Adapting Language Model Responses with Implicit Feedback for Effective Human-AI Collaboration

Jaelle Scheuerman, Shannon McGarry, Ciara Sibley, Noelle Brown
U.S. Naval Research Laboratory
4555 Overlook Ave. S.W.
Washington, DC 20375 USA
jaelle.scheuerman.civ@us.navy.mil

Abstract

To be effective teammates with humans, is necessary for AI systems to deduce user state and context from implicitly collected feedback. In this extended abstract, we explore ways in which user activity and eyetracking systems can provide implicit feedback for the purpose of adapting the AI model to user's changing goals and needs.

Extended Abstract

For AI assistants to be as effective as human teammates, AI models will need to be able to adapt to the user, their current tasks and the environment. This requires that the models are capable of learning and rapidly adapting to a wide range of feedback. Adapting to user feedback presents challenges to many AI systems. Classic machine learning models aim to train models that are similar to the environment in which they are expected to make predictions. This approach provides little opportunity to adapt when changes occur in the environment, often causing model performance to degrade over time. Some efforts have explored how to employ explicit user feedback to incrementally train a model by asking them to validate model results and correct erroneous output. However, this interactive machine learning approach relies on a user taking the time to provide feedback and the model performance may lag until the necessary user feedback arrives.

In this abstract, we explore a potential approach to improving how AI systems adapt to their user's goals and needs while reducing the burden of relying on the user for explicit feedback. Rather than require the user to provide explicit feedback to the model, we explore methods for collecting implicit data passively from the user or environment that is informative about the current user's state and context. For example, the way in which a person's pupil responds to AI's output or actions could inform if the person is on or off task (Unsworth and Robison 2016). 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, guiding its responses to adapt to new tasks using this context as input (Brown et al. 2020). 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, or how many and which applications are open. Such logs can provide enormous context to an AI system relating to what is on the user's screen, what they are doing and how they are doing it. For example, in training scenarios, such information could be used to guide an AI assistant to provide documentation relating to the tools the user is actively working with or highlight best practices that the student may be unaware of. It could further infer factors impacting the user and avoid interrupting them during times of high workload. Many ongoing efforts have explored classifying low-level user activity from logs as high-level events (Rebmann and van der Aa 2024). 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 and Scheuerman 2024), we explored using generated natural language summaries of the activity logs of cybersecurity operators 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.

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 end user. Almost all eye trackers give information about the person's pupil size and scan patterns. For example, pupil size and its evolution, can indicate physical and mental effort, attentional shifts, as well as differences in cognitive abilities (Strauch, Hoogerbrugge, and Ten Brink 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 and Duchowski 2017) . 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 (Poole and Ball 2006). With recent innovations, eye tracking is less invasive and more versatile, mobile, and more cost-effective than ever before and compared to other psychophysiological measures, prompting it for wide-scale use. Research has already shown how eye tracking data can be used in real-time for a range of applications (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 (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.

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 (Wickens et al. 2008). Interpreting user activity logs to differentiate between certain tasks is also a challenge, and interaction patterns may be different 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 not always clear how to synthesize the eye tracking data with these other human-based measures. 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 (Kröger, Lutz, and Müller 2020). 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.

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 for use as language model context. However, if these can be overcome, it will open up many new opportunities for effective human-AI collaboration.

References

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33: 1877–1901.

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

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

Duchowski, A. T.; and Duchowski, A. T. 2017. *Eye tracking methodology: Theory and practice*. Springer.

Klaib, A. F.; Alsrehin, N. O.; Melhem, W. Y.; Bashtawi, H. O.; and 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.; and Müller, F. 2020. What does your gaze reveal about you? On the privacy implications of eye tracking. In *IFIP International Summer School on Privacy and Identity Management*, 226–241. Springer.

Ooms, K.; Coltekin, A.; De Maeyer, P.; Dupont, L.; Fabrikant, S.; Incoul, A.; Kuhn, M.; Slabbinck, H.; Vansteenkiste, P.; and Van der Haegen, L. 2015. Combining user logging with eye tracking for interactive and dynamic applications. *Behavior research methods*, 47: 977–993.

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

Poole, A.; and Ball, L. J. 2006. Eye tracking in HCI and usability research. In *Encyclopedia of human computer interaction*, 211–219. IGI Global Scientific Publishing.

Rebmann, A.; and van der Aa, H. 2024. Recognizing task-level events from user interaction data. *Information Systems*, 124: 102404.

Strauch, C.; Hoogerbrugge, A. J.; and 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.; and Robison, M. K. 2016. Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience*, 16: 601–615.

Wickens, C. D.; McCarley, J. S.; Alexander, A. L.; Thomas, L. C.; Ambinder, M.; and Zheng, S. 2008. Attention-situation awareness (A-SA) model of pilot error. *Human performance modeling in aviation*, 213–239.