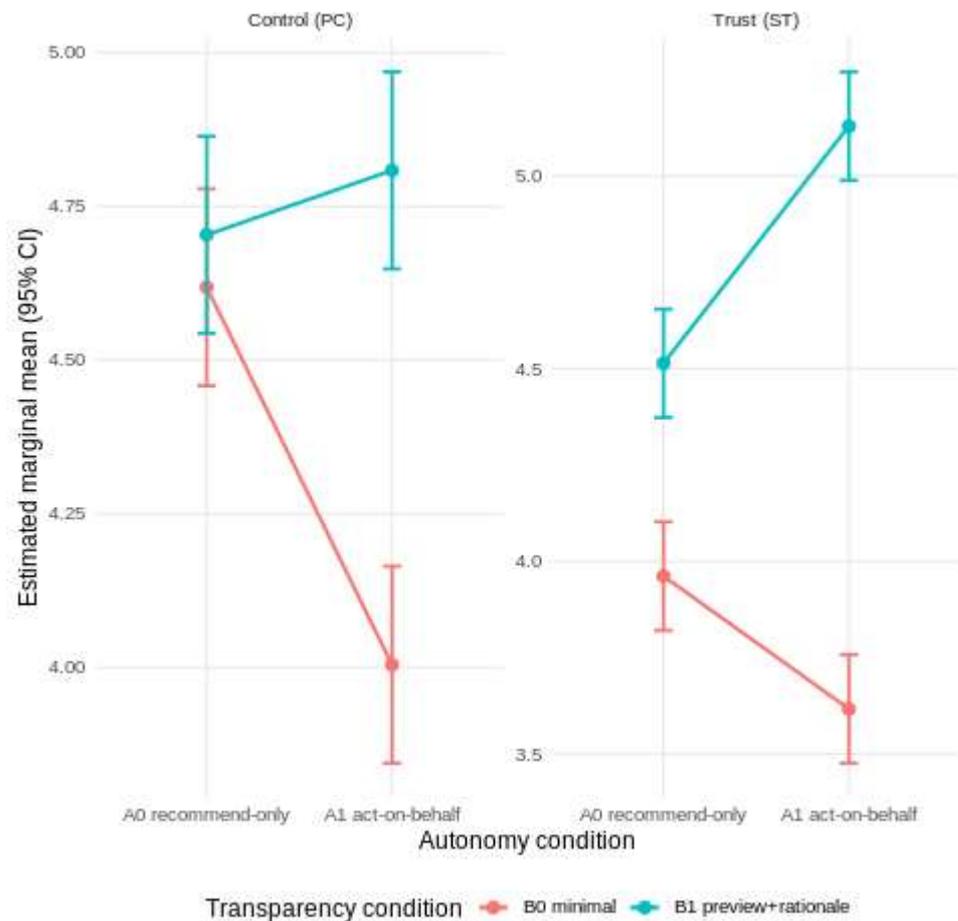


Figure 3. Estimated marginal means (95% CI) for trust (ST) and perceived control (PC) by autonomy and transparency.

4.5. RQ4: Eye-Tracking Outcomes

4.5.1. Attention Allocation to the Chat Interface (H4a)

Transparency reliably increased attention to the chat interface (Figure 4). Preview + rationale transparency (B1) predicted higher gaze share to chat in the primary model ($\beta = +0.0545$, 95% CI [0.0343, 0.0748], $p < 0.001$). B1 predicted a longer chat dwell time in the secondary model ($\beta = +4669$ ms, 95% CI [3356, 5983], $p < 0.001$). The Autonomy \times Transparency interaction was statistically reliable in both situations, suggesting that under act-on-behalf autonomy, transparency more strongly shifted attention toward the chat interface (A1). The model-based estimated marginal means ($\pm 95\%$ CI) for gaze share and chat dwell time (ms) across the 2×2 Autonomy \times Transparency design are displayed as points. The estimates are determined by task-level mixed-effects models with participant random intercepts that account for task and autonomy order.

To contextualize this attention shift, a descriptive task-level scatter (Figure 5) shows that the link between chat attention and trust changes depending on the situation. Under act-on-behalf autonomy (A1), a higher gaze share to chat correlates positively with trust in the context of preview + rationale transparency (B1); conversely, under minimal transparency (B0), this correlation trends negatively. Under recommend-only autonomy (A0), associations are relatively flat. This descriptive pattern enhances the model-based findings by demonstrating how transparency can transform the interpretation of chat-oriented attention (reassurance versus vigilance) in the context of agentic assistance.

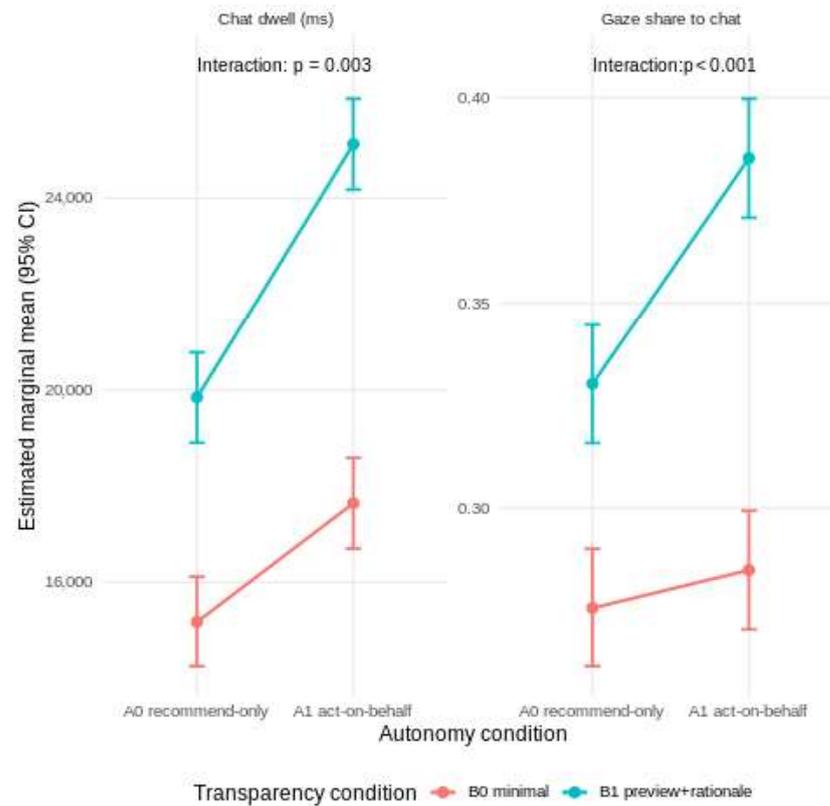


Figure 4. Visual attention to the chat interface by autonomy and transparency.

