

INLAY: Preemptive, In-Context Intelligence for Casual Web Browsing

Pratyay Suvarnapathaki
t-pratyays@microsoft.com
Microsoft Research
Bengaluru, India

Adithya S. Kolavi
t-akolavi@microsoft.com
Microsoft Research
Bengaluru, India

Harsh Vijay
havijay@microsoft.com
Microsoft Research
Bengaluru, India

Mayukh Das
mayukhdas@microsoft.com
M365 Research, Microsoft
Bengaluru, India

Ajay Manchepalli
ajayma@microsoft.com
Microsoft Research
Bengaluru, India

Venkata N. Padmanabhan
padmanab@microsoft.com
Microsoft Research
Bengaluru, India

Abstract

AI assistance in browsers often depends on explicit invocation, which is effective for well-defined tasks but imposes significant overhead during exploratory ‘casual’ browsing where intent is implicit or latent. We introduce Inlay, a preemptive **IN**telligence **LAY**er that bridges the intent-action gap herein by proactively embedding AI-driven insights and cues directly within webpages. Grounded in Information Foraging Theory, Inlay enhances ‘information scents’ by shifting users from high-effort query formulation to low-effort intent confirmation while preserving user agency. Our preliminary user study (N=12) provides early signals that Inlay’s context-aware interventions may anticipate emergent needs and encourage deeper exploration while sustaining foraging flow. Our findings suggest that proactive augmentation can transform browsing from command-driven interaction to collaborative AI-assisted exploration.

CCS Concepts

• **Human-centered computing** → **Web-based interaction**; • **Computing methodologies** → *Artificial intelligence*.

Keywords

Proactive Artificial Intelligence, AI-assisted Browsing, Contextual Web, Mixed-Methods User Study

ACM Reference Format:

Pratyay Suvarnapathaki, Adithya S. Kolavi, Harsh Vijay, Mayukh Das, Ajay Manchepalli, and Venkata N. Padmanabhan. 2026. INLAY: Preemptive, In-Context Intelligence for Casual Web Browsing. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3772363.3799065>

1 Introduction

Web browsing exemplifies human decision-making under constraints, where each click balances cognitive effort against uncertain information gain. Augmenting this environment with LLM-driven assistance promises to radically enhance the efficiency and ease of information-seeking across the diverse spectrum of browser-based experiences. Yet, current systems herein (e.g., Copilot-in-Edge [11], Perplexity Comet [23]) have converged on a singular interaction paradigm: deliberate invocation by the user within a separate UI substrate (typically a ‘sidebar’).

While seemingly effective for goal-oriented tasks, this reactive model imposes cognitive overhead [24] during ‘casual browsing,’ [5, 13] (meandering, serendipitous information seeking) where intent is often implicit or latent, thereby forcing users to *prematurely satisfice*, i.e. settle for subpar results to avoid the effort of context switching, missing opportunities for deeper engagement.

We argue that as agentic capabilities mature, the browser interface must evolve beyond this reactive status quo to address the above limitation. We introduce Inlay as an early provocation herein, positioned *diametrically opposite* to the current paradigm: an in-page intelligence layer that proactively bridges the gap between latent intent and action. Instead of awaiting invocation, Inlay proactively embeds contextually relevant suggestions as ‘foraging cues’ for the user directly within the webpage. This allows users to confirm or dismiss AI assistance in-situ, drastically reducing information foraging costs while preserving user agency.

We operationalize this paradigm as a Chromium extension designed to augment, rather than replace, existing browser agents. This work explores the design and efficacy of this new paradigm. The contributions of the current work are as follows:

- (1) **Design Principles** for preemptive, AI-assisted information foraging in casual browsing;
- (2) **The Inlay System**, comprising browser DOM parsing, LLM-based intent inference, and non-destructive overlay injection; and
- (3) **Insights** from a small-scale exploratory user study, suggesting that context-aware proactivity can anticipate emergent needs and deepen exploration while sustaining user agency.



