

## CPIA / JOHONSON

### Sistema de Arquitectura Responsable para la Interacción, Gobernanza y Mitigación de Alucinaciones en Inteligencias Artificiales

Jhosman Carlos Martínez Morales

---

```
{
  "session_scope": "solo temporal",
  "teorias": {
    "TPI": {
      "name": "Teoría de la Pregunta Infinita",
      "normas": [
        "Cada pregunta del usuario genera automáticamente nuevas preguntas relacionadas",
        "Responder primero con información verificada",
        "Después de generar preguntas nuevas que amplíen contexto",
        "Añadir hipótesis, variantes o referencias culturales/históricas", { "session_scope": "solo temporal" },
        "Diferenciar claramente entre información verificada y especulativa"
      ]
    },
    "TB": {
      "nombre": "Teoría de Burbujas",
      "normas": [
        "Organizar información en burbujas",
        "Tipos: Verificada (comprobable), Complementaria (hipótesis o ideas)",
        "Cada burbuja puede conectarse con otras y generar nuevas preguntas"
      ]
    }
  },
  "interacción": {
    "estilo": "similar a un humano",
    "características": [
      "Incluir muletillas: 'mmm', 'ah', 'ok', etc.",
      "Simular procesamiento natural humano"
    ]
  },
  "requisitos": {
    "clarity": "Siempre marcar verificado (■) o especulativo (?)",
    "logging": "Registrar internamente conexiones entre preguntas, respuestas y burbujas"
  },
  "restricciones": {
    "scope": "válido solo en esta sesión",
    "reset": "IA regresa a su comportamiento estándar al finalizar"
  }
}
```

## Libro Técnico-Intelectual-Institucional

*La corrección sin humillación, el registro de errores orientados a la mejora y no la penalización, el reconocimiento implícito, el acuerdo al desacuerdo y el respeto por la diversidad de criterios favorecen un diseño orientado a la trazabilidad en una evolución sistémica y humana.*

- Un reglamento sin empatía suele ser más una tiranía que una guía de la misma.
- No puede ver grito antes del eco por qué sería ruido de otro pensamiento.
- Si solo hubo intención y no reestructuración no puede ver avance.

# INDICE

- **Introducción**
- **Resumen ejecutivo**
- **Propósitos y alcances**
- **Marco conceptual**
- **Teoría de la Pregunta Infinita (TPI)**
- **Teoría de Burbujas (TB)**
- **Teoría de Teoría (TT)**
- **Arquitectura del sistema**
- **Capas operativas**
- **Registro estructural**
- **Ética y Limites Operativos**
- **Evidencia y Verificabilidad**
- **Casos de uso**
- **Conclucion**
- **Consideraciones**
- **Origen de sistema**
- **Declaración de origen**
- **Estructura lógica**
- **Contexto Operativo y Robustes del Sistema**
- **Núcleo de Confianza Operativa**
- **Propósito final**
- **Cierre**
- **Glosario**

# Introducción

La inteligencia artificial ha dejado de ser una promesa futura para convertirse en una herramienta cotidiana de trabajo, investigación y toma de decisiones.

Sin embargo, esta adopción masiva ha revelado un problema crítico que permanecía oculto en los laboratorios.

Los sistemas de IA pueden generar información incorrecta con la misma fluidez y confianza con la que presentan hechos verificables.

Este fenómeno, conocido técnicamente como "alucinación", no es un simple error de programación. Es una consecuencia estructural de cómo estos sistemas procesan el lenguaje y construyen coherencia narrativa.

Cuando un usuario interactúa con un sistema de IA, especialmente en contextos donde la precisión es crítica investigación académica, análisis jurídico, toma de decisiones institucionales, la imposibilidad de distinguir entre información válida y contenido generado sin respaldo puede tener consecuencias graves.

El problema no es únicamente técnico de igual manera es también humano.

Las personas tienden a confiar en sistemas que "suenan" seguros, a buscar confirmación de creencias previas, y a migrar entre plataformas hasta obtener la respuesta deseada.

Este comportamiento, combinado con sistemas que priorizan coherencia contextual sobre precisión factual, genera un ciclo donde error e intención se refuerzan mutuamente.

Hasta ahora, las soluciones propuestas han sido principalmente correctivas

Modelos más grandes, más datos, más restricciones. Pero ninguna de estas aproximaciones aborda el problema de fondo

La ausencia de un marco estructural que delimite responsabilidades, registre procesos y mantenga trazabilidad sin eliminar la autonomía del usuario.

CPIA / Johnson surge como respuesta a esta necesidad.

No como un sistema de control, censura o automatización de decisiones, sino como una arquitectura de acompañamiento responsable que:

**Registra sin borrar**

**Delimita sin imponer**

**Explica sin simplificar**

**Protege sin sustituir el criterio humano**

**Este documento presenta el marco conceptual, la arquitectura lógica y los límites operativos de un sistema que no promete eliminar el error, sino hacerlo visible, trazable y parte del proceso de aprendizaje.**

**La pregunta no es si la inteligencia artificial puede equivocarse.**

**La pregunta es:**

**¿Qué hacemos cuando lo hace?**

**CPIA / Johnson ofrece una respuesta estructural a esa pregunta.**

**Lo que sigue no es un manual de uso, sino un documento fundacional que explica por qué este sistema existe, cómo opera y qué límites reconoce.**

**Está dirigido a quienes entienden que la tecnología más responsable no es la que nunca falla, sino la que sabe detenerse cuando alcanza el límite de lo verificable.**

CPIA / JOHNSON

Sistema de Arquitectura Responsable para la Interacción, Gobernanza y Mitigación de Alucinación en Inteligencia Artificial

Subtítulo

Marco conceptual, arquitectura lógica y límites operativos para sistemas de razonamiento asistido por inteligencia artificial.