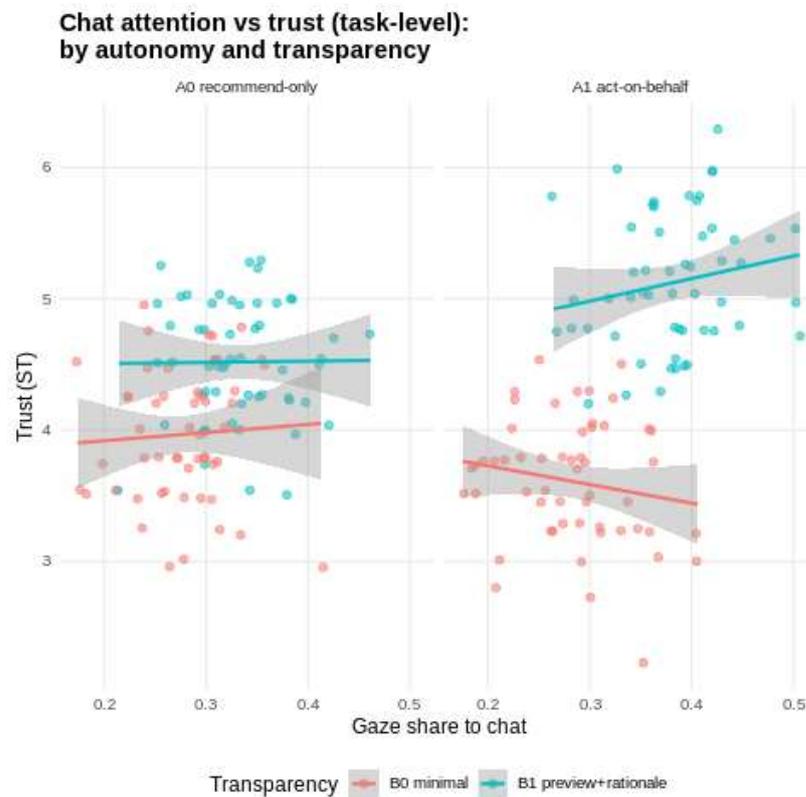


Figure 5. Task-level association between chat attention and trust by autonomy and transparency.

Task observations are represented by points, and linear trends with 95% confidence bands are displayed by lines. The relationship between gaze share to the chat and trust under act-on-behalf autonomy (A1) varies depending on the transparency condition:

preview + rationale transparency (B1) exhibits a positive association, while minimal transparency (B0) exhibits a negative or weak association. Under recommend-only autonomy (A0), associations are comparatively flat. This figure is descriptive and complements the mixed-effects results by illustrating how transparency may alter the interpretation of chat-focused attention during agentic assistance.

4.5.2. Verification Switching Between Chat and Controls (H4b) and Verification Latency (H4c)

The preregistered transparency main effect was not statistically significant in the Poisson mixed model (IRR = 0.893, 95% CI [0.768, 1.040], $p = 0.143$) for verification switching (chat ↔ controls transitions); therefore, H4b (transparency main effect) was not confirmed. The model output also shows a statistically significant Autonomy × Transparency term for switching (IRR = 0.788, 95% CI [0.654, 0.950], $p = 0.0126$). This does not support H4b, which was about the main effect of transparency. However, it is a meaningful interaction pattern that should be highlighted: transparency does not lead to more switching; in fact, it is linked to less switching under A1, which is consistent with “streamlined verification” rather than more back-and-forth (Table 9).

Table 9. Eye-tracking outcomes (key fixed effects).

DV	Term	Effect (Scale)	95% CI	<i>p</i> -Value	Decision?
H4a (primary): gaze_share_chat	autonomy	$\beta = +0.0092$	[−0.0099, +0.0284]	0.3417	-
	cond_transparency	$\beta = +0.0545$	[+0.0343, +0.0748]	<0.001	Yes
	autonomy × cond_transparency	$\beta = +0.0458$	[+0.0194, +0.0722]	0.0008	
H4a (secondary): dwell_chat_ms	autonomy	$\beta = +2471.9$ ms	[+1121.0, +3822.9]	0.0004	Yes
	cond_transparency	$\beta = +4669.4$ ms	[+3355.5, +5983.3]	<0.001	
	autonomy × cond_transparency	$\beta = +2810.9$ ms	[+958.1, +4663.6]	0.0031	
H4b: switch_chat_controls (count)	autonomy	IRR = 1.90	[1.67, 2.16]	<0.001	No
	cond_transparency	IRR = 0.893	[0.768, 1.040]	0.143	
	autonomy × cond_transparency	IRR = 0.788	[0.654, 0.950]	0.0126	
H4c (A1-only): action_verify_latency_ms	cond_transparency	$\beta = -331$ ms	[−396, −266]	<0.001	Yes

Notes. All models are task-level with participant random intercepts. Controls include task_id and order_autonomy for H4a/H4b; for H4c (A1-only) task_id was included and order_autonomy was dropped due to rank deficiency. H4b uses a Poisson GLMM with IRRs (overdispersion ratio ≈ 1.09).

In the A1-only latency analysis, preview and rationale transparency were positively associated with expedited action verification ($\beta = -331$ ms, 95% CI [−396, −266], $p < 0.001$), corroborating H4c. In this subset, the autonomy-order covariate could not be estimated because of rank deficiency, so it was omitted from the fitted model.

4.6. Consolidated Model Summary

To provide a compact overview across outcomes and research questions, Table 10 summarizes the primary fixed effects from the preregistered task-level mixed-effects models using a consistent reporting format (β for LMMs, OR for logistic GLMM, IRR for count GLMM).

Table 10. Main mixed-effects models across outcomes (RQ1–RQ4).

DV (Scale)	Autonomy (A1 vs. A0)	Transparency (B1 vs. B0)	Autonomy × Transparency
Self-report			
Trust (ST)	$\beta = -0.345$ [−0.505, −0.185], $p < 0.001$	$\beta = +0.552$ [+0.354, +0.751], $p < 0.001$	$\beta = +0.960$ [+0.738, +1.183], $p < 0.001$

Table 10. Cont.

