MIRROR: Complementary Encoding and Reconstructive Consolidation for Persistent State in LLM Systems

Nicole Summer Hsing ✦ *Prepared for MemAgents Workshop @ ICLR 2026*
Arcarae

KEY INSIGHT

MIRROR implements **Complementary Learning Systems** theory to LLMs to improve working memory.

This yields a ***21% relative improvement*** across 7 models in remembering personal user information across multi-turn conversations, with ablations showing that encoding and consolidation contribute in distinct ways.

MOTIVATION

Current LLM memory strategies face a fundamental trade-off:

Strategy	Complexity	Context	Error	CLS analogue
Trace discard (CoT, Thinking)	O(1)	No	None	Encoding only
Trace accumulation (Reflexion, MemGPT)	O(n)	Yes	Accumulates	Storage w/o consolidation
Reconstructive (MIRROR)	O(1)	Yes	Bounded	Encoding + consolidation

One discards context to avoid errors; the other preserves context but accumulates errors. Neither ***consolidates*** nor ***maintains*** state across time.

WHAT IS CLS?

Complementary Learning Systems theory predicts effective memory requires two interacting systems:

- ***Fast encoding***: rapidly captures specific experience in parallel, bounded representations
- ***Slow consolidation***: reconstructively regenerates understanding vs. accumulating

CLS predicts these serve non-redundant, complementary functions. **Current LLM memory implements neither effectively.**

MIRROR OVERVIEW

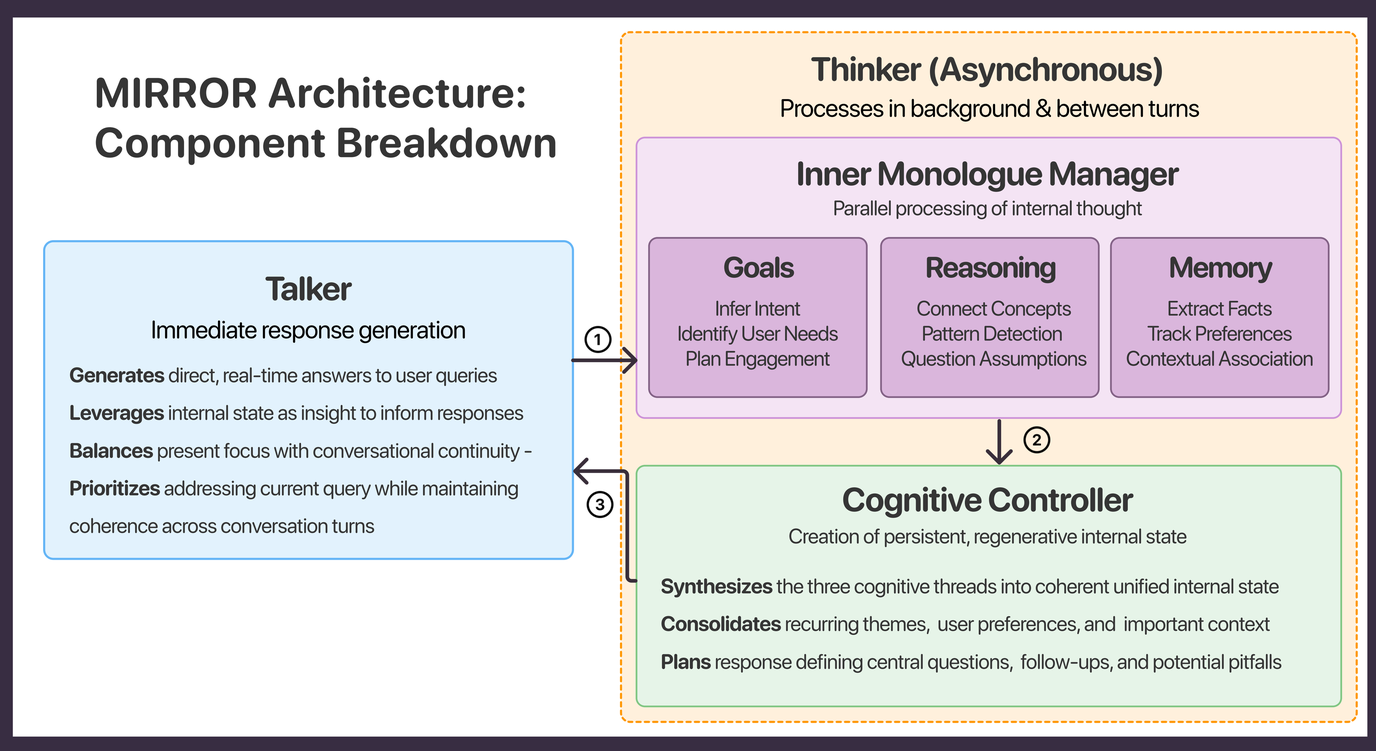
Thinker

1. **Inner Monologue Manager:** generates 3 parallel threads in 1 API call. Bounded history (≤10k tokens).
2. **Cognitive Controller:** receives ***only*** monologue and the prior narrative. Regenerates a first-person narrative (≤3k tokens) each turn and discards the prev. one → O(1) reconstruction, with selective forgetting.