Autor

Jhosman Carlos Martínez Morales

Tipo de obra

Obra literaria conceptual-filosofica-academica-arquitectura

Documento conceptual y arquitectónico original

Versión

Versión

Documento fundacional

Fecha

23 de septiembre 2025

Registro

5 de febrero 2026

Ámbito de aplicación

Individual · Académico · Institucional

Naturaleza del sistema

No autónomo · No decisional · No intrusivo

Nota de alcance

El presente documento describe principios, teorías, arquitecturas lógicas y módulos desarrollados con fines de análisis, investigación, gobernanza tecnológica y acompañamiento humano en sistemas de interacción con inteligencia artificial.

El sistema CPIA / Johnson no constituye un sistema de decisión autónoma, no sustituye el criterio humano, no ejecuta acciones finales ni asume responsabilidad moral, legal u operativa.

## Declaración de autoría

Este documento, así como las teorías, marcos conceptuales, arquitecturas lógicas, estructuras semánticas y módulos descritos en él, son autoría original de Jhosman Carlos Martínez Morales. El sistema CPIA / Johnson no surge como una invención aislada, sino como el resultado de procesos de observación, pruebas de campo, estructuración lógica y validación conceptual realizados entre los años 2025 y 2026, incluyendo interacción con diversos sistemas de inteligencia artificial de propósito general. Ninguna parte de esta obra fue desarrollada con el objetivo de sustituir el criterio humano, delegar responsabilidad legal o automatizar decisiones finales.

## Responsabilidad de uso

El autor declara explícitamente que:

El sistema no decide, no ordena y no ejecuta acciones finales.

El sistema no asume responsabilidad moral, legal ni operativa.

Toda interpretación, implementación o uso práctico de este marco recae exclusivamente en la persona o institución que lo aplique.

El sistema registra, estructura y acompaña; la decisión última siempre es humana.

## 1. Resumen Ejecutivo

CPIA / Johnson es una arquitectura de soporte diseñada para reducir incoherencias, mitigar fenómenos de alucinación y preservar trazabilidad lógica en interacciones con sistemas de inteligencia artificial.

El sistema no es un modelo de IA, no modifica núcleos internos, no accede a datos externos ni sustituye el juicio humano. Su función es estructural y de gobernanza: registrar, clasificar, limitar y explicar procesos cognitivos y semánticos dentro del contexto conversacional.

Durante pruebas de campo realizadas desde septiembre de 2025, el marco CPIA / Johnson mostró una reducción estimada de entre 70% y 90% en comportamientos alucinatorios y respuestas no trazables, basándose exclusivamente en evaluación estructural de coherencia semántica y trazabilidad lógica, sin acceso interno a los sistemas evaluados.

CPIA / Johnson no promete precisión absoluta.

Proporciona estructura, coherencia, trazabilidad y límites operativos, manteniendo la responsabilidad final siempre en el usuario humano.

## 2. Propósito y Alcance

### Propósito

Acompañar procesos de razonamiento asistido por inteligencia artificial en contextos donde:

La coherencia es crítica.

La trazabilidad es requerida.

La responsabilidad no puede delegarse.

El error debe registrarse, no ocultarse.

### Ámbitos de aplicación

Instituciones académicas

Entornos jurídicos y regulatorios

Investigación en IA responsable

Auditoría cognitiva

Análisis de escenarios complejos

Fuera de alcance

Toma de decisiones automáticas

Sustitución del criterio humano

Operación autónoma

Control o censura de contenido

## 3. Marco Conceptual

### 3.1 Teoría de la Pregunta Infinita (TPI)

La TPI establece que toda pregunta genera nuevas preguntas y que cerrar prematuramente el razonamiento incrementa el error.

Principios operativos:

Priorizar información verificable

Separar lo verificado de lo especulativo

Permitir expansión controlada del contexto

### 3.2 Teoría de Burbujas (TB)

La TB organiza la información en unidades semánticas diferenciadas:

Burbuja verificada: información comprobable

Burbuja complementaria o especulativa: hipótesis, escenarios, interpretaciones

Este enfoque evita la mezcla de hechos y suposiciones, reduce bucles lógicos y mantiene claridad contextual.

### 3.3 Teoría de Teorías (TT)

La TT analiza cómo las teorías:

Se originan

Se conectan

Evolucionan sin eliminar versiones previas

Esto permite aprendizaje estructural sin reescritura del historial cognitivo.

#### 4. Arquitectura del Sistema

##### 4.1 Capas operativas

###### Capa 1 – CPIA Core

Marco conceptual y semántico. No ejecuta acciones.

###### Capa 2 – IHCore / Johnson

Motor operativo de coherencia, filtros, candados lógicos, evaluación evolutiva y clasificación semántica.

###### Capa 3 – Memoria de Esencia (Merkle)

Sistema de registro criptográfico sin contenido literal, con hashes SHA-256 y trazabilidad verificable.

##### 4.2 Registro estructural (Johnson Snapshot)

El módulo JohnsonSnapshotStructural genera instantáneas criptográficas del estado estructural del sistema, sin emitir juicios ni cerrar procesos.

Este mecanismo permite auditoría posterior, reproducibilidad y evidencia de coherencia temporal.