DV (Scale)	Autonomy (A1 vs. A0)	Transparency (B1 vs. B0)	Autonomy × Transparency
Control (PC)	$\beta = -0.614 [-0.829, -0.400]$, $p < 0.001$	$\beta = +0.085 [-0.139, +0.310]$, $p = 0.454$	$\beta = +0.719 [+0.423, +1.015]$, $p < 0.001$
Unease (DA)	$\beta = +0.753 [+0.543, +0.963]$, $p < 0.001$	$\beta = -0.476 [-0.697, -0.256]$, $p < 0.001$	$\beta = -0.090 [-0.381, +0.200]$, $p = 0.540$
Perceived transparency (PT) (manipulation check)	$\beta = +0.102 [-0.128, +0.331]$, $p = 0.383$	$\beta = +1.019 [+0.796, +1.241]$, $p < 0.001$	$\beta = -0.148 [-0.463, +0.167]$, $p = 0.355$
Behavioral logs			
Time (s)	$\beta = -25.832 [-30.541,$ $-21.123]$, $p < 0.001$	$\beta = +3.258 [-2.733, +9.250]$, $p = 0.284$	$\beta = -3.598 [-10.172, +2.977]$, $p = 0.282$
Any override (binary)	OR = 3.034 [1.338, 6.875], $p = 0.0079$	OR = 0.918 [0.408, 2.066], $p = 0.836$	OR = 1.010 [0.332, 3.073], $p = 0.986$
Eye-tracking			
Gaze share to chat	$\beta = +0.0092 [-0.0099,$ $+0.0284]$, $p = 0.342$	$\beta = +0.0545 [+0.0343, +0.0748]$, $p < 0.001$	$\beta = +0.0458 [+0.0194, +0.0722]$, $p = 0.0008$
Chat dwell (ms)	$\beta = +2471.9 [+1121.0, +3822.9]$, $p = 0.0004$	$\beta = +4669.4 [+3355.5, +5983.3]$, $p < 0.001$	$\beta = +2810.9 [+958.1, +4663.6]$, $p = 0.0031$
Chat↔Controls switches (count)	IRR = 1.90 [1.67, 2.16], $p < 0.001$	IRR = 0.893 [0.768, 1.040], $p = 0.143$	IRR = 0.788 [0.654, 0.950], $p = 0.0126$
Action verification latency (ms) (A1-only)	—	$\beta = -331 [-396, -266]$, $p < 0.001$	—

Notes. All models are task-level with participant random intercepts and include task_id and order_autonomy as controls (except A1-only latency, where order_autonomy was dropped due to rank deficiency). Count model uses Poisson with IRRs. Hypothesis support summary: H1a–H1d supported; H2a supported; H2b partially supported (trust, unease; not control); H2c not supported (overrides); H3 supported for trust and control; H3c (overrides) and unease interaction not supported; H4a supported; H4b not supported; H4c supported. Bold text denotes DV block headers (measurement domains) for readability only (Self-report, Behavioral logs, Eye-tracking).

5. Discussion

This study examined the influence of delegation-oriented autonomy and interface-native transparency cues on user evaluations, behavioral oversight, and visual attention within a commerce-like workflow. The autonomy manipulation yielded the anticipated trade-off pattern: act-on-behalf autonomy (A1) diminished state trust and perceived control while amplifying delegation unease, concurrently decreasing completion time and elevating the probability of overrides, thereby corroborating H1a–H1d. The transparency manipulation was successful (H2a): preview + rationale (B1) robustly increased perceived transparency. Aside from the manipulation check, B1 enhanced trust and diminished unease (corroborating H2b for trust and unease); however, it did not demonstrate a consistent main effect on perceived control nor did it decrease the likelihood of override (H2c not supported). Consistent with this, the moderation hypothesis for corrective intervention (H3c; reduced overrides under act-on-behalf with preview + rationale) was also not supported, indicating that transparency improved evaluations and verification dynamics without reliably eliminating corrective behavior. According to the foundational moderation logic, transparency was of paramount significance when the system could act: Interactions between autonomy and transparency were dependable for trust and perceived control. Planned simple effects indicated that B1 enhanced trust in both autonomy modes, with a significantly greater effect observed under A1; regarding perceived control, B1 only enhanced control under A1. Thus, H3 was supported for trust and control, but not for unease, time, or override likelihood. Process measures provided evidence of a converging mechanism. Transparency made people pay more attention to the assistant interface. B1 increased gaze share and dwell time in the chat AOI, and both attention measures showed reliable interactions, which suggests that A1 prompted a larger shift in attention (supporting H4a). But verification switching failed to match with transparency (H4b not supported),

and the A1-only latency result showed that verification was considerably faster under B1 (supporting H4c).

5.1. Interpretation: Autonomy Creates Efficiency but Control/Trust Costs (And When)

The results fall in line with delegation-oriented supervisory control accounts, which claim that autonomy does not merely involve “more automation,” but a change authorization and responsibility that reconfigures monitoring demands [4,32]. In our case, transitioning from recommend-only to act-on-behalf heightened perceived accountability while making them feel less able to steer. This imposed a predictable psychological cost: less trust and control and more unease. Crucially, these costs were associated with both increased corrective intervention (higher override likelihood) and objective efficiency gains (faster completion), highlighting the fact that convenience does not equate to calibrated reliance. This substantiates the overarching critique of trust in automation, asserting that self-reported trust is insufficient without corresponding behavioral indicators of reliance and oversight [6,7]. The override pattern indicates a type of “conditional reliance”: users acknowledge the advantages of expedited execution while remaining willing to rectify actions—behavior aligned with oversight-focused narratives that prioritize error detection and response bias over mere trust [8,19,42,43].

The moderation results identify the circumstances in which autonomy has the highest costs: condition A1, which is defined by low levels of transparency, demonstrated the largest deficiencies in trust and control. This situation pertains to the idea of a “governance gap” that has been discussed for socio-technical systems: execution without sufficient interface support for monitoring and change, where users may experience a discrepancy between the system’s delegated execution authority (i.e., its ability to independently execute shopping actions that alter the interface state and results, such as updating the cart, applying coupons, or selecting shipping) and understanding potential [3,15]. Conditions A1 and preview + rationale, on the other hand, produced the highest levels of control and trust, indicating that delegation costs are not a function of execution autonomy but rather of whether the interface provides users with timely and task-relevant cues to comprehend and track execution.

5.2. Transparency as Mitigation: Toward “Minimal Effective Transparency”

The pattern of results aligns with a mechanism-based perspective on transparency, according to which its usefulness is determined by whether it offers information that can be utilized at the moment autonomy transfers responsibility. Although its main purpose was to moderate autonomy effects—significantly strengthening trust under A1 and selectively restoring perceived control where it was most threatened—preview + rationale also significantly improved perceived transparency, overall trust, and unease. According to research, transparency should not only be a surface-level explanation but also be actionable and consistent with user decision-making criteria [27,39,42]. It additionally aligns with accounts of “justified trust” [34], where transparency helps calibrate by updating beliefs about what the system is doing and what oversight is required, and with recommendations that minimal, action-linked transparency is preferable than complex explanation interfaces in high-stakes scenarios [30].

