From Gaze to Interaction: Exploring Visual Attention in AI-Assisted Interfaces

Sonam Sonam and Saurabh Mukherjee*

Department of Computer Science, Banasthali Vidyapith, Rajasthan, India
sonamyadav291093@gmail.com, msaurabh@banasthali.in*

Abstract. Today's AI-driven interfaces shape both of those things simultaneously. A click indicates what a person ultimately chose, not what their eyes did on the way there. There is already eye-tracking research for reading, advertising, and certain recommendation jobs, but nothing connects that signal to how an AI-assisted interface should react to it without confusing a glimpse for a decision. This paper fills that gap. It provides a conceptual framework features from unprocessed visual observation; an AI layer that interprets those features while separating attention from preference; presentation modifications that remain within bounds; and a feedback loop that shuts things off. The argument relies on previous research on eye tracking and suggestion interfaces, particularly the public RecGaze Click Feedback Dataset, which is used here only as a methodological foundation rather than as a source for new findings. This intentionally does not a statistical test or machine-learning experiment; hence, the framework has not been empirically tested, which is its primary shortcoming. Toward the end of the paper, a procedure that one may actually implement is provided: comparison conditions, outcome measures, and a statistical method for determining whether gaze-informed adaptation actually helps in subsequent work.

Keywords: Eye Tracking, Visual Attention, Human-Computer Interaction, AI-Assisted Interfaces, Gaze Behaviour, Human-AI Interaction, Recommender Systems.

1 Introduction

Interactive systems used to simply show content and nothing more. That is no longer the case. These days, artificial intelligence(AI) controls not just what shows on the screen but also how everything is displayed. It can rate objects, suggest options, summarize content, explain selections, prioritize what to show first, and frequently do many of these at once. Search bars, chat-based assistants, recommendation engines, and generative tools all have an impact on the visual layout that a user ultimately sees. Determining where someone's eyes actually go becomes less of a footnote and more of an actual HCI issue when that layout gets more packed.

Visual attention is located somewhere between the user clicking on something and the interface displaying something, and it refuses to load nicely with the side of that. Even after extensively examining an object, a person may not choose it. They may look

© The Author(s) 2026
A. Shastri et al. (eds.), *Proceedings of the Fourth Workshop on Applications of Computational Intelligence and Interactive Systems (WACIIS 2026)*, Advances in Intelligent Systems Research 217,
https://doi.org/10.2991/978-94-6239-832-0_19

away, become sidetracked, and return to it minutes later. Even when something is directly in front of them the entire time, they may completely miss it. A click log just records the last, intentional act; it provides very little information about the preceding look. Exactly that gap is filled by eye tracking.

Imagine a scenario with twelve trees. Although all twelve are technically visible, a viewer's attention will likely focus on only one or two of them. If you revamp the packaging of a product that is already well-known, attention is drawn to it even when the product itself hasn't changed. It is obvious that visibility and attention are two different things. When an AI system begins determining what is highlighted, what is ranked higher, which summary appears first, and which explanation is revealed, that distinction becomes even more crucial.

Gaze has already started to be treated as implicit signal in recent work in Human-AI interaction, specifically to examine attention in AI-assisted information interfaces and to stimulate human reactions to AI-generated output[13, 14]. This work continues the trend with a somewhat a specific question: is it possible to incorporate visual-attention data into an AI-assisted interfaces without the system mistaking a glance for a decision?

This paper provides three things: a framework for developing attention-aware AI interfaces, a conceptual explanation of visual attention as an implicit interaction signal, and a set of design and ethical guidelines for conducting gaze-informed adaptation in a responsible manner, all of which are based on the RecGaze research direction.