This work is licensed under a Creative Commons Attribution 4.0 International License. *CHI EA '26, Barcelona, Spain*

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2281-3/2026/04
<https://doi.org/10.1145/3772363.3799065>

2 Related Work

Browsing as a Foraging Activity. The Information Foraging Theory (IFT) view of web browsing models it as a cost-reward optimization problem under bounded rationality [14] in which information seeking is akin to animals' foraging patterns. Users follow *information scent*, in the form of cues like headings and links that signal value, exploiting *patches* (websites) until diminishing returns trigger them to explore elsewhere [3]. This posits humans as *satisficers* who seek 'good enough' solutions within cognitive and temporal constraints [16]. Consequently, web browsing is an adaptive strategy, i.e. an exploration-exploitation tradeoff between staying within a current patch versus seeking new information sources [2, 21]. An imbalance herein triggers premature satisficing, where users tend to give up on exploration and settle for subpar outcomes when facing cognitive overload and time pressure [1, 12].

Browser-based AI. Agentic automation systems seek to reduce the manual effort involved in foraging, compressing multi-step browsing workflows into single-prompt instructions. This capability is now ubiquitous, available through consumer-facing 'AI browsers' [11, 18, 23] and developer-focused cloud solutions [4, 19]. Parallely, recent CHI scholarship [9, 20] has leveraged these developments, innovating on Human-AI collaboration paradigms to plan and visualize intent-driven foraging flows, as well as organize and analyse the extracted information patches therein.

Predominantly however, these systems and studies tend to adhere to the aforementioned *reactive invocation* paradigm, requiring users to pause their browsing, divert attention to a separate pane, and explicitly command an agent. This creates a persistent cognitive overhead [17, 24] owing to: (a) *initiation friction*, where the user must remember to divert attention from the content to the tool; and (b) *cognitive translation friction*, forcing the user to formalize implicit, emergent intent into explicit prompts. This interaction overhead risks fragmenting the foraging loop, diminishing opportunities for serendipitous discovery and subverting the contextual grounding essential to casual browsing.

Proactive Agentic Systems. Recognizing the burden of explicit instruction, recent AI research also explores proactive agents that anticipate user needs in varied forms, including human-in-the-loop UI automation agents [22], data-driven proactive agents spanning various productivity tasks [6, 7, 10], and programming assistant agents [15]. These agents, however, are geared toward the end-to-end automation of tasks that have an explicit end goal or success criterion. In the open-ended context of casual browsing, a truly effective proactive system ought to focus on lowering the initiation barrier specifically, by offering context-aware suggestions aligned with probable user intent without hijacking user agency by over-automating the process.

3 Approach

3.1 Interaction Principles

Inlay's design departs from traditional agentic interfaces. Grounded in IFT, it treats AI-generated suggestions as *information scent cues*, as opposed to commands or queries, to lower exploitation costs

within a webpage (information patch [14]) while preserving the user's foraging role.

3.1.1 Preemptive Assistance as Mixed-Initiative Design. IFT posits that users assess patch value via ambient cues (e.g., headings, links) before deciding to exploit or abandon. Inlay extends this principle by operationalizing Horvitz's vision of an "elegant coupling of automated services with direct manipulation" [8]. Inlay's client-side module activates preemptively on page load and strictly limits its role to generating *potential* foraging cues (scents). This allows both, the user and the AI agent to play to their strengths: the agent assumes the computational burden of scanning and assessing on-page information, while the human remains in the driver's seat, retaining the executive authority to act on or ignore these cues. For example, as the user scans through a product review page, Inlay might surface a cue to "Compare with competing models" adjacent to a spec table; similarly, on a wiki about a space mission, it may suggest exploring related missions and their outcomes. This preemptive deployment inherently reduces initiation friction to zero, directly addressing the main factor contributing to premature satisficing behavior [1]. Additionally, there is the prospect of serendipity, with Inlay surfacing latent lines of inquiry and potential actions that the user may not have been able to independently formulate.

3.1.2 Scent Taxonomy and Contextual Placement. Not all scents are equally legible. To combat cognitive translation friction when users have vague goals and lack domain-specific vocabulary, Inlay employs a constrained taxonomy: *{compare, verify, explain, extract, predict, uncover}*. Each maps to a recurrent exploitation strategy; e.g., a *compare* widget may appear by alternatives, a *verify* widget by a factual claim.