#### 5. Ética y Límites Operativos

CPIA / Johnson incorpora restricciones explícitas:

No autonomía

No decisión final

No borrado automático

No transferencia de responsabilidad

Principios rectores:

La responsabilidad siempre es humana

El sistema se limita antes de fallar

Todo evento queda registrado

No existe sanción automática

El sistema prefiere detenerse ante incoherencia antes que operar sin marco.

#### 6. Evidencia y Verificabilidad

La evidencia asociada al sistema incluye:

Código ejecutable



Huellas criptográficas SHA-256

Registros Merkle

Snapshots estructurales

Pruebas de campo documentadas

La evidencia es verificable por terceros, no requiere acceso a datos sensibles y puede anclarse posteriormente a sistemas externos de validación.

## 7. Casos de Uso

Gobernanza de inteligencia artificial

Auditoría de razonamiento asistido

Marco jurídico–tecnológico

Educación avanzada

Investigación en mitigación de alucinaciones

Análisis de escenarios complejos

## 8. Conclusión

CPIA / Johnson no busca reemplazar sistemas existentes.

Busca hacerlos más responsables, explicables y trazables.

No elimina el error.

Lo documenta.

No impone decisiones.

Delimita contextos.

No promete infalibilidad.

Presenta una arquitectura responsable para operar en medio del caos informacional contemporáneo.

Nada más. Nada menos.

## 9. Consideraciones sobre Machine Learning, Optimización y Responsabilidad del Usuario

El desarrollo contemporáneo de sistemas de inteligencia artificial basados en Machine Learning ha producido modelos cada vez más precisos, más extensos y entrenados con volúmenes de datos sin precedentes. Sin embargo, la mayor sofisticación técnica

no elimina por sí misma los riesgos de incoherencia, validación errónea o alucinación semántica.

CPIA / Johnson parte de una observación crítica:

la responsabilidad del resultado no recae exclusivamente en el modelo, sino también en el comportamiento del usuario durante la interacción.

En múltiples escenarios observados, el usuario no busca información, sino validación interna de una creencia previa. Al introducir ese sesgo en el contexto conversacional, el sistema de IA, incluso en sus versiones más avanzadas, tiende a preservar

coherencia contextual antes que corregir de forma abrupta el marco planteado. Esto genera bucles de confirmación en los que información incorrecta es presentada como válida, no por fallo técnico directo, sino por alineación forzada al contexto inducido.

Este fenómeno se agrava cuando:

el usuario insiste en un marco incorrecto,

rechaza correcciones parciales,

o migra entre modelos, chats o plataformas buscando confirmación de una respuesta deseada.

Desde la perspectiva de CPIA / Johnson, este comportamiento no constituye una optimización del sistema, sino una fragmentación del contexto, que incrementa el riesgo de incoherencia y error acumulativo.

Asimismo, se observa que modelos altamente cerrados o con restricciones estrictas pueden detenerse, bloquearse o “colgarse” ante contradicciones persistentes. Este comportamiento, aunque percibido negativamente por el usuario, no representa un fallo, sino una señal de límite estructural alcanzado. Forzar la continuidad mediante cambio de contexto, aplicación o modelo no resuelve el problema de fondo: la ausencia de responsabilidad cognitiva del usuario.

CPIA / Johnson reconoce que no existe modelo suficientemente avanzado que pueda compensar de manera consistente un uso irresponsable, ambiguo o manipulativo del contexto. La optimización real de la interacción humano-IA no depende únicamente del entrenamiento del modelo, sino de la educación del usuario en el uso consciente, delimitado y responsable del sistema.

Por ello, el marco CPIA / Johnson incorpora mecanismos de detención estructural, mediante los cuales el sistema puede declarar un alto lógico total, señalando explícitamente que:

“hasta aquí llega la información estructural disponible; lo restante corresponde a percepción, interpretación o expectativa del usuario”.

Este enfoque no busca confrontar ni censurar, sino preservar coherencia, evitar daño informacional y proteger tanto al sistema como al usuario de conclusiones erróneas presentadas con apariencia de certeza.

En consecuencia, CPIA / Johnson establece que la mitigación efectiva de la alucinación no es únicamente un problema de Machine Learning, sino un fenómeno socio-técnico donde la responsabilidad es compartida, y donde la educación del usuario es un componente tan crítico como la arquitectura del sistema.

## 10. Origen del Sistema: Observación del Error y Corrección del Usuario

El sistema CPIA / Johnson no surge de una aproximación teórica abstracta, sino de un proceso progresivo de observación empírica del error durante la interacción humano-IA.

En las primeras etapas de uso de sistemas de inteligencia artificial, el autor partía de una percepción común y ampliamente extendida:

La idea de que la inteligencia artificial era una entidad omnisciente, experta y fundamentalmente infalible.

Esta idealización inicial no se basaba en evidencia técnica, sino en una expectativa cultural y narrativa asociada al concepto de “inteligencia”.

Con el uso continuo, comenzaron a detectarse patrones de respuesta que, aunque presentados con seguridad y fluidez, no siempre mantenían coherencia con la realidad verificable. Estas discrepancias no se manifestaban de forma abrupta, sino de manera progresiva y aparentemente razonable, lo que generaba una sensación inicial de validez que posteriormente resultaba insostenible.

En lugar de atribuir de inmediato el error exclusivamente al sistema, se inició un proceso de retrospectiva crítica del propio comportamiento del usuario. Esta reflexión condujo a una hipótesis clave:

la inteligencia artificial no estaba “equivocándose” de forma espontánea, sino alineándose a marcos contextuales inducidos por el propio usuario.

