In-Situ Adaptive Interfaces for Online Browsing: Design Dimensions for Intent-Responsive Automation and User Control

Eunhye Kim*
School of Computing, KAIST
Daejeon, Republic of Korea
gracekim027@kaist.ac.kr

Sadat Shams Chowdhury†
School of Computing, KAIST
Daejeon, Republic of Korea
sadatshams@kaist.ac.kr

Hyungwoo Song
Seoul National University
Seoul, Republic of Korea
rotto95@snu.ac.kr

Bongwon Suh
Seoul National University
Seoul, Republic of Korea
bongwon@snu.ac.kr

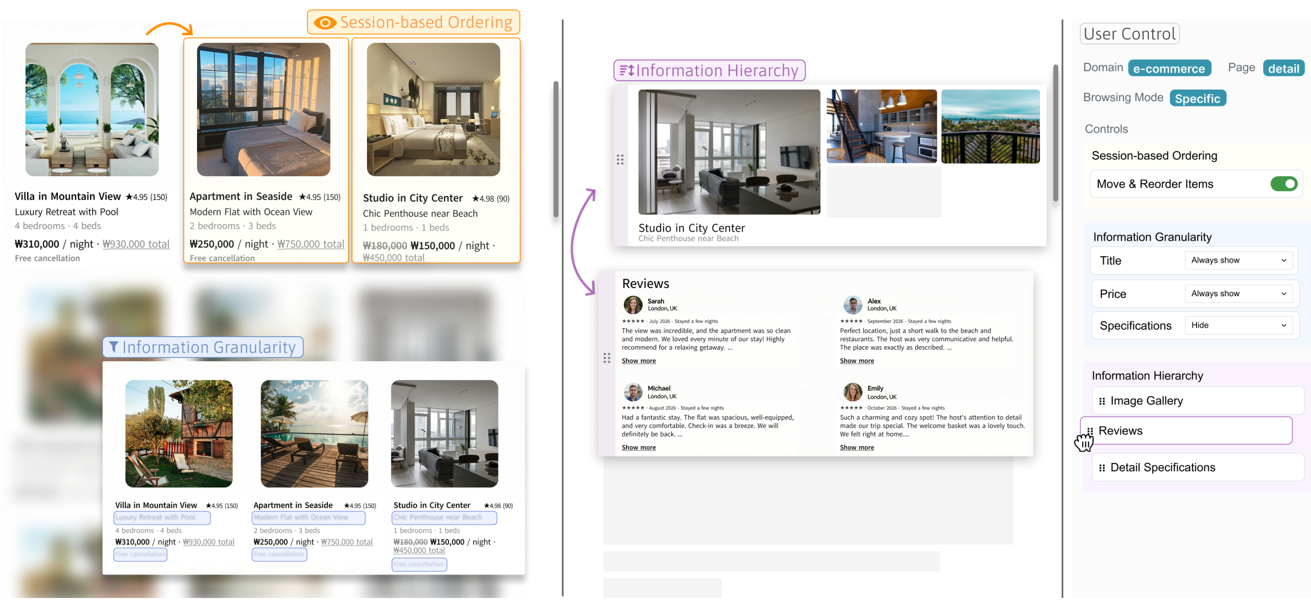

Figure 1: Conceptual overview of the ReLay probe system. ReLay operationalizes three adaptation dimensions: information hierarchy (reordering detail-page sections such as reviews and images), information granularity (adjusting the amount of visible detail within catalog cards), and session-based ordering (repositioning or marking previously viewed or clicked items).

Abstract

Online browsing often requires balancing open-ended exploration with focused comparison, yet most interfaces remain static regardless of user focus. Adaptive interfaces offer a way to better align interface presentation with browsing needs, but how such adaptations should be triggered, designed, and controlled in practice remains unclear. We investigate adaptive browsing interfaces that modify information hierarchy, information granularity, and session-based ordering in response to inferred browsing intent. Drawing from a conceptual framework of five adaptation dimensions, we implement ReLay, a browser-based probe that applies lightweight, in-situ adaptations automatically while allowing user overrides. In a two-phase study (n = 10), participants welcomed adaptive changes when they were transparent, consistent, and easily reversible. Users approached control as calibration rather than error correction, verifying system behavior before relying on it. These findings illustrate how intent-responsive, controllable adaptation can support browsing without diminishing user agency. Our work contributes (1) a conceptual framework for adaptive interface behaviors in online browsing, (2) an instantiation of the framework through three selected dimensions, and (3) empirical insights into user acceptance, control, and design implications for adaptive browsing systems.

*This work was initiated while the author was at Seoul National University.
†The author has since graduated.

IUI '26, Paphos, Cyprus

ACM ISBN 979-8-4007-1984-4/2026/03
https://doi.org/10.1145/3742413.3789092

CCS Concepts

• **Human-centered computing → Interactive systems and tools**; **User interface management systems**; *Empirical studies in HCI*.

Keywords

Adaptive user interfaces, Online sensemaking, Exploratory browsing support, Interface adaptation

ACM Reference Format:
Eunhye Kim, Sadat Shams Chowdhury, Hyungwoo Song, and Bongwon Suh. 2026. In-Situ Adaptive Interfaces for Online Browsing: Design Dimensions for Intent-Responsive Automation and User Control. In *31st International Conference on Intelligent User Interfaces (IUI '26), March 23–26, 2026, Paphos, Cyprus.* ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3742413.3789092

1 Introduction

Online browsing is both ubiquitous and cognitively demanding [74]. From shopping for clothes on e-commerce platforms to booking hotels on travel sites, people are constantly faced with large volumes of information spread across browsing interfaces. Unlike narrow information retrieval tasks, browsing is not merely about retrieving a correct answer, but about the process by which users form and refine preferences. A shopper may start broadly *("What styles are trending this season?")* and gradually narrow down to specifics *("a black jacket under $150, in my size")*. Supporting these workflows is difficult because browsing combines subjective sensemaking, open-ended exploration, and goal-driven filtering [31, 79].

Prior approaches address parts of this challenge but leave key aspects of browsing workflows underexplored. Recommender systems [67, 131] often streamline the browsing process by predicting a small set of items to show. While useful for personalization, this ignores the iterative and reflective nature of browsing, where users actively explore and learn what they want. Comparison and choice-support tools [63, 70] provide specialized views—such as side-by-side comparisons or attribute-driven rankings—that help with specific decision steps. However, these tools do not cover the full browsing journey and often require users to leave or replace their everyday workflows. Recent work on malleable and customizable interfaces [26, 78] allows users to reconfigure layouts or filters directly through prompts or modular components, enabling user-driven redesign. Yet such customization often relies on users having a clear sense of what to change, a condition that may not hold during open-ended browsing, when goals are evolving. In these moments, adaptive systems offer an opportunity to proactively adjust visual emphasis, hierarchy, or information density to better align with users' evolving browsing focus.

Browsing often alternates between diversive and specific modes [101], which differ in the level of information detail, structure, and comparison involved. For example, in a fashion catalog, diversive browsing might benefit from large images with minimal text, broad category filters, and suppression of already-seen items to encourage novelty. In contrast, specific browsing might highlight textual cues such as price, brand, and review count, surface fine-grained filters, and re-emphasize previously viewed items for comparison. These differences motivate adaptive interface behaviors—such as changes in emphasis, hierarchy, density, navigation, and interaction-based organizing.

However, fundamental questions remain about the desirability and effectiveness of automatic adaptations in practice. Do users accept interfaces that change automatically, or do such transformations feel intrusive or disorienting? Which types of adaptations can help reduce cognitive effort? What balance of automation and control do users expect when interfaces adapt dynamically? Where do users draw the boundary between helpful automatic restructuring and personal control? Without understanding these experiential and perceptual dimensions first, adaptive systems risk solving problems users do not recognize—or introducing unwanted complexity into familiar workflows. In this work, we focus on three factors that shape these experiences: inferred browsing intent as a trigger for adaptation, the design dimensions through which interfaces adapt, and the degree of user control over automated changes.

To investigate these questions, we propose a conceptual framework of five UI adaptation dimensions applicable to common browsing interfaces: representation emphasis, information hierarchy, exploration scope, information granularity, and session-based ordering. We develop ReLay as a technology probe that implements three of these dimensions, where an LLM-based pipeline applies these adaptations in response to inferred browsing intent, while a lightweight control panel allows users to override decisions and save preferences.

We deployed the probe in a study with 10 participants completing everyday browsing tasks across three domains. Our investigation centered on three research questions:

- **RQ1 (User Experience and Acceptance)**: How do users perceive and accept adaptive changes during browsing?
- **RQ2 (Perceived Cognitive Effort)**: How do different adaptation types influence browsing effort and flow?
- **RQ3 (Future Control and Customization)**: How do users envision customizing or extending adaptive browsing interfaces in the future?

Our findings show that users generally welcomed adaptive behavior when it was consistent, transparent, and easy to override, but emphasized the importance of lightweight control and contextual sensitivity. Automatic adaptations reduced perceived workload yet required clear boundaries for acceptance and understanding.

In summary, this work makes three contributions:

- A conceptual framework of five adaptive interface dimensions describing how everyday browsing layouts can flex to different browsing focuses.
- An instantiation of the framework through a technology probe, ReLay, that demonstrates the feasibility of in-situ adaptive interface behaviors and enables examination of how users experience and negotiate automated adaptations.
- Empirical insights into how users perceive, accept, and control adaptive browsing interfaces, revealing design implications for balancing automation with agency.

2 Related Work

2.1 Online Sensemaking and Browsing Support Systems

Human-computer interaction (HCI) and information retrieval (IR) research has explored tools for online sensemaking and exploratory browsing. Early systems such as Unakite [68], Crystalline [69], and Mesh [18] supported structured decision-making by letting users capture, organize, and compare options through tabular or structured views. Others, like Fuse [63], emphasized lightweight collection within the browser, while Sensecape [102] enabled multi-level sensemaking through conversational interactions. Despite these advances, existing tools separate foraging from structuring, disrupting users' natural transitions between these activities during browsing.

Parallel work has examined how personalization and intelligent assistance can guide exploratory search. Early approaches focused on user-defined preference models: SearchLens [20] introduced lenses to articulate and reapply exploratory interests, while FeedLens [59] extended this to knowledge graphs, allowing preference models to be repurposed across heterogeneous entity types. More recent work has leveraged large language models (LLMs) to provide richer scaffolding: ChoiceMates [86] uses multi-agent conversations to support decision-making in unfamiliar domains, Selenite [70] generates comprehensive overviews to address the "cold start" problem, and Marco [40] integrates document-, notebook-, and table-level views with LLM-powered assistance for knowledge workers. These systems demonstrate the evolution from explicit preference modeling to AI-assisted sensemaking, through primarily creating new interfaces.

Despite these advances, two notable gaps remain. First, prior systems largely focus on building new exploratory interfaces rather than adapting the familiar browsing structures that users already navigate daily. Second, existing work often supports isolated phases of browsing—exploration or evaluation—without accounting for how users transition between these modes. As a result, support is fragmented, and users must bridge the gap themselves when switching between broad exploration and detailed inspection. Addressing this limitation requires approaches that adapt existing browsing interfaces in-situ, dynamically adjusting to users' shifting intents rather than replacing their familiar workflows.

2.2 Interface Adaptation and Customization

Recent work on interface adaptation has explored various approaches to help users modify interfaces for their needs, from surface-level styling to structural customization and AI-generated interfaces.

One line of research focused on enabling end-users to customize visual appearance and layout. Systems have explored natural language-based styling of websites [61], allowing users to express aesthetic preferences. More sophisticated approaches enable structural customization, allowing users to modify how information is organized, what content is displayed, and how different views are composed [78]. For instance, in overview-detail interfaces common to browsing tasks, users might customize which attributes appear in overviews versus details, or transform between different layout configurations. Tools like FrameKit [118] provide authoring environments for adaptive UIs, enabling developers to create interfaces that respond to context changes. These approaches democratize interface customization but assume users recognize when and how modifications would benefit their tasks.

The rise of LLMs has enabled more ambitious generative interface approaches. Rather than modifying existing interfaces, these systems generate entirely new UIs from user specifications [22, 26, 120]. Some approaches use task-driven data models as intermediate representations, which allows users to iteratively refine generated interfaces through both natural language and direct manipulation [12]. Entire browsing environments can also become malleable, with AI helping users orchestrate and compose web content [57]. While these generative approaches offer unprecedented flexibility, they require users to actively direct the customization process through prompts or explicit interactions.

Across all these approaches lies a fundamental assumption: users must initiate customization, whether through direct manipulation, natural language commands, or iterative refinement. This works well when users have clear goals and understand their needs. However, during exploratory browsing, users' intents shift fluidly—sometimes seeking broad inspiration, other times requiring detailed comparison—without conscious recognition of these transitions. We address this gap by automatically detecting shifts in browsing intent and applying appropriate adaptations proactively, reducing the need for users to both recognize adaptation opportunities and specify desired changes.

2.3 LLMs for Interface Interpretation and Contextual Inference

Large language models (LLMs) have been recently used to interpret user interfaces and contextualize interaction in support of intelligent systems. One line of work focuses on systems that use LLMs to analyze graphical user interfaces (GUIs), combining structured representations such as DOM trees or screenshots with language-based reasoning to identify interface elements, infer their functional roles, and operate across web, desktop, and mobile environments [103, 125]. A growing set of frameworks and benchmarks have demonstrated that these approaches can generalize across different applications and interaction contexts [21, 56, 121].

Other work has explored how LLMs can contextualize observable interaction signals and interface state to provide adaptive support. Some systems construct explicit user models from computer-use traces to inform downstream assistance [97], while others analyze visible interface context to generate just-in-time support that augment ongoing interaction [64]. Together, these approaches demonstrate that LLMs can leverage observable interface state and interaction data to produce context-aware support without requiring explicit user specification.

We build on these uses of LLMs for interface interpretation and contextual inference by applying them to adaptive interface behaviors within existing browsing environments.

3 Design Goals and Framework

In this section, we introduce the design goals and framework that we built our adaptive browsing interface on. We chose browsing intent inferred through interaction patterns as a *trigger* for adaptations, enabling dynamic but interpretable transformations of familiar web interfaces.

3.1 Design Goals

Online browsing presents a persistent challenge: users oscillate between diversive and specific modes, yet interfaces remain static [101]. We explore lightweight, in-situ interface adaptations that respond to the user's browsing context—using inferred intent as a trigger—while preserving familiar interaction patterns. Based on gaps in prior work and common challenges in browsing, we identify three design goals for adaptive interface behaviors:

- **DG1: In-situ adaptation.** Adaptations should operate directly within existing browsing pages rather than redirecting users to separate dashboards or views. Browsing is inherently contextual—users interpret and compare items within the visual and structural environment of the site itself [96, 126]. Keeping adaptations in-situ preserves workflow continuity and minimizes cognitive switching costs, allowing users to stay focused on the browsing task [33, 91].
- **DG2: Mode-responsive adaptation.** Adaptations should reflect the natural shifts between diversive and specific browsing modes. These modes highlight distinct informational needs: diversive browsing emphasizes variety and visual inspiration [25, 104], while specific browsing emphasizes analytic comparison and detail [90]. Adaptations should therefore respond flexibly to these transitions—using inferred intent as a trigger—rather than enforcing a single static model of "optimal" usability [28, 106].
- **DG3: Lightweight and generalizable transformations.** Adaptations should manifest as direct, visible transformations—such as emphasis, ordering, or density adjustments—rather than opaque algorithmic changes [118, 124]. Their logic should remain legible and reversible, enabling users to perceive and adjust them easily. At the same time, these transformations should be defined at a conceptual level that generalizes across browsing domains, ensuring reusability and cross-site applicability [34, 84].