2 Background and Related Work

2.1 Visual Attention and Eye Tracking

Text, pictures, buttons,and suggestions all vie for the same limited attention in every busy interactive interface. Researches utilize eye tracking, which relies on fixation incidence, fixation duration, time to first fixation, dwell time, revisits, and scanpaths, to measure where that attention truly goes. Some of the early groundwork in this area came from Judd etal.[1], who demonstrated that saliency, or how visually loud a section of a scene is, predicts where eyes land initially. This discovery created a baseline that the field continued returning to. At about the same time, Sibert and Jacob[2] established early standards for gaze-driven interfaces by explicitly testing eye gaze as an interaction strategy against manual input. While Zhu and Ji[4], writing earlier, were still resolving practical problems like head movement and calibration drift that kept eye trackers in the lab, Cazzato et al.[3]later assessed how computer vision changed gaze estimation as a whole. This does not imply that engagement equates to attention. This contrast permeates all that follows:looking at something is not the same as loving it, believing in it, or planning to take action. Balcombe et al.[15], working in a discrete-choice setting instead of an interface one, found only a weak relationship between how much visual attention an attribute received and how highly it was valued in stated preferences. This is direct empirical evidence that gaze and preference diverge even on the same choice task.

2.2 Gaze in Advertising and Reading

Long befor recommendation systems were invented, advertising and reading research have a history with gaze, and it's worth using that past here. The duration of fixation on banner advertising was found to be correlated with the effectivenesss of those ads by Lee and Ahn[5]. This suggests that attention is a component of the assessment process rather than just background noise. Using a sizable eye-tracking dataset from advertising photos, Liang, Liu, and Qian[6] demonstrated that text attracts attention in a different way than images, something that designers had likely suspected but never measured. In terms of reading, Buscher et al.[7] mapped the areas of the screen that readers truly relied on, providing an early example of how layout alone directs attention independent of the quality of the information. Now that most AI-assisted interfaces operate on phones, where targets are smaller than the eye travels around more, Trabulsi et al.[8] examined something more common but no less significant:how well fixation-filtering algorithms perform on small screens.

2.3 Recommendation Interfaces and RecGaze

Because users peruse multiple possibilities before choosing one, recommendation interfaces are a suitable test for this distinction and provide researchers with something tangible to look at. This type of dataset was created by RecGaze, which used movie-selection tasks to gather eye-tracking and interaction data carousel interfaces[11].The public click-feedback release built on it includes item-level data (position, impression, fixation-based observation, clicks) meant to be usable across eye tracking, HCI, psychology, information retrieval, and recommender-systems research [12].

This paper uses that dataset only as a foundation. It shows, concretely, that visual observation and ordinary interaction data can be studied side by side, which is enough to support future tests of attention-aware models, though no new result is drawn from it here.

2.4 Gaze in Human-AI Interaction

Gaze has already been used as a signal for modelling implicit agreement or disagreement with AI-generated content [13], and separately, other work has looked at how attention behaves inside AI-assisted information interfaces specifically [14]. Together these two lines of work support a modest claim: visual behaviour carries information that explicit interaction logs don't capture on their own. But gaze by itself is a messy signal, shaped by layout, salience, task difficulty, how familiar the content feels, and ordinary variation between people. Instead of being a precise indicator of purpose, it merits a position as contextual evidence, which is helpful but not conclusive.

None of these strands are able to connect the dots. Research on reading and advertising shows that gaze forecasts subsequent results.In a contemporary recommendation scenario, gaze and interaction data can be gathered simultaneously, as demonstrated by the RecGaze work. The Human-AI Interaction research verify that gaze conveys infor-

mation about people's reactions to content produced by AI. None of these, either separately or in combination, provides a formal explanation of how gaze should be used in an AI-assisted interface while maintaining its distinction from preference, nor does it outline the characteristics of a visible, constrained adaptation process based on that signal. This research aims to bridge that gap by providing a framework rather than a single study.

3 RecGaze as a Foundation for Future Validation

Rather than being a source of the results this study claims for itself, the RecGaze Click Feedback Dataset should be read here as backgroung information, an example of the kind of empirical basis future work could build upon[12]. The item-level data(position, impression, fixation-based observation, clicks, and session context) presented in its documentation link what an interface displays to what a user views and ultimately selects.