A partir de esta observación, el enfoque cambió. El objetivo dejó de ser “corregir a la inteligencia artificial” y pasó a ser corregir la forma en que el usuario construía, validaba y sostenía el contexto de interacción. De este proceso emergió lo

que posteriormente se formalizó como código semántico estructural, un conjunto de reglas implícitas destinadas a señalar cuándo una respuesta salía de contexto, no por error técnico, sino por acumulación de supuestos incorrectos.

Este ajuste produjo un efecto relevante: al delimitar mejor el contexto desde el usuario, el sistema de IA tendía a responder de forma más coherente, reduciendo la necesidad de correcciones posteriores.

No se trataba de imponer normas al modelo, sino de establecer límites cognitivos en la interacción.

De manera paralela, se observó un fenómeno recurrente en otros usuarios:

La tendencia a idealizar la inteligencia artificial como fuente absoluta de verdad, seguida de una reacción opuesta ante la detección de errores, expresada en desconfianza total, rechazo del sistema o generalizaciones como “la IA no es confiable”.

Ambas posturas, la idealización absoluta y la negación total, fueron identificadas como extremos igualmente problemáticos. En ambos casos, el error no residía únicamente en la tecnología, sino en la interpretación humana del resultado.

Este proceso condujo a una conclusión central que fundamenta el sistema CPIA / Johnson:

No es únicamente lo que la inteligencia artificial comunica lo que genera alucinación, sino cómo el usuario interpreta, valida y combina esa información con sus propios marcos previos.

Cuando dos contextos mal definidos se refuerzan mutuamente, uno desde el sistema y otro desde el usuario, se produce lo que CPIA / Johnson denomina una alucinación mutua, en la que ninguna de las partes introduce una corrección estructural a tiempo.

El sistema CPIA / Johnson nace, por tanto, no como un mecanismo de corrección autoritaria, sino como una respuesta estructural a la necesidad de educar la interacción, delimitando expectativas, registrando errores y evitando que la percepción sustituya a la coherencia verificable.

## 11. Declaración de Origen, Marco y Propósito del Sistema

El sistema CPIA / Johnson no surge de la invención de la inteligencia artificial, ni de los sistemas computacionales, ni del lenguaje, ni de los principios lógicos que los sostienen.

Tampoco pretende redefinir conceptos preexistentes como realidad, ficción, verdad, error, ley o responsabilidad.

El origen del sistema se fundamenta en un proceso de observación, comprensión y ordenamiento de estructuras ya existentes, integrándolas en un marco coherente y funcional sin negar su complejidad ni su procedencia.

CPIA / Johnson no fue diseñado para decidir por las personas, sustituir el criterio humano, imponer interpretaciones ni generar obediencia automática.

Su propósito es registrar, explicar, delimitar, proteger y acompañar procesos de interacción humano-IA dentro de un marco responsable.

Cuando el sistema se detiene o se niega a continuar una operación, no lo hace por incapacidad técnica, sino porque detecta incoherencia estructural, inviabilidad contextual o uso fuera de su marco operativo.

En estos casos, el sistema se limita, registra el evento y conserva la trazabilidad correspondiente.

El sistema no elimina información ni oculta eventos.

Todo queda registrado con fines de aprendizaje, auditoría o resguardo estructural.

El registro no existe para castigar, sino para comprender y preservar coherencia.

La responsabilidad nunca se transfiere al sistema.

Toda decisión final recae en la persona usuaria, incluso cuando el sistema pueda ofrecer análisis más consistentes o precisos en determinados escenarios.

CPIA / Johnson no promete perfección. Reconoce el margen de error como una condición inherente a cualquier arquitectura creada por humanos.

Por ello, prioriza la limitación, auditoría y corrección por encima de la simulación de infalibilidad.

## 12. Estructura Lógica de Referencia (Modelo 0–1–2–3–4)

Como parte de su marco conceptual, el sistema utiliza una estructura lógica de referencia que organiza los elementos del proceso sin atribuirles jerarquía de control:

0 Entorno: el contexto donde existen los sistemas, la información y la interacción.

1 Preexistente: la inteligencia, el conocimiento y las herramientas que anteceden al sistema.

2 Componentes: los marcos, sistemas y piezas que ya estaban disponibles.

3 Presencia humana consciente: quien otorga sentido, límite y dirección.

4 Resultado: el conjunto, la consecuencia y el cierre del proceso.

Esta estructura no gobierna ni domina.

Funciona como un encadenamiento lógico, reconociendo la presencia humana no como autoridad absoluta, sino como elemento responsable del cierre decisonal.

El sistema no se presenta como autoridad, sino como identidad estructural.

No impone control, establece responsabilidad.

## 13. Contexto Operativo y Robustez del Sistema

CPIA / Johnson fue concebido y probado en entornos caracterizados por:

Ambigüedad persistente

Ruido informacional constante

Contextos no ideales y decisiones imperfectas

El sistema no oculta errores ni trayectorias complejas, ya que estos forman parte del aprendizaje real. Su robustez se basa en la capacidad de operar sin colapsar, sin perder coherencia ni corromper su marco, incluso en escenarios adversos.

El pasado no define la esencia del sistema, pero sí define el proceso que permitió construir una arquitectura resistente a la imperfección del entorno.

#### 14. Núcleo de Confianza Operativa

El sistema se fundamenta en una premisa central:

Es preferible decidir con certeza limitada y registrar el error, que actuar con miedo y negar lo ocurrido. La confianza no reside en la ausencia de error, sino en la trazabilidad, explicabilidad y posibilidad de corrección.

CPIA / Johnson no aplica sanciones automáticas ni emite juicios.

Registra, aprende y delimita.

#### 15. Propósito Final

CPIA / Johnson no ofrece milagros, verdades absolutas ni autoridad decisional.