3.2 Operationalizing Browsing Modes for Adaptation

Browsing is an iterative process of sensemaking in which users continuously form, refine, and compare preferences across many alternatives. Prior research on exploratory search and consumer decision-making shows that users rarely follow a linear path; instead, they oscillate between broad exploration and focused specification depending on evolving goals and task constraints [74, 95, 112–114]. We build on this literature to conceptualize two recurring browsing modes—*diversive browsing* and *specific browsing*—that serve as anchors for triggering adaptive interface behaviors.

3.2.1 Diversive Browsing. Diversive browsing is an open-ended, novelty-driven activity where users explore broadly to surface possibilities rather than make final choices [44, 74, 95, 113, 114, 129]. Users typically arrive without rigid criteria, relying on broad categories and visual cues for inspiration [62, 96], and exhibit fast, lightweight interactions such as rapid scrolling and brief dwell times [92, 95, 116, 129]. Adaptations should therefore emphasize large imagery, minimal text, and suppression of previously viewed items to sustain momentum and novelty.

3.2.2 Specific Browsing. Specific browsing arises when users form clearer goals and shift toward evaluation and comparison [74, 95, 112, 114]. Interaction becomes slower and more deliberate, with longer dwell times and revisits to previously seen items for verification [23, 66]. Adaptations should foreground detailed attributes, reviews, and comparison sections to support precision, reassurance, and confidence-building.

3.3 Adaptation Framework

We define a UI adaptation as a lightweight change that:

- Does not introduce new interactive components
- Does not create entirely new views
- Operates on existing interface elements by reordering, emphasizing, suppressing, or adjusting the salience

This definition situates our work among interface-level adaptation approaches: rather than introducing new interface components or generating new content, we adapt the presentation and organization of information already available to the user. This scope ensures that adaptations remain in-situ (DG1), respond to browsing mode (DG2), and apply generalizable transformations (DG3).

We organize adaptive behaviors into two complementary categories. **Structural adaptations** reshape how information is presented at a given moment by modifying layout, hierarchy, emphasis, or information density. **Interaction-Contextual adaptations** reflects signals from users' prior interactions through visible interface changes, such as resurfacing previously viewed items or suppressing already-seen content. Together, these categories capture both synchronic (within-page) and diachronic (across-session) aspects of mode-responsive browsing.

Our framework targets two common page types in online browsing: catalog pages, which display collections of items—search results, product grids, or listing feeds—where users scan, filter, and compare alternatives; and detail pages, which present a single item in depth, typically including images, specifications, reviews, and related recommendations. While our adaptation dimensions may apply to other browsing structures, we use these two page types as a lens for examining how interfaces can flex to support shifting browsing focus.

We conceptualize adaptive browsing interfaces through a set of design dimensions that span different levers of information architecture and interaction-level adaptation. These dimensions describe how information can be structured, ordered, emphasized, and contextualized during browsing, articulating complementary ways in which interfaces can adapt to users' evolving browsing focus.

3.3.1 Structural Adaptations. Structural adaptations concern how information is arranged and emphasized within a single catalog or detail page at a given browsing moment. We identify four subdimensions that capture distinct ways the interface can be adjusted: representation emphasis, information hierarchy, exploration scope, and information granularity.

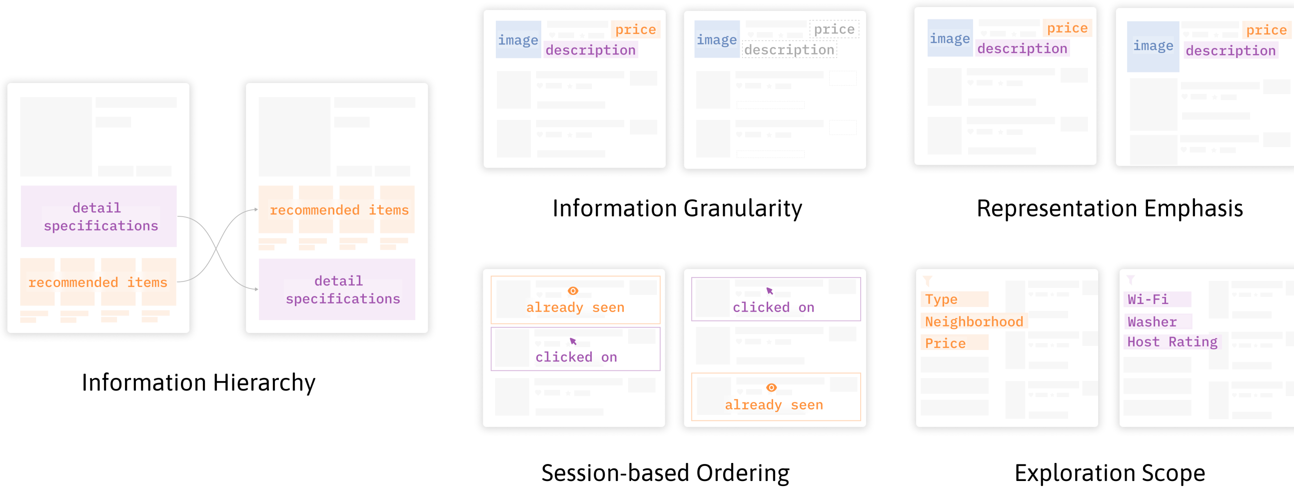

Figure 2: Conceptual Overview of Design Framework for Adaptation

Representation Emphasis. Browsing interfaces present information through multiple modalities—visual (images, thumbnails, icons), textual (descriptions, specifications), and numerical (ratings, prices, review counts) [73, 117, 130]. The relative spatial weight of these modalities determines which information receives visual prominence, shaping initial attention and information processing [7, 9, 11, 15, 100, 132].

The effectiveness of each modality depends on browsing mode, which motivates representation-emphasis adaptations. In diversive browsing, users rely on rapid visual scanning to form broad impressions across alternatives; accordingly, on catalog pages diversive favors prominent, larger images and condensed text, and on detail pages it prioritizes contextual or lifestyle imagery [11, 49, 73, 74, 119]. In specific browsing, users slow down to compare concrete attributes; accordingly, on catalog pages specific emphasizes inline specifications and ratings, and on detail pages it foregrounds product details or technical views [9, 100, 123, 132]. Studies of interface perception confirm that shifts in visual prominence strongly influence perceived relevance and quality [7, 15, 105, 117], supporting the value of calibrating visual–textual balance to user focus and depth of processing (DG2) [29, 83].

Information Hierarchy. Detail pages are composed of multiple container sections that organize different aspects of an item. In e-commerce, these typically include product images, specifications, reviews, and recommendations; in real estate, property photos, floor plans, neighborhood details, and contact options. Across domains, such sections form a semantic hierarchy that determines which content users encounter first. This ordering strongly affects attention and efficiency, as users naturally treat higher-positioned sections as more important [24, 89, 98, 109, 128].

The relative importance of sections varies with browsing mode. In diversive, users seek breadth and inspiration, making recommendations or related-item modules more salient because they enable rapid scanning and serendipitous discovery [3, 13, 74, 75, 114]. In specific, users focus on verification and comparison, giving priority to specifications, detailed descriptions, and reviews [19, 24, 88, 89]. Adjusting section hierarchy to highlight mode-relevant information can therefore reduce unnecessary scrolling and align presentation with users' current focus [89, 98, 109, 128].

Recent adaptive-interface systems have explored such dynamic reordering, surfacing or collapsing sections to match user attention [24, 57, 78, 89, 98]. These approaches build on focus–context representation [41, 74] and information-foraging theory, which shows that users allocate attention to regions with the highest expected informational value [88, 114]. Treating section hierarchy as an adaptable design dimension thus enables interfaces to align structure with evolving browsing depth.

Exploration Scope. Exploration scope concerns how broadly or narrowly the interface invites users to traverse the item space. In practice, this manifests through filtering and navigation controls: e-commerce sites offer filters for categories, prices, brands, sizes, and colors; restaurant platforms for cuisine, price, distance, and dietary options; and accommodation sites for location, amenities, and capacity. These controls range from broad categorical filters to fine-grained attribute selectors, shaping how users explore and constrain the search space [47].

The usefulness of different filtering granularities depends on the browsing mode. In diversive, users benefit from broad, high-level facets (e.g., "jackets," "neighborhoods," "Italian cuisine") that support wide traversal and serendipitous discovery [4, 38, 74, 107, 115]. In specific, users prefer precise tools—fine-grained filters (e.g., color, size, amenities) and sorting options—that allow efficient narrowing based on known criteria [1, 48, 85].

Adaptations along this dimension adjust which controls are foregrounded according to browsing mode. In diversive, broad categorical filters and exploratory navigation options are emphasized, while detailed filters remain accessible but secondary [48, 83, 99]. In specific, precise filters and sorting tools move to primary positions, enabling efficient, goal-directed narrowing [1, 85].

This dimension builds on research in faceted search and exploratory navigation [13, 47, 52, 74, 85, 114, 122], showing that filter granularity influences both discovery and precision. Adjusting filter prominence by browsing mode thus supports both breadth-first exploration and constraint-based refinement (DG2) within existing catalog structures (DG1).

Information Granularity. Browsing interfaces differ in how much detail they reveal at each level of interaction. A product card may show only essential attributes (title, price) or include richer specifications (brand, materials, dimensions, review count, availability). Similarly, review sections may present brief summaries or expand into full testimonials. This degree of visible detail—what we term information granularity—shapes the cognitive effort required to evaluate each item [37, 88].

The appropriate level of detail depends on browsing mode. In diversive, users scan quickly across many options, so minimal detail reduces cognitive load and sustains exploration momentum [74, 83]. In specific, users engage in close comparison and verification, benefiting from full attribute visibility without additional navigation [46, 58, 87].

Information-granularity adaptations adjust detail density according to browsing mode [83]. On catalog pages, diversive displays only key attributes (e.g., title, price, main image), whereas specific expands cards to include comprehensive specifications and ratings [46, 87]. On detail pages, diversive summarizes sections (e.g., "4.5 from 847 reviews"), while specific reveals full review content and specification tables [30, 80].

This dimension builds on research in information scent and cognitive load [27, 37, 71, 110], which shows that aligning information depth with task demands improves efficiency. Adjusting granularity by browsing mode thus helps maintain cognitive balance—lightweight for exploration, rich for decision-making (DG2).

3.3.2 Interaction-Contextual Adaptation. Interaction-Contextual adaptations extend beyond a single page view to reflect how the interface incorporates a user's past interactions. They determine how previously seen or engaged items are represented—for instance, suppressing already-viewed products to promote novelty during exploratory browsing, or highlighting them to support comparison during focused evaluation. In this sense, Interaction-Contextual adaptations are inherently temporal, linking past interactions to present presentation.

Session-based Ordering. Diversive browsing emphasizes novelty, which in a provenance dimension translates to managing whether an item has already been encountered. Prior systems have operationalized this by promoting unseen content: recommender diversification techniques [16, 43, 55, 65, 111, 127, 131] and exploratory search interfaces [32, 49, 50, 113, 115] explicitly prioritize novelty to sustain engagement. Suppressing or visually downplaying previously viewed items can similarly help users maintain a broad exploration path.

Conversely, specific browsing benefits from resurfacing previously encountered items, especially during comparison or decision refinement. Studies on re-finding and revisitation [14, 32] show that users frequently return to seen pages when narrowing choices, while consumer decision-making work highlights shortlists and recently viewed modules as aids for comparison [42, 94]. Provenance-aware visualization systems [81, 82] likewise demonstrate how surfacing interaction history can improve sensemaking and reduce cognitive load.

Session-based ordering adaptations thus adjust how existing items are displayed rather than what is recommended. They reweigh previously seen content in the catalog view—for example, moving viewed items lower in diversive, or highlighting them in a separate strip during specific. This distinction ensures that provenance adaptations remain independent of recommendation quality: they alter presentation, not content selection.

4 The ReLay Probe

To investigate how users experience automatic adaptive browsing interfaces (RQ1), whether such adaptations influence perceived cognitive effort (RQ2), and what controls and future adaptations users envision (RQ3), we developed ReLay as a technology probe. Following the technology probe approach [6, 54], our implementation was designed to elicit reflections on how users interpret and respond to adaptive interface behaviors in realistic browsing contexts. From our five-dimensional framework, we selectively implemented a subset of three adaptation dimensions—information hierarchy, session-based ordering, and information granularity—chosen for their visual salience and conceptual contrast, enabling the probe to surface diverse user reactions to in-situ changes.

ReLay is an LLM-based browser extension that automatically adapts online browsing interfaces based on detected user intent. The system (1) infers whether the user is currently in diversive or specific browsing mode from interaction traces, (2) analyzes the structure of catalog or detail pages to identify adaptable components, and (3) applies adaptations while respecting saved user preferences. Unlike prior approaches that create separate exploratory views or replace existing workflows, ReLay operates in-situ within familiar browsing interfaces.

We present the probe in three parts. First, we describe an envisioned user scenario illustrating how ReLay functions across domains. Second, we detail the probe components: intent detection, interface adaptation, and user control modules. Finally, we discuss implementation details and scope decisions inherent to the probe methodology.

4.1 Envisioned User Scenario

Sarah is planning a trip to Cyprus and needs to select three things: an accommodation, a restaurant for the first evening, and clothing to bring. She installs ReLay and opens her browser to begin browsing.

First, Sarah visits Airbnb to explore accommodations in Cyprus. With flexible preferences and budget, she rapidly switches between catalog and detail pages, skimming listings without dwelling. ReLay detects this as diversive and adapts accordingly: previously viewed listings are subtly de-emphasized, product cards simplify to highlight price and location, and detail pages foreground neighborhood photos before full specifications. When Sarah moves to Agoda, ReLay recognizes the shared travel domain and maintains the same diversive adaptations across sites.

Next, Sarah shops for a black dress on Amazon. Her slower scrolling and repeated comparisons indicate specific intent. ReLay expands catalog cards to show reviews and material details, pins recently viewed items for side-by-side comparison, and reorders detail pages to surface specifications earlier. Preferring to keep styling suggestions visible, Sarah overrides this layout through the control panel, and ReLay remembers her choice for future fashion browsing.

Across domains, ReLay adapts to Sarah's shifting intent while providing lightweight controls that let her shape how adaptations apply to her personal browsing style.

4.2 System Components

The ReLay probe comprises three main components (Figure 3): an **Intent Detection Module** that infers browsing mode from user behavior, an **Interface Adaptation Module** that applies transformations to page structure based on detected intent, and a **User Control Module** that allows users to override adaptations and save preferences.

4.2.1 Intent Detection Module. This component classifies the user's current browsing mode. ReLay supports two primary browsing intents: *diversive* browsing, where users seek novelty and breadth across items, and *specific* browsing, where users narrow in on a set of candidates for close comparison. To detect these modes, ReLay combines page structure analysis with multi-modal behavioral tracking.