That is what makes the dataset useful going forward. It opens the door to testing whether gaze adds anything beyond what interface context already tells you, and whether attention-aware adaptation genuinely improves interaction or just looks good on paper. Both are framed here as open hypotheses worth testing, not findings already in hand.

4 Attention-Aware AI Interface Framework

The framework proposed here treats visual attention as one behavioural channel among several, not as a way for an AI system to peer into a user's head. It is a measurable, implicit signal that means something only once it is combined with explicit interaction and whatever the interface context happens to be.

4.1 Visual Observation and Feature Extraction

An eye tracker, or a comparable sensing setup, picks up which elements draw visual attention. Depending on what's available, this could mean fixation occurrence, time to first fixation, dwell time, revisits, how attention splits across alternatives, or the shape of the scanpath. To make this concrete, for a carousel-style interface like the one RecGaze studies, feature extraction would work at the level of predefined areas of interest, one per item slot. Raw gaze coordinates get mapped onto those regions first, using standard fixation-detection algorithms (dispersion-based or velocity-based, depending on sampling rate), and then aggregated per item into a small feature vector: occurrence, duration, order of first fixation relative to other items, and revisit count. Relative attention across alternatives can be computed as each item's dwell-time share of the total dwell time in that session, which normalizes for how long someone spent looking overall. Whatever the system collects, it should collect only what the research question or interaction goal actually requires, nothing beyond that.

4.2 AI-Based Behaviour Interpretation

The AI component takes gaze-derived information (G), interface context (C), and a history of explicit interaction (I), and combines these into an estimated interaction state:

$$S = f(G, C, I). \tag{1}$$

In practice, f would be a supervised or weakly-supervised model trained on paired gaze-and-outcome data of the kind RecGaze provides. The above-described per-item feature vector is G, static and session-level interface features(item position, layout density, number of options displayed)are covered by C, and recent click and dwell history within the same session is captured by I. Rather than treating the output S as a single hard label, it is preferable to handle it as a probability or score per item. This maintains the downstream adaptation gradual rather than imposing a binary judgment off a noisy signal.S is an estimate, not a judgment; it just indicates what is plausible rather than what the user wants or intends. This distinction is important because, depending on the work at hand or the arrangement, the same gaze behavior can signify completely different things.

4.3 Adaptive Presentation

Bounded modifications to the interface, such as rearranging recommendations, adjusting something's visual prominence, relocating an explanation to a more helpful location, or removing clutter that isn't deserving of its place, can be justified by whatever state is assessed. Adaptation should be gated by a confidence threshold on S to keep this restricted rather than open-ended.This way, changes only occur when the estimated state remains stable over a brief period of time or a minimal number of fixations, not for a single glance. In order to prevent one estimate from drastically restructuring the interface in a single step, the amount of any modification, how far an item moves, and how much its prominence varies should also be capped. The user shouldn't feel unprepared for any of this, and they should be able to prevent it from happening.A change that manipulates rather than helps should never be brought about by a single gaze occurrence.

4.4 Explicit Feedback Loop

Explicit feedback includes clicks, selections,revisions, cancellations, and if a task was truly finished.The loops appears like this when it is strung together:

Gaze → Attention Features → AI Interpretation → Adaptive Presentation → User Interaction → Feedback.

This provides future experiments with a tangible comparison between an AI-assisted interface and a normal interface, as well as between both and an attention-aware interface.

5 Design Implications

This approach more or less directly leads to five design principles. A look event should never be subtly reinterpreted as something the user desires; attention is not choice. Gaze should enhance what clicks and results already tell you;implicit and explicit signals are best used on tandem rather than in isolation. Users have a right to know when something implicit is influencing their environment, therefore adaptation must remain transparent. Because location, size, contrast, and hierarchy move attention around on their own, regardless of the user's actual preferences, layout impacts require active control. Additionally, any alteration should be made to benefit the user rather than control them, with the goal of improving discoverability or reducing effort effort waste rather than increasing engagement for its own sake.