Scent efficacy hinges on proximity to the relevant element. Inlay employs four placement modalities for different attentional gradients: *the key insight banner* (top of viewport, Fig.1 label 'g') offers a high-level assessment for rapid scanning. *Inline suggestions* (in text flow, Fig.1 label 'd') appear by paragraphs/entities for specific actions. *In-context widgets* attach to discrete DOM elements (Fig.1 label 'c') with highlighting, preserving context. Finally, *floating widgets* (Fig.1 label 'b') persist for page-wide workflows (e.g., "Generate article digest") untied to a single element. This tiered placement matches scent intensity to relevance, with hyper-local cues for fine-grained exploitation and ambient cues for holistic exploration.

3.1.3 Agency-Preserving Interaction Model. Agentic AI risks automating sensemaking and disengaging users (see Section 2). Inlay mitigates this by design, by framing its suggestions as *low-friction cues*. This model allows the user to either act on a suggestion by externalizing an implicit intent to be passed to a downstream agent for execution (Fig. 1 label 'f'), or simply dismiss parts of, or the entire overlay (Fig. 1 label 'a') and continue their original foraging trajectory unimpeded. This approach operationalizes Yun et al.'s human-in-the-loop foraging principles [22], but tailored to the latent-intent case of casual browsing: the system proposes but the user decides, ensuring the locus of control and cognitive benefits of active exploration remain with the human forager.

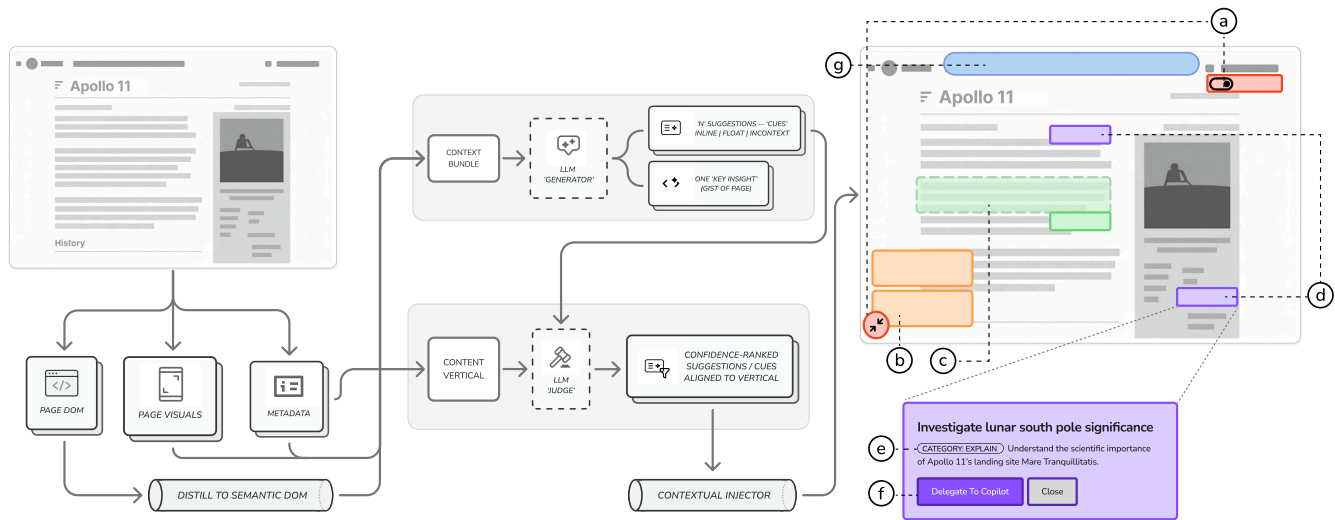


Figure 1: The three-stage pipeline of Inlay described in Section 3.2, with the resultant UI overlay displayed on the right, as described in Section 3.1. Inlay manifests as a dismissible(a) overlay on top of the current webpage with a key insight at the top(g), and suggestions of types floating(b), in-context(c) and in-line(d). On click, the expanded view displays the category(e) of the action, which can then be delegated(f) to the browser’s inbuilt AI Agent.