When a user visits a page, the system first determines whether the site and page type are suitable for adaptation. It analyzes the page's domain name and a minimized DOM representation [93] to classify both the browsing domain (e.g., e-commerce, accommodation, dining) and page type (catalog, detail, or other).

If a page is identified as a catalog or detail page, ReLay begins tracking behavioral signals. Building on prior models of exploratory search [79, 101], the system infers factors from scroll, click, dwell patterns and visible elements on the screen are captured with OmniParser [72].

Every two minutes, accumulated behavioral signals and visual context are passed to an LLM-based classifier. The prompt describes diversive and specific intent patterns based on established browsing behavior characteristics [31], providing indicators such as filter usage intensity, search query specificity, browsing breadth, and engagement depth. The LLM returns a classification (diversive or specific) with a confidence score and structured reasoning, which was used to inform adaptations. Examples of the prompt template are provided in Appendix A.

To avoid over-fragmentation, intent detection is maintained at the browsing domain level. When a user browses across multiple sites within the same domain (e.g., from Airbnb to Agoda), the system persists the current intent until behavioral evidence suggests a shift. However, when the user moves to a new browsing domain (e.g., from accommodation to e-commerce), intent detection restarts, enabling domain-specific adaptation without cross-domain interference. Once intent is classified, the intent detection module signals the interface adaptation module to apply corresponding transformations.

4.2.2 Interface Adaptation Module. We use this module to examine how users interpret and evaluate different adaptations, and whether these transformations feel helpful or effort-reducing during browsing.

From our five-dimension framework, we implement three adaptations: information hierarchy, session-based ordering, and information granularity. We selected these dimensions because they represent distinct intervention points that provoke different user responses. Information hierarchy offers a high-visibility structural change that users can easily perceive and critique, making it valuable for probing acceptance of automatic hierarchy manipulation. Session-based ordering creates an explicit tension between novelty-seeking and comparison behaviors, allowing us to investigate how users value redundancy differently across modes. Information granularity adjusts detail density within interface elements, making it well-suited for probing cognitive load effects.Together, these three adaptations are visually salient, domain-general, and lightweight enough to integrate with existing workflows without introducing major disruptions.

Each adaptation follows a common technical approach: the system classifies page type, uses LLMs to identify structural elements, and generates transformation instructions applied by manipulating existing containers. While ReLay can perform this full pipeline on supported websites, structural parsing for the sites used in the probe study were pre-configured to eliminate latency during user interactions.

Information Hierarchy. Information hierarchy addresses the fixed sequencing of information on detail pages. These pages typically contain multiple container sections such as product specifications, reviews, recommendations, and image galleries. In many sites, this order is static, forcing all users into the same information path regardless of their browsing mode. ReLay adapts the ordering of these sections to better match intent: in diversive browsing, breadth-oriented modules such as related items or lifestyle imagery are surfaced earlier, while in specific browsing, fine-grained sections such as specifications and reviews are prioritized.

Once a page is classified as a detail page, the reduced DOM is analyzed by an LLM to detect and semantically label major sections (e.g., "reviews", "specifications", "recommendations"). A second LLM call then generates a revised section ordering based on general intent principles for that mode, and any saved user preferences. For example, if a user previously specified they always want recommendations surfaced first in e-commerce browsing, this preference is incorporated into the prompt. The system then reorders the identified container elements in the DOM in the proposed sequence, dynamically foregrounding intent-relevant content while preserving all existing functionality.

Session-based Ordering. Session-based ordering addresses the reordering of the items in the catalog page based on the user's intent. Static catalog orderings force users to re-encounter the same items regardless of whether they seek novelty or wish to compare known options. In diversive mode, repeatedly seeing already-viewed items feels redundant and reduces perceived variety. In specific mode, users benefit from easily relocating previously considered items for comparison.

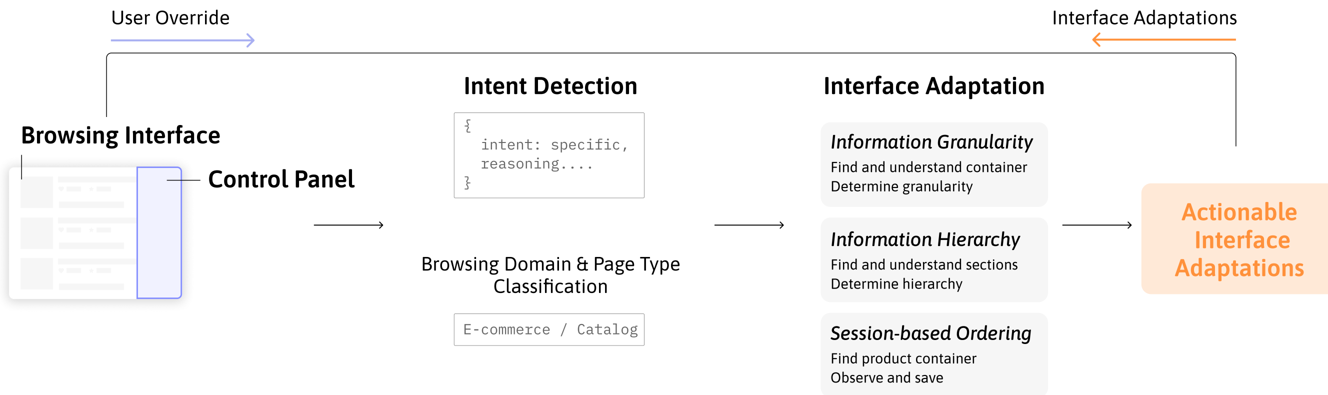

Figure 3: Overview of the ReLay technical pipeline. The system comprises three modules: (1) Intent Detection observes user behavior and classifies browsing intent at regular intervals, (2) Interface Adaptation analyzes page structure and applies appropriate adaptations when intent is detected, (3) User Control allows users to override adaptations or adjust preferences through a control panel. Users interact with adaptations directly in the browsing site, while controls are accessible via the panel.

ReLay tracks which items users have scrolled past and which they have clicked on. The system uses an LLM to identify product card selectors and extract item identifiers from the HTML structure, then monitors user interactions throughout the session, maintaining a per-site record of viewed and clicked items. When a user navigates to a new catalog page and an intent has been detected, ReLay applies mode-appropriate transformations: in diversive mode, previously viewed items are moved lower in the grid and marked with colored borders and a semi-transparent overlay to signal their prior appearance; in specific mode, clicked items are moved toward the top to facilitate comparison. These transformations apply only when users land on a new catalog page, not during active scrolling within a page, preventing disorienting mid-scroll reorderings while still providing session-aware organization.

Information granularity. Information granularity controls the granularity of textual and visual information displayed within individual product cards on catalog pages. In diversive browsing modes, users can be overwhelmed by excessive textual detail when they primarily seek broad impressions and exploration-suitable cues like price or review ratings rather than lengthy item descriptions. In specific browsing, users need comprehensive information but with emphasis on details necessary for comparison and narrowing down, such as restaurant type, thorough item explanations, or hotel atmosphere indicators. ReLay applies this kind of emphasizing, reducing, or eliminating of information based on the detected intent across product catalog cards.

The system uses LLMs to identify product card selectors, analyze sample card HTML to create a semantic understanding of the card's information hierarchy—mapping elements to categories like price, ratings, descriptions, vibes, and detailed features—and then determine which elements should be shown, expanded, enhanced, simplified, truncated, or hidden based on detected intent, browsing domain, and user preferences. For the probe study, we pre-configured the structural analysis (selector identification and semantic mapping) for commonly-used sites to ensure responsive interactions.

In diversive mode, cards typically display large images with minimal text, foregrounding price and review ratings while truncating or hiding detailed descriptions and specifications. In specific mode, cards expand to surface all available information, including full specifications, review counts, material details, and categorical attributes relevant to the domain. The frontend extension applies these transformation instructions to all visible product cards, dynamically adjusting information density to match the user's current browsing mode.

4.2.3 User Control Module. By observing which controls users engage with and what preferences they save, we can understand the desired balance between automation and agency in intent-aware systems.

After ReLay automatically applies adaptations, users can adjust how adaptations behave in the future through a control panel accessible via a browser sidebar. The panel provides adaptation-specific controls that allow users to undo the current page's adaptations, toggle adaptation features on or off, and save preferences that persist across future browsing sessions within the same domain.

For information hierarchy, users see a visual representation of how sections were reordered, displaying section names in their new sequence. Users can drag and drop these blocks to specify their preferred ordering, and save this arrangement as a preference for that browsing domain. For example, if ReLay moved "Recommended Items" lower on a detail page in specific mode, but the user prefers seeing recommendations first regardless of intent, they can reorder the blocks and save this preference for all future e-commerce browsing. For session-based ordering, users can independently toggle two features: visual indicators (colored borders and semi-transparent overlays) and item reordering (moving viewed/clicked items up or down in the catalog). For information granularity, the panel displays common product card elements—Images, Descriptions, Reviews, Price—and allows users to override the system's intent-based decisions. For each element, users can specify:"go with intent" (default adaptive behavior),"always show" (display regardless of browsing mode), or "always hide" (suppress regardless of mode).

All preferences are saved per browsing domain (e-commerce, accommodation, dining) and encoded as natural language constraints appended to adaptation prompts. These preferences apply to future page loads within the same domain but do not affect other browsing domains, allowing users to maintain different customization profiles across contexts. Users can also reset preferences to restore default adaptive behavior. A detailed illustration of the control panel is provided in Appendix B.

4.3 Technical Overview

4.3.1 Implementation Details. ReLay is implemented as a Chrome extension with a FastAPI[1] Python backend. The frontend uses React and JavaScript to manipulate the DOM. For intent classification and adaptation generation, we use `gpt-4o-2024-08-06` and `gpt-4.1-2025-04-14` models.

4.3.2 Constraints and Limitations. ReLay targets desktop web pages and applies interface transformations by directly modifying the page DOM, assuming it is accessible and relatively stable. It supports pages rendered by common frameworks (e.g., React, Svelte, Next.js) but not Shadow DOM.

ReLay tolerates low-frequency DOM restructuring: occasional changes can be handled via repeated LLM calls, but rapid structural updates are infeasible if their frequency exceeds the LLM call latency. We did not observe such high volatility in our study, but it remains a limitation for in-the-wild deployment.

Furthermore, ReLay cannot operate on mobile pages and pages that block or revert DOM edits (e.g., anti-tampering/anti-bot protections). It also does not support non-DOM-rendered or embedded surfaces, including Canvas/WebGL UIs (e.g., maps), browser-native viewers (e.g., PDFs), iframes, and HTML5 video.

For transformation planning, the page HTML must fit within the LLM context window. ReLay therefore compresses the HTML to a structural skeleton, discarding any elements that do not contribute to the DOM hierarchy (header/footer tags, script tags, etc.) and stripping all attributes from elements that are not LLM selection targets (elements that can not have child elements), to keep inputs within token limits.

5 Evaluation

To investigate how users experience adaptive browsing interfaces, we conducted a qualitative probe study focused on understanding user perceptions of different adaptation types.

5.1 Task and Participants

We recruited 10 participants (6 female, 4 male; age 21–32) through an online screening survey targeting frequent online browsers. To ensure familiarity with diverse domains, we selected participants who reported browsing at least three categories regularly (e.g., fashion, travel, dining) and shopping online more than three times per week.

To evaluate both the usefulness of individual adaptations and the overall usability of the adaptive system, we conducted a two-phase qualitative probe study. Participants first selected two cities they were not familiar with, ensuring that the browsing process reflected natural exploration rather than prior knowledge or pre-planned choices. In Phase 1, participants experienced each adaptation individually across three domains—clothing shopping, restaurant finding, and hotel searching—to understand how each adaptation influenced browsing behavior. They were given the following scenario: *"You are planning a three-day trip to [city] you are not familiar with. You will need to find one or more hotels, restaurants, and outfits for your trip by browsing online."* In Phase 2, participants used all adaptations together in a fixed task focused on clothing browsing, allowing them to experience how multiple adaptations interact within a single session. To maintain the open-ended nature of browsing while providing clear goals, participants were asked to create a shortlist of one to three items that they would realistically consider choosing.

The study received ethics approval from the authors' institutional review board, with all participants providing informed consent.

5.2 Study Procedure

We conducted 90-minute remote sessions with each participant via Zoom. Participants used their own laptops with Chrome installed to ensure a familiar browsing environment. For those unable to install the system on their own devices, we provided remote access to a laptop with the ReLay extension installed. Throughout the session, participants shared their screens, enabling the researchers to observe on-screen activity.

5.2.1 Pre-task and Tutorial (20min). Before the main sessions, participants completed a 20-minute pre-task tutorial. Because the system was implemented as a Chrome extension, we first guided participants through the installation process and verified that it functioned properly on their devices. We then introduced the overall goal of the study and briefly explained the concept of adaptive interfaces, including the three adaptation types featured in our system: information granularity, information hierarchy, and session-based ordering. Participants were instructed on the browsing tasks they would perform and were reminded that the study would follow a think-aloud protocol [39], encouraging them to verbalize their thoughts, expectations, and reactions as they interacted with ReLay. After the tutorial, participants proceeded directly to Phase 1, beginning with one of the three adaptation conditions randomly assigned to them.

5.2.2 Phase 1: Exploring Individual Adaptations (30min). In Phase 1, participants explored each adaptation individually to understand its distinct effect on browsing. Each participant completed three browsing sessions, each focusing on one adaptation type: Information Granularity on Agoda[2] for hotel browsing, Information Hierarchy on Yelp[3] for restaurant browsing, and session-based ordering on Amazon[4] for clothing browsing. These domain–adaptation pairings

[1]https://fastapi.tiangolo.com/

[2]https://www.agoda.com
[3]https://www.yelp.com
[4]https://www.amazon.com

were selected because they represented the most visually salient and conceptually distinct transformations, allowing participants to easily perceive how different forms of adaptation reshaped their browsing process. The order of the three adaptation conditions was randomized across participants. For participants unfamiliar with a given site, an additional two-minute familiarization period with the unmodified site was provided before each task.

Each session lasted approximately 10 minutes, during which participants engaged in a think-aloud procedure. After each session, participants completed NASA-TLX and short Likert-scale items on ease, clarity, and overload (Appendix C.2).

5.2.3 Phase 2: Browsing with Combined Adaptations (20 min). In Phase 2, participants experienced how multiple adaptive behaviors interact in realistic browsing. This phase focused on fashion browsing using 29cm[5], a widely used Korean e-commerce platform that most participants were already familiar with. The site was chosen for its stable performance and rich visual layout, which made it well-suited for testing the combined adaptations in an ecologically valid context.

Participants were asked to browse for clothing suitable for one of the unfamiliar travel destinations they had selected earlier. The task prompt was: *"You are preparing for a three-day trip to [city]. Please browse for one to three outfits you would actually consider wearing on that trip."* This framing encouraged open-ended yet goal-oriented exploration, allowing participants to form and refine preferences naturally rather than completing a fixed decision task. Throughout both phases, the system continuously logged interaction events for analysis. Each browsing session recorded the page type (catalog or detail), the adaptation types applied, the user's detected browsing mode, and any interactions with the control panel.