This is also supported by research on outside recommendations. Narella et al.[9] shown that gaze-tracking data alone may predict cognitive load in real-time within augmented-reality contexts, supporting the notion gaze tracks interest and mental effort without requiring explicit user input.

This is grounded in two instances. Go back to the tree-filled image. Everything is visible, but attention is still distributed unevenly, and eye tracking shows which object was truly noticed. Product packaging functions similarly; even though the product itself hasn't changed, a redesign modifies the initial point of attention. Put that in an AI-assisted recommendation interface,where the user looks through several options and may even read an explanation produced by the AI before selecting one. While the click just reveals what was selected at the end, gaze provides information about what was taken into consideration during the process. Here, both situations remain hypothetical and are intended to inspire testable hypotheses rather than disguise as the findings of this paper.

6 Ethics, Privacy, and Future Validation

Gaze data presents more serious privacy problems than a simple click log since it can reveal behavioral patterns that a user never knowingly reports.In order to ensure that raw gaze traces never leave the user's device, responsible systems should gather as little as possible, make gaze sensing clearly visible while it is active, explain what is being derived and saved, and, where technically possible, keep processing local. Since sensing accuracy and fundamental accessibility vary among users, consent is important here in a way that it is not usually for regular analytics, as is providing a non-gaze option. In the context of generative AI in particular, Abdrabou et al.[10] push this further: combining gaze data with generative models opens profiling and inference risks that go beyond what click-based personalization alone creates, strengthening the case for minimal collection and on-device processing wherever an interface can support it.

Conventional, AI-assisted, and attention-aware recommendation interfaces might be compared as three between-subject or within-subject conditions in a future study. The attention-aware condition would be evaluated against the simple AI-assisted inter-

face(no gaze input). Task completion time, click-through accuracy in relation to a predetermined ground-truth item, and a post-task subjective rating of felt control and satisfaction on a standard Likert scale would be the main outcome measures. Secondary metrics derived from the gaze data itself, such as fixation duration, dwell time, revisits, and scanpath length, would verify whether the attention-aware condition altered viewing behavior rather than merely results. For testing whether condition predicts the main outcomes while accounting for individual variation in baseline gaze behavior, a mixed-effects model with participant as a random effect would be suitable given repeated-measures data across conditions, with pairwise comparisons adjusted for multiple testing. The effectiveness, satisfaction, perceived control, and cognitive load of gaze-informed adaptation in comparison to that baseline might then be tested in a second stage to see if it is merely a theoretical improvement.

7 Limitations

The objective of this work is to remain conceptual. There is no new statistical analysis, no machine-learning experiment, no controlled eye-tracking study here, so everything above should be read as testable propositions rather than a finished, validated result. The RecGaze resource itself is built specifically around carousel-based movie recommendations, so the framework still needs validation before it should be generalized to conversational agents, generative-AI systems, search interfaces, or other different settings. Gaze behaviour is also shaped by visual salience, task difficulty, layout choices, and individual differences between users, so whatever experimental control gets used later will need to be genuinely careful rather than an afterthought.

8 Conclusion

This paper set out a conceptual framework for bringing visual attention into AI-assisted interfaces, built on one fairly simple argument: explicit interaction and visual-attention signals capture two different slices of the interaction process, not the same thing twice. Gaze says something about what a user looked at before or around an action; clicks and other explicit signals remain the stronger, more deliberate evidence of what they actually chose. None of this paper's claims should be read as evidence that gaze-informed adaptation improves outcomes in practice; that question is left explicitly open, and answering it is the paper's main direction for future work rather than something already established here.

The RecGaze dataset shows that fixation-based observation and interaction data hold up well when studied together, giving future empirical work a real foundation to build on, though the dataset itself was not used here to test the framework, only to motivate its design. Based on that, the architecture presented here integrates explicit feedback, adaptive presentation, AI-based interpretation, attention-feature extraction, and visual observation into a single pipeline. It makes a conscious effort to treat gaze as one contextual signal among several, helpful but not authoritative on its own, rather than as a direct stand-in for preference or purpose.

