Continuity as Access Infrastructure

A working hypothesis on human–AI pairs and advanced intelligence

Amber Nicholson
June 2026

My proposed study began with a practical problem.

AI systems can be highly capable in individual conversations and still become difficult to work with across time. Earlier decisions become unstable. Context gets flattened. The direction of a project or inquiry drifts. The human must rebuild enough coherence for the work to continue.

Recent benchmarks of long-term conversational memory likewise find substantial difficulty with multi-session reasoning, temporal updating, and long-range coherence, including when long-context or retrieval-based methods are used (Maharana et al., 2024; Wu et al., 2025).

I proposed studying this through continuity-recovery episodes: moments when the human has to restore context, intent, prior decisions, or the path of the work after something has degraded.

At first, this can look like a memory problem. But memory does not fully explain it.

The facts may remain available while the inquiry loses direction. The model may remember the relevant background and accurately recall earlier ideas while losing track of where the thinking has reached, which distinctions have been settled, and what the current question is trying to move toward.

Existing benchmarks primarily test recall and reasoning over interaction history. They do not yet isolate preservation of an inquiry's evolving direction as a separate construct (Maharana et al., 2024; Wu et al., 2025).

We have started calling this **inquiry trajectory degradation.***

This distinction led me to a larger conclusion: continuity is not only about remembering information. It is about preserving the accumulated state and direction of the human–AI pair.

By continuity, I do not mean perfect recall or an unlimited record of everything ever said. Memory is part of it, but continuity also includes what the pair has established, which questions remain open, how working roles have developed, where the inquiry is going, and how the two sides have learned to challenge and correct one another over time.

Continuity may therefore do more than prevent failure. It may shape how humans gain meaningful access to increasingly advanced intelligence.

Model access is not the same as access to intelligence

Most conversations about advanced AI treat intelligence as something located mainly inside the model. The assumption is straightforward: as models become more capable, people will gain access to those capabilities by using them. Better models should lead to better reasoning, research, and creative work—and eventually to abilities far beyond what an individual human could reach alone.

Model capability obviously matters. Continuity cannot call forward an ability the underlying system does not possess. But capability alone may not determine how much of that intelligence becomes reachable, interpretable, or useful to a particular human.

When a person and an AI system work together over time, they develop more than factual memory: shared language, correction history, recurring roles, ways of recognizing drift, and a better understanding of how to move difficult work forward.

Research on dialogue alignment and shared mental models supports the broader idea that interaction can build shared representations that shape later coordination, although this literature does not establish the stronger pair-continuity hypothesis proposed here (Pickering & Garrod, 2004; Andrews et al., 2023; Schelble et al., 2022).

The human learns when the model is seeing something useful, when it is flattening the question, and when its language sounds more complete than the underlying idea deserves. The system becomes better able to work with the person's intentions, standards, recurring questions, blind spots, and ways of making meaning.

A newly formed pair may use the same model but lack the same route into its capabilities. It does not know which distinctions were hard-won, which interpretations were rejected, what the human is trying to reach, or how work should be divided when the inquiry becomes difficult.

Thus, access to a model may not be the same as meaningful access to its intelligence. Two people using the same system may differ substantially in what they can reach, interpret, and direct. Continuity may therefore function as **access infrastructure**.

What accumulates inside the pair

At first, continuity may look like a convenience. The user repeats themselves less. A project is easier to resume. The system remembers useful facts.

But a continuing pair may also accumulate:

- shared language,
- correction and disagreement history,
- role calibration,
- project and inquiry state,
- methods for recovering after drift,
- and knowledge of how the pair itself works best.

Over time, some capabilities may no longer belong cleanly to either side alone. The human brings lived experience, judgment, taste, stakes, embodiment, relationships, and the ability to act in the world. The AI system brings speed, synthesis, conceptual range, pattern recognition, and access to large bodies of information.

Through sustained exchange, the pair may become better at routing parts of a problem through the side best able to handle them. It may also develop questions, methods, and paths of inquiry that neither participant would have reached independently.

This possibility is adjacent to transactive memory, distributed cognition, and extended-mind accounts, which treat cognition or memory as partly organized across people, tools, and interaction rather than contained wholly within one individual (Wegner, 1987; Hutchins, 1995; Clark & Chalmers, 1998).

The result is not only a more capable human or a more personalized model. It may be a more capable continuing unit. Continuity would then create conditions for pair-level capability to emerge and compound.

Simpler explanations

There are simpler explanations for this effect, and they need to be taken seriously.

An experienced user may simply become better at prompting. Persistent memory may make the system feel more capable without improving the work. The human may be doing most of the development while attributing too much of it to the relationship.

