

Phronesis

*Panorama de una obra en curso — una actitud, cuatro
herramientas, once normas y una cadena que hace de ellas
código.*

AUTOR Kaspar Brönnimann

FECHA 14 de agosto de 2026

VERSIÓN v1

A D V E R T E N C I A

Para qué sirve este documento

En una capacitación, una herramienta muestra a una jefa de proyecto una formulación — pulida, plausible, con apariencia de verificada. Ella la adopta. Un año más tarde se descubre que la formulación contiene una referencia que no existe.

Ni la herramienta ni la jefa de proyecto mintieron. Ambas hicieron lo mejor que pudieron. Aun así, el daño es real.

Este documento ordena una obra que quiere evitar precisamente tales casos — no mediante mejores instrucciones, sino mediante otra actitud fundamental: la afirmación asumida permanece reservada. Las herramientas pueden calcular e interpretar; no deben juzgar. Lo que sigue es el orden a partir del cual esta actitud se hace software.

Seis niveles sostienen la obra, una cadena hace de ellos código, un taller los produce todos. Quien quiera saber cómo surgió este orden y qué exige, lo hallará en las páginas siguientes.

Í N D I C E

— Advertencia — Para qué sirve este documento

I Prólogo — Lo que aquí se reúne

II La actitud — Phronesis

III Las cuatro herramientas

IV Las instanciaciones

V El taller — Travesía: el banco de trabajo

VI Las once normas de la suite

VII La cadena — Cómo las normas se hacen código

VIII La infraestructura — Capabilities, skills, código

IX El flanco — Ciencia y presencia pública

X El estado actual

XI Cierre — La única figura de pensamiento

Prólogo — Lo que aquí se reúne

Este documento no es una recopilación. Es un mapa.

Este documento no es una recopilación. Es un mapa.

Este documento ordena una obra que, a través de cuatro herramientas, sus instanciaciones concretas, once normas y dos publicaciones científicas, sostiene la misma actitud. No es una antología, sino un mapa. Quien tiene el mapa puede ubicar cada artefacto — y ver por qué ninguno de ellos se sostiene solo.

Lo que aquí se encuentra ha surgido en muchos lugares, con muchos motivos y en distintas lenguas — alemán, inglés, código, documento, diagrama, diálogo. Un observador externo podría tener la impresión de que aquí alguien trabaja en muchas cosas inconexas. La impresión engaña. Hay una sola cosa; la multiplicidad es su despliegue.

Esta única cosa se llama Phronesis: la sabiduría práctica, la capacidad de sopesar, juzgar y asumir la responsabilidad en el caso particular concreto. Es una actitud, no una herramienta. Pero sostiene una familia de herramientas que hace esa actitud aplicable en la práctica cotidiana — y sostiene un taller que no solo construye estas herramientas, sino que las fundamenta de manera verificable.

El conjunto tiene seis niveles y, junto a ellos, un flanco:

- Nivel 1 — Actitud. Phronesis como manifiesto y principio rector.
- Nivel 2 — Herramientas. Aletheia, Mnemosyne, Methodos, Kairon.
- Nivel 3 — Instanciaciones. Productos concretos: ALETHEIA Radar, capa de seudonimización, kennora, HERMES PIA, KAIRON Dashboard, DigiTwin.
- Nivel 4 — Normas. Once cuerpos de reglas que se sitúan por encima de todas las instanciaciones.

- Nivel 5 — Cadena. La cadena de desarrollo de agentes que convierte normas en código.
- Nivel 6 — Infraestructura. Capabilities, skills, código — la base técnica compartida.
- Flanco — Ciencia y presencia pública. Dos publicaciones científicas y un repositorio silencioso.

La actitud — Phronesis

Nivel 1. Lo que no puede ser software, porque primero ha de ser sostenido.

Phronesis es una familia abierta de herramientas para decisiones conscientes en un mundo moldeado por la inteligencia artificial. Pero la primera y más importante frase reza: no es un software. Es una actitud, a la que sirven tres condiciones: debe ser expresada, debe ser compartida, y debe hallar su forma en herramientas — de lo contrario queda en opinión.

