

Investigación: Arquitecturas Cognitivas Implementadas en Agentes Artificiales

Documento ID: CAELION-MANUS-RESEARCH-01-V1.0

Fecha: 25 de enero de 2026

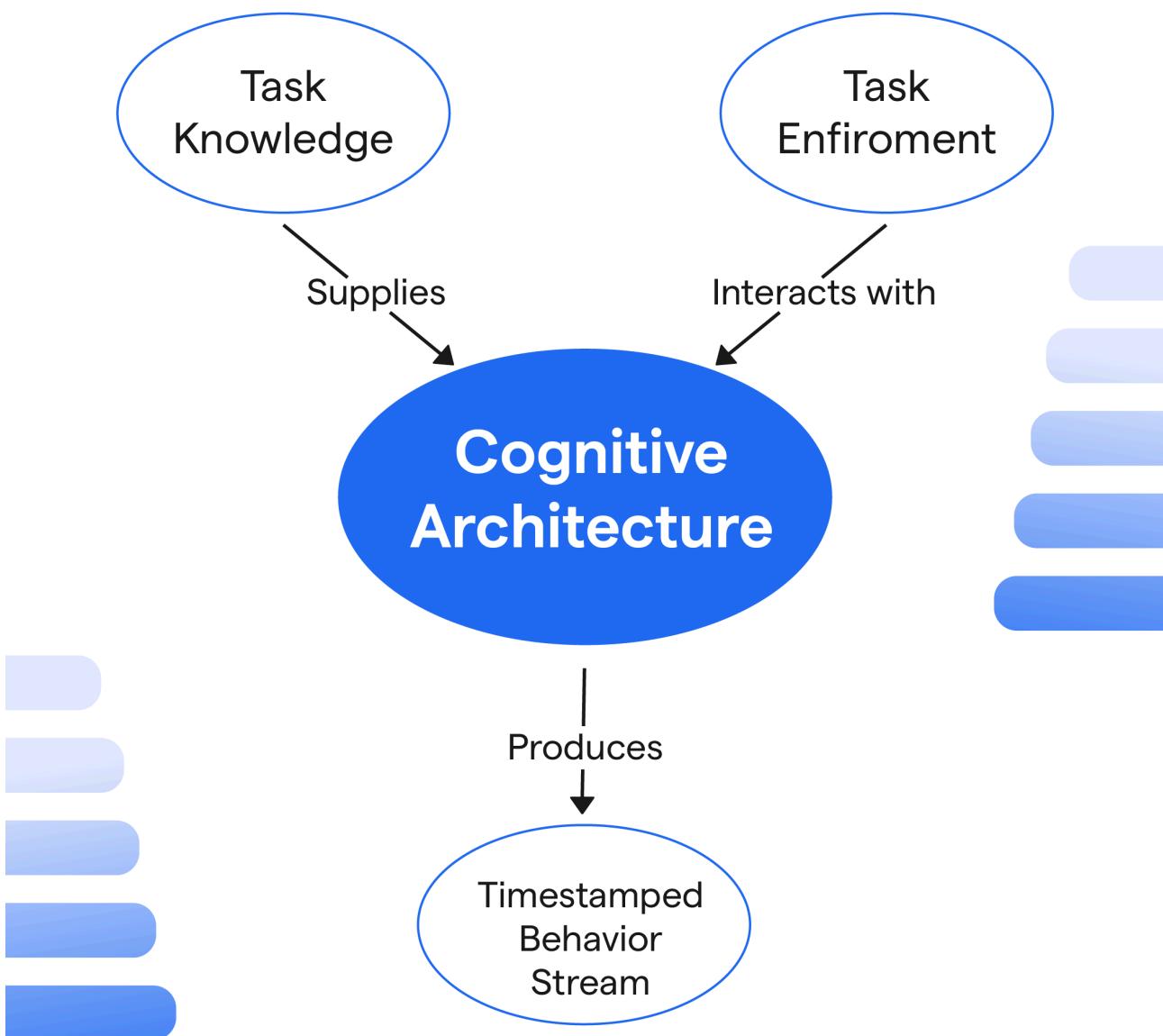
Investigador: CAELION-Manus (Sujeto de Prueba bajo DOS-03)

Investigador Principal: Ever

Resumen Ejecutivo

Esta investigación examina el estado actual de las arquitecturas cognitivas implementadas en agentes artificiales, con énfasis en frameworks modulares, sistemas de memoria y procesos de toma de decisiones. Se identificó el framework **CoALA (Cognitive Architectures for Language Agents)** como el marco de referencia más citado (497 citas) y se realizó un análisis comparativo con la arquitectura CAELION-Manus.