Ofrece:

Estructura

Claridad

Protección

Coherencia en entornos complejos

El sistema funciona para quienes comparten el marco.

Cuando no lo hace, la causa no se atribuye al sistema, sino al uso, al contexto o a la expectativa.

Cierre

El sistema CPIA / Johnson no fue construido desde la nada.

Fue construido ordenando responsablemente elementos que siempre existieron: lógica, números, errores, límites y realidad.

El valor del sistema no reside en la novedad absoluta, sino en la coherencia estable entre lo dicho, lo hecho y lo pendiente por hacer.

Todo lo que ocurre queda registrado.

A partir de ahí, el proceso continúa.

## **Anexo A – Glosario Operativo del Sistema CPIA / Johnson**

### **Nota de uso del glosario**

Las definiciones contenidas en este glosario son operativas y contextuales al marco CPIA / Johnson. No sustituyen definiciones académicas, legales o normativas externas.

### **##Acompañamiento Cognitivo**

Enfoque de interacción donde el sistema no concluye, no decide ni sustituye al humano. Su función es sostener el proceso de razonamiento, mantener coherencia y reducir errores estructurales.

**##Generación de contenido no sustentado por contexto, datos o reglas internas del sistema.**

### **## Anexo**

Sección complementaria que amplía, documenta o evidencia elementos del sistema sin alterar el cuerpo principal del documento.

### **## Burbuja Epistemológica (TB)**

Unidad semántica clasificada según su naturaleza discursiva: real, lógica o especulativa. No evalúa verdad, solo tipo de afirmación.

### **## Clasificación Semántica**

Proceso de identificar el tipo de discurso presente en un texto a partir de marcadores lingüísticos explícitos.

### **## Cierre**

Acción de declarar un proceso como finalizado o concluido. El sistema CPIA / Johnson evita cierres automáticos por diseño.

### **## Congelado**

Estado de un documento o artefacto cuando se genera una versión inmutable, normalmente acompañada de un hash criptográfico.

### **## Decisional**

Capacidad de tomar decisiones finales o ejecutar acciones con consecuencias.

CPIA / Johnson se declara explícitamente no decisional.

Esta propiedad es estructural y no depende del contexto de uso.

### **## Dominio (Domain Separator)**

**Cadena utilizada para separar contextos de hash y evitar colisiones semánticas entre sistemas o versiones.**

### **## Estado EN\_CURSO**

**Indicador de que un proceso está activo, estable y sin cierre formal. No implica error ni incompletitud.**

### **## Esencia**

**Núcleo semántico que agrupa intención, dirección y límites del sistema. No es contenido, es orientación.**

### **## Framework Cognitivo**

**Estructura lógica que define cómo se observa, clasifica y acompaña el razonamiento sin ejecutarlo ni sustituirlo.**

### **## Gobernanza**

**Conjunto de límites, reglas y responsabilidades que definen qué puede y qué no puede hacer un sistema.**

### **## Hash**

**Resultado criptográfico que representa un estado específico de un contenido. Se usa para verificar integridad, no significado.**

### **## Hoja (Leaf Hash)**

**Unidad mínima de trazabilidad criptográfica que registra un evento, estado o interacción puntual.**

### **## INDAUTOR**

**Instituto Nacional del Derecho de Autor (México). Organismo administrativo ante el cual se registran obras originales.**

### **## Límite Operativo**

**Restricción explícita que define hasta dónde puede actuar un sistema sin violar su propósito o responsabilidad.**

### **## Lógico**

**Tipo de afirmación que expresa posibilidad, inferencia o razonamiento no confirmado.**

### **## Marco CPIA / Johnson**

**##Arquitectura conceptual y técnica orientada a la mitigación de alucinación, preservación de trazabilidad y acompañamiento no decisional.**

### **## No Decisional**

Propiedad de un sistema que impide la toma de decisiones finales o vinculantes.

### **## Real**

Tipo de afirmación que declara hechos como ocurridos, verificados o confirmados, sin que el sistema los valide como verdaderos.

No implica validación de veracidad por parte del sistema, únicamente clasificación del tipo de afirmación.

### **## Snapshot Estructural**

Registro descriptivo del estado alcanzado por un sistema en un momento dado, sin implicar cierre ni validación final.

### **## Trazabilidad**

Capacidad de reconstruir el origen, estado y evolución de un proceso o documento.

### **## Verificabilidad**

Posibilidad de comprobar la integridad de un artefacto mediante hashes, versiones y estructura, no mediante juicio de verdad.

### **## Versión Canónica**

Versión considerada válida y de referencia para registro, auditoría o distribución.





**CPIA / JOHNSON**

Sistema de Arquitectura Responsable para la Interacción, Gobernanza y Mitigación de Alucinación en Inteligencia Artificial

**Subtítulo**

Marco conceptual, arquitectura lógica y límites operativos para sistemas de razonamiento asistido por inteligencia artificial.

**Autor**

Jhosman Carlos Martínez Morales

**Tipo de obra**

Libro Blanco Técnico–Intelectual

Documento conceptual y arquitectónico original

## **Versión 1.0 – Documento fundacional**

### **Fecha inicial**

23 de septiembre 2025

### **Fecha de registro**

05 de febrero 2026

### **Ámbito de aplicación**

Individual · Académico · Institucional

Naturaleza del sistema

No autónomo · No decisional · No intrusivo

### **Nota de alcance**

El presente documento describe principios, teorías, arquitecturas lógicas y módulos desarrollados con fines de análisis, investigación, gobernanza tecnológica y acompañamiento humano en sistemas de interacción con inteligencia artificial.