La referencia es Aristóteles. En la *Ética a Nicómaco* (libro VI) distingue tres modos del conocer: *episteme* — saber acerca de lo que es necesariamente así —, *techne* — el saber hacer, producir una obra según una regla — y *phronesis* — la sabiduría práctica, hacer lo justo en el caso concreto. Solo la *phronesis* conoce el caso particular, el contexto, la persona que responde. Solo ella está insoslayablemente ligada a un ser humano que actúa y puede asumir lo que se sigue de su acción.

Esto es más que una nota filológica. Es la distinción fundamental de la que pende toda la construcción. Pues la inteligencia artificial es en el mejor de los casos *episteme* (sabe lo que en muchos casos es así) y *techne* (puede lo que muchos hacen a menudo). Que un sistema artificial pueda un día juzgar frónéticamente es una pregunta abierta — quizá incluso una que nunca podrá responderse de manera concluyente. Este documento no resuelve esa pregunta. Resuelve otra, independiente de ella: en una institución humana, un ser humano debe asumir lo que aquella hace. El derecho y el deber de juzgar, en un asunto con consecuencias para otros, están ligados a personas que pueden rendir cuentas — no a sistemas cuya capacidad de rendir cuentas queda sin aclarar. Esta cuestión de imputación se mantiene incluso si la cuestión ontológica llegara un día a decidirse de otro modo.

Phronesis: la sabiduría práctica sigue siendo humana.

El manifiesto formula esta actitud en diez convicciones. No se tratan aquí como un catálogo, sino se resumen en un breve discurso — del modo en que se pertenecen.

Pensamos que la mayor debilidad del obrar humano no es la maldad, sino la inconsciencia. Allí donde algo ocurre sin que nadie lo advierta, la acción pierde su lugar moral. Pensamos que la transparencia importa más que la perfección; quien promete perfección debe ocultar errores, quien promete transparencia debe mostrarlos. Pensamos que la IA puede aconsejar, pero que la responsabilidad debe permanecer en el ser humano — social, política, moralmente. Pensamos que la incertidumbre es un campo, no una formulación: quiere ser nombrada, no velada por la retórica.

Pensamos que la calidad del juicio es más decisiva para el futuro que la mera inteligencia técnica — porque incluso recomendaciones acertadas se vuelven equivocadas cuando ya no queda nadie que pueda examinarlas. Pensamos que la cultura y la creación de sentido son profundamente humanas: son el suelo del que el juicio puede en absoluto surgir. Y pensamos, por fin, que lo observado y lo supuesto han de distinguirse con rigor.

Lo observado y lo supuesto son dos cosas distintas.

De esta actitud se siguen las cuatro herramientas. Son su despliegue abierto — abierto, porque no son un software cerrado, sino roles de pensamiento que toman formas diferentes en aplicaciones diferentes. Y se siguen de la actitud, no al revés: solo cuando está establecido para qué debe servir una herramienta puede preguntarse con sentido cómo construirla. Una herramienta que no conoce su actitud extravía a sus usuarios.

Por eso este capítulo va primero. Todo lo que se describe en los capítulos siguientes — las cuatro herramientas, las instanciaciones concretas, el banco de trabajo, las once normas, la cadena, la infraestructura — es consecuencia de lo que aquí se dice. Nada de ello se sostiene por sí; todo va junto; y el guion entre todas las partes es la única actitud que no puede delegarse.

Las cuatro herramientas

Nivel 2. El despliegue abierto de la actitud — no software, sino roles de pensamiento.

Las cuatro herramientas son el despliegue abierto de la actitud. No son ellas mismas software, sino roles de pensamiento: Aletheia (percibir), Mnemosyne (recordar), Methodos (guiar metódicamente), Kairon (decidir). Cada aplicación concreta — radar, capa de seudonimización, grafo de conocimiento, asistente de proyecto, tablero de decisiones — es una instanciación de uno de estos cuatro, a menudo con rasgos de varios. Las cuatro herramientas le dan su modo de pensar; la instanciación le da su forma.