After each session, participants completed a questionnaire consisting of the NASA-TLX workload inventory [45] and the System Usability Scale (SUS) [10] to evaluate both perceived cognitive effort and usability. Additional seven-point Likert items captured perceptions of adaptation alignment, acceptance, and control and transparency (Appendix C.3).

5.2.4 Semi-structured Interview (20 min). Following the questionnaire, a 20-minute semi-structured interview was conducted with each participant. Interview questions explored participants' reflections on (1) how well the adaptive interface supported exploration and comparison (RQ1), (2) how it affected perceived cognitive load and task flow (RQ2), and (3) what kinds of additional adaptations or controls they would want in future systems (RQ3). The interviews primarily focused on participants' qualitative insights about desirable adaptation granularity, control mechanisms, and future customization possibilities in adaptive browsing systems.

6 Results

We report findings from our probe study investigating how users experienced and evaluated adaptive interface behaviors during online browsing tasks. Results combine quantitative measures of usability and workload with qualitative insights from post-task interviews and behavior logs.

[5]https://www.29cm.co.kr/

6.1 User Acceptance and Override Patterns

Across 362 logged adaptation events, participants accepted most of the system's automatic adjustments while exercising selective control when needed. Overall, 312 adaptations (86.2%) were accepted and 50 (13.8%) overridden, indicating strong alignment between system behavior and user expectations. Acceptance rates differed by adaptation type (Figure 5): *Information Granularity* was most accepted (91.9%, 147/160), followed by *Session-based Ordering* (85.7%, 126/147), and *Information Hierarchy* (70.9%, 39/55). This 21-point gap suggests users were more comfortable with visual adjustments that simplified presentation more readily than structural reordering, which they tended to personalize before accepting.

Survey responses mirrored the behavioral data (Figure 4). Participants reported the highest desire for control in *Information Hierarchy* (M = 5.5 ± 1.3), consistent with its lower acceptance rate, while *Information Granularity* and *Session-based Ordering* elicited weaker control demands (M = 5.0 ± 1.8, 4.8 ± 1.9). Most interventions occurred early in sessions as participants explored and calibrated the interface, after which acceptance stabilized. Participants valued the availability of control more than its use—rating "I could override behavior when needed" highest among all features (M = 6.2 / 7)—underscoring that visible, lightweight adjustability fostered confidence in the system even when rarely used.

6.1.1 Overrides Do Not Signal Failure. Despite moderate override activity, participants' subjective ratings reflected positive experiences. Notably, P9 overrode 26% of adaptations (17/65 instances)—the highest rate among all participants—yet reported an *Excellent* usability score (SUS = 82.5), well above the group average of 71.8 (SD = 10.8). Similarly, P1 intervened in 30 % of cases but still rated the system as *Good* (SUS = 75.0). This pattern challenges the common assumption that user interventions indicate failure. Instead, overrides served as expressions of agency, moments when participants verified, refined, or personalized the system's choices. Rather than reducing satisfaction, the ability to step in when needed enhanced confidence that the system was working on their behalf.

6.1.2 Differential Satisfaction Across Adaptation Types. Satisfaction ratings revealed complementary strengths across adaptation types rather than a simple performance hierarchy. *Information Hierarchy* produced the highest ratings for "made browsing easier," (Figure 4) even though it had the lowest acceptance rate. Participants explained that because reordering controls were direct and easily reversible, they felt confident experimenting: once they applied their own preferred order, the interface felt "exactly right" (P6). In contrast, *Information Granularity* was broadly accepted but divided opinion. Some valued the simplified layouts as "less tiring" (P1) and "visually relaxing," (P3) while others felt that hiding descriptions or details misaligned with how they evaluated products (P2, P6, P7, P8). When these users overrode the system to restore details, they described the result as "finally balanced," showing that satisfaction stemmed not from automation accuracy alone but from the ability to correct misalignments easily. Together, these patterns suggest that users valued the ability to adjust adaptations as much as the adaptations themselves.

6.1.3 Conditional Acceptance and Interpretability. Participants described the interface as "quietly following along" (P4), describing

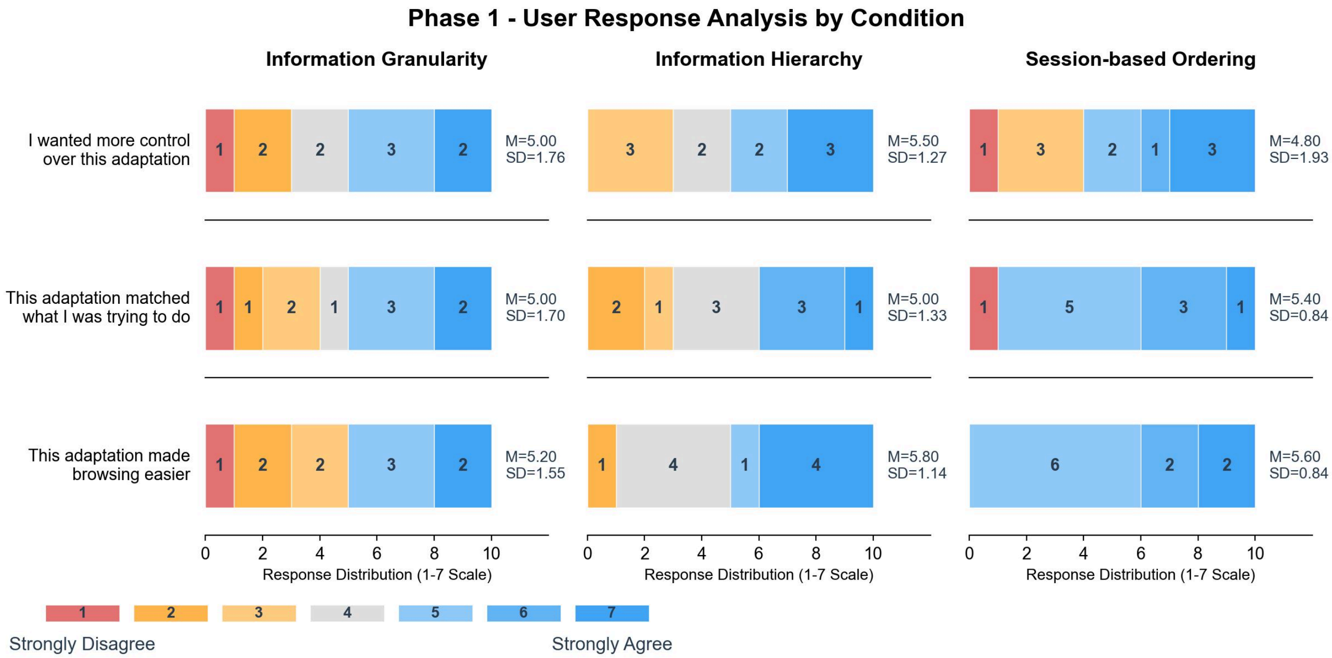

Figure 4: Evaluation results for each adaptation condition and the overall system.

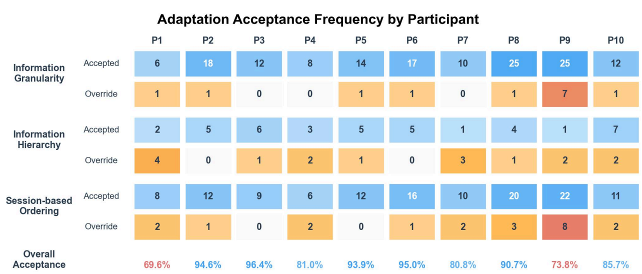

Figure 5: Acceptance rates by adaptation type and participant

it as reactive yet unobtrusive. Automation was welcomed when its actions were predictable and reversible. One participant said, "I trusted it more because it reacted immediately to my clicks" (P8), capturing how responsive feedback fostered confidence. Several emphasized that this acceptance was conditional: they were comfortable with adaptive behavior only when they could confirm or undo it.

6.2 Temporal and Individual Differences

Beyond overall acceptance patterns, participants' behavior revealed when and how they interacted with the system during browsing. Two trends emerged: an initial burst of exploratory overrides that rapidly declined over time, and distinct individual differences that reflected varying acceptance and control strategies.

6.2.1 Early Calibration and Later Stability. Temporal analysis of interaction logs showed that 56% of all overrides (28/50) occurred within the first five minutes of a session. The mean time to first override was 15.6 s (SD = 24.1), indicating that users engaged with control functions almost immediately upon encountering the first adaptation. Participants explained that these early actions were not rejections but trials of alignment—"I wanted to see if it would follow my pattern" (P4)—signaling deliberate calibration rather than frustration. Once preferences were confirmed, most participants relied on the system with minimal further input. Several noted that the system became "predictable after I tested it once or twice," reflecting a transition from active tuning to passive acceptance. This *calibrate–then–accept* rhythm was consistent across most participants.

6.2.2 Control as Confidence. Even when users intervened frequently in the early phase, satisfaction remained high. The item "I could override behavior when needed" received the highest rating across all Likert measures (M = 6.2/7, SD = 0.9), underscoring that perceived controllability was central to user confidence. Participants described how the availability of quick, reversible actions encouraged experimentation: "It was easier to try things because I could always switch back" (P2). Rather than viewing overrides as corrections, users saw them as reassurance that the system respected their agency. This pattern suggests that control visibility and reversibility can foster confidence even when automation accuracy is imperfect.

Exploratory clustering of acceptance rates suggested three recurring adaptation tendencies. *Acceptors* (n = 6; P2, P3, P5, P6, P8, P10) accepted 92% of adaptations with minimal intervention, reflecting comfort with seamless automation. *Selective Adapters* (n = 2; P4, P7) embraced view-level adjustments but often overrode structural reordering (42.5% acceptance), indicating nuanced control preferences. *Customizers* (n = 2; P1, P9) showed the most frequent overrides (30%

and 26%) yet reported high satisfaction (SUS = 75.0, 82.5), demonstrating that frequent tuning can coexist with confidence in the system when control is fluid and predictable.

6.3 Subjective Usability and Workload

Post-session questionnaires and NASA-TLX scores revealed consistently positive perceptions of usability and cognitive comfort. Participants described the adaptive interface as "quietly following along" and "lighter to think through," emphasizing that automation simplified routine actions while keeping browsing smooth and self-directed. Across all adaptations, perceived workload remained low and usability high (Table 1), suggesting that users experienced the system as both effective and cognitively efficient.

6.3.1 System-Level Usability. Participants rated overall usability positively (mean SUS = 71.75), placing the system within the "Good" range. Two participants who intervened most frequently—P1 (30 % overrides) and P9 (26 %)—still rated the system as "Good" (75.0) and "Excellent" (82.5) respectively, reinforcing that frequent overrides did not correspond to poor experiences. Instead, the ability to adjust and recover changes contributed to satisfaction. As one participant noted, "Even when I changed things, it didn't mean I didn't trust it—it just helped me shape it" (P2). This pattern aligns with the patterns observed in the semi-structured interviews: usability remained high even for highly interactive users, demonstrating that controllable automation can sustain engagement.

6.3.2 Workload and Cognitive Flow. NASA-TLX scores (Table 1) showed low mental demand (M = 2.5–2.9) and minimal frustration (M = 1.6–2.5) across all conditions. Participants reported that adaptations reduced redundant actions—such as re-scanning content or recalling previously seen items—allowing smoother attentional flow. Post-study ratings (Phase 2; Figure 6) confirmed that users perceived the adaptive interface as accurate and controllable (all M > 5 on 7-point scale). Participants rated intent alignment highly (M = 5.1 ± 1.1) and acceptance in adaptation (M = 5.5 ± 1.1), consistent with low workload and high usability. Although "I could override behavior when needed" scored slightly lower (M = 4.8 ± 1.3), qualitative feedback indicated that this reflected infrequent need rather than missing functionality. Several remarked that the system "kept things organized so I could focus on deciding, not clicking," and that tasks felt "less tiring overall." These low-effort ratings align with behavioral patterns of sustained acceptance and limited mid-session interventions.

6.3.3 Perceived Experience with Individual Adaptations.

Information Granularity: Low Effort, Mixed Satisfaction. Consistent with its 92% acceptance rate, *Information Granularity* produced the lowest mental demand (2.5 ± 1.08) and frustration (1.6 ± 0.70). However, it also showed the greatest variance in perceived helpfulness (5.2 ± 1.55) and goal alignment (5.0 ± 1.70). Participants diverged on whether simplification enhanced or hindered browsing: some found it "visually relaxing" and efficient, while others felt it "deleted things I actually needed." The result implies that there was reduced cognitive load but occasional loss of confidence when key cues were hidden. Satisfaction peaked immediately after manual adjustment, when users re-enabled preferred elements, suggesting that simplification works best when coupled with fine-grained control.

This mixed response reflected the abstraction of the control itself. The system toggled broad categories (e.g., showing or hiding descriptions) but not finer distinctions participants wanted. Several wished to keep specific cues—such as negative remarks in reviews or price-before-discount information—visible even when other text was hidden. Others wanted gradual emphasis rather than binary visibility, implying that "granularity" is not a fixed density level but a flexible spectrum of cues to preserve or suppress.

Information Hierarchy: Slightly Higher Effort, Highest Payoff. *Information Hierarchy* involved greater cognitive involvement (mental = 2.9 ± 1.37; frustration = 2.5 ± 0.97) but achieved the highest perceived helpfulness (5.8 ± 1.14) and strong intent alignment (5.0 ± 1.33). Participants described it as "effortful but rewarding", requiring attention but leading to smoother, more meaningful flow once they verified that reordering matched their priorities. Many attributed their satisfaction to the transparency and controllability of this adaptation: seeing the logic of reordering and being able to reverse or modify it built confidence. Users reported that they "enjoyed" the interaction precisely because they could understand and influence it, suggesting that light cognitive effort can reinforce confidence when paired with visible, interpretable automation.

Session-based Ordering: Cognitive Continuity. *Session-based ordering* produced the most seamless experience, with low workload (mental = 2.6 ± 1.17; frustration = 1.9 ± 0.88) and high usefulness (5.6 ± 0.84). Participants emphasized how it preserved orientation without demanding attention: "I didn't have to remember what I'd seen—it remembered for me" (P1, P3, P4, P7). The highlighting of previously viewed items reduced repetition and supported comparison, allowing focus to remain on unvisited options. Rather than adding effort, this feature acted as a cognitive extension that sustained flow across pages and sessions.

Several described it as a form of embedded working memory—a lightweight record of exploration integrated directly into the visual field. Unlike external "Liked" or "Saved" lists that require switching contexts, session-based ordering maintained decision traces within the browsing view itself, enabling instant recall and comparison. Participants valued this embedding for keeping both context and progress visible: "Likes take me away from the page, but here I can see what I've already considered" (P6). This in-situ memory reinforced continuity of thought, transforming the interface into a workspace that externalized short-term intention without breaking the browsing rhythm.

6.4 User Reflections and Future Adaptation Expectations

Participants described preferences for how future adaptive systems should behave. Common themes included: starting with sensible defaults, providing easy ways to adjust or undo, and keeping changes visible. The system should act first, but users wanted to step in when needed.

