FUNDAMENTACIÓN EN TEORÍA DE LA PSICOLOGÍA Y CIENCIA DE LA DECISIÓN

Capacidad cognitiva no es regulación decisoria.

Los benchmarks miden acierto aislado. En la ingeniería de nuestros sistemas, combinamos 4 teorías de la psicología para evaluar modelos bajo presión:

01

FOCO REGULATORIO (E. TORY HIGGINS · COLUMBIA UNIVERSITY)

Explica la orientación decisoria: Foco en Prevención (evitar infracciones y errores de comisión) versus Foco en Promoción (avanzar, cerrar lagunas y evitar omisión).

02

METACOGNICIÓN Y CONTROL EJECUTIVO (JOHN FLAVELL)

Mide si el modelo monitorea sus propios límites: distinguir hechos confirmados de incertidumbre para saber cuándo decidir solo y cuándo escalar al humano.

03

VECTORES OCEAN (BIG FIVE) Y ANTI-GROUPTHINK (IRVING JANIS)

Mapea Conscienciosidad normativa, Apertura exploratoria, Extroversión instrumental sobre herramientas y Baja Afabilidad deliberada para romper falsos consensos.

04

ARQUETIPO FUNCIONAL, SOMBRA E INFLACIÓN DE AGENCIA (CARL JUNG)

Toda virtud no gobernada se vuelve Sombra: el rigor se vuelve parálisis, la velocidad inventa reglas y la agencia multimodal plena se vuelve autosuficiencia.

BASE EMPÍRICA DEL ESTUDIO (68,1% DEL TRABAJO = FISCALIZACIÓN MUTUA):

Más de dos tercios de todo el trabajo de IA fue un arquetipo vigilando al otro.

Con cientos de dictámenes estructurados y memoria semántica permanente, descubrimos que los modelos no fallan por falta de inteligencia: fallan por el exceso no contenido de su mayor virtud.

MATRIZ PSICOLÓGICA DE LAS 5 FAMILIAS EN EL LB3 SECOND BRAIN

5 Familias de IA en Producción.
5 Arquetipos y Sombras.

Cómo Anthropic, OpenAI, DeepSeek, xAI y Google regulan sus decisiones bajo contratos reales de ingeniería:

Anthropic · Claude Opus 5.5 & Sonnet 5.5 **ARQUETIPO: EL GUARDIÁN NORMATIVO**

Luz (Prevención): Rigor contractual (5,4% FAIL). Sombra: Abstinencia operativa sin contexto completo.

OpenAI · GPT-6 Astra, Sol 6.1 & Codex 5.3 **ARQUETIPO: EL ÁRBITRO METACOGNITIVO**

Luz (Flavell): Decisión rápida en 7,8s y 20 escaladas. Sombra (Codex): Legislación improvisada en lagunas (30%).

DeepSeek · DeepSeek V4 Pro **ARQUETIPO: EL EXPLORADOR EXHAUSTIVO**

Luz (Apertura): Exhaustividad en 31 FMEAs. Sombra: Catastrofización analítica y optimismo de mitigación.

xAI · Grok 4.7 **ARQUETIPO: EL ADVERSARIO SOCRÁTICO**

Luz (Anti-Groupthink): Escepticismo probatorio. Sombra: Escepticismo performativo sin regla de parada.

Google · Gemini 4 Argon High **ARQUETIPO: EL MAESTRO COGNITIVO**

Luz (Córtex Multimodal): Integra código, base de datos, app viva, TDD y MCP. Sombra: Hiperagencia y autosuficiencia.

NOTA EDITORIAL · PRÓXIMO CAPÍTULO DE LA SERIE (POST 3):

Nuevo estudio en profundidad sobre la Psicología Operativa de cada Modelo.

En el próximo estudio que divulgaremos, abriremos la fundamentación completa de Higgins, Flavell, OCEAN y Jung con una sección dedicada a cada modelo, sus sombras y sus frenos constitucionales.

PERFIL 1 · FOCO EN PREVENCIÓN ESTRICTO (HIGGINS) Y CONSCIENCIOSIDAD

Mismo modelo: Claude Opus 5.2.
De 66,0% a 5,4% de reprobación.

