Evaluating Personalised Generative User Interfaces on an AI Mediated Web

JOSEPH O'HAGAN, University of Glasgow, Scotland

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**.

Additional Key Words and Phrases: AI, AI Mediation, Generative AI, Research Methods

ACM Reference Format:
Joseph O'Hagan. 2026. Evaluating Personalised Generative User Interfaces on an AI Mediated Web. In *Proceedings of CHI '26: "What does Generative UI mean for HCI Practice?" Workshop (CHI '26).* ACM, New York, NY, USA, 3 pages.

1 Introduction

In this position paper I reflect on the workshop's provocation from my experience developing a generative AI tool *WebPerceptor* [6, 10] and research I have conducted on context aware dynamic UIs [2, 4, 7–9].

WebPerceptor [6] is an open source, client-side Chromium plugin which, for any webpage, identifies text content, relays this to a local or cloud-based LLM with a given user-defined prompt and then modifies the existing text (to replace, append, or remove it) based on the LLM response. In effect, *WebPerceptor* employs LLM-driven text generation and in-browser DOM manipulation to produce a personalized, in-line, real-time remixing of web content, presented seamlessly within the browser such that, from the user's perspective, the mediated version appears as the published web page itself. This approach represents a shift from the static publication model of the web to one of dynamic mediation, where web pages cease to be fixed documents that are authored once and distributed identically to all readers. Instead, web content is made mutable, co-authored in real time by the model and prompts, offering unprecedented personalisation whilst destabilising the notion of a common, shared web experience. Replacing the web browsing experience with an "AI Mediated Web", an automatic, personalised, in-line, real-time remixing of web content by an LLM based on given prompts and presented in-browser such that the user perceives the end result as a published web page. I open by describing *WebPerceptor* to highlight that the potential for AI models to generate UIs is not only limited to the framing and presentation of content (e.g. the position of UI elements, scales, etc) but also the data/information presented through a UI, e.g. the text content of a webpage, to a user.

This concept of an "AI Mediated Web", and such modification of web content, provides a lens to consider different architectures for implementing generative UIs and to reflect on the challenges this presents HCI researchers evaluating designs. Note: in the following text I use the modification of text content of a web page (be it a header, post, long-form article, etc) in my examples, but the ideas can be applied more broadly to other generative UI designs.

At a high level, tools like *WebPerceptor* can built using server-side and client-side architectures. If we consider a web page being published, at the point of committing the authoritative, canonical page, a server-side LLM (or other generative AI system) could execute a series of prompts to generate and cache multiple rewritten variants ready to

Author's Contact Information: Joseph O'Hagan, joseph.ohagan@glasgow.ac.uk, University of Glasgow, Glasgow, Scotland.

Manuscript submitted to ACM

serve to specific users or user groups, each aligned with a specific ideological, demographic, or stylistic profile. For example, a news article [5] might be rewritten with both moderate or extreme left/right-wing variants to be served based on the tracked behavioural or declared profile of the visitor. This approach centralises control to the publisher of the web page (or system being used). For HCI practitioners, many of the existing methods we use are applicable, i.e. the designer has a target a user group, can conduct usability studies to measure the effectiveness of the generated content, can monitor engagement and make adjustments, etc, although crucially is able to do so for deployed content, in real-time, at a much larger scale. This is a designer controlled use of generative UI within a system, and is arguably an evolution of algorithmic approaches to content selection and serving [1, 14], where the UI and content itself is adapted.

If instead we consider the point a web page is rendered in a user's browser, on page load a client-side program can identify and extract the pertinent DOM elements, hide it from rendering, and batch feed this to a local or cloud-based LLM (or other generative AI system) alongside a specified prompt. The resulting output can then be inserted back into the DOM, replacing or appending the original page view. In a designer controlled approach, external providers (i.e. the creator/distributor of the generative system) may specify the prompts and content to target which they may or may not make transparent to the user. This approach, similar to the client-side architecture described above, is one where a HCI practitioner can utilise existing practices to build and evaluate a designed user experience for the system, with users having no control over what content is targeted and how it is modified. However, a user controlled approach is also possible, where users can specify the content to target and prompts to modify the targeted content with. This provides each user with complete control over when, where, and how modification occurs, enabling a personalised re-framing of their experience using the system. This presents a difficult evaluation challenge for HCI practitioners. While the scope of targets or modifications could be controlled by a designer to be a fixed set of evaluated use cases, this severely limits the potential of such tools.