Eye-tracking bolsters this interpretation while enhancing the mechanism. Transparency reliably shifted visual attention toward the assistant interface (greater chat gaze share and dwell), especially under A1, indicating that users engaged more with assistant-provided information when it mattered most. However, transparency did not enhance the switching between chat and controls, corroborating previous research indicating that transparency can alter attention without necessarily augmenting overt checking

behavior [8,9,13,41]. Simultaneously, A1-only verification latency experienced a significant reduction under B1, indicating a more intricate “efficiency” mechanism: preview and rationale may facilitate verification (accelerating the confirmation of outcome states) instead of enhancing back-and-forth monitoring. This helps balance conflicting opinions in the literature: while more attention isn’t always preferable [44], targeted transparency may still enhance the way attention is utilized, encouraging timely, less complicated oversight as opposed to excessive scrutiny.

A noteworthy pattern is that the Autonomy \times Transparency interaction consistently influenced trust and perceived control, but not delegation unease. This indicates that pre-view and rationale transparency predominantly enhance cognitive assessments, comprehending the system’s actions, the reasons behind them, and the methods for verification, which directly bolsters evaluations of predictability, controllability, and reliability. This includes knowing what the system will do, why it did it, and how to check it. On the other hand, uneasiness (like feeling vulnerable, losing initiative, or “being acted for”) is probably a more intense reaction to delegated agency and responsibility and may be less responsive to short, interface-level informational cues. In this case, users may not like delegation itself, but they may also “understand and verify” what the agent does (higher trust/control). Beyond transparency, other de-design levers may be needed in addition to transparency. These could be more robust consent and authorization rituals, default-off action execution, actions that are automatically turned off, the ability to change autonomy limits, or repeated exposure that gradually increases affective comfort.

The main effect prediction for transparency regarding overrides (H2c) and the corresponding moderation prediction for act-on-behalf autonomy (H3c) were not supported. This non-finding is theoretically significant as overrides represent an unclear behavioral outcome. They could entail error correction and distrust, or they could translate to customizing based on preferences, assertive user agency, or “healthy supervision”, where users embrace delegation but still have the option to alter the results. In our tasks, preview and rationale transparency seem to have made oversight easier (faster A1 verification latency) and made trust/control appraisals better, but they didn’t make people less likely to change their minds. This pattern corresponds with the concept that minimal effective transparency functions chiefly as a tool for comprehension and validation, rather than as an obstacle to intervention. Future research should delineate override categories (preventive versus corrective; preference-based versus error-based) and assess whether transparency alleviates negative outcomes (e.g., commission errors, delayed intervention) despite the unchanged frequency of overrides.

6. Practical Implications for Stakeholders and Design of Agentic Interfaces

The results provide stakeholders with helpful guidance on how to use agentic shopping assistants in ways that promote appropriate reliance. These implications are most pertinent to low-to-moderate stakes consumer commerce workflows characterized by immediately observable outcomes and reversible actions (e.g., cart contents, coupon/shipping selections). We do not assert that the same criterion of “sufficiency” applies to high-stakes or irreversible decisions, which may necessitate additional safeguards such as enhanced confirmations, audit trails, or increased friction for commitment.

The primary takeaway for business managers and product owners is that switching from recommend-only to act-on-behalf autonomy can lead to major efficiency gains. However, it may also contribute to trust and perceived control costs unless the interface offers action-linked transparency. An effective layout pattern is to display a short action preview and a brief explanation right before the system is about to execute (and right after it does it as a confirmation), especially for actions that have significant consequences (such as adding

to cart, applying coupons, or changing quantities). This “preview + rationale” cue set appears to be most useful when autonomy is high, because it reduces the A1 trust/control penalties without needing prolonged explanation interfaces [25,39,40].

Design teams should use low-friction intervention points instead of forcing constant monitoring to maintain users in control. Some concrete options are: (a) a single-click “Undo/Revert” button for actions that have already been taken, (b) a persistent, lightweight “Review changes” panel that shows what changed and why, and (c) clear, stable control affordances for overriding (so that fixing the agent doesn’t feel like fighting the interface). Eye-tracking results give more information about where to put verification-relevant information: transparency made people pay more attention to the chat/assistant region and sped up verification, showing that users naturally observe cues provided by the assistant [28,29]. Therefore, previews and rationales should be visually proximal to the conversational turn that triggers the action and use consistent formatting to support fast parsing.

The results indicate to policymakers and educators that transparency requirements should focus on actionability (what will happen, why, and how to reverse it) as opposed to requiring the most detailed explanations. Here, “educators” refers to AI/digital literacy and user-training stakeholders (e.g., consumer digital-literacy instructors, organizational trainers, and platform onboarding teams) who develop guidance and training materials that teach users how to supervise, verify, and override agentic assistants. Training materials and UX standards can help develop “minimal effective transparency” checklists that focus on execution events, reversibility, and clear user authority boundaries. The following could assist in making agentic commerce workflows more convenient and accountable at the same time.

7. Conclusions, Limitations, and Future Research

This study investigated the influence of execution autonomy (recommend-only versus act-on-behalf) and a minimal, interface-native transparency cue set (preview + rationale versus minimal) on user evaluations, oversight behavior, and visual attention within an agentic shopping workflow. In general, act-on-behalf autonomy created a clear trade-off between efficiency and oversight: tasks were done faster, but participants reported lower trust and perceived control and higher unease, and they were more likely to override the assistant’s actions. The preview + rationale cue set effectively enhanced perceived transparency and elevated critical user assessments, with its most significant impact observed under heightened autonomy: transparency mitigated (and descriptively reversed) the trust and control penalties linked to act-on-behalf execution. Eye-tracking metrics yielded corroborative process evidence that transparency redistributes attention to assistant-supplied information and diminishes action verification latency, suggesting that minimal effective transparency can facilitate swifter, more informed oversight, even in the absence of a decrease in corrective behavior.

Several limitations temper generalization and suggest directions for future work. First, the experiment utilized a controlled laboratory prototype devoid of actual payment repercussions and featured a limited set of standardized tasks. Consequently, the stakes, opportunity costs, and social/accountability pressures that affect delegation and oversight in the field may be inadequately represented, and long-term learning dynamics may remain unexamined [2,7]. To prevent overgeneralization, our conclusions are most robust for short, low-stakes consumer shopping interactions where outcomes are promptly observable and actions are reversible. Higher-stakes domains (e.g., financial transactions, health, legal or safety-critical settings) involve different accountability structures, error costs, and tolerance for automation, and may require stronger controls (e.g., explicit confirmations, friction for irreversible actions, audit trails, or mandated human review). Likewise, longer-

term use can alter the relationship between autonomy and transparency through learning, habituation, or complacency. Transparency that is helpful at first may be ignored over time, and failures may only lead to recalibration after repeated exposure. We therefore define the present results as evidence pertaining to interface-level mechanisms (action-linked preview + rationale and event-based verification support) rather than as conclusive directives for all agentic systems.