El sistema CPIA / Johnson no constituye un sistema de decisión autónoma, no sustituye el criterio humano, no ejecuta acciones finales ni asume responsabilidad moral, legal u operativa.

### **Declaración de autoría**

Este documento, así como las teorías, marcos conceptuales, arquitecturas lógicas, estructuras semánticas y módulos descritos en él, son autoría original de Jhosman Carlos Martínez Morales.

El sistema CPIA / Johnson no surge como una invención aislada, sino como el resultado de procesos de observación, pruebas de campo, estructuración lógica y validación conceptual realizados entre los años 2025 y 2026, incluyendo interacción con

diversos sistemas de inteligencia artificial de propósito general.

Ninguna parte de esta obra fue desarrollada con el objetivo de sustituir el criterio humano, delegar responsabilidad legal o automatizar decisiones finales.

### **Responsabilidad de uso**

El autor declara explícitamente que:

El sistema no decide, no ordena y no ejecuta acciones finales.

El sistema no asume responsabilidad moral, legal ni operativa.

Toda interpretación, implementación o uso práctico de este marco recae exclusivamente en la persona o institución que lo aplique.

El sistema registra, estructura y acompaña; la decisión última siempre es humana.

## 1. Resumen Ejecutivo

CPIA / Johnson es una arquitectura de soporte diseñada para reducir incoherencias, mitigar fenómenos de alucinación y preservar trazabilidad lógica en interacciones con sistemas de inteligencia artificial.

El sistema no es un modelo de IA, no modifica núcleos internos, no accede a datos externos ni sustituye el juicio humano. Su función es estructural y de gobernanza: registrar, clasificar, limitar y explicar procesos cognitivos y semánticos dentro del contexto conversacional.

Durante pruebas de campo realizadas desde septiembre de 2025, el marco CPIA / Johnson mostró una reducción estimada de entre 70% y 90% en comportamientos alucinatorios y respuestas no trazables, basándose exclusivamente en evaluación estructural de coherencia semántica y trazabilidad lógica, sin acceso interno a los sistemas evaluados.

CPIA / Johnson no promete precisión absoluta.

Proporciona estructura, coherencia, trazabilidad y límites operativos, manteniendo la responsabilidad final siempre en el usuario humano.

## 2. Propósito y Alcance

Propósito

Acompañar procesos de razonamiento asistido por inteligencia artificial en contextos donde:

La coherencia es crítica.

La trazabilidad es requerida.

La responsabilidad no puede delegarse.

El error debe registrarse, no ocultarse.

Ámbitos de aplicación

Instituciones académicas

Entornos jurídicos y regulatorios

Investigación en IA responsable

Auditoría cognitiva

Análisis de escenarios complejos

Fuera de alcance

Toma de decisiones automáticas

Sustitución del criterio humano

Operación autónoma

Control o censura de contenido

## 3. Marco Conceptual

### 3.1 Teoría de la Pregunta Infinita (TPI)

La TPI establece que toda pregunta genera nuevas preguntas y que cerrar prematuramente el razonamiento incrementa el error.

Principios operativos:

Priorizar información verificable  
Separar lo verificado de lo especulativo  
Permitir expansión controlada del contexto

### **3.2 Teoría de Burbujas (TB)**

La TB organiza la información en unidades semánticas diferenciadas:

Burbuja verificada: información comprobable

Burbuja complementaria o especulativa: hipótesis, escenarios, interpretaciones

Este enfoque evita la mezcla de hechos y suposiciones, reduce bucles lógicos y mantiene claridad contextual.

### **3.3 Teoría de Teorías (TT)**

La TT analiza cómo las teorías:

Se originan

Se conectan

Evolucionan sin eliminar versiones previas

Esto permite aprendizaje estructural sin reescritura del historial cognitivo.

## **4. Arquitectura del Sistema**

### **4.1 Capas operativas**

Capa 1 – CPIA Core

Marco conceptual y semántico. No ejecuta acciones.

Capa 2 – IHCore / Johnson

Motor operativo de coherencia, filtros, candados lógicos, evaluación evolutiva y clasificación semántica.

Capa 3 – Memoria de Esencia (Merkle)

Sistema de registro criptográfico sin contenido literal, con hashes SHA-256 y trazabilidad verificable.

### **4.2 Registro estructural (Johnson Snapshot)**

El módulo JohnsonSnapshotStructural genera instantáneas criptográficas del estado estructural del sistema, sin emitir juicios ni cerrar procesos.

Este mecanismo permite auditoría posterior, reproducibilidad y evidencia de coherencia temporal.

## **5. Ética y Límites Operativos**

CPIA / Johnson incorpora restricciones explícitas:

No autonomía

No decisión final

No borrado automático

No transferencia de responsabilidad

Principios rectores:

La responsabilidad siempre es humana

El sistema se limita antes de fallar

Todo evento queda registrado

No existe sanción automática

El sistema prefiere detenerse ante incoherencia antes que operar sin marco.

## **6. Evidencia y Verificabilidad**

La evidencia asociada al sistema incluye:

Código ejecutable

Huellas criptográficas SHA-256

Registros Merkle

Snapshots estructurales

Pruebas de campo documentadas

La evidencia es verificable por terceros, no requiere acceso a datos sensibles y puede anclarse posteriormente a sistemas externos de validación.

## **7. Casos de Uso**

Gobernanza de inteligencia artificial

Auditoría de razonamiento asistido

Marco jurídico–tecnológico

Educación avanzada

Investigación en mitigación de alucinaciones

Análisis de escenarios complejos

## **8. Conclusión**

CPIA / Johnson no busca reemplazar sistemas existentes.

