# Shared Cognitive State in Human–LLM Interaction

## Failure Modes, Risk Surfaces, and Structural Implications

**Author:** Alan Friar
**Organization:** EnBra Group
**Version:** 1.0
**Date:** 2026-02-01

---

## Purpose

This document defines and analyses the structural risks that emerge when conversational context is shared across multiple human participants and one or more Large Language Models (LLM's).

It focuses on failure modes introduced by *shared cognitive state itself*, independent of model architecture, training methodology, or implementation detail.

The intent of this document is descriptive rather than prescriptive. It does not propose solutions, enforcement mechanisms, hierarchies, products, or implementation strategies.

It exists to establish the problem space clearly and independently.

---

## Scope

This document applies to interaction environments where:

- Conversational context is persistent across time
- Multiple humans may contribute to or consume that context
- One or more LLM's participate using the same or derived context
- Outputs from LLM's may influence human decision-making

The document applies equally to synchronous and asynchronous interaction models.

---

# Explicit Non-Goals

This document does not:

- Define behavioral contracts
- Specify decision authority or governance
- Propose hierarchical roles for humans or LLM's
- Describe enforcement or mediation layers
- Address legal liability or compliance frameworks
- Provide architectural or implementation guidance

These exclusions are intentional.

---

# Definition: Shared Cognitive State

A *shared cognitive state* exists when conversational context is:

- Persistent beyond a single interaction
- Accessible to multiple participants
- Capable of influencing subsequent reasoning or decisions

In such systems, no single participant—human or artificial—can be assumed to possess complete authorship, intent, or responsibility for the evolving state.

---

# Failure of Single-User Assumptions

Most LLM interaction models implicitly assume:

- A single human user
- A single locus of intent
- A single chain of reliance
- Clear attribution between prompt and response

These assumptions fail immediately once conversational state is shared. Context becomes collective, intent becomes ambiguous, and responsibility diffuses across participants.

---

## Authorship and Attribution Collapse

Within shared conversational state:

- Prompts are influenced by prior unseen exchanges
- Outputs reflect layered human and machine contributions
- Conclusions emerge without a clear author

This makes post-hoc attribution unreliable and undermines accountability for both correctness and misuse.

---

## Temporal Asymmetry and Late Entry

Participants may enter a shared conversation after critical reasoning has occurred.

Late-arriving participants inherit conclusions without exposure to the uncertainty, assumptions, or exploratory reasoning that produced them. This temporal asymmetry increases confidence without increasing understanding.

---

## Confidence Laundering

Shared cognitive state enables a structural failure mode in which:

- Speculative outputs are restated
- Restated outputs are summarized
- Summarized outputs are treated as established conclusions

Confidence increases through repetition rather than verification. The presence of multiple participants or multiple LLM's does not inherently mitigate this effect.

---

## Correlated Reasoning Across Multiple LLM's

When multiple LLM's operate over the same conversational context, their outputs are not independent.

Shared context couples their reasoning, leading to correlated assumptions rather than true cross-validation.

Multiplicity without enforced separation amplifies coherence, not correctness.

---

## Error Persistence and Propagation

Errors introduced into shared cognitive state tend to persist.

Because LLM outputs are conditioned on prior context, early inaccuracies propagate forward and influence subsequent reasoning without explicit detection or correction.

---

## Audit and Replay Ambiguity

In shared conversational environments:

- Message order alone is insufficient for audit
- State boundaries are ill-defined
- Replay ability does not reliably reconstruct reasoning

This complicates incident analysis and undermines post-decision review.

---

## Human Reliance Under Shared Context

Humans interacting with shared AI outputs often assume:

- Collective validation
- Implicit safeguards
- Prior review by others

These assumptions are frequently incorrect and lead to over-reliance on outputs whose provenance and certainty are unclear.

---

## Structural Implications

The failure modes described in this document arise from shared cognitive state itself.

They are not resolved by improved models, better prompts, or increased transparency alone.

Any system that enables shared human–LLM interaction must treat shared cognitive state as a first-class risk surface rather than an incidental feature.

## Relationship to Other Work

This document does not define behavioral operating contracts or enforcement mechanisms.

Those concerns are addressed separately in:

Friar, A. (2026). *The Missing Engineer-Grade AI: Operating Contracts for Real-World Assistance.*

Agreement with the problem definition presented here does not require adoption of any specific solution framework.

## Status

This document represents **version 1.0** of the paper.
Future revisions, if any, will be versioned explicitly and released as separate documents.

## Citation

If referencing this work, please cite as:

Friar, A. (2026). *Shared Cognitive State in Human–LLM Interaction: Failure Modes, Risk Surfaces, and Structural Implications.*
GitHub repository (canonical source). Version 1.0.