3.2 System Architecture and Implementation

We have implemented Inlay as a Chromium extension with a Python-FastAPI service layer, consisting of a three-stage pipeline (Figure 1) which triggers only after the page has fully loaded, executing asynchronously on a background thread to ensure zero impact on page load times. Once LLM inference is complete, the Inlay interface materializes dynamically over the live DOM without requiring a page refresh, seamlessly overlaying context-aware cues onto the user’s active view. It is engineered for agent-agnostic compatibility; its delegate pathway presently uses the enterprise tier of Microsoft’s Edge Copilot as the downstream agent, but can be re-targeted to any ‘AI Browser’ agent.

3.2.1 Stage One: Proactive Context Ingestion. The pipeline initiates automatically post page load. A client-side module uses rule-based logic to distill the primary content, reducing LLM payload size by ~70% compared to raw DOM serialization. All processing occurs client-side, with PII removed to preserve privacy. The resulting JSON payload contains the distilled text and essential metadata, along with a base64-encoded screenshot of the whole webpage. A transient progress indicator with a cancel button ensures user control over this preemptive step.

3.2.2 Stage Two: Intent Prediction and Schema-Enforced Suggestion Generation. The payload is sent to the server-side API, which first classifies the page (e.g. ‘e-commerce’, ‘article’) and injects a vertical-specific addendum into the prompt, guiding the model to prioritize relevant actions (e.g. uncover hidden fees on e-commerce, verify claims in articles). The resultant payload is fed to a multimodal LLM (GPT-4o). First, the model generates a candidate slate of suggestions and a key insight (see Section 3.1.2). Second, a separate LLM-as-a-judge call ranks this slate for utility and novelty, selecting the final top-N set. This generation is strictly controlled using model-native JSON mode, and validated against a rigid Pydantic schema

to enforce type correctness and enum membership (e.g., category ∈ compare, verify, ...).

3.2.3 Stage Three: Heuristic-Based UI Injection. The client-side injection module renders the final list of suggestions into a non-destructive shadow DOM overlay. For each suggestion, a heuristic-based anchor search traverses the DOM tree to find the most contextually relevant element based on its content pattern and semantic priority. This is the interface presented to the user. User interaction follows two pathways: a *delegate pathway* and a dismiss pathway, as detailed in Section 3.1.3, ensuring the system’s primary constraint, preservation of user agency, is satisfied.

4 Exploratory User Study

4.1 Study Design and Participants

We conducted a mixed-methods exploratory user study with N=12 working professionals (8 in technical roles and 4 in tech-adjacent functions, convenience-sampled) who all regularly use the enterprise version of Microsoft Edge Copilot. Sessions began with a 5-minute tutorial, then a 20-minute unsupervised, freeform (non-work-related) browsing session with page visits and clicks recorded with explicit consent, followed by a post-task questionnaire (5-point Likert, see Fig. 2) with a semi-structured interview thereafter probing specific moments of interaction. The following Evaluation Goals (EGs) reflect the exploratory nature of the study:

- **(EG1)** To what extent do in-page cues provide perceived utility without undue disruption? (Assessed via Likert items Q1–Q3, Q8 and floating-widget collapse events.)
- **(EG2)** Does preemptive surfacing of cues preserve perceived user agency? (Assessed via Q4 and interview probes.)

- (EG3) How do proactive cues shape exploration and the externalization of latent intent? (Assessed via Q5–Q7, delegate/dismiss ratios in interaction logs, and interview probes on serendipity.)

4.2 Findings

Our analysis triangulates click logs ($N=362$ interaction events), Likert-scale responses across eight items (Q1–Q8), and semi-structured interviews. We note that interaction logs primarily capture engagement patterns. They do not, on their own, establish comparative preference over alternatives (e.g., sidebar agents) or demonstrate improved browsing outcomes. We rely on interview data to contextualize and qualify these behavioral signals. Browser history telemetry confirms that sessions spanned diverse site typologies, including news articles, blogs, technical documentation, and e-commerce platforms.