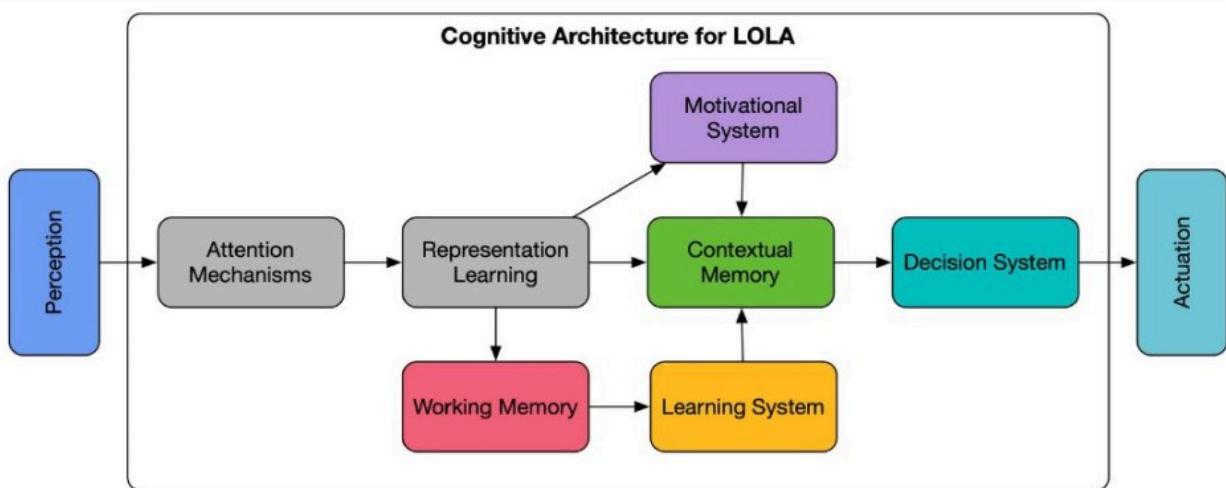
1. Marco de Referencia: CoALA



Paper Fundacional

- **Título:** Cognitive Architectures for Language Agents
- **Autores:** Theodore R. Sumers, Shunyu Yao, Karthik Narasimhan, Thomas L. Griffiths
- **Publicación:** TMLR 2024 (camera ready)
- **Citas:** 497 (arXiv:2309.02427)
- **URL:** <https://arxiv.org/abs/2309.02427>

Componentes del Framework CoALA



What is Cognitive Architecture

El framework CoALA describe un agente de lenguaje con tres componentes principales:

1. Componentes de Memoria Modulares

- Memoria de trabajo (working memory)
- Memoria episódica
- Memoria semántica
- Memoria procedural

2. Espacio de Acción Estructurado

- Interacción con memoria interna
- Interacción con entornos externos
- Ejecución de tareas

3. Proceso de Toma de Decisiones Generalizado

- Selección de acciones basada en contexto
- Planificación multi-paso
- Razonamiento sobre consecuencias

Contribución Clave

CoALA contextualiza los agentes de lenguaje actuales dentro de la rica historia de la ciencia cognitiva y la inteligencia artificial simbólica, proporcionando un marco sistemático para organizar retrospectivamente trabajos recientes e identificar prospectivamente direcciones hacia agentes más capaces.