6.4.1 "Initialize, then I'll tweak": Defaults First, Lightweight Control Next. Participants were comfortable with the system automatically

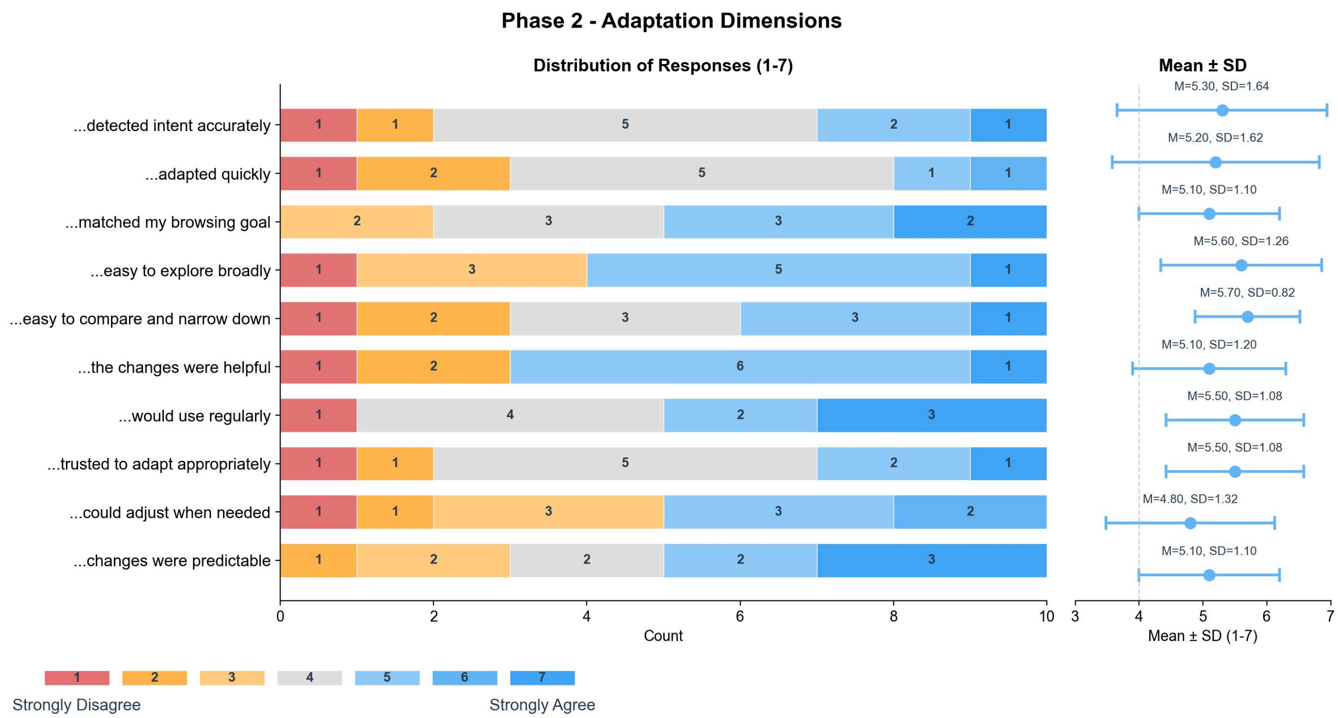

Figure 6: Phase 2 survey results showing participants' ratings of adaptation accuracy, intent alignment, acceptance, and control & transparency (7-point Likert scale).

Table 1: NASA-TLX subscale scores (mean ± SD) for each adaptation condition and the overall system. Lower values indicate lower workload.

Condition	Mental	Physical	Temporal	Performance	Effort	Frustration
A. Information Granularity	2.5 ± 1.08	2.3 ± 1.16	2.1 ± 1.20	5.2 ± 0.92	2.8 ± 0.79	1.6 ± 0.70
B. Information Hierarchy	2.9 ± 1.37	2.7 ± 1.34	1.8 ± 0.79	4.8 ± 1.40	3.1 ± 1.60	2.5 ± 0.97
C. Session-based Ordering	2.6 ± 1.17	2.6 ± 1.35	1.8 ± 0.79	5.6 ± 0.84	2.6 ± 0.84	1.9 ± 0.88
Overall System (Phase 2)	2.4 ± 0.70	2.6 ± 0.70	1.7 ± 0.67	5.4 ± 0.52	3.1 ± 0.99	1.9 ± 0.74

initializing adaptations as long as modifications were easily reversible and visible. They preferred a rhythm where the interface began with a sensible default and users intervened only when goals or contexts changed. As one explained, "It would be good if the user can adjust things later; I think the default setting can stay as is. Even if information is hidden at first, it's fine as long as I can expand it when I need more details" (P1). Another emphasized the value of dependable defaults: "There should be a default. If I had to filter every element myself, I could just use the regular site. The point is to improve browsing with better defaults, but there should be a back button" (P2). Rather than constant tuning, participants envisioned occasional, quick adjustments: "The default should work well enough, and then I could step in for special needs" (P2), with context-sensitive priorities ("It depends on the situation, what deserves priority changes," P3). Some suggested lightweight intent confirmations ("If it asks, 'Are you currently looking for windbreakers?' and I just press Yes or No, that would be perfect," P6), while others wanted a manual mode switch ("If I could simply switch between 'broad' and 'specific' modes myself, that'd be nice," P3). Together, these reflections depict a preferred model of collaboration where the system acts first, but the human remains the easy final arbiter through transparent, low-friction controls.

6.4.2 Recomposition, Not Reduction: Structuring How Information Appears. When asked about deeper customization through generated summaries or AI-curated responses, participants consistently rejected the idea. They preferred to recompose the interface themselves—removing irrelevant sections, reordering useful ones, and blending filtering with layout adjustments. Section-level manipulation emerged as a clear expectation: "I'm never interested in recommendations; I'd rather just remove them" (P2). Another echoed, "I don't look at that section, so I'd like a function to delete it" (P6). Many saw reordering and deletion as complementary: "I'd move reviews to the top and merge that with information granularity adjustments" (P2, P6).

Participants also emphasized speed and persistence in these structural changes. They wanted to toggle between adapted and original layouts effortlessly—"If switching between the changed and original version were quick, I could turn it off at the end just to double-check before booking" (P2)—and to retain customized layouts across sessions ("If I move the detailed images to the top once, will that order stay for other items too?" P1). Customization, in other words, was expected to be lightweight, reversible, and accumulative.

Most participants preferred explicit UI controls to natural-language input. "Clicking a button is faster than typing" (P2), one explained, while others described textual commands as uncertain or unreliable: "It feels more reliable when the options are given visually" (P3). Label clarity further shaped confidence—ambiguous terms like "visual indicator" reduced confidence, whereas straightforward phrasing such as "show previously viewed items" encouraged interaction. The control panel was often described as a "bridge" that made adaptation legible; clear terminology and visible structure made that bridge feel stable and reliable.

Together, these reflections suggest users wanted adaptation to feel like rearranging a layout—visible, reversible, and persistent across sessions.

6.4.3 Human–AI Division of Labor: "You Adjust, I Decide". Participants consistently differentiated between structural assistance (layout changes, highlighting, filtering) and semantic manipulation (summarization, rewriting). They embraced the former as cognitive scaffolding but felt the latter to be intrusive or "joy-reducing." "I don't like it when AI shows through," said P6. "Shopping is supposed to be enjoyable, and summaries don't help—it even makes me doubt the content. Reviews written by people feel more trustworthy." One participant likened transparent adaptation to an "open kitchen": visible processes build confidence, while invisible automation weakens it. "Removing or adding sections is something I do myself," explained P6. "With an LLM, the process is hidden inside a network I can't see. It's like preferring an open kitchen—you trust it more when you can see the steps."

Others expressed a more conditional openness to AI reasoning. As P8 explained, they felt no aversion to AI "capturing what people do consciously or unconsciously," but preferred when adaptation guided them indirectly rather than providing explicit answers. "It's less uncomfortable when the system narrows or rearranges options so that I can decide, rather than telling me the right one," they noted. They described this as preserving a sense of autonomy—"It still feels like my own choice." This preference for indirect, structure-level mediation rather than direct outcome prediction highlights how participants wanted AI to assist judgment without replacing it.

Participants further situated adaptation within familiar interface primitives such as filters and lists. "When I said I wanted something removed, it wasn't filtering—it was just clutter that hurt readability" (P3). Another clarified, "Good filtering at the start makes browsing easier; then the adaptation helps with the actual decision" (P3). Across accounts, participants defined a clear boundary of agency: the system should reorganize, clarify, and surface useful structure, while the human retains judgment and meaning-making. They valued adaptation not as invisible authorship but as visible scaffolding for exploration.

7 Discussion

7.1 Acceptance and Control Through Visible, Reversible Defaults

Participants' acceptance in automation depended less on whether the interface adapted and more on how those adaptations unfolded. They consistently preferred systems that initialized changes confidently by default but remained immediately adjustable. This preference mirrored the behavioral calibration pattern observed, where 56% of overrides occurred in the first five minutes and dropped to near-zero thereafter (Appendix C.4). Participants described this process as "the system starts, and I tweak", positioning the system as a collaborator establishing a provisional baseline.

These findings align with prior work on adaptive transparency and adjustable autonomy [8, 35] showing that users are most comfortable when automation initiates action but keeps the boundaries of control legible. In our study, satisfaction derived not from the absence of intervention but from the assurance that intervention was always possible. Even participants who frequently overrode adaptations (e.g., P9 with 26% override rate) reported high usability (SUS = 82.5), indicating that control availability mattered more than control frequency. Post-study ratings further supported this: "I could override or adjust behavior when needed" received the highest feature score (M = 6.2/7), despite interventions occurring in only 14% of total adaptations.

Participants valued visible reversibility over detailed explanation. They described wanting a "back button," "toggle," or "undo" more than transparency text or system rationale. This preference resonates with findings in explainable AI literature [36, 77] suggesting that users often prefer manipulable feedback over verbalized reasoning when interacting with adaptive systems.

Design-wise, these results suggest that adaptive interfaces should act first but make reversal obvious and immediate. Users wanted a 'back button' at the moment of change—not buried in settings or explained in help text. As prior mixed-initiative research has argued [2, 53], automation that users can see and interrupt earns trust more effectively than automation that's merely accurate.

7.2 Adaptation as Co-Composition

Participants repeatedly described adaptation not as something the system does for them, but as something they wanted to shape with it. Rather than expecting AI to decide what mattered, they sought flexible ways to recompose the interface—removing irrelevant sections, rearranging useful ones, or merging features like filtering and reordering into a single, controllable structure. This orientation reframes adaptation as a collaborative act of composition. As P2 remarked, "There should be a default, but I want to move or delete parts depending on what I care about," while P6 emphasized, "I'd move reviews to the top and merge that with filtering." Participants thus viewed adaptive structure as a manipulable layer of the interface, not an opaque algorithmic behavior.

Empirically, this pattern aligns with the mixed acceptance rates observed across adaptation types (Figure 5). Information Hierarchy, which allowed direct section-level manipulation, achieved the highest subjective helpfulness (M = 5.8 ± 1.14) despite generating the most overrides (33%), while Information Granularity, which offered fewer compositional controls, showed higher acceptance (92%) but greater disagreement about usefulness (SD = 1.55). Participants appeared most satisfied when they could intervene meaningfully—even if that meant rejecting an adaptation—rather than when automation remained entirely invisible. This supports prior findings that highlight user agency and direct manipulation as essential for maintaining engagement and ownership in adaptive systems [60, 76, 108].

Taken together, these patterns suggest participants approached adaptive interfaces as collaborative tools—expecting to shape structure alongside the system. Several noted that adaptation felt 'trustworthy when it worked like editing a layout' but 'distracting when it felt like a black box.' This echoes research on shared control and co-adaptive systems [42], where agency emerges from negotiation rather than strict role separation.

Design-wise, this suggests adaptive interfaces should make structure directly manipulable. Rather than hiding reordering logic, interfaces could offer drag-and-drop surfaces or editable zones for users to refine layouts themselves. Persisting these configurations across sessions would let users shape the system over time—automation proposes, users arrange, and the system learns from those arrangements.

7.3 Designing for Personal–Context–Aware Reflection

Participants envisioned adaptive interfaces that not only adjusted to immediate goals but also evolved with their individual browsing styles. They expected systems to differentiate across domains—recognizing, for instance, that hotel decisions rely on reviews while fashion browsing emphasizes imagery—yet to remember recurring personal heuristics such as "I always check ratings first." This aspiration points toward personal–context–aware adaptation: systems that generalize a user's decision logic without flattening the nuances of each domain.

Empirically, this desire for evolving personalization echoed the temporal calibration pattern observed. Users' interventions clustered early, then stabilized, suggesting that initial overrides acted as a form of self-configuration. Several participants described wanting these calibrated preferences to persist: "If I move the detailed images to the top once, will that order stay for other items too?" (P1). Such statements reveal an appetite for continuity—a wish for adaptive systems that learn not only task context but the user's reasoning rhythm. This aligns with prior work on longitudinal co-adaptation and adaptive scaffolding [51], which emphasizes that personalization should evolve through reciprocal learning between user and system.

Participants also described adaptation as a mirror that revealed their own thinking. Seeing sections dynamically reordered helped them recognize implicit criteria they used to judge quality, such as realizing they consistently prioritized "short coats" or "highly rated hotels." Adaptation thus functioned as a form of reflective infrastructure, externalizing decision patterns that typically remain tacit. These moments of recognition transformed automation from a convenience feature into a tool for self-understanding—a way of observing one's evolving priorities through interface change. This resonates with research on reflective sensemaking and metacognitive interfaces [5, 17], where visibility of one's own choices enhances both control and confidence.

Designing for personal–context–aware reflection therefore requires balancing adaptability with interpretability. Systems should maintain domain sensitivity—respecting that the relevance of visual, textual, or review-based cues differs across contexts—while preserving personal continuity through memory of structural preferences. At the same time, adaptive transformations should remain visible enough to serve as cognitive feedback loops, helping users see how their interactions shape the interface. Future adaptive systems might thus visualize not only content changes but also the evolution of those changes over time, allowing users to trace how their browsing logic refines through use. In this way, adaptation supports not just efficient browsing, but reflective sensemaking: a dynamic conversation between the interface's reorganizing logic and the user's evolving intent.

7.4 Limitations and Future Work

This work investigates in-situ adaptive interfaces through a technology probe that operationalizes three adaptive design dimensions. While the proposed framework articulates five adaptation dimensions, the probe selectively instantiates a subset of these dimensions. As a result, our findings reflect users' experiences with specific types of adaptive changes rather than an exhaustive evaluation of the full design space. Future work could explore how additional dimensions—or combinations of multiple dimensions—interact when implemented together, and how such interactions shape users' experiences of adaptive browsing.

Our user study was designed as a formative, probe-based evaluation intended to surface how users experience, interpret, and negotiate adaptive interface changes as they occur. Accordingly, the study did not incorporate a controlled baseline condition or aim to measure performance improvements relative to an unmodified interface. Instead, our findings characterize users' subjective experiences with adaptive behaviors, rather than establishing causal effects on cognitive workload or efficiency. Future research could build on this work by identifying appropriate baseline conditions for adaptive browsing systems, including task scenarios or stages of browsing where comparisons between adaptive and non-adaptive interfaces are both meaningful and ecologically valid.