How then might we evaluate such systems when used by users resulting in infinite permutations of possible use? From my own work evaluating usage of context aware, dynamic AR/VR UIs we have effective methods of identifying [11] and utilising use case scenarios to identify user preferences for how known UI designs might adapt based on a given context, e.g. customising the disruptiveness of a message notification based on the message content [2], or adapting the external awareness of a VR user relative to their interaction with a bystander [9]. We also have effective methods of evaluating in-the-wild use of systems, e.g. identifying patterns of in-situ UI placements for AR interfaces [3] or using large scale Reddit scraping and thematic analysis to understand how users use XR devices whilst travelling (what they use it for, what works, and what doesn't work) [12]. It may be that we must rely increasingly on such methods, e.g. incorporating evaluations of use case scenarios into traditional usability studies or regularly conducting large-scale analysis of user accounts of system use and changing systems in response to such feedback. Developers might even build automated systems to monitor the discourse of their users, e.g. on forums, through user reviews, or through public posts on Reddit, and automatically flag when sentiment shifts significantly (e.g. in response to a system update or change) or topics emerge which are actively being discussed by many users [13].

Or is an entirely new approach needed? What if, for example, many of the methods we use (or variations of them) to "improve usability" are embedded within the systems users use. By doing this, and automatically analysing the captured data and making adjustments to the generated UIs, we can develop personalised generative UIs tailored to the specific needs, preferences, and use of individual users. Such a shift in approach would transform both the design and creation workflow of a system's creators and the experience of users, where every user experiences a personalised variation of the system design, and is an exciting prospect I am keen to discuss further.

Acknowledgments

Funded by UK Research and Innovation (UKRI) under the UK's Horizon Europe funding guarantee [EP/Z000068/1] (AUGSOC).

References

[1] Engin Bozdag. 2013. Bias in algorithmic filtering and personalization. *Ethics and information technology* 15, 3 (2013), 209–227.

[2] Hyunsung Cho, Drew Edgar, David Lindlbauer, and Joseph O'Hagan. 2025. Evaluating Dynamic Delivery of Audio+Visual Message Notifications in XR. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 277–287. doi:10.1109/VR59515.2025.00052

[3] Hyunsung Cho, Yukang Yan, Kashyap Todi, Mark Parent, Missie Smith, Tanya R. Jonker, Hrvoje Benko, and David Lindlbauer. [n. d.]. MineXR: Mining Personalized Extended Reality Interfaces. 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 609, 17 pages. doi:10.1145/3613904.3642394

[4] Diego Drago, Joseph O'Hagan, Fahim Kawsar, and Stephen Anthony Brewster. 2025. Exploring Avatar Seat Allocation Strategies in Social AR. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 239, 7 pages. doi:10.1145/3706599.3719867

[5] Sarah Kreps, R Miles McCain, and Miles Brundage. 2022. All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of experimental political science* 9, 1 (2022), 104–117.

[6] Joseph O'Hagan. 2026. *WebPerceptor Release v1.0.0.* doi:10.5281/zenodo.18724545

[7] Joseph O'Hagan and Julie R. Williamson. 2020. Reality aware VR headsets. In *Proceedings of the 9TH ACM International Symposium on Pervasive Displays* (Manchester, United Kingdom) *(PerDis '20)*. Association for Computing Machinery, New York, NY, USA, 9–17. doi:10.1145/3393712.3395334

[8] Joseph O'Hagan, Julie R. Williamson, Mohamed Khamis, and Mark McGill. 2022. Exploring Manipulating In-VR Audio To Facilitate Verbal Interactions Between VR Users And Bystanders. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces* (Frascati, Rome, Italy) *(AVI '22)*. Association for Computing Machinery, New York, NY, USA, Article 35, 9 pages. doi:10.1145/3531073.3531079

[9] Joseph O'Hagan, Julie R. Williamson, Florian Mathis, Mohamed Khamis, and Mark McGill. 2023. Re-Evaluating VR User Awareness Needs During Bystander Interactions. 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 876, 17 pages. doi:10.1145/3544548.3581018

[10] Joseph O'Hagan, Graham Wilson, and Mark McGill. 2026. WebPerceptor: An Open Source Chromium Plugin for Real-Time LLM-Based In-Line, In-Browser Re-Writing of Website Content. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '26)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3772363.3798730

[11] Joseph O'Hagan, Julie R. Williamson, Mark McGill, and Mohamed Khamis. 2021. Safety, Power Imbalances, Ethics and Proxy Sex: Surveying In-The-Wild Interactions Between VR Users and Bystanders. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 211–220. doi:10.1109/ISMAR52148.2021.00036

[12] Katharina M. T. Pöhlmann, Graham Wilson, Laura Bajorunaite, Mark McGill, and Joseph O'Hagan. 2025. "Everyone Knows You're Watching P*rn": Reflecting on First-Hand User Accounts on the Use and Public Perception of in-Transit Passenger XR. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 560–570. doi:10.1109/ISMAR67309.2025.00066

[13] Angelo Singh and Joseph O'Hagan. 2024. Exploring Topic Modelling of User Reviews as a Monitoring Mechanism for Emergent Issues Within Social VR Communities. arXiv:2406.03994 [cs.HC] https://arxiv.org/abs/2406.03994

[14] Zeynep Tufekci. 2015. Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colo. Tech. LJ* 13 (2015), 203.