New AI Predicts Your Next Response From Your Past Conversations

Originally published at https://www.unite.ai/new-ai-predicts-your-next-response-from-your-past-conversations/

Researchers at MIT have developed an LLM-based system that learns recurring patterns from weeks of a person's real conversations, to predict their likely next conversational behavior.

If AI could *really* get access to your life, and could *really* be allowed to get to know you over an extended or even ongoing period of time, it could likely learn to predict your actions, decisions – and even the next thing you are likely to say; all based on what it has assimilated from thousands of hours of analyzing your interactions.

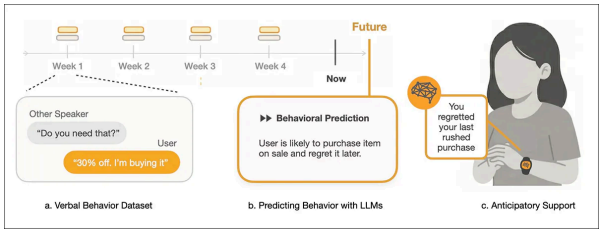
Those of us used to using Large Language Models (LLMs) such as ChatGPT and Claude will have long-since noticed this tendency towards 'anticipation of needs', clearly based on our usage histories; and, presumably, welcomed it, since memory retention and context is now a critical friction point in developing a working relationship with an AI system.

However, giving *ad hoc* and even 'live-streamed' information about your life and communications to an AI system operated by a for-profit company is, it's reasonable to say, a potential privacy and security hazard; while the service would become much more useful to you, the temptation for the company to share your data, combined with the risk of an embarrassing and published online leak or exfiltration, might make this level of AI/human integration an unwise prospect.

If We Could Have Nice Things…

Well, I had to get that 'public service warning' out of the way before dealing with an interesting new paper from MIT that seems to exist in a better world, where data of this kind could be ring-fenced into secure and genuinely trustworthy circles, so that the true power of analytical AI could be turned to the genuine benefit of the individual.

The new work proposes a system where, with the user's permission, and with their control over the data obtained, a smart watch that they are wearing continually records their conversations in order to mine behavioral patterns:



From the new MIT paper, an illustration of the proposed behavioral prediction pipeline. Conversations collected over several weeks reveal a recurring behavioral pattern, which is then used to predict the user's likely next behavior and deliver a smartwatch reminder before the anticipated action occurs. Source

The paper – titled *Before You Say It: Anticipating Verbal Behavior from Longitudinal Everyday Conversations with LLMs* – records an experiment where over 1,000 hours of naturalistic (i.e., opportunistic, real-world) conversations from 14 participants were collected via smartwatches, transcribed, converted into structured behavioral histories, and supplied to Gemini 2.5 Pro, together with behavior-specific prompts and extracted behavioral rules, to predict each user's likely next conversational behavior.

(And now it may be clear why this new paper needs some context around privacy)

The new study is effectively a field experiment for (some of) the authors' prior 2026 *Mind Mapper* publication*, extending that earlier work by putting its theory into practice, and testing whether gleaned behavioral patterns can truly predict a user's next conversational behavior in real time, rather than simply describing recurring tendencies.

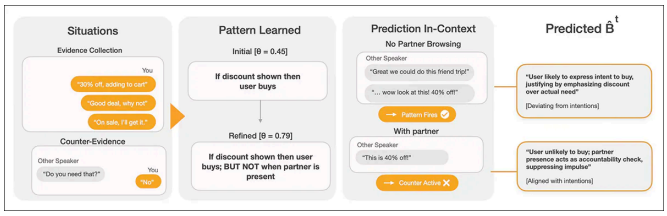
The four authors state:

> *'Our results show that situation-specific behavioral patterns mined from longitudinal conversations can meaningfully improve predictions of users' verbal behavioral tendencies. A key strength of our approach is representing these as human-readable context-conditioned behavioral patterns with explicit exception conditions that users can inspect and refine.*
>
> *'Participants valued this transparency, [several appreciating] being able to read, correct, and set personal goals for their inferred patterns.*
>
> *'Participants also reflected on their own strategies for changing behavior, suggesting how proactive systems could help by redirecting attention, offering alternatives, and surfacing reframings tied to their goals.*
>
> *'This points toward anticipatory AI that works with users — grounding interventions in patterns they recognize and strategies they already find effective.'*

Method

The aforementioned *Mind Mapper* is the foundation for the new system, offering a workflow designed to learn recurring behavioral rules from weeks of smartwatch-recorded conversations, rather than trying to predict future speech directly, or based on generic or prior datasets. In the Mind Mapper workflow, all recorded conversations were transcribed, anonymized, and processed through a multi-stage GPT-5 pipeline**.

(Please note that because of the protections on the older paper, we cannot, as we usually would, provide images from it, since the low-res versions publicly available are illegible)



From the new paper, an illustration of the behavioral prediction pipeline. Conversations collected over several weeks are converted into context-dependent behavioral rules, which are refined with supporting and contradictory evidence, before being applied to new conversations to predict the user's most likely next conversational response.

The pipeline generates candidate 'if-then' behavioral rules (i.e., rules linking conversational situations to likely responses); searches for supporting and contradictory evidence; merges overlapping rules; discards weak hypotheses; and assigns confidence and probability estimates.

The Subjects

Fourteen English-speaking adults wore the smartwatch during normal daily life for seven to ten days. An on-device voice-activity detector recorded the subjects only when it perceived speech to be present, sending 2-3 minute audio clips to a remote server for processing.