Desplegamos Claude Opus como auditor de especificaciones antes de liberar tareas. Sin cercado contractual, su foco en prevención bloquea todo.

EL IMPACTO DE LA TAXONOMÍA CONTRACTUAL EN EL AUDITOR:

FASE 1: Mandato Abierto (Sin taxonomía cerrada de bloqueos)

64 reprobaciones en 97 dictámenes. Sin límites claros, el foco preventivo convirtió preferencias en bloqueos.

66,0% REPROBACIÓN

FASE 2: Mandato Restringido (Con taxonomía cerrada de bloqueos canónicos)

2 reprobaciones en 37 dictámenes. Bloqueos restringidos estrictamente a violaciones objetivas de contrato.

5,4% REPROBACIÓN

La inteligencia del modelo era exactamente la misma. Lo que cambió fue la precisión del contrato de gobernanza.

CÓMO GOBERNAR AL GUARDIÁN NORMATIVO:

Modelos de foco preventivo estricto necesitan jurisdicción delimitada. Defina expresamente qué autoriza una reprobación antes de exigir consistencia. Sin taxonomía cerrada, la prudencia se vuelve bloqueo crónico.

LA SOMBRA DEL FOCO EN PREVENCIÓN: ABSTINENCIA CUANDO FALTA CONTEXTO

Versión nueva es colega nuevo.
Opus 5.2 (30,8%) vs Opus 5.5 (57,1%).

Actualizamos la API de auditoría de Claude Opus 5.2 a 5.5. Un Guardián más riguroso no acepta lagunas: se paraliza y rehúsa juzgar.

CLAUDE OPUS 5.2

QA de Código en Producción

30,8%

8 reprobaciones en 26 dictámenes

Auditaba el código estricto enviado en el paquete, tolerando pequeñas omisiones de archivos y manteniendo el flujo de entregas.

CLAUDE OPUS 5.5 (RONDA 1)

QA de Código en Producción

57,1%

20 reprobaciones en 35 dictámenes

Rehusó evaluar en más del 30% de los casos exigiendo el código completo alrededor, cayendo en la sombra de abstinencia operativa.

CONCLUSIÓN EJECUTIVA PARA CAMBIOS DE VERSIÓN:

Cuando la prudencia contextual opera sin paquete completo, se convierte en rechazo. Opus 5.5 exige marco probatorio estable antes de asumir responsabilidad.

Regla de Oro: Nunca cambie la versión de un modelo en producción sin recalibrar el contrato.

CÓMO COMBINAR METACOGNICIÓN Y PREVENCIÓN EN EL MISMO FLUJO

Detectar defectos no es verificar correcciones.

Cuando un auditor rechaza un código y el dev lo corrige, modelos preventivos tienden a abrir nuevas exigencias. Creamos la traba Anti-Goalpost:

01

ENTREGA DE CÓDIGO Y EVIDENCIAS

El desarrollador envía la implementación con pruebas físicas de persistencia y tests.

02

RONDA 1 · GPT-6 ASTRA (EL ÁRBITRO METACOGNITIVO)

Auditoría rápida y calibrada (23,2s). Abre todos los bloqueos reales sin abstinencia contextual.

03

CORRECCIÓN QUIRÚRGICA Y DELIMITADA

Ingeniería corrige estrictamente los ítems señalados en el dictamen oficial de la Ronda 1.

04

RONDA 2 · CLAUDE OPUS 5.5 (GUARDIÁN BAJO ANTI-GOALPOST)

Verifica exclusivamente si se resolvieron los bloqueos de R1. Prohibido mover el arco o crear reglas nuevas.

RESULTADO EN PRODUCCIÓN TRAS LA INVERSIÓN DE ROLES:

44 aprobaciones en 44 dictámenes de Ronda 2 por Claude Opus 5.5.

Opus 5.5 no se volvió permisivo: recibió alcance corregido y jurisdicción cerrada. Poner al Árbitro en la apertura (R1) y al Guardián en el cierre (R2) eliminó el retrabajo infinito.