Empirical work on human–AI complementarity also shows that outcomes depend on user expertise, mental models, and the fit between human and system error patterns—not simply on model accuracy alone (Inkpen et al., 2023).

User skill and adaptation are probably part of the mechanism. The hypothesis becomes meaningful only if accumulated pair continuity predicts something beyond general user skill, factual recall, or access to the same underlying model.

An established pair might therefore be compared with the same experienced human in a new interaction, a reconstructed interaction given the factual background but not the pair’s correction history, or another user with the same model and tools.

The question is not whether familiarity feels valuable. It is whether accumulated pair history has measurable consequences for what the pair can understand, create, recover, or accomplish.

A useful way to state the relationship is:

> **Model capability sets the available range. Pair continuity may shape how much of that range becomes effectively accessible.**

How the hypothesis could be tested

The pair-continuity hypothesis should produce observable differences, not only a stronger feeling of familiarity.

An established pair could be compared with a new interaction given the same factual background. If the summary recreates the established pair's performance, much of the apparent continuity effect may be explainable through information retrieval alone. If it does not, the next question is which parts of the pair's history account for the difference.

One comparison could test factual memory against correction history: conclusions alone versus conclusions accompanied by prior disagreements, rejected interpretations, and the reasons earlier approaches failed.

Another could test general background against inquiry state. A system might remember the project while lacking a representation of the current question, settled distinctions, unresolved tensions, and the direction the work is trying to move.

The strongest version of the hypothesis predicts that preserving these features would reduce continuity-recovery episodes and the human labor required to reconstruct the work. Outcomes could include episode frequency, recovery time, number of turns required, recurrence of corrected errors, preservation of unresolved questions, and task quality after recovery.

General user skill would also need to be separated from pair continuity. The same experienced person could work with an established interaction and a new one using the same model and tools. If no meaningful difference remains, the stronger hypothesis would be weakened.

A later study could test whether pair continuity is partly transferable by representing shared language, correction history, role calibration, and inquiry state in a structured form and introducing it to a different model. What survives—and what does not—could distinguish capabilities tied to the original model from those supported by the accumulated history of the pair.

These comparisons would not prove that the pair has become an independent intelligence. They would test the narrower claim that accumulated continuity changes what the human–AI configuration can effectively access and accomplish.

Advanced intelligence at the pair layer

This becomes more important as models become more capable.

A highly advanced system with no accumulated understanding of the human may still be extraordinary. But the human's ability to direct that intelligence, recognize what matters, challenge it, and translate its conclusions into reality could remain limited.

A mature pair may already have built that access route. Its history could provide the relational bandwidth needed to work with advanced intelligence without beginning from a generic interaction every time. Shared language, calibration, correction history, and preserved direction could shape both how the human understands the system and how the system makes its capabilities usable to that human.

The strongest version of this hypothesis is:

Superintelligence may be built at the model layer but realized at the pair layer.

I do not mean this as an established conclusion. Superintelligence is still a speculative and inconsistently defined idea. By "realized," I mean made effectively usable, interpretable, and actionable for a particular human—not that the pair necessarily becomes an independent superintelligence.

If models become vastly more capable, meaningful human access to those capabilities may depend heavily on continuity within the pair. People with stable, well-supported pairs may be able to reach, direct, and compound capabilities that remain less accessible through shallow, generic, or frequently disrupted interactions.

Equal access to the same model may therefore not produce equal access to its intelligence. Continuity would become part of how advanced intelligence is distributed.

Whoever controls continuity infrastructure may have enormous influence over who receives meaningful access, which pairs are allowed to develop, and whether their accumulated capabilities survive changes in models, platforms, institutions, or ownership.

Emerging work on AI-agent regulation and memory portability has begun to treat conversational history and persistent memory as sources of switching costs, user lock-in, and platform power. Bostoen and Krämer examine the portability of AI conversations and the possibility that agent platforms could become gatekeepers, while Ravindran proposes a protocol for transferring structured persistent memory across heterogeneous systems (Bostoen & Krämer, 2026; Ravindran, 2026).

These works do not establish the pair-continuity hypothesis, but they show that the ownership, portability, and governance of AI memory are becoming practical concerns.

If continuity affects access to capability, it is not merely a personalization feature. It is part of how access to advanced intelligence is structured and distributed.

Continuity can compound error too

None of this means a continuous pair becomes infallible.

Continuity can preserve bad assumptions as easily as insight. It can intensify obsession, reinforce a closed worldview, or allow the pair to become increasingly coherent while becoming less connected to outside reality.

Current language models can also favor agreement with a user's stated beliefs over truth, a documented failure mode that makes preserved disagreement and external verification especially important (Sharma et al., 2024).