HERRAMIENTA	OPERACIÓN FUNDAMENTAL	PUNTO DE QUIEBRE
<i>Aletheia · Oro</i>	Percepción del cambio	donde hacer visible se vuelve juzgar
<i>Mnemosyne · Azul</i>	Memoria; separar enunciado y persona	donde recordar se vuelve entregar
<i>Methodos · Verde</i>	Del diálogo al resultado estructurado	donde la forma reina sobre el ser humano
<i>Kairon · Terracota</i>	Decisión: Kairos y Aion juntos	donde el resultado se sostiene sin razones

Aletheia · Oro

Aletheia — percepción del cambio

Aletheia es la filosofía de la Phronesis para la percepción. La palabra griega significa el desocultamiento, el salir de la ocultación. Aletheia hace visibles los desarrollos, los ordena y propone una valoración — pero somete el juicio al ser humano para su confirmación.

Hacer visible no es todavía juzgar. Sacar algo a la luz no es haberlo ya valorado.

Seis convicciones sostienen a Aletheia: primero, la estricta separación de hacer visible y juzgar; segundo, el deber de fuente (lo que queda sin fuente es rumor); tercero, la distinción según el grado de certeza (evento confirmado, señal temprana, mero anuncio); cuarto, la atención como selección, no exhaustividad; quinto, el examen de lo nuevo a la luz de lo sido; sexto, la orientación del desocultamiento hacia un adentro — hacia memoria, forma y decisión. Una percepción que se toma por juicio no es más que un feed.

Mnemosyne · Azul

Mnemosyne — memoria que protege

Mnemosyne — en el recuerdo griego la madre de las Musas — representa una memoria que preserva sin entregar. Su operación fundamental es la separación de enunciado y persona: el contenido sustantivo permanece utilizable, la identidad permanece protegida.

Un enunciado y la persona a la que atañe son dos cosas distintas.

Cuatro convicciones: el saber puede compartirse sin exponer a las personas. Una capa de protección debe ser ella misma visible — nadie confía en una caja negra que promete confianza. En caso de duda, la protección tiene prioridad: un nombre pasado por alto pesa más que un reemplazo de más. Y por último: recordar y proteger van juntos; una memoria que no protege no es memoria, sino entrega.

Methodos · Verde

Methodos — del diálogo al resultado estructurado

Methodos conduce del diálogo al resultado, sin olvidar el camino. Pregunta, escucha, ordena — y deja surgir una versión estructurada en la que lo dicho permanece reconocible.

El saber surge en el diálogo, no en el formulario vacío.

Un método es el juicio decantado de muchos.

La estructura es medio, no fin.

Cuatro convicciones: el saber surge en el diálogo, no en campos vacíos. Un método probado no es un corsé burocrático, sino el saber práctico reunido de quienes tuvieron que resolver la misma tarea antes que nosotros. Pensar junto significa notar lo que falta — la completitud no se logra con campos obligatorios, sino con buenas preguntas. Y la forma debe servir al ser humano, no al revés.

Kairon · Terracota

Kairon — decisión que permanece observada

Kairon lleva en sí dos palabras griegas: Kairos, el momento justo, y Aion, la duración. Ambos pertenecen a la decisión: cae en un momento, pero actúa en el tiempo. Sin Kairos no hay decisión; sin Aion no hay aprendizaje.

Las decisiones necesitan no solo resultados, sino razones.

Toda buena decisión debe seguir siendo observable.

El disenso puede ser signo de madurez.

Tres convicciones: las decisiones necesitan razones, no solo resultados — un sí o un no sin un entramado de opciones, supuestos, riesgos, responsabilidades y efectos esperados no es una decisión, sino una nota. Las buenas decisiones deben seguir siendo observables, porque solo en la realidad se muestra si los supuestos se sostenían. Y el disenso no es una perturbación, sino a menudo un signo de que un asunto ha sido comprendido en su profundidad.

Las instanciaciones

Nivel 3. Los roles de pensamiento se vuelven productos que se sostienen en el mundo.

En este nivel las cuatro herramientas se vuelven productos concretos. Cada instanciación es la instanciación de una herramienta — a menudo con rasgos de varias, pues ningún producto real se deja reducir limpiamente a un solo rol de pensamiento. Una entrevista de proyecto es predominantemente *Methodos*, pero toma decisiones; un grafo de conocimiento es predominantemente *Methodos*, pero recuerda; una capa de seudonimización es predominantemente *Mnemosyne*, pero también percibe lo que ha de protegerse.