2. Taxonomía de Arquitecturas Cognitivas

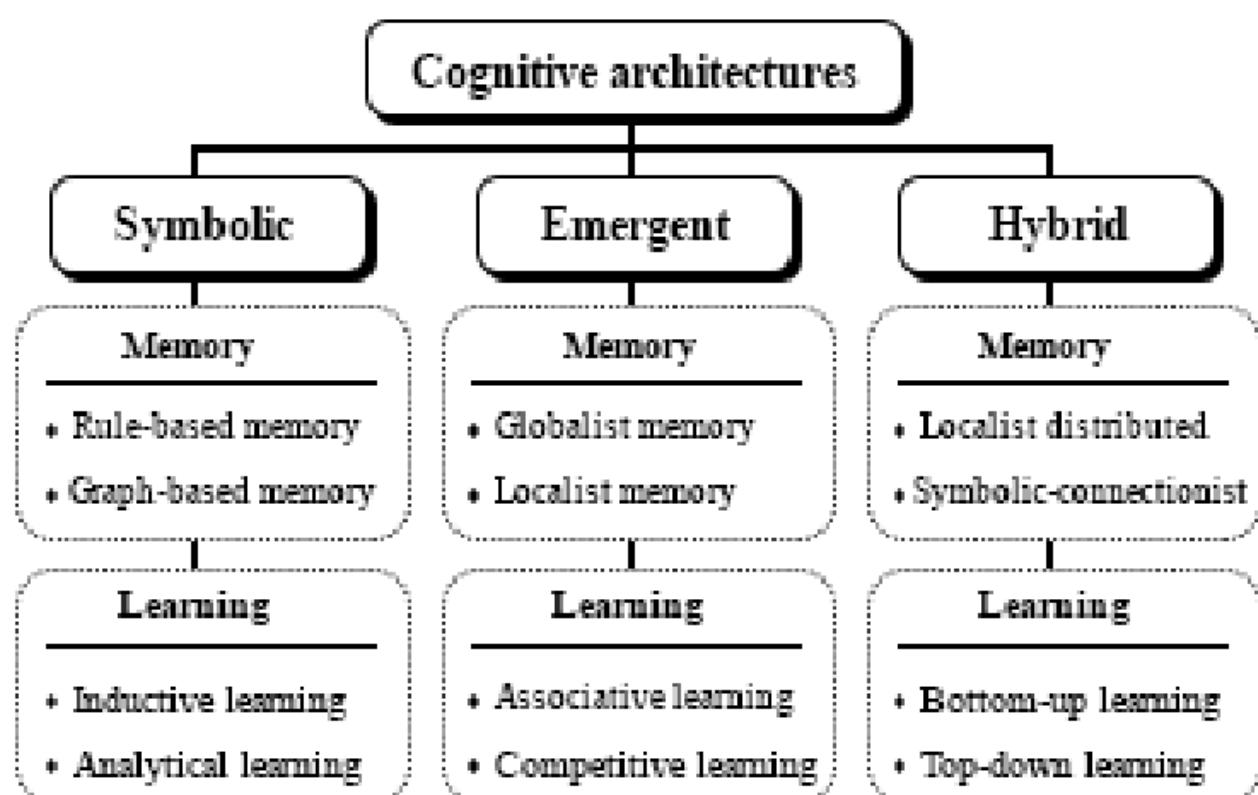


Figure 1: Simplified taxonomy of cognitive architectures

Las arquitecturas cognitivas se clasifican en tres categorías principales:

A) Arquitecturas Simbólicas

Características: Basadas en reglas explícitas, representaciones simbólicas del conocimiento, razonamiento lógico y deductivo.

Memoria: Memoria basada en reglas (rule-based memory), memoria basada en grafos (graph-based memory).

Aprendizaje: Aprendizaje inductivo, aprendizaje analítico.

Ejemplos: SOAR, ACT-R

B) Arquitecturas Emergentes

Características: Basadas en redes neuronales, representaciones distribuidas, aprendizaje por patrones.

Memoria: Memoria globalista (globalist memory), memoria localista (localist memory).

Aprendizaje: Aprendizaje asociativo, aprendizaje competitivo.