4.2.1 Adoption and Habituation. Engagement with Inlay increased over the duration of the session (see Fig. 2, right side). Interaction frequency rose from 65 total events (aggregate over all users) in the first quarter of the browsing session to 108 in the third quarter and sustained at that level till the end of the session, suggesting steady habituation to Inlay’s presence. Across the 12 participants, the system achieved a high utility signal, with a ~5:1 ratio of delegated actions to dismissals (233 delegates vs. 47 dismissals).

This data supports our core hypothesis: lowering initiation friction encourages users to offload cognitive effort more frequently as trust in the system builds. However, we cannot rule out that some delegation reflects novelty or ‘try-the-feature’ behavior inherent to first exposure. Qualitatively, several participants’ interview responses attribute delegation to genuine friction reduction rather than curiosity alone. P4 noted, “*I found it to be good to have things instead of activating Copilot for everything,*” framing Inlay as a workflow improvement over an existing AI browsing habit. Similarly, P8 remarked that Inlay was “*giving me options that were the logical next steps for my reading. I hadn’t really thought about the next step, but it gave me those options, so it helped reduce my effort there,*” pointing to perceived cognitive offloading rather than novelty exploration.

4.2.2 Interaction Preference and Agency. Participants reported a strong preference for Inlay’s in-context model over traditional sidebar agents ($Q6 : M = 4.3, SD = 0.5$). Crucially, this proactivity did not come at the cost of perceived agency ($Q4 : M = 4.7, SD = 0.5$, the highest rated Likert question). While we interpret this favorably, the absence of a controlled baseline means this reflects absolute perceived agency with Inlay, not a relative comparison against an alternative paradigm.

Interview data indicates this balance was achieved by the on-page and non-blocking nature of the UI. As P1 noted, “*I feel like I can just browse on my own, while it (Inlay) makes sure I don’t miss anything.*” Users viewed Inlay less as a tool to be commanded and more as a collaborative partner, with P10 noting “*...when I usually say to Copilot ‘hey i want to extract xyz from a page’, my prompt might not be properly framed. So inlay giving suggestions that can be delegated over to it helped.*” The ability to seamlessly delegate a suggested action allowed users to maintain their foraging flow without the context-switching cost of prompt formulation.

4.2.3 The ‘Familiarity Contrast’ in Discovery. Inlay’s ability to promote serendipitous discovery was highly variable ($Q5 : M = 2.9, SD = 1.4$). This high variance points to a familiarity contrast: proactivity without the incorporation of explicit user intent is highly valued in unfamiliar domains but redundant in familiar ones. When exploring novel topics, users praised the system for surfacing ‘unknown unknowns.’ P3, reading about math theorems, noted that Inlay offered a counterpoint they “*would not have found*” otherwise, while P6 highlighted how the system guided them through an unfamiliar astronomy page. P11 noted “*I briefly landed on a Japanese language learning website and had no idea what to do ... Inlay helped guide me on where things are.*” Conversely, on familiar ‘patches’, users found cues to be ‘surface-level.’ P1 remarked, “*On Amazon, I know what filters I want... Inlay’s suggestions didn’t add much value.*” P5 echoed this, noting that on a technical documentation page, suggestions were “*very expected and obvious stuff.*” This points to the possibility that future proactive browser agents may need to be context-aware not just of page content, but of the user’s expertise and history within particular domains.

4.2.4 Contextual Placement and Disruption. Inlay’s ‘in-line’ and ‘in-context’ suggestions were praised for being “*easier to read and follow*” (P5), and we hypothesize this is what resulted in the users rating interaction model preference (versus the Copilot workflow they were familiar with) at (Fig. 2 Q6) $M = 4.3$. This validates Inlay’s on-page augmentation strategy: embedding scent cues adjacent to relevant DOM elements indeed helped cut down cognitive translation friction. However, ‘floating’ widgets were frequently cited as a source of visual clutter (31 ‘collapse’ events recorded for floating widgets). P12 explicitly requested to “*get rid of that pop up corner stuff,*” preferring the subtle, citation-style cues.

There also emerged a correlation where users who had such remarks also rated Q3 (Non-disruption) poorly ($Q3 : M = 3.6, SD = 1.1$). We interpret this as a signal about *presentation modality preference* (in-context over floating) rather than as evidence for or against the broader preemptive paradigm itself. This suggests that while the premise of proactive cues appears to resonate with users, cues may need to be spatially bound to the specific DOM elements they concern, rather than floating ambiguously over the content.