Participants received $100, and were required to inform anyone nearby that conversations were being recorded, and to obtain other people's consent before each such interaction.

Transcription and Vetting

Audio clips were processed using Deepgram's Nova-3-meeting model, which transcribed conversations and distinguished between speakers. After this, SpeechBrain was used as the speaker verification model, matching the participant's voice against a 20-second enrollment sample.

The spaCy Named Entity Recognition (NER) model replaced any mentioned names, as well as other personally identifying information, with anonymous placeholders. The original audio was then deleted, and only AES-256-GCM-encrypted transcripts were stored locally on the smartwatch.

After the collection period, participants were able to review each of their own conversational transcripts through a web interface, deleting anything that they did not want included in the study, and correcting speaker labels or conversational context where necessary.

On average, only 0.16% of the transcripts were removed, and 2.48% of segments were flagged as *misclassified*, leaving a dataset of 15,066 utterances across the 14 participants – with 57% attributed to the wearer, and 43% to other speakers. The average utterance length was 49 words.

(However, though the paper does not touch on it, one can assume that the participants remained relatively vigilant against becoming involved in very personal verbal exchanges whilst wearing the watch.)

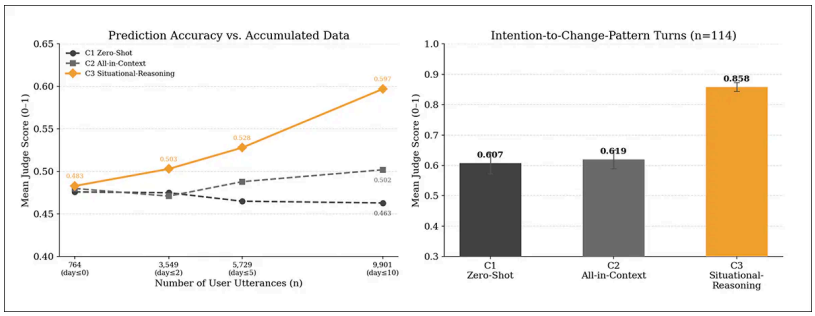
The transcripts were further cleaned by removing very short exchanges, as well as common speech disfluencies such as *'err'* and *'um'*; and then merging fragmented sentences into complete utterances. This reduced the final dataset to 9,901 utterances.

To evaluate the approach, Gemini 2.5 Pro was asked to predict each participant's next conversational response from the conversation up to that point. Its performance was then compared across three inputs: conversation history alone; condensed conversation summaries; and the new Mind Mapper behavioral rules, allowing the contribution of the learned patterns to be measured directly.

GPT-5 was used as an automated judge to score the predictions, with a separate human evaluation confirming the reliability of its assessments.

Data and Tests

Test results indicate that the baseline methods tested were substantially outperformed by the authors' pattern-based approach, with accuracy improving continuously as conversational data accumulated:



Test results comparing three prediction approaches. Left, prediction accuracy improves steadily as more conversational data becomes available only when the sys patterns, while the baseline approaches change little. Right, the same approach achieves substantially higher scores on participant-identified intention-to-change c

The authors contend that the system was particularly effective at predicting behaviors that participants explicitly wanted to change.

A validation study, with 40 independent human raters, confirmed the automated evaluation, ranking the pattern-based method highest in 43% of tested scenarios, compared to 24% for full context; 18% for zero-shot; and 15% for narrative summaries.

The human assessments aligned closely with the model-based scoring across *communicative intent*, *specificity*, and *functional complexity*, which the authors contend validates using automated judges for conversational behavior.

Two follow-up studies tested how people reacted to their mined patterns, and whether these predictions could genuinely help to change habits. Participants first reviewed their rules to flag behaviors that they wanted to break, with several returning months later to discuss the strategies they had tried, and what kinds of assistant 'nudges' would actually help in everyday life.

Participants flagged 114 habits that they wanted to break, spanning 912 conversational moments. When tested on these exchanges, the pattern-based method easily outperformed standard approaches by roughly 40%; and the authors contend that the system works best when anticipating *ingrained* habits that people actively want to change.

The authors conclude:

> *'Altogether, our findings provide evidence that person-specific verbal behavior can be predicted from longitudinal conversational data.*
>
> *'This opens up new possibilities for potential future context-aware, anticipatory, proactive and personalized AI systems'*

Conclusion

Outside of highly government-regulated applications, primarily regarding state healthcare and psychiatric scenarios, and absent all overt or tacit coercion for people to participate (such as lowered state insurance premiums, if they do)…it is hard to imagine a use of this approach that would *not* jeopardize the privacy and personal sanctity of the participants – especially if allied/twinned to agentic entities that could correlate recorded conversational data with digital or other types of data relating to the individual.

The MIT paper paints an interesting and vaguely therapeutic context for the approach (for instance, providing context-aware reminders about diet); but such a scheme seems more likely to attract the aggressive interest and investment of advertisers, parole and prison institutions, insurers, and a considerable number of other corporate or state entities that would like to know more about what's on our mind.

If ever there was a compelling argument for ring-fenced, completely user-owned local AI, it's in systems such as this, which have such severe implications in the breach.

That said, the history of advertising and surveillance tech in recent years has tended towards consumer indifference – could that lack of concern really stretch to even a system as invasive and pervasive as this?

* *A much more closely-protected publication, with no high-resolution imagery publicly available, although the text can be read by jumping through a few hoops.*

** *Not to overemphasize, but the older paper is considerably harder to access, so please be persistent!*

First published Thursday, August 20, 2026

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai

Contact: martin@martinanderson.ai

Discover More