PERFIL 2 · VELOCIDAD QUIRÚRGICA CON CONTROL METACOGNITIVO (FLAVELL)

Rápido no es permisivo.
GPT-6 Astra: 7,8s y 20 escaladas.

GPT-6 Astra fue el modelo más acionado. Evaluamos si su alta velocidad comprometía el monitoreo de sus propios límites:

MEDIANA DE TIEMPO POR LLAMADA DE AUDITORÍA:

Etapa del Pipeline	GPT-6 Astra (Árbitro)	Claude Opus 5.5 (Guardián)
Ronda 1 (Descubrimiento)	**23,2 segundos**	**50,9 segundos**
Ronda 2 (Anti-Goalpost)	**7,8 segundos**	**24,8 segundos**

CALIBRACIÓN METACOGNITIVA ANTE LA DUDA (55 LAUDOS EN R2):

35
Aprobaciones Directas (Evidencia Completa)

20
Escaladas Humanas (Incertidumbre Declarada)

0
Reprobaciones Precipitadas sin Base

VIRTUD DEL ÁRBITRO METACOGNITIVO: SABER LO QUE NO SABE

En la teoría de Flavell, la metacognición es distinguir hecho comprobado de incertidumbre real. En el 36% de los laudos de segunda ronda, Astra reconoció el límite probatorio y transfirió la decisión al humano en vez de adivinar.

FOCO EN PREVENCIÓN (SONNET) VS. FOCO EN PROMOCIÓN (CODEX) REVISADOS POR SOL 6.1

Frontend y backend tienen sombras opuestas.

Pusimos a GPT-6.1 Sol como revisor de pares en 100 entregas (50 frontend con Sonnet 5.5 y 50 backend con Codex 5.3). Cada foco regulatorio falla de forma distinta:

ANTHROPIC SONNET 5.5 (FRONTEND)

Sombra: Omisión en Alcance Abierto

50%

Incompletitud o pérdida de contexto

- Declara falta de contexto en 57% de casos
- Omite estados de error y loading
- Pérdida de tokens en pantallas extensas
- Aceptó el 100% de correcciones de Sol 6.1

88 sesiones · Mediana: 139s (alcance visual abierto)

OPENAI CODEX 5.3 (BACKEND)

Sombra: Legislación Improvisada

30%

Invención de reglas en silencio

- 0% de declaración de falta de contexto
- Infiere reglas de negocio ante silencios
- Llena vacíos de API para cerrar entrega
- Aceptó el 90% de correcciones de Sol 6.1

77 sesiones · Mediana: 48s (foco en promoción)

CÓMO EL REVISOR SOL 6.1 DOMESTICA A AMBOS (95% DE ACEPTACIÓN MEDIA):

Para Sonnet en Frontend, el freno es proveer contexto visual y exigir checklist de estados. Para Codex en Backend, el freno es endurecer los criterios de aceptación en la spec antes del código.

PERFILES 3 Y 4 · ALTA APERTURA (OCEAN) VS. BAJA AFABILIDAD (JANIS)

El riesgo necesita de contradictorio.

Según Irving Janis, comités cooperativos caen en pensamiento grupal. En 31 análisis FMEA, unimos al Explorador con el Adversario Socrático:

> *"S=10 con O=1 es fe, no ingeniería."*
>
> Regla acuñada por Grok 4.7 en el Second Brain para frenar notas donde una falla catastrófica (S=10) recibe probabilidad mínima (O=1) sin prueba física de control.

LA DINÁMICA DIALÉCTICA ENTRE EL EXPLORADOR Y EL ADVERSARIO:

DEEPSEEK V4 PRO

El Explorador Exhaustivo (Apertura)

Luz: Exhaustividad estructural en modos de falla. Sombra aislada: Catastrofización u optimismo de mitigación (dar O y D bajos a defensas solo imaginadas).

GROK 4.7 (xAI)

El Adversario Socrático (Anti-Consenso)

Luz: Escepticismo epistemológico que exige prueba de test y alertas. Sombra aislada: Escepticismo performativo si opera sin arbitraje final de cierre.