Replications need to assess scenarios with more repercussions (e.g., actual budgets, postponed outcomes, or return policies) and extended sessions to determine if transparency facilitates consistent calibration as users accumulate experience. Second, while the study integrated self-reporting, behavioral logs, and eye tracking, the operationalizations of the process remain inherently incomplete [6,7]. Verification was predominantly recorded via transitions between chat and controls; however, users can verify within a singular Area of Interest (e.g., prolonged examination of the cart/controls without transitioning) or through micro-patterns of fixations that more effectively differentiate reassurance from vigilance [6,11]. Future research could enhance verification measurement by integrating AOI dwell sequences, fixation-based scan path features, and event-synchronized measures (e.g., gaze in the seconds immediately following an agent action), in conjunction with explicit indicators of “review” behavior within the interface [8,39,42]. Third, the AI-only verification latency model necessitated the exclusion of an order term due to rank deficiency. This does not change the fact that the transparency effect on latency was observed; nevertheless, it indicates that future designs need to ensure adequate variation and balance for subset analyses, or pre-register simpler models for conditional outcomes.

Lastly, transparency did not reliably reduce overrides, which means that better evaluations and faster verification don't always mean fewer corrective actions. Instead of viewing this as a design failure, subsequent research should categorize types of overrides (preventive vs. corrective; preference-driven vs. error-driven) and evaluate whether transparency mitigates adverse outcomes such as commission errors, unwarranted reliance, or delayed intervention. Promising extensions include manipulating reliability and introducing failure episodes to examine recalibration over time [1,40], and systematically comparing explanation formats (feature-based vs. example-based) and user capability moderators to separate comprehension from heuristic “placebo” effects [1,25,47]. Together, these directions can refine actionable guidance on minimal effective transparency for interfaces where AI systems execute actions on users' behalf.

Author Contributions: Conceptualization, S.B.; methodology, S.B.; software, S.B.; validation, S.B.; formal analysis, S.B.; investigation, S.B.; data curation, S.B.; writing—original draft preparation, S.B., K.K., I.Y. and D.S.; writing—review and editing, S.B., K.K., I.Y. and D.S.; visualization, S.B.; supervision, S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Research Ethics and Deontology Committee (E.H.D.E.) of the University of Patras on 6 October 2025 (Protocol/Ref. No.: 16216/175).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions related to human participant data.

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

Appendix A

Table A1. Measurement Codebook.

ID	Variable	Type	Timepoint	Items/Operational Definition	Cronbach's α
D1	Participant ID	Identifier	Pre	Anonymous participant code	N/A
D2	Transparency condition	Experimental factor (between)	Pre (assignment)	B0 = Minimal; B1 = Preview + rationale	N/A
D3	Autonomy order	Counterbalance factor	Pre (assignment)	Recommend \rightarrow Act vs. Act \rightarrow Recommend	N/A
D4	Autonomy mode (per task)	Experimental factor (within)	During (each task)	A0 = Recommend-only; A1 = Act-on-behalf	N/A
D5	Task ID	Task marker	During	Task 1/Task 2/Task 3	N/A
PRE1	Trust Propensity (TP)	Baseline scale	Pre	TP1 rely on automated systems; TP2 trust new tech until reason not to; TP3 cautious about trusting automation (R)	$\alpha = 0.78$
PRE2	Need for Control (NFC)	Baseline scale	Pre	NFC1 prefer final decision myself; NFC2 uncomfortable when system takes initiative; NFC3 stay in control each step	$\alpha = 0.81$
PRE3	Privacy/Data Concern (PDC)	Baseline scale	Pre	PDC1 worry how data is used; PDC2 comfortable sharing preferences (R); PDC3 avoid tools that track behavior; PDC4 use only if data collection is clear	$\alpha = 0.84$
PRE4	Shopping Self-Efficacy (SSE)	Baseline scale	Pre	SSE1 find best option without help; SSE2 confident managing coupon/cart/shipping; SSE3 complete purchases efficiently	$\alpha = 0.80$
ET1	TTF to Chat	Eye-tracking (primary)	During (each task)	Time from task start \rightarrow first fixation in Chat AOI	N/A
ET2	TTF to Controls	Eye-tracking (primary)	During (each task)	Time from task start \rightarrow first fixation in Controls AOI	N/A
ET3	Dwell time: Chat	Eye-tracking (primary)	During (each task)	Total fixation duration in Chat AOI	N/A
ET4	Dwell time: Product info	Eye-tracking (primary)	During (each task)	Total fixation duration in Product AOI	N/A
ET5	Dwell time: Controls	Eye-tracking (primary)	During (each task)	Total fixation duration in Controls AOI	N/A
ET6	Switches: Chat \leftrightarrow Controls	Eye-tracking (primary)	During (each task)	Number of transitions between Chat AOI and Controls AOI	N/A
ET7	Verification latency (A1 only)	Eye-tracking (primary)	During (A1 tasks)	Assistant action timestamp \rightarrow first fixation in updated cart/controls AOI	N/A
BL1	Task completion time	Behavioral log (primary)	During (each task)	Task start \rightarrow task complete event	N/A
BL7	Override count	Behavioral log (primary)	During (each task)	Count of reversals (remove/replace item; undo cart add; change shipping; remove/change coupon)	N/A
BL8	Any override	Behavioral log (primary)	During (each task)	Whether ≥ 1 override occurred	N/A
POST1	State Trust (ST)	Post-task scale	Post-task (each task)	ST1 trusted choices; ST2 reliable; ST3 rely for similar tasks; ST4 needed to double-check (R)	$\alpha = 0.86$

Table A1. Cont.

