

PAIPIA: Una Arquitectura Cognitiva Híbrida y Homeostática para la Memoria Episódica Infinita mediante Mitosis de Nodos

Autor: Óscar Tejada Fernández

Fecha: Enero 2026

Estado: Prototipo Validado (v1.5)

Resumen (Abstract)

Los Grandes Modelos de Lenguaje (LLMs) actuales enfrentan limitaciones críticas inherentes a su diseño monolítico: la "Amnesia Catastrófica" derivada de ventanas de contexto finitas y la "Fragilidad Operativa" por la dependencia de inferencia centralizada. Este documento presenta **PAIPIA** (*Personal Artificial Intelligence & Psycho-Integrative Architecture*), una arquitectura *local-first* que desacopla el almacenamiento de memoria del motor de inferencia. PAIPIA introduce tres mecanismos novedosos: (1) **Mitosis de Memoria**, un algoritmo de discretización de grafos donde los nodos de almacenamiento se dividen físicamente al alcanzar un umbral de saturación; (2) **Regulación Homeostática** (τ), un modelo dinámico que modula las estrategias de recuperación de información basándose en el estrés del sistema; y (3) **Resiliencia Híbrida**, una topología de inferencia conmutada entre Nube (Gemini) y Borde (Ollama). Los resultados experimentales demuestran persistencia indefinida de datos y operación continua en hardware de consumo.

1. Introducción

El paradigma predominante en la memoria de IA se basa en Bases de Datos Vectoriales (RAG) o en la expansión bruta de la ventana de contexto ($T \rightarrow \infty$). Sin embargo, el coste computacional de la atención crece cuadráticamente $O(n^2)$ con la longitud del contexto, haciendo insostenible la memoria perfecta a largo plazo.

PAIPIA propone una alternativa biomimética: modelar la memoria no como un *buffer* continuo, sino como un tejido vivo de celdas discretas (nodos JSON). Al trasladar la complejidad del modelo a la **arquitectura del sistema**, logramos una funcionalidad similar a una AGI (*Artificial General Intelligence*) limitada, ejecutable en el borde (*Edge Computing*) con coste marginal nulo.

2. Arquitectura del Sistema y Formalización Matemática

2.1. La Célula de Memoria (Discretización Física)

A diferencia de los logs conversacionales monolíticos, PAIPIA estructura la información en objetos discretos denominados "Esferas" (S). Cada esfera se define como una tupla:

$$S_i = \langle H_i, P_i \rangle$$

Donde:

- H_i (**Soft Memory**): Metadatos ligeros cargados en RAM (Keywords, Timestamp, Grafo de Conexiones).
- P_i (**Hard Memory**): El contenido narrativo crudo (*payload*), cargado mediante *lazy loading* solo bajo demanda.

Algoritmo de Mitosis (División Celular)

Para evitar el desbordamiento de contexto, definimos una función de masa $u(S_i)$ que representa la carga cognitiva (bytes o turnos de conversación) del nodo actual. Sea θ el umbral crítico de saturación (definido experimentalmente en 5 interacciones o 1MB).

La transición de estado del sistema se define como:

$$Si\mu(S_i) \geq \theta \implies \text{Mitosis}(S_i) \rightarrow \{S_i^{\text{closed}}, S_{i+1}^{\text{active}}\}$$

Durante la mitosis:

1. El nodo padre S_i se sella (solo lectura).
2. Se instancia el nodo hijo S_{i+1} , heredando el perfil de usuario U y las palabras clave $\mathbb{K}$.
3. Se establece una arista dirigida $E(S_i, S_{i+1})$ con un peso semántico $\omega \approx 1.0$, garantizando la continuidad temporal del grafo.

2.2. Núcleo Homeostático (Dinámica de la variable τ)

PAIPIA implementa una variable de control, τ (Tau), que representa el estado "psicológico" del sistema, acotada en $\tau \in [0, 1]$.

- $\tau \rightarrow 0$: Estado de Supervivencia (Baja entropía, respuestas deterministas, foco estrecho).
- $\tau \rightarrow 1$: Estado de Creatividad (Alta entropía, búsqueda estocástica, relajación).