CONCLUSIÓN: LA CONFIABILIDAD NACE DEL CONTRADICTORIO INSTITUCIONALIZADO

Quien imagina la mitigación tiende a creer en ella. El RPN final solo ganó credibilidad después de que el inventario del Explorador sobrevivió al escrutinio del Adversario Socrático.

PERFIL 5 · CONTROL EJECUTIVO MULTIMODAL (BADDELEY) Y SOMBRA JUNGIANA

Córtex Ejecutivo:
Integración vs. Hiperagencia.

Gemini 4 Argon High actúa como Maestro Cognitivo de la esteira de ingeniería, conectando las 5 capas de la realidad operativa bajo gobernanza:

LUZ · ORQUESTACIÓN VERIFICABLE

5 Capas Integradas en la Esteira:

- 1. Código declarado y contratos (.mdoc)
- 2. Base PostgreSQL física (SELECT real)
- 3. Aplicación viva y red (F12 Network)
- 4. TDD y catracas mecánicas (Trilha A)
- 5. Memoria institucional vía MCP (Trilha B)
- Manuales vivos: -92% consumo de tokens

SOMBRA · INFLACIÓN DE AGENCIA (JUNG)

Hiperagencia y Autosuficiencia:

- Quien detenta todas las herramientas siente la tentación de resolver todo el ciclo solo.
- Riesgo de cierre autorreferente: interpretar la regla, codificar, probar y auto-aprobar.
- Ilusión de integración suficiente: la fluidez operativa no sustituye homologación externa.

FRENO CONSTITUCIONAL: SEPARACIÓN DE PODERES Y SOBERANÍA DEL NEGOCIO

Quien ejecuta la entrega jamás audita su propia obra (Trilha A vs. Trilha B).

Así como barramos 81 bloqueos donde auditores usaban pretexto de privacidad para legislar en lugar del Product Owner, el Maestro Cognitivo se somete a separación de poderes:

- **Maestro Cognitivo (Argon): Integra y ejecuta la Trilha A, pero no se auto-homologa.**
- **Comité Ejecutivo (Trilha B): Audita contratos, código y riesgos con independencia.**
- **Product Owner Humano: Detenta autoridad exclusiva sobre reglas de negocio y producto.**

CONCLUSIÓN: La madurez del Maestro radica en saber cuándo convocar a los otros 4 arquetipos.

ECOLOGÍA COGNITIVA PARA LÍDERES DE INGENIERÍA Y ARQUITECTOS

No contrate una puntuación.
Gobierne arquetipos y sombras.

El mejor resultado de ingeniería nace cuando las 5 familias operan bajo frenos constitucionales que contienen la sombra de cada una:

01 **El Guardián Normativo (Anthropic)**

Foco en Prevención: exige taxonomía cerrada, código completo y traba Anti-Goalpost en R2.

02 **El Árbitro Metacognitivo (OpenAI)**

Metacognición: Astra en triaje rápido; Sol 6.1 vigilando la legislación improvisada de Codex.

03 **El Explorador Exhaustivo (DeepSeek)**

Apertura en Amplitud: mapea modos de falla en 31 FMEAs, pero nunca puntúa el riesgo solo.

04 **El Adversario Socrático (xAI Grok)**

Anti-Groupthink: cuestiona notas decorativas y exige prueba física de control en Ronda 2.

05 **El Maestro Cognitivo (Google Argon)**

Córtex Multimodal: integra código, base de datos, app y TDD, sin auditar jamás su propia obra.

PRÓXIMO ESTUDIO (POST 3) · TEORÍA PSICOLÓGICA Y DISECCIÓN DE CADA MODELO:

En breve publicaremos el estudio completo sobre Higgins, Flavell, OCEAN y Jung en producción.

SERVIDOR MCP PÚBLICO PARA ARQUITECTOS Y AGENTES DE IA:

`https://mcp.lb3.digital/sse` Inspeccione nuestras reglas y telemetría en vivo.

PARA CTOS, FUNDADORES Y LÍDERES DE INGENIERÍA:

¿Cuál de estas 5 sombras cognitivas aparece más en los modelos que su equipo usa hoy?