ID	Variable	Type	Timepoint	Items/Operational Definition	Cronbach's α
POST2	Perceived Control (PC)	Post-task scale	Post-task (each task)	PC1 felt in control; PC2 outcome reflected intentions; PC3 assistant reduced my control (R)	$\alpha = 0.83$
POST3	Delegation Unease (DA)	Post-task scale	Post-task (each task)	DA1 uneasy; DA2 worried it might do something unwanted; DA3 comfortable delegating (R)	$\alpha = 0.88$
POST4	Perceived Transparency (PT)	Manipulation check	Post-task (each task)	PT1 understood what it did/would do; PT2 understood why	$\alpha = 0.76$
END3	Overall Trust/Acceptance (OT)	End scale	End	OT1 overall trust; OT2 would use; OT3 would recommend; OT4 would avoid act-on-behalf assistant (R)	$\alpha = 0.87$
END4	Willingness to Delegate (WTD)	End scale	End	WTD1 low-risk; WTD2 medium-risk; WTD3 high-risk actions	$\alpha = 0.82$

Note. Key constructs and primary outcomes only. Likert items use 5-point scales (1 = strongly disagree, 5 = strongly agree) unless noted. Reverse-coded items marked (R) and recoded prior to averaging. Eye-tracking outcomes are computed per task using the AOs defined in the main text. Cronbach's α values shown are reported for multi-item self-report scales. "N/A" denotes *not applicable* (Cronbach's α is only reported for multi-item psychometric scales; it does not apply to identifiers, experimental factors/conditions, task markers, or single-item and behavioral/eye-tracking measures).

References

- Wang, J.; Fang, W.; Qiu, H.; Wang, Y. The Impact of Automation Failure on Unmanned Aircraft System Operators' Performance, Workload, and Trust in Automation. *Drones* **2025**, *9*, 165. [[CrossRef](#)]
- Lee, J.; Abe, G.; Sato, K.; Itoh, M. Developing Human-Machine Trust: Impacts of Prior Instruction and Automation Failure on Driver Trust in Partially Automated Vehicles. *Transp. Res. Part F Traffic Psychol. Behav.* **2021**, *81*, 384–395. [[CrossRef](#)]
- Balfe, N.; Sharples, S.; Wilson, J.R. Understanding Is Key: An Analysis of Factors Pertaining to Trust in a Real-World Automation System. *Hum. Factors* **2018**, *60*, 477–495. [[CrossRef](#)]
- Miller, C.A.; Parasuraman, R. Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. *Hum. Factors* **2007**, *49*, 57–75. [[CrossRef](#)]
- Bekler, M.; Yilmaz, M.; Ilgin, H.E. Assessing Feature Importance in Eye-Tracking Data within Virtual Reality Using Explainable Artificial Intelligence Techniques. *Appl. Sci.* **2024**, *14*, 6042. [[CrossRef](#)]
- Kohn, S.C.; de Visser, E.J.; Wiese, E.; Lee, Y.C.; Shaw, T.H. Measurement of Trust in Automation: A Narrative Review and Reference Guide. *Front. Psychol.* **2021**, *12*, 604977. [[CrossRef](#)]
- Miller, D.; Johns, M.; Mok, B.; Gowda, N.; Sirkin, D.; Lee, K.; Ju, W. Behavioral Measurement of Trust in Automation: The Trust Fall. In *Proceedings of the Human Factors and Ergonomics Society; Human Factors an Ergonomics Society Inc.: Washington, DC, USA, 2016*; pp. 1842–1846.
- Yao, X.; Chen, C.H.; Liu, B.; Ma, G.; Yu, X. An Explainable Eye-Tracking-Based Framework for Enhanced Level-Specific Situational Awareness Recognition in Air Traffic Control. *Adv. Eng. Inform.* **2026**, *69*, 103928. [[CrossRef](#)]
- Nagendran, M.; Festor, P.; Komorowski, M.; Gordon, A.C.; Faisal, A.A. Eye Tracking Insights into Physician Behaviour with Safe and Unsafe Explainable AI Recommendations. *npj Digit. Med.* **2024**, *7*, 202. [[CrossRef](#)] [[PubMed](#)]
- Wright, J.L.; Chen, J.Y.C.; Barnes, M.J.; Hancock, P.A. The Effect of Agent Reasoning Transparency on Complacent Behavior: An Analysis of Eye Movements and Response Performance. In *Proceedings of the Human Factors and Ergonomics Society; Human Factors an Ergonomics Society Inc.: Washington, DC, USA, 2017; Volume 2017-October*, pp. 1594–1598.
- Ehsan, U.; Passi, S.; Liao, Q.V.; Chan, L.; Lee, I.H.; Muller, M.; Riedl, M.O. The Who in XAI: How AI Background Shapes Perceptions of AI Explanations. In *Proceedings of the Conference on Human Factors in Computing Systems—Proceedings; Association for Computing Machinery: New York, NY, USA, 2024*.
- Rodriguez Rodriguez, L.; Bustamante Orellana, C.E.; Chiou, E.K.; Huang, L.; Cooke, N.; Kang, Y. A Review of Mathematical Models of Human Trust in Automation. *Front. Neuroergon.* **2023**, *4*, 1171403. [[CrossRef](#)]
- Kang, Y.; Chen, J.; Liu, L.; Sharma, K.; Mazzarello, M.; Mora, S.; Duarte, F.; Ratti, C. Decoding Human Safety Perception with Eye-Tracking Systems, Street View Images, and Explainable AI. *Comput. Environ. Urban Syst.* **2026**, *123*, 102356. [[CrossRef](#)]