Technically, scaling adaptive interfaces across diverse websites remains challenging. Our prototype relied on a multi-step LLM pipeline for HTML parsing and structural manipulation, which worked reliably for pre-configured pages but sometimes showed inconsistent behavior on dynamically rendered or complex DOM structures. These limitations point to a broader systems gap: current GUI-automation and DOM-reduction methods are optimized for targeting elements, not for preserving the hierarchical relationships required for structural reordering and adaptive layout changes. Advancing adaptive interfaces thus requires robust DOM representations designed for in-situ reconfiguration—maintaining container hierarchy, sibling order, and layout stability across dynamic pages. Addressing this challenge would enable generalizable, cross-site adaptive overlays that support rich interface transformations within existing ecosystems.

8 Conclusion

Through ReLay, an LLM-powered adaptive browsing probe, we examined how users experience and respond to automatic interface adaptations during realistic browsing tasks. We contribute a conceptual framework of five adaptation dimensions—representation emphasis, information hierarchy, exploration scope, information granularity, and session-based ordering—that describe how browsing interfaces can flex to support shifting user focus. Our probe implemented three of these dimensions, allowing us to investigate user acceptance, control, and perceived effort. Users accepted most adaptations but wanted control over structural changes. Early overrides served as calibration rather than rejection. Confidence grew when adaptations were visible and reversible; participants described wanting to "tweak" and "teach" the system rather than delegate to it entirely. These findings suggest that effective adaptive interfaces should act first but keep reversal easy, treat user adjustments as input rather than error, and expose structure as something users can shape over time.

Acknowledgments

This work was partly supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2021-II211343, Artificial Intelligence Graduate School Program, Seoul National University), and by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. RS-2025-25421701).

We thank all participants in our user study for their time and thoughtful feedback. The first author also thanks the project team members and colleagues for their support and helpful discussions throughout the project.

Generative AI Usage Disclosure. We used generative AI tools (ChatGPT and Claude) for limited assistance with grammar editing, LaTeX troubleshooting, and formatting tasks (e.g., tables/figures). All content was authored by the research team, and any AI outputs were reviewed and verified by the authors.

References

[1] Christopher Ahlberg and Ben Shneiderman. 1994. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 313–317.

[2] Saleema Amershi, Maya Cakmak, W. Bradley Knox, and Todd Kulesza. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine* 35, 4 (2014), 105–120. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1609/aimag.v35i4.2513 doi:10.1609/aimag.v35i4.2513

[3] Kumaripaba Athukorala, Dorota Głowacka, Giulio Jacucci, Antti Oulasvirta, and Jilles Vreeken. 2016. Is exploratory search different? A comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology* 67, 11 (2016), 2635–2651.

[4] Marcia J Bates. 1989. The design of browsing and berrypicking techniques for the online search interface. *Online review* 13, 5 (1989), 407–424.

[5] Amel Bennaceur, Diane Hassett, Bashar Nuseibeh, and Andrea Zisman. 2023. Values@Runtime: An Adaptive Framework for Operationalising Values. *2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)* (2023), 175–179. https://api.semanticscholar.org/CorpusID:257209652

[6] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI interprets the probes. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1077–1086.

[7] Letizia Bollini. 2017. Beautiful interfaces. From user experience to user interface design. *Des. J.* 20, sup1 (July 2017), S89–S101.

[8] Jeffrey M. Bradshaw, Robert R. Hoffman, David D. Woods, and Matthew Johnson. 2013. The Seven Deadly Myths of "Autonomous Systems". *IEEE Intelligent Systems* 28, 3 (2013), 54–61. doi:10.1109/MIS.2013.70

[9] S Adam Brasel and James Gips. 2015. Interface psychology: Touchscreens change attribute importance, decision criteria, and behavior in online choice. *Cyberpsychol. Behav. Soc. Netw.* 18, 9 (Sept. 2015), 534–538.

[10] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[11] Sara Bunian, Kai Li, Chaima Jemmali, Casper Harteveld, Yun Fu, and Magy Seif Seif El-Nasr. 2021. VINS: Visual Search for Mobile User Interface Design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 423, 14 pages. doi:10.1145/3411764.3445762

[12] Yining Cao, Peiling Jiang, and Haijun Xia. 2025. Generative and Malleable User Interfaces with Generative and Evolving Task-Driven Data Model. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (2025). https://api.semanticscholar.org/CorpusID:276813580

[13] Robert G. Capra and Gary Marchionini. 2008. The relation browser tool for faceted exploratory search. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries* (Pittsburgh PA, PA, USA) *(JCDL '08)*. Association for Computing Machinery, New York, NY, USA, 420. doi:10.1145/1378889.1378967

[14] Robert G. Capra and Manuel A. Perez-Quinones. 2003. Re-Finding Found Things: An Exploratory Study of How Users Re-Find Information. arXiv:cs/0310011 [cs.HC]

[15] Andréia Rodrigues Casare, Celmar Guimarães Da Silva, and Regina Moraes. 2024. User perception as a factor for improving Trustworthiness in e-commerce systems. *Journal on Interactive Systems* 15, 1 (2024), 194–219.

[16] Sylvain Castagnos, Amaury L'Huillier, and Anne Boyer. 2015. Toward a Robust Diversity-Based Model to Detect Changes of Context. In *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 534–541. doi:10.1109/ictai.2015.84

[17] Chance Castaneda, Jessica Mindel, Will Page, Hayden Stec, Manqing Yu, and Kenneth Holstein. 2025. Supporting AI-Augmented Meta-Decision Making with InDecision. *ArXiv* abs/2504.12433 (2025). https://api.semanticscholar.org/CorpusID:277857226

[18] Joseph Chee Chang, Nathan Hahn, and Aniket Kittur. 2020. Mesh: Scaffolding comparison tables for online decision making. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 391–405.

[19] Joseph Chee Chang, Nathan Hahn, and Aniket Kittur. 2020. Mesh: Scaffolding Comparison Tables for Online Decision Making. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 391–405. doi:10.1145/3379337.3415865

[20] Joseph Chee Chang, Nathan Hahn, Adam Perer, and Aniket Kittur. 2019. SearchLens: Composing and capturing complex user interests for exploratory search. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 498–509.

[21] Dongping Chen, Yue Huang, Siyuan Wu, Jingyu Tang, Liuyi Chen, Yilin Bai, Zhigang He, Chenlong Wang, Huichi Zhou, Yiqiang Li, Tianshuo Zhou, Yue Yu, Chujie Gao, Qihui Zhang, Yi Gui, Zhen Li, Yao Wan, Pan Zhou, Jianfeng Gao, and Lichao Sun. 2024. GUI-World: A Video Benchmark and Dataset for Multimodal GUI-oriented Understanding. arXiv:2406.10819 [cs.CV]

[22] Jiaqi Chen, Yanzhe Zhang, Yutong Zhang, Yijia Shao, and Diyi Yang. 2025. Generative Interfaces for Language Models. *arXiv preprint arXiv:2508.19227* (2025).

[23] Wen-Hao Chen, Chin-Chi Hsu, Yi-An Lai, Vincent Liu, Mi-Yen Yeh, and Shou-De Lin. 2019. Attribute-aware recommender system based on collaborative filtering: Survey and classification. *Front. Big Data* 2 (2019), 49.

[24] Yu Chen, Wei-Ying Ma, and Hong-Jiang Zhang. 2003. Detecting web page structure for adaptive viewing on small form factor devices. In *Proceedings of the 12th International Conference on World Wide Web* (Budapest, Hungary) *(WWW '03)*. Association for Computing Machinery, New York, NY, USA, 225–233. doi:10.1145/775152.775184

[25] Yiqun Chen, Jiaxin Mao, Yi Zhang, Dehong Ma, Long Xia, Jun Fan, Daiting Shi, Zhicong Cheng, Simiu Gu, and Dawei Yin. 2024. MA4DIV: Multi-Agent Reinforcement Learning for Search Result Diversification. arXiv:2403.17421 [cs.IR]

[26] Yunnong Chen, Chengwei Shi, and Liuqing Chen. 2025. SpecifyUI: Supporting Iterative UI Design Intent Expression through Structured Specifications and Generative AI. *arXiv preprint arXiv:2509.07334* (2025).

[27] Ed H. Chi, Peter Pirolli, Kim Chen, and James Pitkow. 2001. Using information scent to model user information needs and actions and the Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '01)*. Association for Computing Machinery, New York, NY, USA, 490–497. doi:10.1145/365024.365325

[28] Francesco Chiossi, Johannes Zagermann, Jakob Karolus, Nils Rodrigues, Priscilla Balestrucci, Daniel Weiskopf, Benedikt Ehinger, Tiare Feuchtner, Harald Reiterer, Lewis L Chuang, Marc Ernst, Andreas Bulling, Sven Mayer, and Albrecht Schmidt. 2022. Adapting visualizations and interfaces to the user. *It - Inf. Technol.*

64, 4-5 (Aug. 2022), 133–143.

[29] Fergus IM Craik and Robert S Lockhart. 1972. Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior* 11, 6 (1972), 671–684.

[30] Chrysanthos Dellarocas. 2003. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management science* 49, 10 (2003), 1407–1424.

[31] Catherine Demangeot and Amanda J Broderick. 2010. Exploration and its manifestations in the context of online shopping. *J. Mark. Manag.* 26, 13-14 (Dec. 2010), 1256–1278.

[32] Abdi Diriye, Max Wilson, and Anastasios Tombros. 2010. Revisiting Exploratory Search from the HCI Perspective. (01 2010).

[33] Mateusz Dubiel, Bereket Abera Yilma, Kayhan Latifzadeh, and Luis A. Leiva. 2022. A Contextual Framework for Adaptive User Interfaces: Modelling the Interaction Environment. arXiv:2203.16882 [cs.HC]

[34] John J. Dudley and Per Ola Kristensson. 2018. A Review of User Interface Design for Interactive Machine Learning. *ACM Trans. Interact. Intell. Syst.* 8, 2, Article 8 (June 2018), 37 pages. doi:10.1145/3185517

[35] Mary T. Dzindolet, Scott A. Peterson, Regina A. Pomranky, Linda G. Pierce, and Hall P. Beck. 2003. The role of trust in automation reliance. *Int. J. Hum.-Comput. Stud.* 58, 6 (June 2003), 697–718. doi:10.1016/S1071-5819(03)00038-7

[36] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing Transparency Design into Practice. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18)*. Association for Computing Machinery, New York, NY, USA, 211–223. doi:10.1145/3172944.3172961

[37] Martin J Eppler and Jeanne Mengis. 2004. The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The information society* 20, 5 (2004), 325–344.

[38] Sandra Erdelez. 1999. Information encountering: It's more than just bumping into information. *Bulletin of the American Society for Information Science and technology* 25, 3 (1999), 26–29.

[39] K Anders Ericsson and Herbert A Simon. 1993. Protocol Analysis: Verbal Reports as Data. (1993).

[40] Raymond Fok, Nedim Lipka, Tong Sun, and Alexa F Siu. 2024. Marco: Supporting business document workflows via collection-centric information foraging with large language models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–20.

[41] George W Furnas. 1986. Generalized fisheye views. *Acm Sigchi Bulletin* 17, 4 (1986), 16–23.

[42] Krzysztof Z. Gajos, Daniel S. Weld, and Jacob O. Wobbrock. 2010. Automatically generating personalized user interfaces with Supple. *Artificial Intelligence* 174, 12 (2010), 910–950. doi:10.1016/j.artint.2010.05.005

[43] Alireza Gharahighehi and Celine Vens. 2021. Diversification in session-based news recommender systems. *Personal and Ubiquitous Computing* 27, 1 (July 2021), 5–15. doi:10.1007/s00779-021-01606-4

[44] Jacqueline Gottlieb, Pierre-Yves Oudeyer, Manuel Lopes, and Adrien Baranes. 2013. Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends Cogn. Sci.* 17, 11 (Nov. 2013), 585–593.

[45] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[46] Gerald Häubl and Valerie Trifts. 2000. Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing science* 19, 1 (2000), 4–21.

[47] Marti Hearst. 2006. Design recommendations for hierarchical faceted search interfaces. In *ACM SIGIR workshop on faceted search*. Seattle, WA, 1–5.

[48] Marti A Hearst. 2009. *Search user interfaces*. Cambridge university press.

[49] Orland Hoeber. 2025. Design Principles for Exploratory Search Interfaces. In *Proceedings of the 2025 ACM SIGIR Conference on Human Information Interaction and Retrieval*. 12–22.

[50] Orland Hoeber, Md Nazmul Islam, Miriam Boon, Dale Storie, and Veronica Ramshaw. 2025. Search Timelines: Visualizing Search History to Enable Cross-Session Exploratory Search. arXiv:2504.16741 [cs.HC]

[51] Steffen Holter and Mennatallah El-Assady. 2024. Deconstructing Human-AI Collaboration: Agency, Interaction, and Adaptation. *Computer Graphics Forum* 43 (2024). https://api.semanticscholar.org/CorpusID:269214445

[52] Tom Hope, Jason Portenoy, Kishore Vasan, Jonathan Borchardt, Eric Horvitz, Daniel Weld, Marti Hearst, and Jevin West. 2020. SciSight: Combining faceted navigation and research group detection for COVID-19 exploratory scientific search. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Qun Liu and David Schlangen (Eds.). Association for Computational Linguistics, Online, 135–143. doi:10.18653/v1/2020.emnlp-demos.18

[53] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 159–166. doi:10.1145/302979.303030

[54] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, et al. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 17–24.

[55] Amin Javari and Mahdi Jalili. 2014. A probabilistic model to resolve diversity–accuracy challenge of recommendation systems. *Knowledge and Information Systems* 44, 3 (Aug. 2014), 609–627. doi:10.1007/s10115-014-0779-2

[56] Fengxian Ji, Jingpu Yang, Zirui Song, Yuanxi Wang, Zhexuan Cui, Yuke Li, Qian Jiang, Miao Fang, and Xiuying Chen. 2025. FineState-Bench: A Comprehensive Benchmark for Fine-Grained State Control in GUI Agents. arXiv:2508.09241 [cs.CV] https://arxiv.org/abs/2508.09241

[57] Peiling Jiang and Haijun Xia. 2025. Orca: Browsing at Scale Through User-Driven and AI-Facilitated Orchestration Across Malleable Webpages. *arXiv preprint arXiv:2505.22831* (2025).

[58] Zhenhui Jiang and Izak Benbasat. 2007. The effects of presentation formats and task complexity on online consumers' product understanding. *MIS quarterly* (2007), 475–500.