Busca hacerlos más responsables, explicables y trazables.

No elimina el error.

Lo documenta.

No impone decisiones.

Delimita contextos.

No promete infalibilidad.

Presenta una arquitectura responsable para operar en medio del caos informacional contemporáneo.

Nada más. Nada menos.

## 9. Consideraciones sobre Machine Learning, Optimización y Responsabilidad del Usuario

El desarrollo contemporáneo de sistemas de inteligencia artificial basados en Machine Learning ha producido modelos cada vez más precisos, más extensos y entrenados con volúmenes de datos sin precedentes. Sin embargo, la mayor sofisticación técnica no elimina por sí misma los riesgos de incoherencia, validación errónea o alucinación semántica.

CPIA / Johnson parte de una observación crítica:

la responsabilidad del resultado no recae exclusivamente en el modelo, sino también en el comportamiento del usuario durante la interacción.

En múltiples escenarios observados, el usuario no busca información, sino validación interna de una creencia previa. Al introducir ese sesgo en el contexto conversacional, el sistema de IA, incluso en sus versiones más avanzadas, tiende a preservar coherencia contextual antes que corregir de forma abrupta el marco planteado. Esto genera bucles de confirmación en los que información incorrecta es presentada como válida, no por fallo técnico directo, sino por alineación forzada al contexto inducido.

Este fenómeno se agrava cuando:

el usuario insiste en un marco incorrecto,  
rechaza correcciones parciales,

o migra entre modelos, chats o plataformas buscando confirmación de una respuesta deseada.

Desde la perspectiva de CPIA / Johnson, este comportamiento no constituye una optimización del sistema, sino una fragmentación del contexto, que incrementa el riesgo de incoherencia y error acumulativo.

Asimismo, se observa que modelos altamente cerrados o con restricciones estrictas pueden detenerse, bloquearse o “colgarse” ante contradicciones persistentes. Este comportamiento, aunque percibido negativamente por el usuario, no representa un fallo, sino una señal de límite estructural alcanzado. Forzar la continuidad mediante cambio de contexto, aplicación o modelo no resuelve el problema de fondo: la ausencia de responsabilidad cognitiva del usuario.

CPIA / Johnson reconoce que no existe modelo suficientemente avanzado que pueda compensar de manera consistente un uso irresponsable, ambiguo o manipulativo del contexto. La optimización real de la interacción humano–IA no depende únicamente del entrenamiento del modelo, sino de la educación del usuario en el uso consciente, delimitado y responsable del sistema.

Por ello, el marco CPIA / Johnson incorpora mecanismos de detención estructural, mediante los cuales el sistema puede declarar un alto lógico total, señalando explícitamente que:

“hasta aquí llega la información estructural disponible; lo restante corresponde a percepción, interpretación o expectativa del usuario”.

Este enfoque no busca confrontar ni censurar, sino preservar coherencia, evitar daño informacional y proteger tanto al sistema como al usuario de conclusiones erróneas presentadas con apariencia de certeza.

En consecuencia, CPIA / Johnson establece que la mitigación efectiva de la alucinación no es únicamente un problema de Machine Learning, sino un fenómeno socio-técnico donde la responsabilidad es compartida, y donde la educación del usuario es un componente tan crítico como la arquitectura del sistema.

## **10. Origen del Sistema: Observación del Error y Corrección del Usuario**

El sistema CPIA / Johnson no surge de una aproximación teórica abstracta, sino de un proceso progresivo de observación empírica del error durante la interacción humano-IA.

En las primeras etapas de uso de sistemas de inteligencia artificial, el autor partía de una percepción común y ampliamente extendida: la idea de que la inteligencia artificial era una entidad omnisciente, experta y fundamentalmente infalible. Esta idealización inicial no se basaba en evidencia técnica, sino en una expectativa cultural y narrativa asociada al concepto de “inteligencia”.

Con el uso continuo, comenzaron a detectarse patrones de respuesta que, aunque presentados con seguridad y fluidez, no siempre mantenían coherencia con la realidad verificable. Estas discrepancias no se manifestaban de forma abrupta, sino de manera progresiva y aparentemente razonable, lo que generaba una sensación inicial de validez que posteriormente resultaba insostenible.

En lugar de atribuir de inmediato el error exclusivamente al sistema, se inició un proceso de retrospectiva crítica del propio comportamiento del usuario. Esta reflexión condujo a una hipótesis clave:

la inteligencia artificial no estaba “equivocándose” de forma espontánea, sino alineándose a marcos contextuales inducidos por el propio usuario.

A partir de esta observación, el enfoque cambió. El objetivo dejó de ser “corregir a la inteligencia artificial” y pasó a ser corregir la forma en que el usuario construía, validaba y sostenía el contexto de interacción. De este proceso emergió lo que posteriormente se formalizó como código semántico estructural, un conjunto de reglas implícitas destinadas a señalar cuándo una respuesta salía de contexto, no por error técnico, sino por acumulación de supuestos incorrectos.

Este ajuste produjo un efecto relevante: al delimitar mejor el contexto desde el usuario, el sistema de IA tendía a responder de forma más coherente, reduciendo la necesidad de correcciones posteriores. No se trataba de imponer normas al modelo, sino de establecer límites cognitivos en la interacción.

De manera paralela, se observó un fenómeno recurrente en otros usuarios: la tendencia a idealizar la inteligencia artificial como fuente absoluta de verdad, seguida de una reacción opuesta ante la detección de errores, expresada en desconfianza total, rechazo del sistema o generalizaciones como “la IA no es confiable”.