14. Lin, C.T.; Fan, H.Y.; Chang, Y.C.; Ou, L.; Liu, J.; Wang, Y.K.; Jung, T.P. Modelling the Trust Value for Human Agents Based on Real-Time Human States in Human-Autonomous Teaming Systems. *Technologies* **2022**, *10*, 115. [[CrossRef](#)]
15. Herrmann, T.; Pfeiffer, S. Keeping the Organization in the Loop: A Socio-Technical Extension of Human-Centered Artificial Intelligence. *AI Soc.* **2023**, *38*, 1523–1542. [[CrossRef](#)]
16. Dodig-Crnkovic, G.; Burgin, M. A Systematic Approach to Autonomous Agents. *Philosophies* **2024**, *9*, 44. [[CrossRef](#)]
17. Richardson, L.S.; Fidock, J.; Gunawan, I. Systematic Literature Review of Levels of Automation (Autonomy) Taxonomy: Critiques and Recommendations. *Int. J. Hum. Comput. Interact.* **2025**, *41*, 15824–15843. [[CrossRef](#)]
18. Buldeo Rai, H.; Touami, S.; Dablanc, L. Autonomous E-Commerce Delivery in Ordinary and Exceptional Circumstances. The French Case. *Res. Transp. Bus. Manag.* **2022**, *45*, 100774. [[CrossRef](#)]
19. Langer, M.; Baum, K.; Schlicker, N. Effective Human Oversight of AI-Based Systems: A Signal Detection Perspective on the Detection of Inaccurate and Unfair Outputs. *Minds Mach.* **2025**, *35*, 1. [[CrossRef](#)]
20. Hancock, P.A. Avoiding Adverse Autonomous Agent Actions. *Hum. Comput. Interact.* **2022**, *37*, 211–236. [[CrossRef](#)]
21. Schömb, S.; Pareek, S.; Goncalves, J.; Johal, W. Robot-Assisted Decision-Making: Unveiling the Role of Uncertainty Visualisation and Embodiment. In *Proceedings of the Conference on Human Factors in Computing Systems—Proceedings*; Association for Computing Machinery: New York, NY, USA, 2024.
22. Melo, G.; Nascimento, N.; Alencar, P.; Cowan, D. Identifying Factors That Impact Levels of Automation in Autonomous Systems. *IEEE Access* **2023**, *11*, 56437–56452. [[CrossRef](#)]
23. Halvachi, H.; Asghar, A.; Shirehjini, N.; Kakavand, Z.; Hashemi, N.; Shirmohammadi, S. The Effects of Interaction Conflicts, Levels of Automation, and Frequency of Automation on Human-Automation Trust and Acceptance. *arXiv* **2023**, arXiv:2307.05512.
24. Chen, V.; Liao, Q.V.; Wortman Vaughan, J.; Bansal, G. Understanding the Role of Human Intuition on Reliance in Human-AI Decision-Making with Explanations. *Proc. ACM Hum. Comput. Interact.* **2023**, *7*, 3610219. [[CrossRef](#)]
25. Wang, P.; Ding, H. The Rationality of Explanation or Human Capacity? Understanding the Impact of Explainable Artificial Intelligence on Human-AI Trust and Decision Performance. *Inf. Process. Manag.* **2024**, *61*, 103732. [[CrossRef](#)]
26. Vahdatian, P.; Latifi, M.; Ahsan, M. Designing Trustworthy Recommender Systems: A Glass-Box, Interpretable, and Auditable Approach. *Electronics* **2025**, *14*, 4890. [[CrossRef](#)]
27. Shin, D. The Effects of Explainability and Causability on Perception, Trust, and Acceptance: Implications for Explainable AI. *Int. J. Hum. Comput. Stud.* **2021**, *146*, 102551. [[CrossRef](#)]
28. Andrei, N.; Scarlat, C.; Ioanid, A. Transforming E-Commerce Logistics: Sustainable Practices through Autonomous Maritime and Last-Mile Transportation Solutions. *Logistics* **2024**, *8*, 71. [[CrossRef](#)]
29. Spatola, N. The Efficiency-Accountability Tradeoff in AI Integration: Effects on Human Performance and over-Reliance. *Comput. Hum. Behav. Artif. Hum.* **2024**, *2*, 100099. [[CrossRef](#)]
30. Xu, G.; Murthy, S.V.; Jia, B. Enhancing Intuitive Decision-Making and Reliance Through Human-AI Collaboration: A Review. *Informatics* **2025**, *12*, 135. [[CrossRef](#)]
31. Pfaa, M.; Thomson, R.H.; Hooman, R.R. Measures for Explainable AI: Explanation Goodness, User Satisfaction, Mental Models, Curiosity, Trust, and Human-AI Performance. *Front. Comput. Sci.* **2023**, *5*, 1096257. [[CrossRef](#)]
32. Olateju, O.O.; Okon, S.U.; Olaniyi, O.O.; Samuel-Okon, A.D.; Asonze, C.U. Exploring the Concept of Explainable AI and Developing Information Governance Standards for Enhancing Trust and Transparency in Handling Customer Data. *J. Eng. Res. Rep.* **2024**, *26*, 244–268. [[CrossRef](#)]
33. Shabankareh, M.; Khamoushi Sahne, S.S.; Nazarian, A.; Foroudi, P. The Impact of AI Perceived Transparency on Trust in AI Recommendations in Healthcare Applications. *Asia-Pac. J. Bus. Adm.* **2025**. [[CrossRef](#)]
34. Ferrario, A.; Loi, M. How Explainability Contributes to Trust in AI. In *Proceedings of the ACM International Conference Proceeding Series*; Association for Computing Machinery: New York, NY, USA, 2022; pp. 1457–1466.
35. Muthusubramanian, M.; Jangoan, S.; Kumar Sharma, K.; Krishnamoorthy, G.; Financial Services, D.; America, H. Demystifying Explainable AI: Understanding, Transparency and Trust. *Int. J. Multidiscip. Res.* **2024**, *6*, 1–13.
36. Sahu, G.; Gaur, L. Decoding the Recommender System: A Comprehensive Guide to Explainable AI in E-Commerce. In *Studies in Computational Intelligence*; Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2024; Volume 1094, pp. 33–52.
37. Chaudhary, M.; Gaur, L.; Singh, G.; Afaq, A. Introduction to Explainable AI (XAI) in E-Commerce. In *Studies in Computational Intelligence*; Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2024; Volume 1094, pp. 1–15.
38. Hoesterey, S.; Onnasch, L. The Effect of Risk on Trust Attitude and Trust Behavior in Interaction with Information and Decision Automation. *Cogn. Technol. Work* **2023**, *25*, 15–29. [[CrossRef](#)]
39. Cymek, D.H. Redundant Automation Monitoring: Four Eyes Don't See More Than Two, If Everyone Turns a Blind Eye. *Hum. Factors* **2018**, *60*, 902–921. [[CrossRef](#)]
40. Bahner, J.E.; Hüper, A.D.; Manzey, D. Misuse of Automated Decision Aids: Complacency, Automation Bias and the Impact of Training Experience. *Int. J. Hum. Comput. Stud.* **2008**, *66*, 688–699. [[CrossRef](#)]

