Artificial Intelligence and Social Computing, Vol. 163, 2025, 177–183
https://doi.org/10.54941/ahfe1006049

Beyond Explicit Instruction: Enhancing Human-AI Collaboration With Implicit User Feedback

Jaelle Scheuerman[1], Shannon McGarry[2], Ciara Sibley[2], and Noelle Brown[2]

[1]US Naval Research Laboratory, Stennis Space Center, MS 39529, USA
[2]US Naval Research Laboratory, Washington, DC 20375, USA

ABSTRACT

Successful human-AI teamwork depends on AI systems that can adjust to the evolving needs and situations of users. Rather than relying on explicit instructions from the user, an adaptable agent can make use of implicit feedback from end-users to infer user's behavioral and situational needs. Implicit information, such as user activity and eye tracking data, can help infer behavioral patterns that uncover the user's desires, requirements, and mental states. This method allows AI systems to deliver more tailored, proactive and wholistic assistance, which not only minimizes user's real-time workload, but also serves to add redundancy to human-error, much like a beneficial human teammate. While this approach offers several potential advantages, there are practical difficulties in gathering and interpreting the data. Upcoming efforts to deduce high-level actions from low-level data will need to tackle these challenges to facilitate intuitive human-AI interactions and improve the efficacy of collaborative systems.

Keywords: Human-machine teams, Adaptive systems, Implicit feedback, Eye-tracking, User activity, Language models

INTRODUCTION

Many research efforts have worked towards a goal of artificially intelligent agents that can work with humans, leveraging the combined strengths of each to improve overall performance. For AI assistants to participate effectively in a human-machine team, they must incorporate models that can learn and adapt to feedback from their human teammates, their tasks and the current environment.

Adapting to feedback presents challenges to many AI systems. Classic machine learning methods aim to train models with data that is similar to the operational environment, but require retraining when that environment changes. This approach provides little opportunity to provide the aforementioned assistance and redundancy, and eventually causes model performance to degrade over time. Online learning and active learning methods have been developed to overcome this challenge by providing a means for the machine learning model to learn from new inputs and more rapidly reach model convergence or respond to changes in the environment

(Botou, 1998; Aggarwal et al., 2014). By employing online and active learning techniques it is possible to incrementally train a model by asking users to validate model results and correct erroneous output (Michael, 2019). Recently, the zero shot capabilities of large language models (LLMs) enable them to respond to many different situations. This has led to the proliferation of chatbots that seem to be able to respond to user prompts about a wide range of topics.

Both interactive machine learning and chat interfaces are generally reliant on users taking the time to provide explicit feedback and the model performance may lag until that occurs. In this vision paper, we explore the potential of using implicit, rather than explicit, feedback for improving how large language model-based interfaces adapt to their user's goals and needs. To accomplish this, we identify methods for passively collecting implicit data from the user or environment that is informative about the user's current state and context. For example, the way in which a person's pupil responds to an AI system's output or actions could inform if the person is on or off task (Unsworth & Robison, 2016). Alternatively, a user's activity logs can inform a generative model about their expertise. We further capitalize on this implicit feedback by recognizing high level concepts about the user state from streams of low-level data. The inferred behavior can be provided to a language model as contextual information, which these models are often missing. Context can be critical to not only the efficacy of the language model's response, but also serve to prevent unintended consequences, which is a continually observed outcome in human-autonomy interaction (Brown et al., 2020; Endsley et al., 2023).

IMPLICIT FEEDBACK FROM USER ACTIVITY

Collecting user activity involves logging information about what the user is doing in the context of some interface or software system. This can include things like mouse tracking, clicks, how many and which applications are open, their search queries, etc. Such logs can provide enormous context to an AI system relating to what is on the user's screen, what they are doing, how they are doing it, and even suggest their level of domain knowledge (Zhang et al., 2011). In training scenarios, user activity information could be used to implicitly guide an AI assistant to provide documentation relating to the tools the user is actively working with or highlight best practices that a novice user may be unaware of. It could further infer factors impacting the user and avoid unnecessarily interrupting them during times of high workload.