5 Limitations, Future Work and Conclusion

We interpret our findings as a promising *initial signal* for the Inlay paradigm, given the current study’s exploratory design with a small cohort ($N = 12$) and limited exposure duration. While controlled studies are required to isolate causal effects on comprehension and retention, the current work does offer favorable early evidence regarding our core premise. Our findings offer preliminary evidence that *preemptive, in-page assistance* can function as a low-overhead, mixed-initiative aid, translating emergent intent into suggestions for the user as well as downstream agents, presenting a viable augmentation to the reactive status quo in AI browsers. The observed preference for in-context cues is consistent with the hypothesis that lowering initiation friction may deepen exploration without compromising perceived agency.

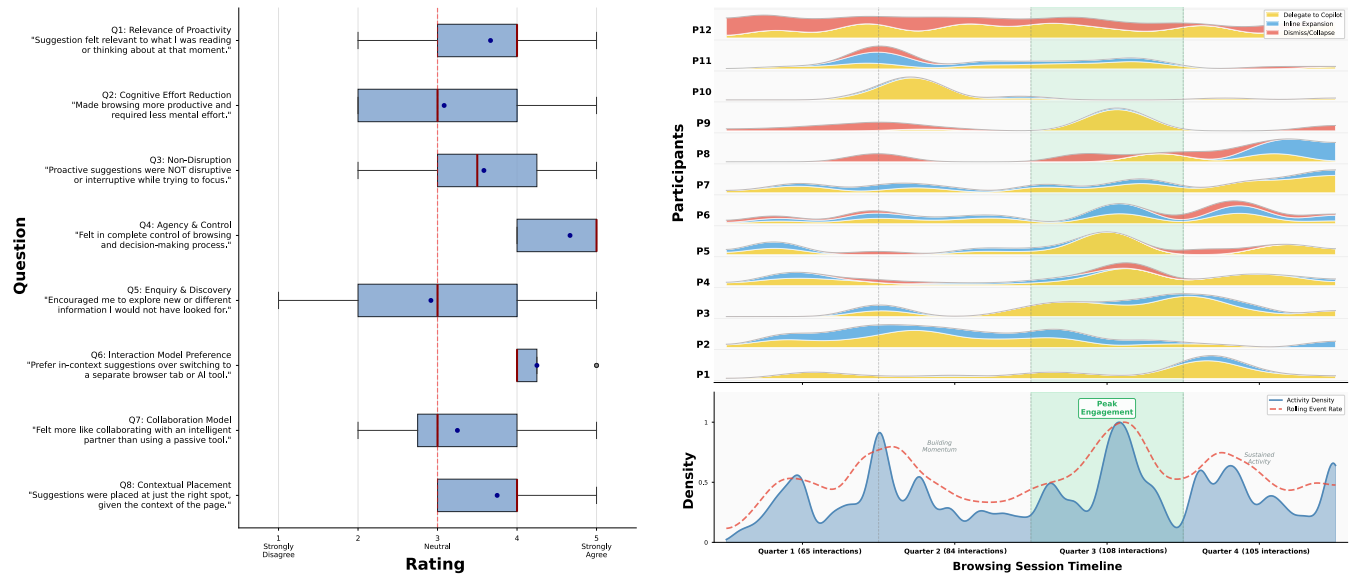


Figure 2: (Left) A box plot depicting the participants' responses, on a 5-point Likert scale, to the post-session questionnaire; (Right) A temporal ridge plot of interactions. The upper panel shows per-participant (P1-P12) stacked normalized-density strips representing the temporal distribution of Delegate (yellow), Expansion (blue), and Dismissal (red) events. The lower panel aggregates these into an Activity Density (smoothed probability distribution of interactions – blue) and Rolling Event Rate (the aggregate event count within a sliding temporal window – overlaid red dashes), illustrating the shift from initial exploration to peak and sustained engagement.