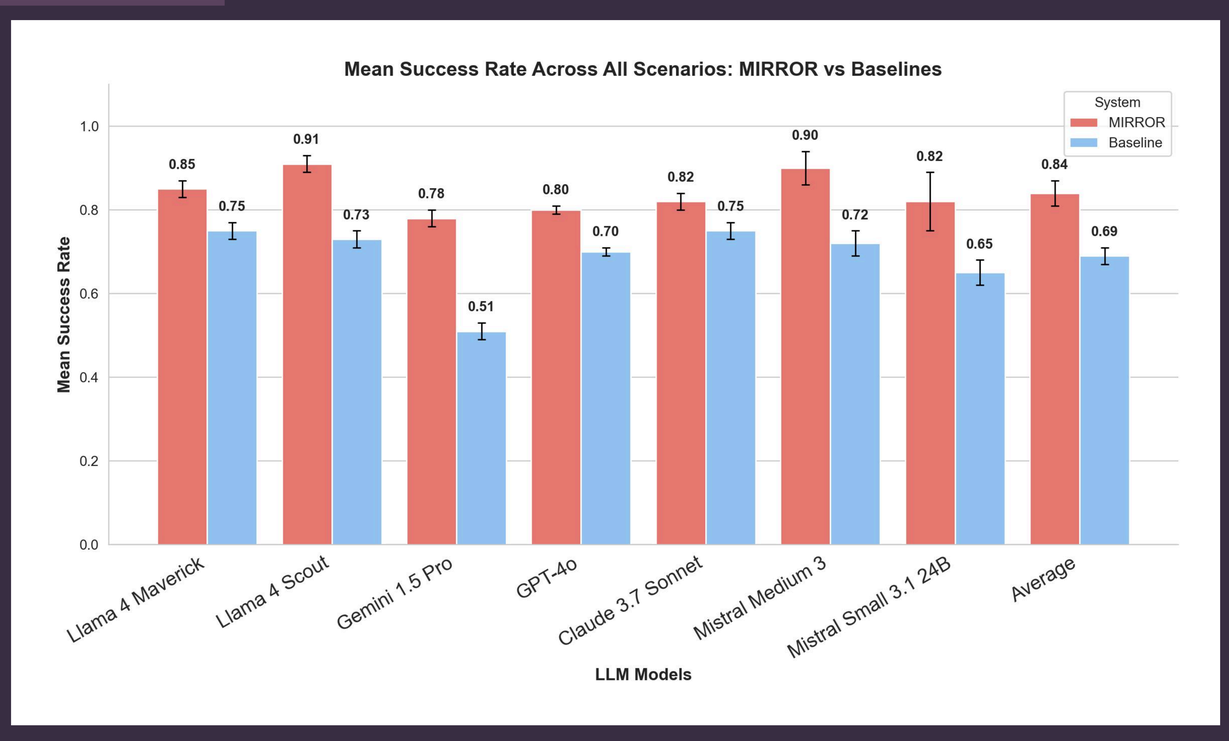
Talker

Generates the response with the most recent consolidated narrative.

Consolidation occurs asynchronously during conversation pauses, allowing immediate contextual response without delay.



RESULTS



MIRROR improves every model tested *(+21% relative on avg)*.

ABLATIONS

Consolidation alone lifts every model *(+5–20%)*; the full system beats either component alone for 6 of 7; supports the CLS prediction that encoding and consolidation are complementary.

Model	Baseline	Threads only	Controller only	Full MIRROR	Synergy
Llama 4 Scout	73%	79%	83%	91%	+8%
Llama 4 Maverick	75%	79%	84%	85%	+1%
Mistral Small 3.1	65%	65%	75%	82%	+7%
Mistral Medium 3	72%	83%	89%	90%	+1%
Gemini 1.5 Pro	51%	72%	71%	78%	+6%
Claude 3.7 Sonnet	75%	78%	87%	82%	−5%
GPT-4o	70%	71%	75%	80%	+5%

Claude 3.7 Sonnet is the sole exception; highly capable models may already implement effective implicit encoding.

TAKEAWAYS

1. ***Principled memory architecture*** provides measurable, **architecture-agnostic** gains
2. CLS complementarity prediction **holds empirically** in LLM-based systems
3. ***O(1)*** bounded reconstruction enables production deployment (~460ms added latency, $0.003–0.13/turn)
4. **Domain-agnostic principle** → extensible to agentic tasks, planning, multi-agent settings

Project page **Contact Me!**