Many ongoing efforts have explored classifying low-level user activity from logs as high-level events (Rebmann & van der Aa, 2024; Asghari et al., 2020). A downside of these approaches is that they can require large amounts of labeled data to train the activity recognition models and creating these such datasets is time consuming. In recent work (Ortego & Scheuerman, 2025), we explored activity logs of cybersecurity operators by using generated natural language summaries that could be used as contextual information to a large language model. By providing the summaries of user activity data and related domain knowledge derived from a taxonomy of cybersecurity workflows, the language model was able to appropriately infer the current task from the user activity logs 88% of the time.

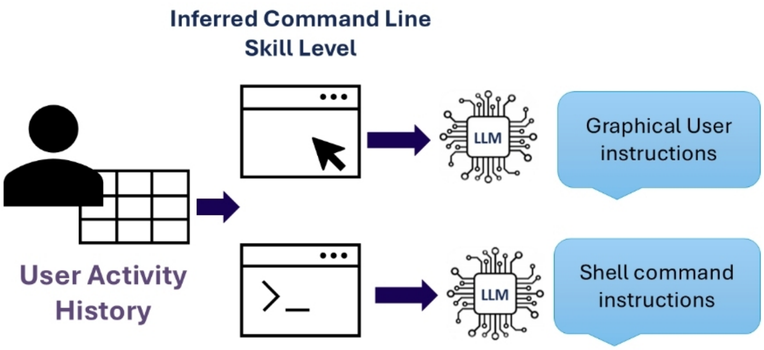

Figure 1: User activity logs can provide information about inferred skill level or preferences that could be used to adapt a language model's output.

IMPLICIT FEEDBACK FROM EYE TRACKING

Another possible source of implicit feedback is eye tracking data. Similar to how we converted user activity logs into high level natural language summaries, we now plan to explore the use of eye tracking data to infer higher-level behaviors and even states of the user. Almost all eye trackers give information about the person's pupil size and scan patterns. For example, pupil size and how it fluctuates can indicate physical and mental effort, attentional shifts, as well as differences in cognitive abilities (Robison et al., 2024; Strauch et al., 2024), whereas studying a person's scan patterns has been found to provide "objective and quantitative evidence of the user's visual, overt attentional processes" (Duchowski, 2017, p. 247).

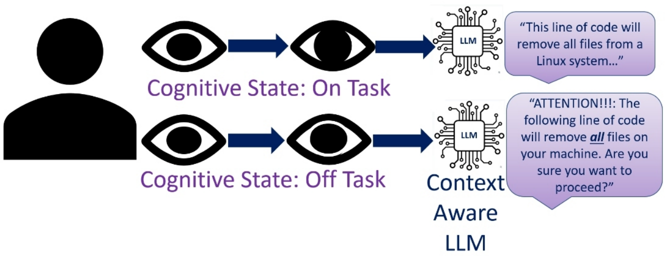

Figure 2: Data gathered through eye tracking can provide information about covert and overt attention shifts of the user. Therefore information could be used to adjust the tone and style of language model output. Above is a schematic of how the language modle may provide output to a user who is on vs off task, as defined byt the task evoked pupillary response of the user.

To do this unobtrusively, remote, or off-the-head, eye trackers are equipped with the ability to emit infrared light and use high-resolution cameras to capture the diameter of pupil (in units of pixels of in the camera image) and track the center of the person's pupil(s) to measure where they are foveating with the corneal reflection technique (see Poole & Ball, 2006, p. 212 for more detail). With recent innovations, eye tracking is less invasive and more versatile, mobile, and cost-effective than ever before and compared to other psychophysiological measures (e.g., EEG; Dorneich et al., 2008), prompting it for wide-scale use (Krafka et al., 2016;).

Research has already shown how eye tracking data can be used in real-time for a range of applications (see review in Duchowski, 2018). Further, inferring user activity, state, and traits from eye tracking data is not new, as several machine learning techniques have been used to infer a range of cognitive states such as fatigue, mental workload, confusion, and intended actions. These methods have also been able to infer traits about the human like the presence of dyslexia (see review in Klaib et al., 2021). Given the broad range of eye tracking data applications, employing it implicit measure to language models is a logical and promising next step