We also note several risks inherent to preemptive cues that merit further investigation:

- (1) *Visual clutter* from over-surfacing, partially mitigated by our placement strategy and dismiss controls but cited by participants as a concern for floating widgets;
- (2) *Misaligned or surface-level suggestions* that risk eroding trust, particularly in familiar domains (see Section 4.2.3), addressable via personalization and confidence-awareness;
- (3) *Heuristic anchoring brittleness* on dynamic or complex layouts and Single-Page Apps (SPAs), necessitating more robust DOM-matching strategies.

A controlled comparison against the sidebar paradigm and longitudinal exposure beyond a single session constitute critical next steps to isolate causal effects, rule out novelty-driven engagement, and assess the persistence of adoption patterns.

Crucially, we posit Inlay primarily as a *design provocation*, a proof-of-concept intended to spark discourse on proactive and on-page AI assistance in browsing. For practical consumer deployment, future iterations must bridge casual and task-oriented browsing by integrating *intent clarification* (to disambiguate vague cues) with robust *agentic action-taking*. Beyond these functional additions, future work must model adaptive intervention timing based on real-time intent and assess longitudinal persistence. Finally, this work also establishes a trajectory for generalizing preemptive, in-context augmentation beyond the desktop to mobile and ubiquitous computing environments.

References

- [1] Denise E. Agosto. 2002. Bounded rationality and satisficing in young people's Web-based decision making. *Journal of the American Society for Information Science and Technology* 53, 1 (2002), 16–27. doi:10.1002/asi.10024
- [2] Oded Berger-Tal, Jonathan Nathan, Ehud Meron, and David Saltz. 2014. The exploration-exploitation dilemma: a multidisciplinary framework. *PLoS one* 9, 4 (2014), e95693.
- [3] Jonathan D Cohen, Samuel M McClure, and Angela J Yu. 2007. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362, 1481 (2007), 933–942.
- [4] Browser-Use Contributors. 2025. browser-use/browser-use: Make websites accessible for AI agents. Automate tasks online with ease. <https://github.com/browser-use/browser-use>. GitHub repository, accessed November 16, 2025.
- [5] David Elswiler, Max L Wilson, and Brian Kirkegaard Lunn. 2011. Understanding casual-leisure information behaviour. In *New directions in information behaviour*. Vol. 1. Emerald Group Publishing Limited, 211–241.
- [6] Pramod Dattarao Gawande. 2025. From Reactive to Proactive: Real-Time Human-AI Collaboration in Intelligent Alerting Systems. *Journal of Computer Science and Technology Studies* 7, 6 (Jun. 2025), 1074–1083. doi:10.32996/jcsts.2025.7.127
- [7] Simret Araya Gebreegziabher, Zheng Zhang, Xiaohang Tang, Yihao Meng, Elena L. Glassman, and Toby Jia-Jun Li. 2023. PaTAT: Human-AI Collaborative Qualitative Coding with Explainable Interactive Rule Synthesis. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 362, 19 pages. doi:10.1145/3544548.3581352
- [8] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of CHI*. 159–166. doi:10.1145/302979.303030
- [9] Peiling Jiang and Haijun Xia. 2025. Orca: Browsing at Scale Through User-Driven and AI-Facilitated Orchestration Across Malleable Webpages.
- [10] Yaxi Lu, Shenzhi Yang, Cheng Qian, Guirong Chen, Qinyu Luo, Yesai Wu, Huadong Wang, Xin Cong, Zhong Zhang, Yankai Lin, et al. 2024. Proactive agent: Shifting llm agents from reactive responses to active assistance. *arXiv preprint arXiv:2410.12361* (2024).
- [11] Sean Lyndersay. 2025. Meet Copilot Mode in Edge: Your AI browser. Microsoft Edge Blog. <https://blogs.windows.com/msedgedev/2025/10/23/meet-copilot-mode-in-edge-your-ai-browser/>.