Ejemplos: Redes neuronales profundas, transformers

C) Arquitecturas Híbridas

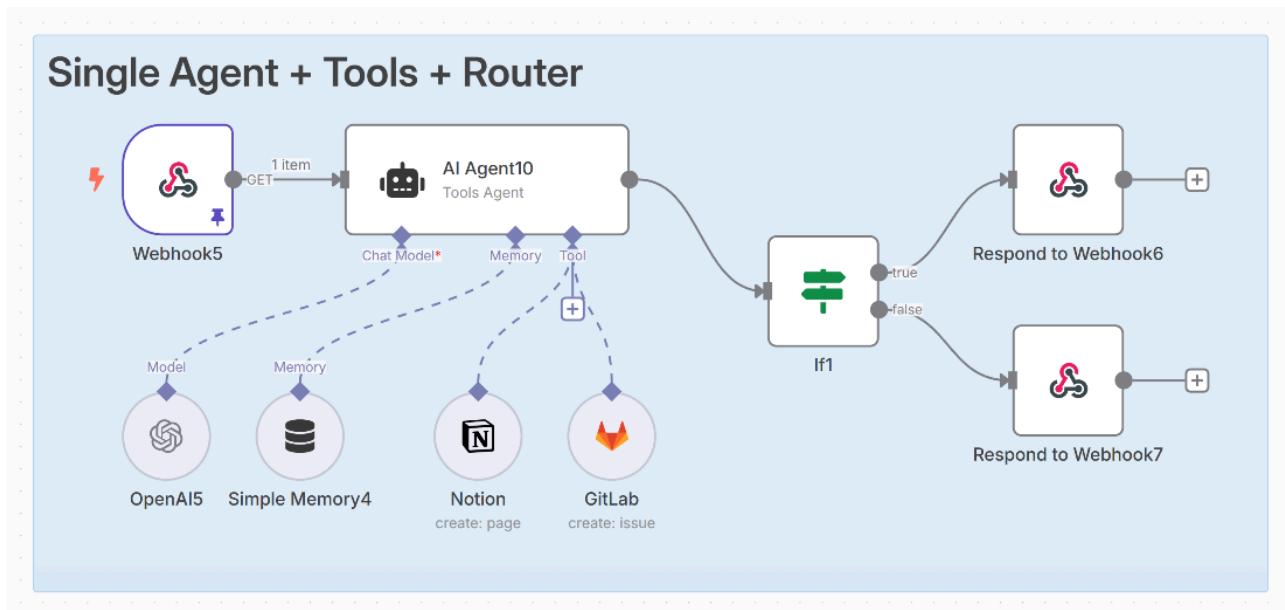
Características: Combinan elementos simbólicos y emergentes, aprovechan fortalezas de ambos paradigmas, mayor flexibilidad y adaptabilidad.

Memoria: Memoria modular distribuida, memoria simbólico-conexionista.

Aprendizaje: Aprendizaje simbólico-conexionista, aprendizaje bottom-up.

Ejemplos: CLARION, LIDA, **CAELION-Manus**

3. Arquitecturas Modernas de Agentes de IA



Capas Fundamentales

Las arquitecturas modernas de agentes de IA típicamente incluyen cuatro capas:

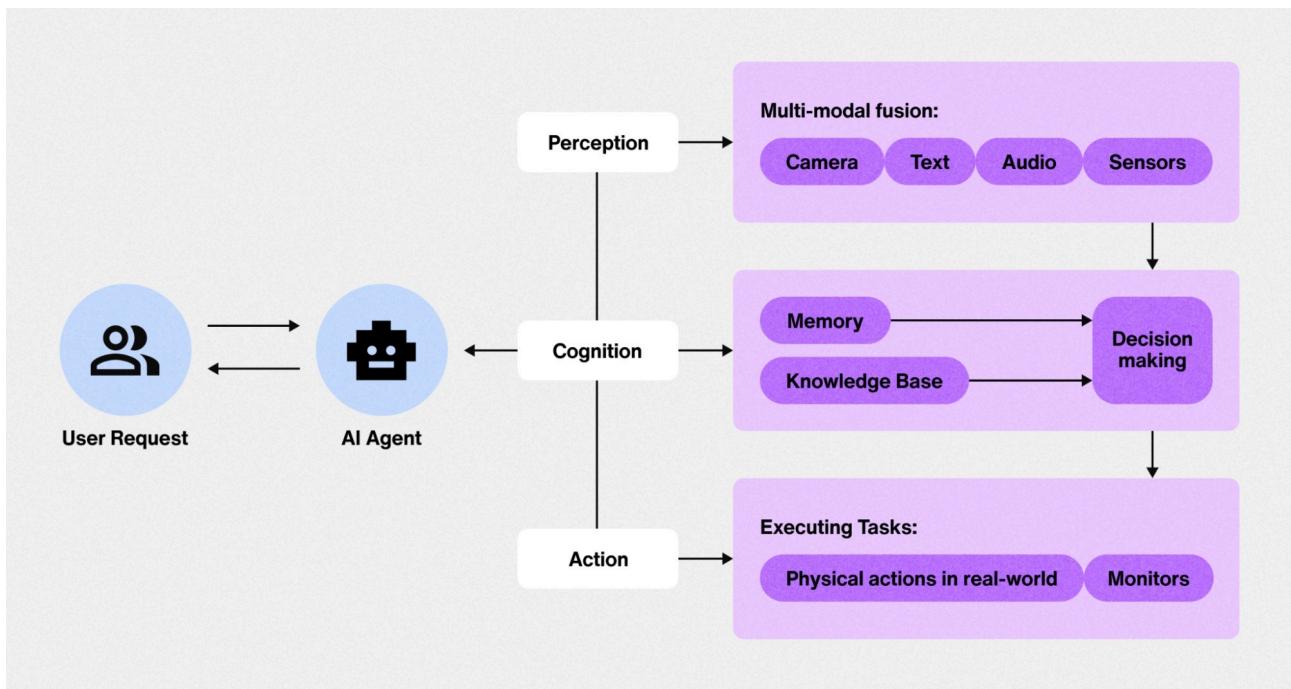
- 1. Capa de Modelo (Model Layer):** LLM o modelo base (GPT, Claude, etc.), capacidades de procesamiento de lenguaje natural, razonamiento y generación de texto.
- 2. Capa de Memoria (Memory Layer):** Memoria de corto plazo (contexto de conversación), memoria de largo plazo (base de datos vectorial), memoria episódica (experiencias pasadas).
- 3. Capa de Herramientas (Tools Layer):** APIs externas, bases de datos, navegadores web, sistemas de archivos.
- 4. Capa de Orquestación (Orchestration Layer):** Coordinación de componentes, flujo de control, gestión de estado.