CHALLENGES IN IMPLICIT FEEDBACK

Both user activity data and eye tracking data present several challenges to be overcome before they can be fully integrated into a language model assistant. For example, the interpretation of both user activity data and eye tracking data is not always consistent and can be very context driven. For example, the simple measure of the length of time visual attention is foveated on one area of the screen could mean the person is very interested in this stimulus or it could indicate that an individual is experiencing cognitive tunneling that may lead to a loss of situational awareness and the increased potential for errors (Marois et al., 2020, Wickens et al., 2008). Interpreting user activity logs to differentiate between certain tasks is also a challenge, and interaction patterns may be differ greatly between experts and non-expert users (Cole et al., 2015). Context collected with user activity logs, combined with eye tracking movements may be able to disentangle the interpretation of a task (Ooms et al., 2015), but it remains challenging as it is unclear how to synthesize the eye tracking data with these other human-based measures (Ries et al., preprint).

Along a similar vein, there are also structural and logistical questions that need to be answered, such as how to integrate, analyze, and interpret eye tracking data for language in real-time. Finally, privacy is an important consideration for both user activity logs and physiological data. Kröger, Lutz, & Müller (2020) discuss how ubiquity of eye tracking will lead to having data sets on how thousands of people completed the task, their cognitive state, demographic data, biometric signatures, diagnosed physical and mental illness, etc. Future research and policy needs to explore how to study this data so that is specifically and only used for its intended purpose, i.e., infer high-level behavior of the end user, and not used to further propagate demographic biases and/or be applied for other purposes.

CONCLUSION AND FUTURE WORK

In conclusion, there are many benefits to providing implicit user feedback to AI assistants as high-level behavior concepts inferred from low-level user data. For example, user activity logs can ensure that the AI assistant responses are tuned to the user's workflow and the tools that they are currently using. Further, having high level conceptual information about the user's behavior could allow the language to adapt how its response is presented. For example, if the user is distracted or under heavy workload, the response could be adapted to be shorter or even delay the response as needed. If the user is not engaged and a particular important alert is about to be presented, the language model can take extra steps to adapt its response to capture their attention, e.g., presenting the response in bold font.

Many open research challenges remain in operationalizing implicit data collection. Future work must explore how implicit feedback from multiple sources can jointly be used to improve inference of the current user state. Language models need to be evaluated when input context contains semantic cues about user behavior. Such evaluation may require new benchmarks or techniques to ensure that the output is appropriate for the user's current state. Overcoming these challenges to implicit feedback will lead to new opportunities for effective human-AI collaboration.

REFERENCES

Aggarwal, C. C.; Kong, X.; Gu, Q.; Han, J.; Philip, S. Y. (2014). Active learning: A survey. In Data Classification; Chapman and Hall/CRC; 599–634.

Asghari, P., Soleimani, E., & Nazerfard, E. (2020). Online human activity recognition employing hierarchical hidden Markov models. Journal of Ambient Intelligence and Humanized Computing, 11, 1141–1152.

Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. Vision Research, 46(18), 2824–2833.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P.,... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877–1901.

Bottou, L. (1998). Online learning and stochastic approximations. On-line learning in neural networks, 17(9):142.

Cole, M. J., Hendahewa, C., Belkin, N. J., & Shah, C. (2015). User activity patterns during information search. ACM Transactions on Information Systems (TOIS), 33(1), 1–39.

Devlin, S. P., Byham, J. K., & Riggs, S. L. (2021). Does what we see shape history? Examining workload history as a function of performance and ambient/focal visual attention. Transactions on Applied Perception, 18(2), 1–17.

Dorneich, M. C., Mathan, S., Ververs, P. M., & Whitlow, S. D. (2008). Cognitive state estimation in mobile environments. Augmented cognition: A practitioner's guide, 75-111.

Duchowski, A. T. (2017). Eye tracking methodology: Theory and practice (3rd ed.). Springer International Publishing AG.

Duchowski, A. T. (2018). Gaze-based interaction: A 30 year retrospective. Computers and Graphics, 73, 59–69.

Dunn, M. J., Alexander, R. G., Amiebenomo, O. M., Arblaster, G., Atan, D., Erichsen, J. T., Ettinger, U., Giardini, M. E., Gilchrist, I. D., Hamilton, R., Hessels, R. S., Hodgins, S., Hooge, I. T. C., Jackson, B. S., Lee, H., Macknik, S. L., Martinez-Conde, S., Mcilreavy, L., Muratori, L. M., … Sprenger, A. (2024). Minimal reporting guideline for research involving eye tracking (2023 edition). Behavior Research Methods, 56.