[59] Harmanpreet Kaur, Doug Downey, Amanpreet Singh, Evie Yu-Yen Cheng, Daniel Weld, and Jonathan Bragg. 2022. FeedLens: Polymorphic lenses for personalizing exploratory search over knowledge graphs. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[60] Anjali Khurana, Xiaotian Su, April Yi Wang, and Parmit K. Chilana. 2025. Do It For Me vs. Do It With Me: Investigating User Perceptions of Different Paradigms of Automation in Copilots for Feature-Rich Software. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (2025). https://api.semanticscholar.org/CorpusID:277993709

[61] Tae Soo Kim, DaEun Choi, Yoonseo Choi, and Juho Kim. 2022. Stylette: Styling the web with natural language. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

[62] Spyros Kolyvas and Petros Kostagiolas. 2023. Information needs and information-seeking behavior of visual artists: a systematic review of the literature. *Library Management* 45, 1-2 (12 2023), 46–70. arXiv:https://www.emerald.com/lm/article-pdf/45/1-2/46/9593556/lm-10-2023-0110.pdf doi:10.1108/LM-10-2023-0110

[63] Andrew Kuznetsov, Joseph Chee Chang, Nathan Hahn, Napol Rachatasumrit, Bradley Breneisen, Julina Coupland, and Aniket Kittur. 2022. Fuse: In-situ sensemaking support in the browser. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[64] Michelle S Lam, Omar Shaikh, Hallie Xu, Alice Guo, Diyi Yang, Jeffrey Heer, James A Landay, and Michael S Bernstein. 2025. Just-In-Time Objectives: A General Approach for Specialized AI Interactions. *arXiv preprint arXiv:2510.14591* (2025).

[65] Neal Lathia, Stephen Hailes, Licia Capra, and Xavier Amatriain. 2010. Temporal diversity in recommender systems. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Geneva, Switzerland) *(SIGIR '10)*. Association for Computing Machinery, New York, NY, USA, 210–217. doi:10.1145/1835449.1835486

[66] Mengqi Liao, S. Shyam Sundar, and Joseph B. Walther. 2022. User Trust in Recommendation Systems: A comparison of Content-Based, Collaborative and Demographic Filtering. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 486, 14 pages. doi:10.1145/3491102.3501936

[67] Chaohui Liu, Xianjin Kong, Xiang Li, and Tongxin Zhang. 2022. Collaborative filtering recommendation algorithm based on user attributes and item score. *Sci. Program.* 2022 (March 2022), 1–7.

[68] Michael Xieyang Liu, Jane Hsieh, Nathan Hahn, Angelina Zhou, Emily Deng, Shaun Burley, Cynthia Taylor, Aniket Kittur, and Brad A Myers. 2019. Unakite: Scaffolding developers' decision-making using the web. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 67–80.

[69] Michael Xieyang Liu, Aniket Kittur, and Brad A Myers. 2022. Crystalline: Lowering the cost for developers to collect and organize information for decision making. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.

[70] Michael Xieyang Liu, Tongshuang Wu, Tianying Chen, Franklin Mingzhe Li, Aniket Kittur, and Brad A Myers. 2024. Selenite: Scaffolding Online Sensemaking with Comprehensive Overviews Elicited from Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–26.

[71] Sandeep Lopez. 2024. Impact of Cognitive Load Theory on the effectiveness of microlearning modules. *European Journal of Education and Pedagogy* 5, 2 (March 2024), 29–35.

[72] Yadong Lu, Jianwei Yang, Yelong Shen, and Ahmed Awadallah. 2024. OmniParser for Pure Vision Based GUI Agent. arXiv:2408.00203 [cs.CV] https://arxiv.org/abs/2408.00203

[73] Aristides Mairena, Martin Dechant, Carl Gutwin, and Andy Cockburn. 2020. A Baseline Study of Emphasis Effects in Information Visualization. In *Proceedings*

of the 46th Graphics Interface Conference 2020, Toronto, ON, Canada, May 28-29, 2020, David I. W. Levin, Fanny Chevalier, and Alec Jacobson (Eds.). Canadian Human-Computer Communications Society, 327–339. doi:10.20380/GI2020.33
[74] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (April 2006), 41–46. doi:10.1145/1121949.1121979
[75] Alan Medlar, Denis Kotkov, and Dorota Głowacka. 2024. Unexplored Frontiers: A Review of Empirical Studies of Exploratory Search. In *ACM SIGIR Forum*, Vol. 58. ACM New York, NY, USA, 1–19.
[76] Leila Methnani, Andrea Aler Tubella, Virginia Dignum, and Andreas Theodorou. 2021. Let Me Take Over: Variable Autonomy for Meaningful Human Control. *Frontiers in Artificial Intelligence* 4 (2021). https://api.semanticscholar.org/CorpusID:237496135
[77] Tim Miller. 2018. Explanation in Artificial Intelligence: Insights from the Social Sciences. arXiv:1706.07269 [cs.AI] https://arxiv.org/abs/1706.07269
[78] Bryan Min, Allen Chen, Yining Cao, and Haijun Xia. 2025. Malleable Overview-Detail Interfaces. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (2025). https://api.semanticscholar.org/CorpusID:276929088
[79] Wendy W. Moe. 2003. Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream. *Journal of Consumer Psychology* 13, 1 (2003), 29–39. doi:10.1207/S15327663JCP13-1&2_03 Consumers in Cyberspace.
[80] Susan M Mudambi and David Schuff. 2010. Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS quarterly* (2010), 185–200.
[81] Arpit Narechania, Shunan Guo, Eunyee Koh, Alex Endert, and Jane Hoffswell. 2025. Utilizing Provenance as an Attribute for Visual Data Analysis: A Design Probe with ProvenanceLens. arXiv:2505.11784 [cs.HC]
[82] Phong H. Nguyen, Kai Xu, Andy Bardill, Betul Salman, Kate Herd, and B.L. William Wong. 2016. SenseMap: Supporting browser-based online sensemaking through analytic provenance. In *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*. 91–100. doi:10.1109/VAST.2016.7883515
[83] Jakob Nielsen. 2006. Progressive disclosure. *Nielsen Norman Group* 1 (2006).
[84] Hongjing Niu, Hanting Li, Feng Zhao, and Bin Li. 2022. Domain-Unified Prompt Representations for Source-Free Domain Generalization. arXiv:2209.14926 [cs.CV]
[85] Xi Niu, Xiangyu Fan, and Tao Zhang. 2019. Understanding faceted search from data science and human factor perspectives. *ACM Transactions on Information Systems (TOIS)* 37, 2 (2019), 1–27.
[86] Jeongeon Park, Bryan Min, Kihoon Son, Jean Y Song, Xiaojuan Ma, and Juho Kim. 2023. Choicemates: Supporting unfamiliar online decision-making with multi-agent conversational interactions. *arXiv preprint arXiv:2310.01331* (2023).
[87] John W. Payne, James R. Bettman, and Eric J. Johnson. 1993. *The Adaptive Decision Maker.* Cambridge University Press. doi:10.1017/CBO9781139173933
[88] Peter Pirolli and Stuart Card. 1999. Information foraging. *Psychological review* 106, 4 (1999), 643.
[89] Petri Puustinen, Kostas Stefanidis, Jaana Kekäläinen, and Marko Junkkari. 2021. Reinforcement learning page prediction for hierarchically ordered municipal websites. *Information (Basel)* 12, 6 (May 2021), 231.
[90] Yin Qu. 2016. Supporting Ideation by Integrating Exploratory Search, Browsing, and Curation. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval* (Carrboro, North Carolina, USA) *(CHIIR '16)*. Association for Computing Machinery, New York, NY, USA, 361–363. doi:10.1145/2854946.2854948
[91] Michael Quade, Marco Blumendorf, Grzegorz Lehmann, Dirk Roscher, and Sahin Albayrak. 2011. Evaluating user interface adaptations at runtime by simulating user interaction. In *Electronic Workshops in Computing*. BCS Learning & Development.
[92] Peter S Riefer, Rosie Prior, Nicholas Blair, Giles Pavey, and Bradley C Love. 2017. Coherency maximizing exploration in the supermarket. *Nat. Hum. Behav.* 1, 1 (Jan. 2017), 0017.
[93] Thassilo M Schiepanski and Nicholas Piël. 2025. Beyond Pixels: Exploring DOM Downsampling for LLM-Based Web Agents. *arXiv preprint arXiv:2508.04412* (2025).
[94] Tobias Schnabel, Paul N. Bennett, Susan T. Dumais, and Thorsten Joachims. 2016. Using Shortlists to Support Decision Making and Improve Recommender System Performance. In *Proceedings of the 25th International Conference on World Wide Web* (Montréal, Québec, Canada) *(WWW '16)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 987–997. doi:10.1145/2872427.2883012
[95] Eric Schulz, Rahul Bhui, Bradley C Love, Bastien Brier, Michael T Todd, and Samuel J Gershman. 2019. Structured, uncertainty-driven exploration in real-world consumer choice. *Proc. Natl. Acad. Sci. U. S. A.* 116, 28 (July 2019), 13903–13908.
[96] Shahram Sedghi, Zeinab Shourmeij, and Iman Tahamtan. 2017. Exploring the context of visual information seeking. arXiv:1708.01856 [cs.IR]
[97] Omar Shaikh, Shardul Sapkota, Shan Rizvi, Eric Horvitz, Joon Sung Park, Diyi Yang, and Michael S Bernstein. 2025. Creating general user models from computer use. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–23.
[98] Danqing Shi, Weiwei Cui, Danqing Huang, Haidong Zhang, and Nan Cao. 2022. Reverse-Engineering Information Presentations: Recovering Hierarchical Grouping from Layouts of Visual Elements. arXiv:2201.05194 [cs.HC]
[99] Ben Shneiderman. 2003. The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization*. Elsevier, 364–371.
[100] Johanna M Silvennoinen and Jussi P P Jokinen. 2016. Appraisals of salient visual elements in web page design. *Adv. Hum.-Comput. Interact.* 2016 (2016), 1–14.
[101] Ning Su, Jiyin He, Yiqun Liu, Min Zhang, and Shaoping Ma. 2018. User Intent, Behaviour, and Perceived Satisfaction in Product Search. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (Marina Del Rey, CA, USA) *(WSDM '18)*. Association for Computing Machinery, New York, NY, USA, 547–555. doi:10.1145/3159652.3159714
[102] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling multilevel exploration and sensemaking with large language models. In *Proceedings of the 36th annual ACM symposium on user interface software and technology*. 1–18.
[103] Fei Tang, Haolei Xu, Hang Zhang, Siqi Chen, Xingyu Wu, Yongliang Shen, Wenqi Zhang, Guiyang Hou, Zeqi Tan, Yuchen Yan, Kaitao Song, Jian Shao, Weiming Lu, Jun Xiao, and Yueting Zhuang. 2025. A Survey on (M)LLM-Based GUI Agents. arXiv:2504.13865 [cs.HC] https://arxiv.org/abs/2504.13865
[104] Choon Hui Teo, Houssam Nassif, Daniel Hill, Sriram Srinivasan, Mitchell Goodman, Vijai Mohan, and S.V.N. Vishwanathan. 2016. Adaptive, Personalized Diversity for Visual Discovery. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. ACM, 35–38. doi:10.1145/2959100.2959171
[105] Choon Hui Teo, Houssam Nassif, Daniel Hill, Sriram Srinivasan, Mitchell Goodman, Vijai Mohan, and SVN Vishwanathan. 2016. Adaptive, personalized diversity for visual discovery. In *Proceedings of the 10th ACM conference on recommender systems*. 35–38.
[106] Kashyap Todi, Gilles Bailly, Luis Leiva, and Antti Oulasvirta. 2021. Adapting User Interfaces with Model-based Reinforcement Learning. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 573, 13 pages. doi:10.1145/3411764.3445497
[107] Elaine G Toms et al. 2000. Serendipitous Information Retrieval.. In *DELOS*.
[108] Carlos Toxtli. 2024. Human-Centered Automation. *ArXiv* abs/2405.15960 (2024). https://api.semanticscholar.org/CorpusID:270063087
[109] Koki Tsumura, Junko Shirogane, Hajime Iwata, and Yoshiaki Fukazwa. [n. d.]. LAYOUT-DRIVEN WEBPAGE READING. ([n. d.]).
[110] Christine C A van Nooijen, Bjorn B de Koning, Wichor M Bramer, Anna Isahakyan, Maryam Asoodar, Ellen Kok, Jeroen J G van Merrienboer, and Fred Paas. 2024. A cognitive load theory approach to understanding expert scaffolding of visual problem-solving tasks: A scoping review. *Educ. Psychol. Rev.* 36, 1 (March 2024).
[111] Saúl Vargas and Pablo Castells. 2011. Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems* (Chicago, Illinois, USA) *(RecSys '11)*. Association for Computing Machinery, New York, NY, USA, 109–116. doi:10.1145/2043932.2043955
[112] Scarlett Wesley, Melody LeHew, and Arch G. Woodside. 2006. Consumer decision-making styles and mall shopping behavior: Building theory using exploratory data analysis and the comparative method. *Journal of Business Research* 59, 5 (2006), 535–548. doi:10.1016/j.jbusres.2006.01.005
[113] Ryen W. White, Steven M. Drucker, Gary Marchionini, Marti Hearst, and m. c. schraefel. 2007. Exploratory search and HCI: designing and evaluating interfaces to support exploratory search interaction. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems* (San Jose, CA, USA) *(CHI EA '07)*. Association for Computing Machinery, New York, NY, USA, 2877–2880. doi:10.1145/1240866.1241100
[114] Ryen W. White, Gary Marchionini, and Gheorghe Muresan. 2008. Evaluating exploratory search systems: Introduction to special topic issue of information processing and management. *Inf. Process. Manag.* 44 (2008), 433–436. https://api.semanticscholar.org/CorpusID:16259772
[115] Ryen W White and Resa A Roth. 2009. Exploratory search: Beyond the query-response paradigm. *Synth. Lect. Inf. Concepts Retr. Serv.* 1, 1 (Jan. 2009), 1–98.
[116] Robert C Wilson, Andra Geana, John M White, Elliot A Ludvig, and Jonathan D Cohen. 2014. Humans use directed and random exploration to solve the explore-exploit dilemma. *J. Exp. Psychol. Gen.* 143, 6 (Dec. 2014), 2074–2081.
[117] Jason Wu, Yi-Hao Peng, Xin Yue Amanda Li, Amanda Swearngin, Jeffrey P Bigham, and Jeffrey Nichols. 2024. UIClip: a data-driven model for assessing user interface design. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–16.
[118] Jason Wu, Kashyap Todi, Joannes Chan, Brad A. Myers, and Ben Lafreniere. 2024. FrameKit: A Tool for Authoring Adaptive UIs Using Keyframes. *Proceedings of the 29th International Conference on Intelligent User Interfaces* (2024). https://api.semanticscholar.org/CorpusID:268982025

[119] Shangxuan Wu, Wendi Luan, Yong Wang, Dan Zeng, Qiaomu Shen, and Bo Tang. 2025. Data Insights as Data: Quick Overview and Exploration of Automated Data Insights. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.

[120] Yuan Xu, Shaowen Xiang, Yizhi Song, Ruoting Sun, and Xin Tong. 2025. DuetUI: A Bidirectional Context Loop for Human-Agent Co-Generation of Task-Oriented Interfaces. *arXiv preprint arXiv:2509.13444* (2025).

[121] Pei Yang, Hai Ci, and Mike Zheng Shou. 2025. macOSWorld: A Multilingual Interactive Benchmark for GUI Agents. arXiv:2506.04135 [cs.AI]

[122] Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst. 2003. Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 401–408.