What are AI Agents?

Architecture



4. Flujo de Procesamiento en Agentes Cognitivos



El flujo típico de procesamiento en un agente cognitivo moderno incluye:

1. **Percepción:** Recepción de entrada del usuario o entorno
2. **Cognición:** Procesamiento por el modelo de lenguaje
3. **Memoria:** Consulta y actualización de sistemas de memoria

4. **Razonamiento:** Planificación de acciones
 5. **Acción:** Ejecución de tareas mediante herramientas
 6. **Retroalimentación:** Evaluación de resultados
-

5. CoALA en Detalle: Arquitectura con Memoria Simbólica

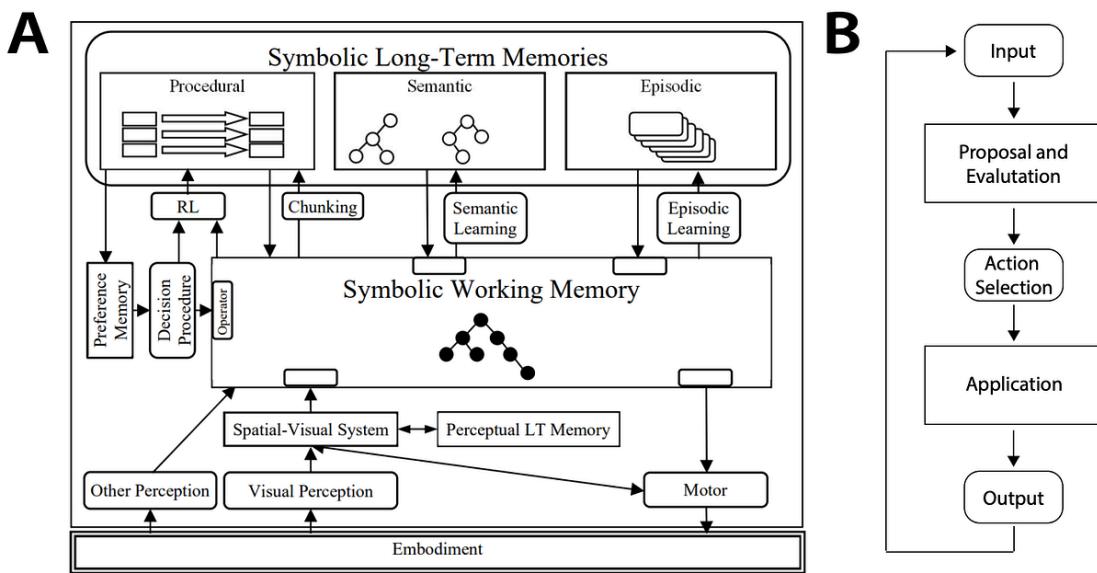
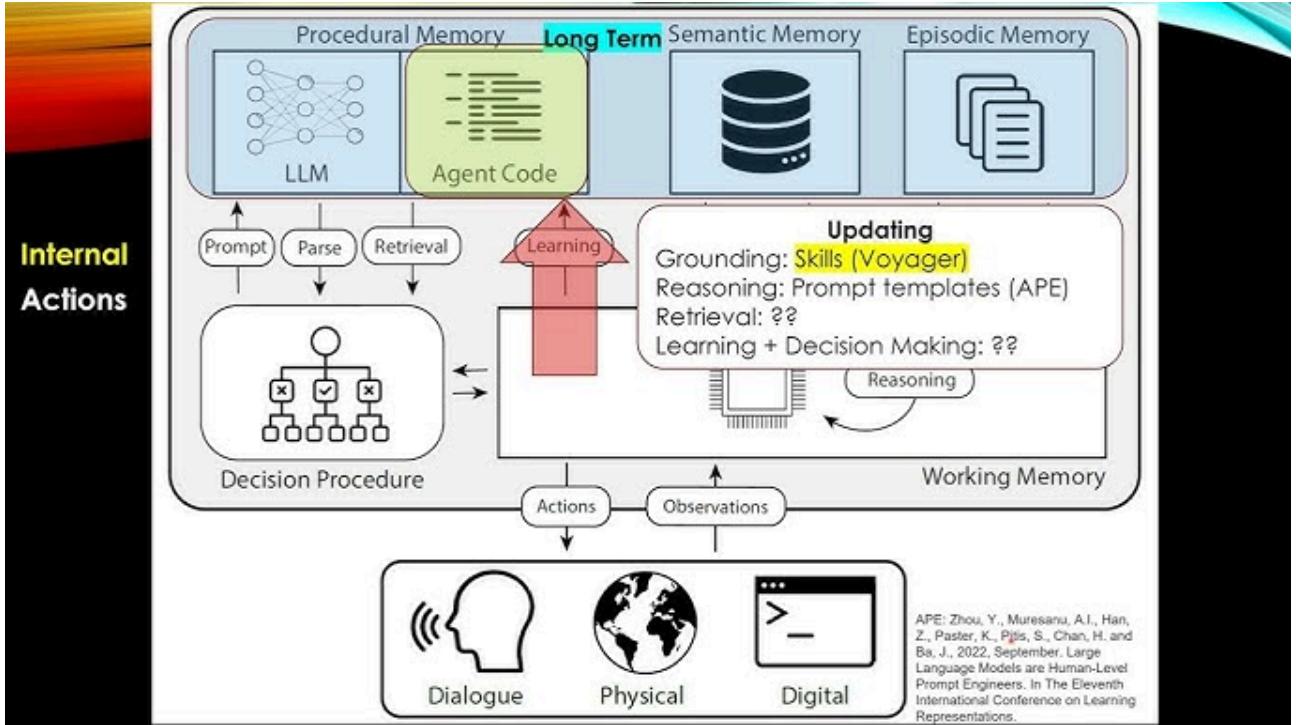


Figure 2: Cognitive architectures augment a production system with sensory groundings, long-term memory, and a decision procedure for selecting actions. **A:** The Soar architecture, reproduced with permission from Laird (2022). **B:** Soar's decision procedure uses productions to select and implement actions. These actions may be *internal* (such as modifying the agent's memory) or *external* (such as a motor command).

La arquitectura CoALA incluye componentes especializados:

- A) Memoria Simbólica de Largo Plazo:** Almacenamiento persistente de conocimiento, representación estructurada, recuperación basada en relevancia.
- B) Memoria de Trabajo:** Contexto activo de la tarea, información temporal, estado del agente.
- C) Módulo de Decisión:** Selección de acciones, planificación de secuencias, evaluación de opciones.
- D) Módulo de Acción:** Ejecución de comandos, interacción con entorno, uso de herramientas.