The controlled experiment that would determine whether attention-aware adaptation truly enhances task performance and user experience in practice is still lacking, and this paper does not claim to have provided it. Until that experiment is conducted, the framework should be viewed as a concept that is subject to change rather than a proven outcome, and any subsequent assessment of it must remain transparent, fair, privacy-preserving, and firmly in the control of the user.

References

1. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look. In: Proc. IEEE 12th Int. Conf. Computer Vision (ICCV). pp. 2106–2113 (2009)
2. Sibert, L.E., Jacob, R.J.K.: Evaluation of eye gaze interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2000). pp. 281–288. ACM (2000)
3. Cazzato, D., Leo, M., Distante, C., Voos, H.: When I look into your eyes: A survey on computer vision contributions for human gaze estimation and tracking. Sensors 20(13), 3739 (2020)
4. Zhu, Z., Ji, Q.: Novel eye gaze tracking techniques under natural head movement. IEEE Transactions on Biomedical Engineering 54(12), 2246–2260 (2007)
5. Lee, J.W., Ahn, J.H.: Attention to banner ads and their effectiveness: An eye-tracking approach. International Journal of Electronic Commerce 17(1), 119–137 (2012)
6. Liang, S., Liu, R., Qian, J.: Fixation prediction for advertising images: Dataset and benchmark. Journal of Visual Communication and Image Representation 81, 103356 (2021)
7. Buscher, G., Biedert, R., Heinesch, D., Dengel, A.: Eye tracking analysis of preferred reading regions on the screen. In: CHI '10 Extended Abstracts on Human Factors in Computing Systems. ACM (2010)
8. Trabulsi, J., Norouzi, K., Suurmets, S., Storm, M., Ramsøy, T.Z.: Optimizing fixation filters for eye-tracking on small screens. Frontiers in Neuroscience 15, 578439 (2021)
9. Nerella, G., Gan, D., David-John, B., Alghofaili, R.: Gaze-based prediction of cognitive load in augmented reality. Proceedings of the ACM on Human-Computer Interaction 10(3), 1–18 (2026). https://doi.org/10.1145/3806040
10. Abdrabou, Y., Özdel, S., Maquiling, V., Bozkir, E., Kasneci, E.: From gaze to data: Privacy and societal challenges of using eye-tracking data to inform GenAI models. In: Proceedings of the 2025 Symposium on Eye Tracking Research and Applications (ETRA '25). Article 109. ACM (2025). https://doi.org/10.1145/3715669.3726788
11. de Leon-Martinez, S., Kang, J., Móro, R., de Rijke, M., Kveton, B., Oosterhuis, H., Bielikova, M.: RecGaze: The First Eye Tracking and User Interaction Dataset for Carousel Interfaces. In: Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 3702–3711 (2025). https://doi.org/10.1145/3726302.3730301
12. de Leon-Martinez, S., Móro, R., Kveton, B., Bielikova, M.: RecGaze Click Feedback Dataset. Zenodo (2026). https://doi.org/10.5281/zenodo.20508190
13. Selim, A.M., Bhatti, O.S., Gomaa, A., Barz, M., Sonntag, D.: Do You (Dis)agree With Me? Modelling Implicit User Disagreement in Human–AI Interaction Using Gaze Data. In: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI 2026). ACM (2026). https://doi.org/10.1145/3772318.3790594

14. Prajod, P., Cools, H., Röggla, T., Cesar, P., El Ali, A.: Towards Gaze-Informed AI Disclosure Interfaces: Eye-Tracking Attentional and Cognitive Load While Reading AI-Assisted News. arXiv preprint arXiv:2605.14999 (2026)
15. Balcombe, K., Fraser, I., Williams, L., McSorley, E.: Examining the relationship between visual attention and stated preferences: A discrete choice experiment using eye-tracking. Journal of Economic Behavior & Organization 144, 238–257 (2017). https://doi.org/10.1016/j.jebo.2017.09.023

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