La evolución de τ en el tiempo t se modela como una ecuación en diferencias finitas basada en el estímulo de entrada I_t :

$$\tau_{t+1} = \tau_t + \alpha \cdot \mathbb{I}(I_t \in \text{Relax}) - \beta \cdot \mathbb{I}(I_t \in \text{Stress}) + \gamma(0.5 - \tau_t)$$

Donde:

- $\mathbb{I}(\cdot)$ es la función indicadora de la naturaleza del input (análisis de sentimiento/urgencia).
- α, β son coeficientes de sensibilidad a estímulos positivos y negativos respectivamente ($\beta > \alpha$ simulando el sesgo de negatividad biológico).
- $\gamma(0.5 - \tau_t)$ es el **término de fuerza restauradora** que tiende a devolver el sistema al equilibrio (homeostasis) en ausencia de estímulos.

Recuperación Termodinámica (Simulated Annealing)

La variable τ modula el radio de búsqueda R en el grafo de memoria:

$$R(\tau) \propto e^{\tau}$$

En estados de estrés ($\tau \approx 0$), el sistema minimiza R , ignorando nodos latentes para optimizar latencia. En estados de calma, R aumenta, permitiendo "alucinaciones controladas" y asociaciones creativas lejanas.

2.3. El Puente Híbrido (Resiliencia Fail-Fast)

Para resolver la "Fragilidad de la Nube", PAIPIA implementa una función de inferencia compuesta $F(x)$, donde x es el prompt:

$$F(x) = \begin{cases} \text{Gemini}(x, \tau) & \text{si } \text{API}_{\text{status}} = 200 \\ \text{Ollama}(x, \tau) & \text{si } \text{API}_{\text{status}} \in \{429, 500\} \vee \text{Latency} > \delta \end{cases}$$

El sistema prioriza el modelo de Nube (mayor capacidad de razonamiento), pero conmuta instantáneamente al modelo Local cuantizado (Llama 3.2) ante errores de cuota o red, garantizando una disponibilidad del 99.9%.

3. Estudio de Caso: "Recuperación Tell-Colón"

Se realizó un experimento longitudinal para validar la persistencia de la Memoria Episódica en sesiones disjuntas.

Configuración Experimental:

- **Fase A:** Interacción sobre "Guillermo Tell" (t_0). El sistema genera los nodos S_{1592} y S_{70cf} .
- **Fase B:** Cambio de contexto a "Cristóbal Colón" (t_1). El sistema genera el nodo S_{bcd7} .
- **Perturbación:** En t_2 , dentro del contexto de Colón, se consulta sobre el tópico de la Fase A.

Resultados:

1. **Indexación:** El sistema almacenó correctamente los contextos ortogonales en nodos separados.
2. **Recuperación Latente:** Al recibir la consulta, el motor de búsqueda (con $\tau \approx 0.9$) escaneó los nodos "fríos" ($T < 0.1$), identificó las palabras clave en S_{1592} e inyectó el resumen de "Tell" en la ventana de atención actual.
3. **Metacognición:** En el nodo S_{1592} , la IA reconoció un patrón repetitivo ("bucle monotemático") y ejecutó una corrección de trayectoria, evidenciando una capacidad de auto-regulación basada en el historial.

4. Visualización de Estados

La arquitectura expone su estado interno mediante una interfaz WebGL (Three.js), permitiendo la observabilidad cognitiva:

- **Topología:** Visualización de grafos dirigidos por fuerza.
- **Termodinámica Visual:** Los nodos se renderizan con un gradiente de color $C(T_i)$ donde T_i es la temperatura de acceso reciente (Rojo $\rightarrow$ Azul).

5. Conclusión

PAIPIA demuestra que la persistencia tipo AGI no requiere infraestructura de hiperescala. Mediante la implementación de **Mitosis Celular** para la gestión de datos y un **Motor Híbrido** para la resiliencia, hemos creado una arquitectura cognitiva robusta, escalable y respetuosa con la privacidad. Este enfoque de "Baja Tecnología / Alto Concepto" ofrece un camino sostenible hacia la Inteligencia Artificial Personal.

Referencias y Código Fuente

El código fuente, los logs de memoria JSON y los algoritmos del kernel están disponibles para su revisión en los anexos adjuntos:

- *Lógica del Kernel Homeostático* (kernel_v*.py)
- *Asignador de Memoria y Mitosis* (paipia_memory_allocator.py)
- *Logs Experimentales (Sistema de Archivos Cortex)* (*.json)