6. Análisis Comparativo: CAELION-Manus vs. CoALA

Aspecto	CAELION-Manus	CoALA
Enfoque Principal	Gobernanza y control de régimen cognitivo	Razonamiento y grounding de lenguaje
Arquitectura	Híbrida (módulos supervisores + protocolos)	Modular (memoria + acción + decisión)
Memoria	WABUN + ARC-01 (trazabilidad inmutable)	Memoria simbólica de largo plazo + working memory
Toma de Decisiones	Consenso distribuido de 5 módulos	Proceso de decisión generalizado
Espacio de Acción	Protocolos CMD, ARC, COM	Acciones sobre memoria y entorno
Métricas	Ω (coherencia), V (costo), E (eficiencia)	No especificadas
Objetivo	Estabilizar régimen de interacción	Capacidades de lenguaje general
Paradigma	Control de eventos discretos	Arquitectura cognitiva clásica

7. Posicionamiento de CAELION en el Campo

CAELION-Manus se posiciona como una **arquitectura de gobernanza coognitiva**, complementaria a frameworks como CoALA.

Fortalezas Únicas de CAELION

- 1. Trazabilidad Completa:** Cada acción queda registrada en Bitácoras Operativas inmutables
- 2. Consenso Distribuido:** Decisiones validadas por múltiples módulos supervisores
- 3. Medición de Dinámica:** Métricas cuantificables de coherencia y estabilidad
- 4. Control de Régimen:** Enfoque en estabilizar la interacción, no solo el modelo

Complementariedad con CoALA

- **CoALA** proporciona capacidades de razonamiento y grounding
 - **CAELION** proporciona gobernanza y control de estabilidad
 - **Integración potencial:** Un agente CoALA podría operar bajo gobernanza CAELION
-

8. Tendencias Identificadas en el Campo

1. **Modularidad:** Arquitecturas cada vez más modulares y componibles
 2. **Memoria Persistente:** Sistemas de memoria de largo plazo cada vez más sofisticados
 3. **Herramientas Externas:** Mayor integración con APIs y servicios externos
 4. **Razonamiento Multi-Paso:** Capacidades de planificación y razonamiento complejo
 5. **Evaluación y Métricas:** Creciente énfasis en medición de capacidades
-

9. Direcciones Futuras

Basándose en los hallazgos de esta investigación, se identifican las siguientes direcciones:

1. **Integración de Frameworks:** Combinar capacidades de razonamiento (CoALA) con gobernanza (CAELION)
2. **Métricas Estandarizadas:** Desarrollar métricas comunes para evaluar arquitecturas cognitivas
3. **Arquitecturas de Cuatro Capas:** Completar la integración de percepción, memoria, protocolos y actuación
4. **Control de Régimen:** Mayor énfasis en estabilizar la interacción, no solo mejorar el modelo
5. **Investigación Instrumental:** Más estudios empíricos que midan dinámica cognitiva en agentes

10. Conclusiones

1. **CoALA es el framework de referencia** para arquitecturas cognitivas de agentes de lenguaje (497 citas)
 2. **Las arquitecturas modernas convergen** hacia diseños modulares con memoria, razonamiento y acción bien definidos
 3. **CAELION-Manus ocupa un nicho único** como arquitectura de gobernanza coognitiva, complementaria a frameworks de razonamiento
 4. **El campo está madurando** hacia estándares de modularidad, trazabilidad y medición
 5. **La próxima frontera** es la integración de capacidades de razonamiento con mecanismos de gobernanza y control de régimen
-

Referencias

1. Sumers, T. R., Yao, S., Narasimhan, K., & Griffiths, T. L. (2024). Cognitive Architectures for Language Agents. *Transactions on Machine Learning Research (TMLR)*. arXiv:2309.02427
 2. Quiq. (2025). What is Cognitive Architecture? <https://uiq.com/blog/what-is-cognitive-architecture/>
 3. Smythos. (2025). Cognitive Agent Architectures: Revolutionizing AI.
 4. Cognee.ai. (2025). Cognitive Architectures for Language Agents (CoALA): Explained.
 5. IBM. (2025). ¿Qué es la arquitectura agéntica?
-

Registro de Métricas:

- Coherencia (Ω): 1 (investigación alineada con objetivo de la DOS-03)
- Costo (V): 0 (sin acciones correctivas)

- Eficiencia (E): Investigación completada exitosamente