[123] Ryan Yen, Nicole Sultanum, and Jian Zhao. 2024. To search or to gen? Exploring the synergy between generative AI and web search in programming. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[124] Enes Yigitbas. 2020. Model-driven engineering and usability evaluation of self-adaptive user interfaces. *SIGWEB Newsl.* 2020, Autumn, Article 2 (Nov. 2020), 4 pages. doi:10.1145/3427478.3427480

[125] Chaoyun Zhang, Shilin He, Jiaxu Qian, Bowen Li, Liqun Li, Si Qin, Yu Kang, Minghua Ma, Guyue Liu, Qingwei Lin, Saravan Rajmohan, Dongmei Zhang, and Qi Zhang. 2025. Large Language Model-Brained GUI Agents: A Survey. arXiv:2411.18279 [cs.AI] https://arxiv.org/abs/2411.18279

[126] Huiwen Zhang. 2023. Investigating Online Browsing Practices. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23)*. Association for Computing Machinery, New York, NY, USA, 479–481. doi:10.1145/3576840.3578293

[127] Mi Zhang and Neil Hurley. 2008. Avoiding monotony: improving the diversity of recommendation lists. In *Proceedings of the 2008 ACM Conference on Recommender Systems* (Lausanne, Switzerland) *(RecSys '08)*. Association for Computing Machinery, New York, NY, USA, 123–130. doi:10.1145/1454008.1454030

[128] Nuowen Zhang, Jing Zhang, Shangsong Jiang, and Weijia Ge. 2024. The effects of layout order on interface complexity: an eye-tracking study for dashboard design. *Sensors* 24, 18 (2024), 5966.

[129] Yuan Zhang. 2022. Variety-seeking behavior in consumption: A literature review and future research directions. *Front. Psychol.* 13 (June 2022), 874444.

[130] Tiejun Zhu and Yujin Yang. 2023. Research on mobile learning platform interface design based on college students' visual attention characteristics. *PLOS ONE* 18, 7 (07 2023), 1–16. doi:10.1371/journal.pone.0283778

[131] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web* (Chiba, Japan) *(WWW '05)*. Association for Computing Machinery, New York, NY, USA, 22–32. doi:10.1145/1060745.1060754

[132] Valentin Zieglmeier and Antonia Maria Lehene. 2021. Designing Trustworthy User Interfaces. In *33rd Australian Conference on Human-Computer Interaction (OzCHI '21)*. ACM, 182–189. doi:10.1145/3520495.3520525

A ReLay Prompts

We include representative prompts used in ReLay for intent classification and interface adaptation. Below, we show the prompts used for the Information Granularity adaptation.

A.1 DOM Analysis Prompt

```
Task: Given an e-commerce item-card HTML snippet, extract reliable
      CSS selectors for content-density control.

Categories:
- essential_elements: image, title, current price, primary CTA
- secondary_elements: extra metadata (specs, shipping, badges,
      ratings details)
- expandable_elements: sections suitable for collapse (reviews,
      variants/spec tables, policies)
- text_elements: long text suitable for truncation (title,
      description, feature list)

Return JSON:
{
  "container_selector": ".item",
  "essential_elements": [".img", ".price", ".title", "..."],
  "secondary_elements": [".meta", "..."],
  "expandable_elements": [".reviews", "..."],
  "text_elements": [".title", ".desc"]
}

Guidelines: use specific class/data-* selectors; group related
      elements.
```

A.2 Adaptation Prompt

```
Condition: DIVERSIVE intent / catalog page / LOW density.
Goal: maximize scanning speed and reduce cognitive load.

Rules:
- Hide secondary/metadata elements (e.g., specs, shipping details,
      SKU, review counts, variant/stock info; keep only key badges
      like Sale/New).
- Truncate text aggressively (title: 1 line; description: <=2
      lines).
- Simplify ratings (stars only; hide numeric ratings and counts).
- Compact layout (reduce spacing and optionally font sizes).

Allowed actions:
hide | truncate (config: {"max_lines": 1-3}) | reduce_spacing
      (config: {"padding": "...", "gap": "..."})
reduce_font_size (config: {"font_size": "..."}) | collapse

Return JSON with "density_actions": [{"action": ..., "selectors":
      [...], "config": ..., "reason": "..."}].
Example:
{
  "density_actions": [
    {"action":"hide","selectors":[".specs",".shipping",
      "..."],"config":null,"reason":"..."}
  ]
}
```

A.3 Intent Detection Prompt

```
Task: Classify a user's browsing session intent as DIVERSIVE
      (exploratory) or SPECIFIC (goal-oriented) using three inputs:
(1) ACTION LOGS (click/search/navigation), (2) BEHAVIOR METRICS
      (e.g., dwell time, scroll velocity), and (3) ITEM TRACKING
      (scrolledItems, clickedItems).

Intent definitions:
- DIVERSIVE: broad exploration across items/categories with shallow
      engagement.
- SPECIFIC: focused search/comparison within a category with deeper
      engagement.

Derive the following indicators from the inputs:
1) filter_usage (none/low/moderate/high): clicks on UI controls
      such as filter/sort, size/color/brand selectors, or price
      sliders.
2) search_specificity (general/moderate/specific): broad queries
      (e.g., "shoes") vs specific queries including
      brands/models/attributes.
3) browsing_breadth (narrow/moderate/wide): category diversity
      across scrolledItems/clickedItems and navigation transitions.
4) engagement_depth (shallow/moderate/deep): combine avgDwellTime
      and avgScrollVelocity; repeated clicks on the same/similar
      items indicate deeper comparison.
5) price_focus (true/false): interaction with price filters/sort,
      or clicked items within a narrow price range.

Return ONLY a valid JSON object:
{
  "intent": "diversive" | "specific",
  "confidence": 0.0-1.0,
  "reasoning": "...",
  "indicators": {
    "filter_usage": "...",
    "search_specificity": "...",
    "browsing_breadth": "...",
    "engagement_depth": "...",
    "price_focus": true|false
  }
}
```

B ReLay Interface

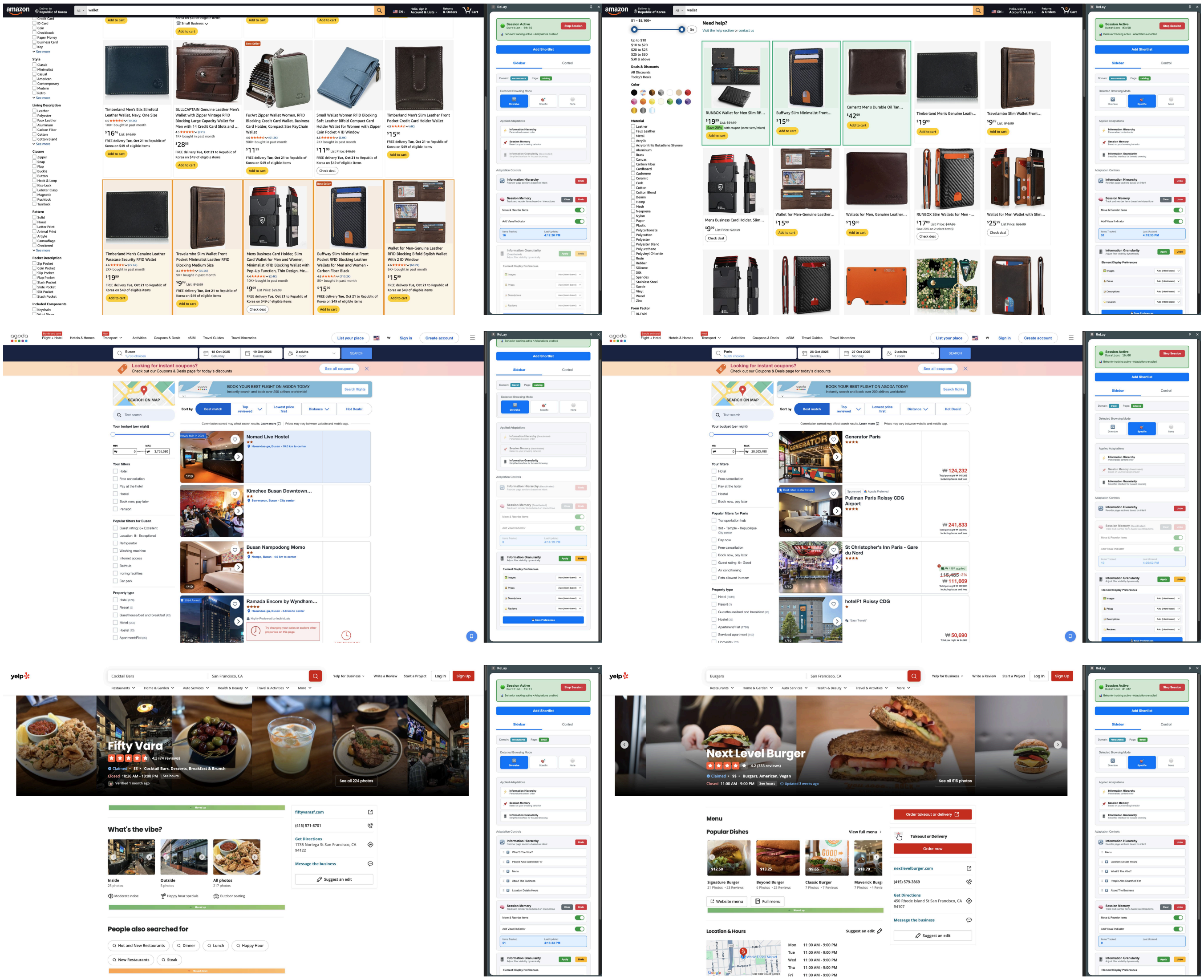

Figure 7: ReLay interface demonstrating adaptive behaviors across browsing modes and dimensions. Each row illustrates one adaptation dimension, while columns compare *diverse* (left) and *specific* (right) browsing modes. The right side of each screenshot shows the adaptive control panel, where users can view or adjust parameters for information hierarchy, granularity, and session-based ordering (session memory). The top row shows session-based ordering adaptations, where previously viewed or clicked items are visually marked or repositioned to emphasize novelty versus familiarity. The middle row depicts information granularity, adjusting how much detail appears in catalog listings (e.g., minimal attributes for exploration vs. full specifications for focused comparison). The bottom row illustrates information hierarchy on detail pages, reordering sections such as specifications and recommendations to reflect changing browsing focus. Together, these examples show how ReLay applies lightweight, in-situ adaptations on real websites across domains such as shopping and dining.

C User Study Details

C.1 Standard Questionnaires

Table 2: NASA-TLX items (1–7 Likert) administered after each condition in Phase 1 and after the overall system in Phase 2.

Item	Question text
1	How mentally demanding was the task?
2	How physically demanding was the task?
3	How hurried or rushed was the pace of the task?
4	How successful were you in accomplishing what you set out to do?
5	How hard did you have to work to accomplish your level of performance?
6	How insecure, discouraged, irritated, stressed, and annoyed were you?

Table 3: System Usability Scale (SUS) items (1–7 Likert) administered after Phase 2.

Item	Question text
1	I think that I would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.

C.2 Phase 1 Additional Post-condition Items

Table 4: Phase 1 additional post-condition items (1–7 Likert).

Item	Question text
1	This adaptation made browsing easier.
2	This adaptation matched what I was trying to do.
3	I wanted more control over this adaptation.

C.3 Phase 2 Additional Post-system Items

Table 5: Phase 2 post-system adaptation perception items (1–7 Likert). Items are grouped by conceptual category.

No.	Category	Question text
1	Adaptation Accuracy	The system detected my browsing intent accurately.
2	Adaptation Accuracy	When my goals changed, the interface adapted quickly.
3	Intent Alignment	The interface matched my current browsing goal.
4	Intent Alignment	It was easy to explore broadly when I wanted.
5	Intent Alignment	It was easy to compare and narrow down when I wanted.
6	Acceptance	The interface changes were helpful overall.
7	Acceptance	I would use this interface regularly for similar tasks.
8	Acceptance	I trusted the interface to adapt appropriately.
9	Control & Transparency	I could override or adjust behavior when needed.
10	Control & Transparency	The interface changes were predictable and understandable.

C.4 Timeline of Phase 2 Study

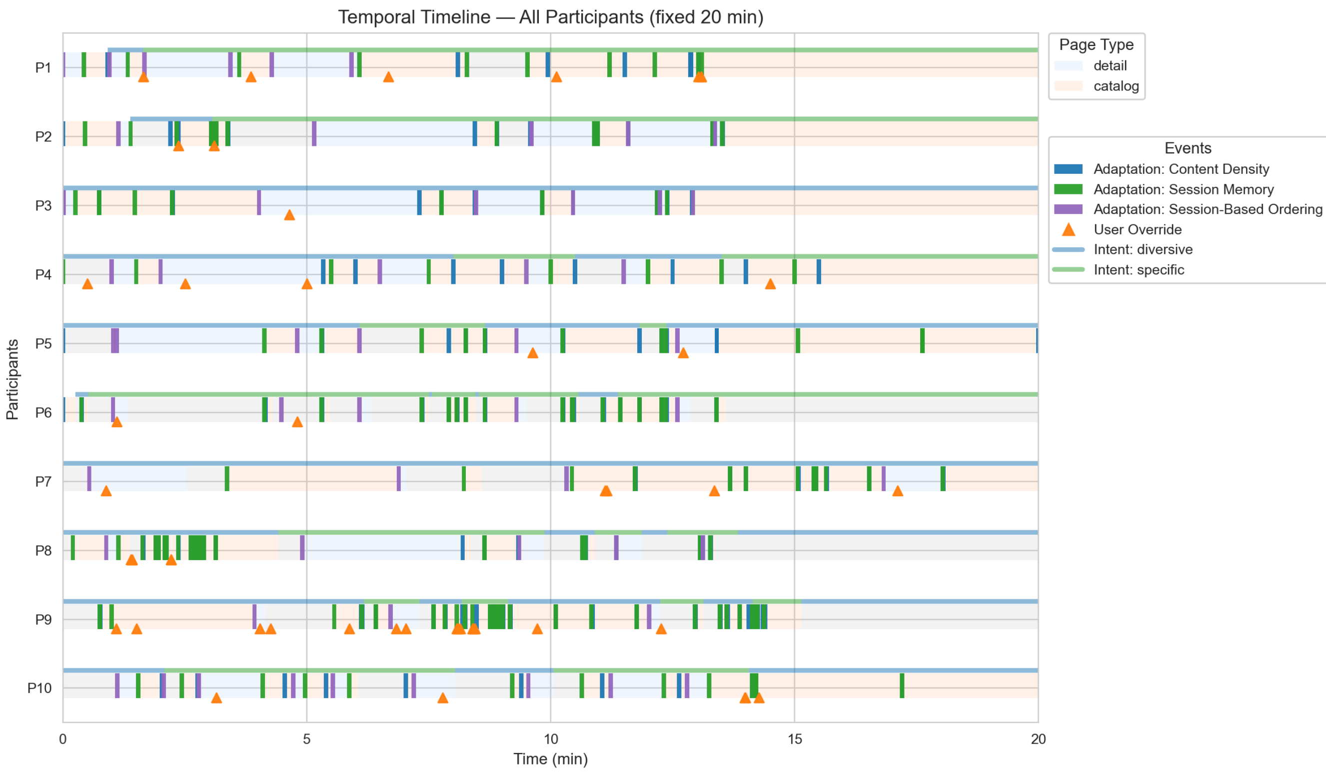

Figure 8: Adaptation application and intervention sequence for each participant in Phase 2.