Endsley, M. R. (2023). Ironies of artificial intelligence. Ergonomics, 66(11), 1656-1668.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: an overview. In G. Underwood (Ed.), Eye guidance in reading and scene perception (pp. 269–293). Elsevier.

Klaib, A. F., Alsrehin, N. O., Melhem, W. Y., Bashtawi, H. O., & Magableh, A. A. (2021). Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies. Expert Systems with Applications, 166, 114037.

Kröger, J. L., Lutz, O. H. M., & Müller, F. (2020). What does your gaze reveal about you? On the privacy implications of eye tracking. In M., Friedewald, M., Önen, E., Lievens, S., Krenn, & S. Fricker (Eds.), Privacy and Identity Management. Data for Better Living: AI and Privacy (Vol. 576, pp. 226–241). Springer, Cham.

Marois, A., Lafond, D., Williot, A., Vachon, F., & Tremblay, S. (2020). Real-time gaze-aware cognitive support system for security surveillance. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 64(1), 1145-1149. https://doi.org/10.1177/1071181320641274

Michael, C. J.; Dennis, S. M.; Maryan, C.; Irving, S.; Palmsten, M. L. (2019). A general framework for human-machine digitization of geographic regions from remotely sensed imagery. In Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems; pp. 259–268.

Ooms, K., Coltekin, A., De Maeyer, P., Dupont, L., Fabrikant, S., Incoul, A.,... & Van der Haegen, L. (2015). Combining user logging with eye tracking for interactive and dynamic applications. Behavior research methods, 47, 977–993.

Ortego, A., & Scheuerman, J. (2024). Informing Cybersecurity Decision-Making: Parsing, Summarizing, and Visualizing User Activity Data, Naval Applications of Machine Learning 2025.

Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. In Encyclopedia of human computer interaction (pp. 211-219). IGI Global Scientific Publishing.

Raptis, G. E., Fidas, C. A., & Avouris, N. M. (2017). On implicit elicitation of cognitive strategies using gaze transition entropies in pattern recognition tasks. Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17, 1993–2000.

Raptis, G., Katsini, C., Belk, M., Fidas, C., Samaras, G., & Avouris, N. (2017). Using eye gaze data and visual activities to infer human cognitive styles: Method and feasibility studies. Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, 164–173.

Ries, A. J., Aroca-Ouellette, S., Roncone, A., & de Visser, E. J. (n.d.). Gaze-informed signatures of trust and collaboration in human-autonomy teams. Retrieved February 4, 2025.

Rebmann, A., & van der Aa, H. (2024). Recognizing task-level events from user interaction data. Information Systems, 124, 102404.

Robison, M. K., Miller, A. L., Wiemers, E. A., Ellis, D. M., Unsworth, N., Redick, T. S., & Brewer, G. A. (2024). What makes working memory work? A multifaceted account of the predictive power of working memory capacity. Journal of Experimental Psychology: General, 153(9), 2193.

Startsev, M., & Zemblys, R. (2023). Evaluating eye movement event detection: A review of the state of the art. Behavior Research Methods, 55(4), 1653–1714.

Steinhauer, S. R., Bradley, M. M., Siegle, G. J., Roecklein, K. A., & Dix, A. (2022). Publication guidelines and recommendations for pupillary measurement in psychophysiological studies. Psychophysiology, 59(4). https://doi.org/10.1111/psyp.14035

Strauch, C., Hoogerbrugge, A. J., & Ten Brink, A. F. (2024). Gaze data of 4243 participants shows link between leftward and superior attention biases and age. Experimental Brain Research, 242(6), 1327-1337

Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. Cognitive, Affective and Behavioral Neuroscience, 16(4), 601–615.

Wickens, C. D., Mccarley, J. S., Alexander, A. L., Thomas, L. C., Ambinder, M., Zheng, S., & Field, M. (2008). Attention-situation awareness (A-SA) model of pilot error. In D. Foyle & N. Hooey (Eds.), Human performance modeling in aviation (pp. 214–239).

Zhang, X., Cole, M., Belkin, N. (2011). Predicting users' domain knowledge from search behaviors. In Proceedings of the 34th international ACM SIGIR Conference on Research and Development in Information Retrieval.