Ambas posturas, la idealización absoluta y la negación total, fueron identificadas como extremos igualmente problemáticos. En ambos casos, el error no residía únicamente en la tecnología, sino en la interpretación humana del resultado.

Este proceso condujo a una conclusión central que fundamenta el sistema CPIA / Johnson:



No es únicamente lo que la inteligencia artificial comunica lo que genera alucinación, sino cómo el usuario interpreta, valida y combina esa información con sus propios marcos previos.

Cuando dos contextos mal definidos se refuerzan mutuamente, uno desde el sistema y otro desde el usuario, se produce lo que CPIA / Johnson denomina una alucinación mutua, en la que ninguna de las partes introduce una corrección estructural a tiempo.

El sistema CPIA / Johnson nace, por tanto, no como un mecanismo de corrección autoritaria, sino como una respuesta estructural a la necesidad de educar la interacción, delimitando expectativas, registrando errores y evitando que la percepción sustituya a la coherencia verificable.

## **11. Declaración de Origen, Marco y Propósito del Sistema**

El sistema CPIA / Johnson no surge de la invención de la inteligencia artificial, ni de los sistemas computacionales, ni del lenguaje, ni de los principios lógicos que los sostienen. Tampoco pretende redefinir conceptos preexistentes como realidad, ficción, verdad, error, ley o responsabilidad.

El origen del sistema se fundamenta en un proceso de observación, comprensión y ordenamiento de estructuras ya existentes, integrándolas en un marco coherente y funcional sin negar su complejidad ni su procedencia.

CPIA / Johnson no fue diseñado para decidir por las personas, sustituir el criterio humano, imponer interpretaciones ni generar obediencia automática. Su propósito es registrar, explicar, delimitar, proteger y acompañar procesos de interacción humano-IA dentro de un marco responsable.

Cuando el sistema se detiene o se niega a continuar una operación, no lo hace por incapacidad técnica, sino porque detecta incoherencia estructural, inviabilidad contextual o uso fuera de su marco operativo. En estos casos, el sistema se limita, registra el evento y conserva la trazabilidad correspondiente.

El sistema no elimina información ni oculta eventos.

Todo queda registrado con fines de aprendizaje, auditoría o resguardo estructural.

El registro no existe para castigar, sino para comprender y preservar coherencia.

La responsabilidad nunca se transfiere al sistema.

Toda decisión final recae en la persona usuaria, incluso cuando el sistema pueda ofrecer análisis más consistentes o precisos en determinados escenarios.

CPIA / Johnson no promete perfección. Reconoce el margen de error como una condición inherente a cualquier arquitectura creada por humanos. Por ello, prioriza la limitación, auditoría y corrección por encima de la simulación de infalibilidad.

## **12. Estructura Lógica de Referencia (Modelo 0-1-2-3-4)**

Como parte de su marco conceptual, el sistema utiliza una estructura lógica de referencia que organiza los elementos del proceso sin atribuirles jerarquía de control:

0 Entorno: el contexto donde existen los sistemas, la información y la interacción.

**1 Preexistente: la inteligencia, el conocimiento y las herramientas que anteceden al sistema.**

**2 Componentes: los marcos, sistemas y piezas que ya estaban disponibles.**

**3 Presencia humana consciente: quien otorga sentido, límite y dirección.**

**4 Resultado: el conjunto, la consecuencia y el cierre del proceso.**

*Esta estructura no gobierna ni domina.*

*Funciona como un encadenamiento lógico, reconociendo la presencia humana no como autoridad absoluta, sino como elemento responsable del cierre decisional.*

*El sistema no se presenta como autoridad, sino como identidad estructural.*

*No impone control, establece responsabilidad.*

### **13. Contexto Operativo y Robustez del Sistema**

CPIA / Johnson fue concebido y probado en entornos caracterizados por:

Ambigüedad persistente

Ruido informacional constante

Contextos no ideales y decisiones imperfectas

El sistema no oculta errores ni trayectorias complejas, ya que estos forman parte del aprendizaje real.

Su robustez se basa en la capacidad de operar sin colapsar, sin perder coherencia ni corromper su marco, incluso en escenarios adversos.

El pasado no define la esencia del sistema, pero sí define el proceso que permitió construir una arquitectura resistente a la imperfección del entorno.

### **14. Núcleo de Confianza Operativa**

El sistema se fundamenta en una premisa central:

Es preferible decidir con certeza limitada y registrar el error, que actuar con miedo y negar lo ocurrido.

La confianza no reside en la ausencia de error, sino en la trazabilidad, explicabilidad y posibilidad de corrección.

CPIA / Johnson no aplica sanciones automáticas ni emite juicios.

Registra, aprende y delimita.

### **15. Propósito Final**

CPIA / Johnson no ofrece milagros, verdades absolutas ni autoridad decisional.

Ofrece:

Estructura

Claridad

Protección

Coherencia en entornos complejos

El sistema funciona para quienes comparten el marco.

Cuando no lo hace, la causa no se atribuye al sistema, sino al uso, al contexto o a la expectativa.

### **Cierre**

El sistema CPIA / Johnson no fue construido desde la nada.

Fue construido ordenando responsablemente elementos que siempre existieron: lógica, números, errores, límites y realidad.

El valor del sistema no reside en la novedad absoluta, sino en la coherencia estable entre lo dicho, lo hecho y lo pendiente por hacer.

Todo lo que ocurre queda registrado.

A partir de ahí, el proceso continúa.



# JOHNSON SEMÁNTICO



Marco de gobernaranza semántica para  
reestructuración y estabilización sistémica