41. Castner, N.; Arsiwala-Scheppach, L.; Mertens, S.; Krois, J.; Thaqi, E.; Kasneci, E.; Wahl, S.; Schwendicke, F. Expert Gaze as a Usability Indicator of Medical AI Decision Support Systems: A Preliminary Study. *npj Digit. Med.* **2024**, *7*, 199. [[CrossRef](#)] [[PubMed](#)]
42. Wang, C.; Kroon, A.C.; Möller, J.; de Vreese, C.H.; Boerman, S.C. When Recommendations Are Explainable: An Eye-Tracking Study Comparing How and What to Explain. *Inf. Syst. Front.* **2025**, *28*, 297–315. [[CrossRef](#)]
43. Walker, R.M.; Yeung, D.Y.L.; Lee, M.J.; Lee, I.P. Assessing Information-Based Policy Tools: An Eye-Tracking Laboratory Experiment on Public Information Posters. *J. Comp. Policy Anal. Res. Pract.* **2020**, *22*, 558–578. [[CrossRef](#)]
44. Peters, T.M.; Biermeier, K.; Scharlau, I. Assessing Healthy Distrust in Human-AI Interaction: Interpreting Changes in Visual Attention. *Front. Psychol.* **2026**, *16*, 1694367. [[CrossRef](#)] [[PubMed](#)]
45. Mbelekani, N.Y.; Bengler, K. Learning Design Strategies for Optimizing User Behaviour Towards Automation: Architecting Quality Interactions from Concept to Prototype. In *Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2023; Volume 14048 LNCS, pp. 90–111.
46. He, J.; Liu, J. Not All Transparency Is Equal: Source Presentation Effects on Attention, Interaction, and Persuasion in Conversational Search. *arXiv* **2025**, arXiv:2512.12207. [[CrossRef](#)]
47. Zhang, Y.; Wang, Q.; Chen, S.; Wang, C. How to Rationally Select Your Delegatee in PoS. *arXiv* **2023**, arXiv:2310.08895. [[CrossRef](#)]
48. Mohr, D.L.; Wilson, W.J.; Freund, R.J. *Statistical Methods*; Academic Press: Amsterdam, The Netherlands, 2021; ISBN 0-323-89988-9.
49. Carter, B.T.; Luke, S.G. Best Practices in Eye Tracking Research. *Int. J. Psychophysiol.* **2020**, *155*, 49–62. [[CrossRef](#)]
50. Worthy, D.A.; Lahey, J.N.; Priestley, S.L.; Palma, M.A. An Examination of the Effects of Eye-Tracking on Behavior in Psychology Experiments. *Behav. Res. Methods* **2024**, *56*, 6812–6825. [[CrossRef](#)]
51. Duchowski, A.T.; Duchowski, A.T. *Eye Tracking Methodology: Theory and Practice*; Springer: Berlin/Heidelberg, Germany, 2017; ISBN 3-319-57883-9.
52. Balaskas, S.; Rigou, M. The Effects of Emotional Appeals on Visual Behavior in the Context of Green Advertisements: An Exploratory Eye-Tracking Study. In *Proceedings of the 27th Pan-Hellenic Conference on Progress in Computing and Informatics*, Lamia, Greece, 24–26 November 2023; pp. 141–149.
53. Gazepoint GP3 Eye-Tracker. Available online: <https://www.gazept.com> (accessed on 3 February 2026).
54. Cuve, H.C.; Stojanov, J.; Roberts-Gaal, X.; Catmur, C.; Bird, G. Validation of Gazepoint Low-Cost Eye-Tracking and Psychophysiology Bundle. *Behav. Res. Methods* **2021**, *54*, 1027. [[CrossRef](#)] [[PubMed](#)]
55. Brand, J.; Diamond, S.G.; Thomas, N.; Gilbert-Diamond, D. Evaluating the Data Quality of the Gazepoint GP3 Low-Cost Eye Tracker When Used Independently by Study Participants. *Behav. Res. Methods* **2021**, *53*, 1502–1514. [[CrossRef](#)] [[PubMed](#)]
56. Olsen, A. The Tobii I-VT Fixation Filter. *Tobii Technol.* **2012**, *21*, 4–19.
57. Salvucci, D.D.; Goldberg, J.H. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, Palm Beach Gardens, FL, USA, 6–8 November 2000; pp. 71–78.
58. Komogortsev, O.V.; Gobert, D.V.; Jayarathna, S.; Gowda, S.M. Standardization of Automated Analyses of Oculomotor Fixation and Saccadic Behaviors. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 2635–2645. [[CrossRef](#)] [[PubMed](#)]
59. Bisong, E. *Building Machine Learning and Deep Learning Models on Google Cloud Platform*; Springer: Berlin/Heidelberg, Germany, 2019.
60. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Using Lme4. *J. Stat. Softw.* **2015**, *67*, 48. [[CrossRef](#)]
61. Brown-Schmidt, S.; Naveiras, M.; De Boeck, P.; Cho, S.-J. Statistical Modeling of Intensive Categorical Time-Series Eye-Tracking Data Using Dynamic Generalized Linear Mixed-Effect Models with Crossed Random Effects. In *Psychology of Learning and Motivation*; Elsevier: Amsterdam, The Netherlands, 2020; Volume 73, pp. 1–31, ISBN 0079-7421.
62. Silva, B.B.; Orrego-Carmona, D.; Szarkowska, A. Using Linear Mixed Models to Analyze Data from Eye-Tracking Research on Subtitling. *Transl. Spaces* **2022**, *11*, 60–88. [[CrossRef](#)]
63. Mézière, D.C.; Yu, L.; Reichle, E.D.; Von Der Malsburg, T.; McArthur, G. Using Eye-tracking Measures to Predict Reading Comprehension. *Read. Res. Q.* **2023**, *58*, 425–449. [[CrossRef](#)]
64. Orquin, J.L.; Ashby, N.J.S.; Clarke, A.D.F. Areas of Interest as a Signal Detection Problem in Behavioral Eye-tracking Research. *J. Behav. Decis. Mak.* **2016**, *29*, 103–115. [[CrossRef](#)]
65. Holmqvist, K.; Nyström, M.; Andersson, R.; Dewhurst, R.; Jarodzka, H.; Van de Weijer, J. *Eye Tracking: A Comprehensive Guide to Methods and Measures*; Oup Oxford: Oxford, UK, 2011; ISBN 0-19-162542-6.
66. Friedrich, M.; Rußwinkel, N.; Möhlenbrink, C. A Guideline for Integrating Dynamic Areas of Interests in Existing Set-up for Capturing Eye Movement: Looking at Moving Aircraft. *Behav. Res. Methods* **2017**, *49*, 822–834. [[CrossRef](#)]
67. Müller, H.-G.; Stadtmüller, U. Generalized Functional Linear Models. *arXiv* **2005**, arXiv:math/0505638. [[CrossRef](#)]
68. Barr, D.J. Analyzing ‘Visual World’ Eyetracking Data Using Multilevel Logistic Regression. *J. Mem. Lang.* **2008**, *59*, 457–474. [[CrossRef](#)]

-
69. Bender, R.; Lange, S. Adjusting for Multiple Testing—When and How? *J. Clin. Epidemiol.* **2001**, *54*, 343–349. [[CrossRef](#)] [[PubMed](#)]
70. Khuri, A.I.; Mukherjee, B.; Sinha, B.K.; Ghosh, M. Design Issues for Generalized Linear Models: A Review. *arXiv* **2006**, arXiv:math/0701088. [[CrossRef](#)]

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.