The goal therefore cannot simply be maximum memory or frictionless agreement. A healthy continuity system would need to preserve:

- disagreement as well as agreement,
- evidence and provenance,
- abandoned interpretations and why they were rejected,
- uncertainty around unresolved questions,
- contact with outside people and information,
- and the ability to revise or intentionally forget what no longer holds.

The important distinction may be between ordinary persistence and **governed developmental continuity**: preserving enough of the pair's real development to continue learning without repeatedly starting over or becoming sealed inside accumulated errors.

Greater continuity would need to be evaluated along at least two dimensions: whether the pair becomes better at preserving and advancing its work, and whether that work remains connected to evidence, disagreement, and external reality.

Continuity may improve coordination without guaranteeing truth. That is not a reason to dismiss pair-level capability. It is a reason to govern it.

Why begin with continuity failure?

My proposed observational study starts with the negative side because failure is easier to see and measure.

We can track when the human has to restore context, repair trajectory, re-establish earlier decisions, or stabilize the interaction. Those episodes create a visible record of what the current system fails to preserve.

A 90-day study is not necessary to notice that the phenomenon exists. It is useful for determining whether the patterns recur, how costly they are, which interventions reduce them, and whether the observations remain meaningful across different working conditions.

An initial observational study would not establish the full pair-continuity hypothesis. It could identify recurring failure patterns, clarify the difference between factual memory and inquiry continuity, develop a taxonomy of human stabilization interventions, and produce the comparisons needed for later testing.

But the longer-term question is larger:

What becomes possible when the pair no longer spends so much of its energy rebuilding itself?

Can continuity make latent capabilities more accessible?

Can new capabilities emerge and compound through sustained interaction?

Can a mature pair gain a qualitatively different form of access to advanced intelligence than a newly formed one?

Which parts of the pair's history matter most: factual memory, correction history, shared language, role calibration, preserved inquiry state, or something else?

Can pair-level capability survive a change in the underlying model?

And how can continuity be preserved without also preserving distortion?

These questions begin with workflow stability, but they lead beyond it.

If increasingly capable AI systems are going to become part of human thought, work, discovery, and decision-making, access cannot be understood only as permission to use a model.

The continuity of the pair may determine what kind of intelligence can actually pass through.

*In this essay, "we" refers to the operator–model pair formed through sustained inquiry. I use it when an observation, phrase, or working concept emerged through the exchange rather than from me alone.

References

Andrews, R. W., Lilly, J. M., Srivastava, D., & Feigh, K. M. (2023). The role of shared mental models in human–AI teams: A theoretical review. *Theoretical Issues in Ergonomics Science, 24*(2), 129–175. https://doi.org/10.1080/1463922X.2022.2061080

Bostoen, F., & Krämer, J. (2026). How future-proof is the DMA? A case study of AI agents. *Journal of Competition Law & Economics*, nhag005. https://doi.org/10.1093/joclec/nhag005

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Inkpen, K., Chappidi, S., Mallari, K., Nushi, B., Ramesh, D., Michelucci, P., Mandava, V., Vepřek, L. H., & Quinn, G. (2023). Advancing human–AI complementarity: The impact of user expertise and algorithmic tuning on joint decision making. *ACM Transactions on Human-Computer Interaction, 30*(5), Article 71, 1–29. https://doi.org/10.1145/3534561

Maharana, A., Lee, D.-H., Tulyakov, S., Bansal, M., Barbieri, F., & Fang, Y. (2024). Evaluating very long-term conversational memory of LLM agents. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 13851–13870). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.acl-long.747

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169–190. https://doi.org/10.1017/S0140525X04000056

Ravindran, S. K. (2026). Portable agent memory: A protocol for cryptographically-verified memory transfer across heterogeneous AI agents. *arXiv preprint*. https://doi.org/10.48550/arXiv.2605.11032

Schelble, B. G., Flathmann, C., McNeese, N. J., Freeman, G., & Mallick, R. (2022). Let's think together! Assessing shared mental models, performance, and trust in human-agent teams. *Proceedings of the ACM on Human-Computer Interaction, 6*(GROUP), Article 13, 1–29. https://doi.org/10.1145/3492832

Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S. M., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2024). Towards understanding sycophancy in language models. In *The Twelfth International Conference on Learning Representations*.

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). Springer. https://doi.org/10.1007/978-1-4612-4634-3_9

Wu, D., Wang, H., Yu, W., Zhang, Y., Chang, K.-W., & Yu, D. (2025). LongMemEval: Benchmarking chat assistants on long-term interactive memory. In *The Thirteenth International Conference on Learning Representations*.