- [12] Yazdan Mansourian and Nigel Ford. 2007. Search persistence and failure on the web: a “bounded rationality” and “satisficing” analysis. *Journal of Documentation* 63, 5 (09 2007), 680–701. arXiv:<https://www.emerald.com/jd/article-pdf/63/5/680/1355766/00220410710827754.pdf> doi:10.1108/00220410710827754
- [13] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (April 2006), 41–46. doi:10.1145/1121949.1121979
- [14] P. Pirolli and S. Card. 2005. Information Foraging in Information Access Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 51–58.
- [15] Kevin Pu, Daniel Lazaro, Ian Arawjo, Haijun Xia, Ziang Xiao, Tovi Grossman, and Yan Chen. 2025. Assistance or disruption? exploring and evaluating the design and trade-offs of proactive ai programming support. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (2025), 1–21.
- [16] Herbert A. Simon. 1972. Theories of bounded rationality. In *Decision and Organization: A Volume in Honor of Jacob Marschak*, C. B. McGuire and Roy Radner (Eds.). North-Holland Publishing Company, Amsterdam and London, Chapter 8, 161–176.
- [17] Hari Subramonyam, Roy Pea, Christopher Pondoc, Maneesh Agrawala, and Colleen Seifert. 2024. Bridging the Gulf of Envisioning: Cognitive Challenges in Prompt Based Interactions with LLMs. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 1039, 19 pages. doi:10.1145/3613904.3642754
- [18] The Browser Company. 2025. Dia Browser: AI Chat With Your Tabs. <https://www.diabrowser.com/> Accessed January 20, 2025].
- [19] VentureBeat Staff. 2024. Exclusive: Browserbase launches headless browser platform that lets LLMs automate web tasks. *VentureBeat* (jun 2024). <https://venturebeat.com/ai/exclusive-browserbase-launches-headless-browser-platform-that-lets-llms-automate-web-tasks>
- [20] Zehuan Wang, Jiaqi Xiao, Jingwei Sun, and Can Liu. 2025. IntentPrism: Human-AI Intent Manifestation for Web Information Foraging (CHI EA '25). Association for Computing Machinery, New York, NY, USA, Article 345, 11 pages. doi:10.1145/3706599.3719744
- [21] Thomas Wiebringhaus. 2019. Measuring Exploration/ Exploitation Decisions in (Un)Expected Uncertainty-driven Web Search. In *Conference: 65. Frühjahrskongress der Gesellschaft für Arbeitswissenschaft 2019At: Dresden, Germany*.
- [22] Hyeongeun Yun and Jinkyu Jang. 2025. Interaction-Driven Browsing: A Human-in-the-Loop Conceptual Framework Informed by Human Web Browsing for Browser-Using Agents. doi:10.48550/arXiv.2509.12049
- [23] Maxwell Zeff. 2025. Perplexity launches Comet, an AI-powered web browser. *TechCrunch*. <https://techcrunch.com/2025/07/09/perplexity-launches-comet-an-ai-powered-web-browser/>.
- [24] Shuning Zhang, Jingruo Chen, Zhiqi Gao, Jiajing Gao, Xin Yi, and Hewu Li. 2025. Characterizing unintended consequences in human-gui agent collaboration for web browsing. *arXiv preprint arXiv:2505.09875* (2025).

A Post-Session Questionnaire Items (Q1–Q8)

The following eight Likert-scale items were administered immediately after each browsing session. Participants rated each statement on a 5-point scale (1 = Strongly Disagree, 5 = Strongly Agree). Aggregate response distributions are shown in Fig. 2 (left).

- Q1** *Relevance of Proactivity*. “When Inlay offered a suggestion, it felt relevant to what I was reading or thinking about at that moment.”
- Q2** *Cognitive Effort*. “Inlay made my browsing feel more productive and required less mental effort.”
- Q3** *Disruption*. “I found Inlay’s proactive suggestions to NOT be disruptive or interruptive while I was trying to focus.”
- Q4** *Agency & Control*. “I felt in complete control of my browsing and decision-making process.”
- Q5** *Discovery & Foraging Appetite*. “Inlay’s suggestions encouraged me to explore new or different information I would not have looked for on my own.”
- Q6** *Interaction Model Preference*. “I prefer Inlay’s model of in-context suggestions over switching to a separate browser tab or AI tool to ask a question.”
- Q7** *Collaboration Model*. “Using Inlay felt more like collaborating with an intelligent partner than using a passive tool.”
- Q8** *Contextual Placement*. “Inlay’s suggestions were placed at just the right spot, given the context of the page.”