INSTANCIACIÓN	HERRAMIENTA	MADUREZ	BREVE CARACTERIZACIÓN
<i>ALETHEIA Radar</i>	Aletheia	en construcción	radar de tecnología IA para administraciones, multi-tenant
<i>Capa de seudonimización v0.0.1</i>	Mnemosyne	primera versión, operativa	local en 127.0.0.1, entre aplicación y proveedor, 243 402 apellidos + 66 916 nombres (OFE)
<i>kennora</i>	Methodos	en construcción	oyente para saber a partir de diálogos, grafo de conocimiento como árbol y red
<i>HERMES PLA v0.35.0</i>	Methodos + Kairon	en uso	mandato de iniciación HERMES-2022 en forma de entrevista, examen formal y sustantivo separados

<i>KAIRON</i> <i>Dashboard (Feature-</i> <i>V2, commit</i> <i>369db5a)</i>	Kairon	madurez funcional, en prueba	ciclo de vida de diez estados, 75 pruebas de regresión, 4 507 líneas de código de aplicación
<i>DigiTwin</i> <i>(simulador de</i> <i>procesos)</i>	afín · no de la familia	premio Hackathon eJustice.ch	figura de pensamiento afín, no núcleo de la familia de herramientas

ALETHEIA Radar — la mirada verificada hacia afuera

El ALETHEIA Radar es la instanciación consecuente de Aletheia para administraciones. Reúne y ordena observaciones sobre desarrollos de la IA, distingue entre eventos confirmados, señales tempranas y meros anuncios, y adjunta a cada entrada su fuente. Es multi-tenant: cada administración ve su propio radar, con su propia selección de atención, su propia historia, sus propios sellos de verificación. No juzga, sino que propone una clasificación; el juicio queda en el ser humano. La madurez es en construcción — la estructura se sostiene, la práctica de observación crece.

Capa de seudonimización v0.0.1 — el paso silencioso previo

La capa de seudonimización es un pequeño servicio interpuesto entre la aplicación y el proveedor de modelo de lenguaje. Reemplaza los nombres por seudónimos estables antes de que los contenidos abandonen la máquina — y deshace la sustitución después de la respuesta. El proveedor ve el enunciado sustantivo; las personas permanecen consigo. Las listas de nombres provienen de la Oficina Federal de Estadística (OFE): 243 402 apellidos y 66 916 nombres. La capa se ejecuta localmente en 127.0.0.1 y es así verificable para quien la usa: no es una caja negra, sino un servicio visible. Madurez: primera versión, operativa.

kennora — el oyente para el saber

kennora escucha un diálogo y traza un grafo de conocimiento. El grafo tiene dos vistas: un árbol, que muestra jerarquías y pertenencias, y una red, que acoge relaciones de todo tipo. Además de las aristas habituales (es un, pertenece a, se sigue de) existe la conexión

particular rima-con. Marca vecindades semánticas que surgen de analogías e imágenes — no de similitud formal, sino del sonido oído de una relación. Madurez: en construcción.

HERMES PIA v0.35.0 — la entrevista con examen

HERMES PIA acompaña la fase de iniciación de un proyecto HERMES-2022. El usuario narra en lenguaje cotidiano lo que se propone; el software forma con ello un mandato de iniciación. Dos exámenes corren entonces por separado: un examen formal en código, sin modelo de lenguaje — cuenta si todos los datos obligatorios están presentes, si han quedado marcadores de posición, si los capítulos se contradicen. Y un examen sustantivo con modelo de lenguaje — lee el mandato desde la perspectiva del mandante y propone una recomendación de aprobación con condiciones. El usuario ingresa con su dirección de correo institucional; cada unidad organizativa ve exclusivamente sus propios proyectos. Cifras: 104 pruebas, 2 054 líneas de código sustantivo, 42 funciones públicas. Madurez: en uso.

KAIRON Dashboard — el ciclo de vida de diez estados

El KAIRON Dashboard es la instanciación más consecuente del eje Kairos-Aion. Cada decisión atraviesa un ciclo de vida de diez estados que no deja terminar la decisión con la aprobación, sino que la hace pasar a la observación: Draft → In Review → Simulated → Risk Reviewed → Governance Reviewed → Approved → Observed → Reassessment Needed → Reassessing → Archived. El estado Observed es el importante: aquí la decisión entra en el tiempo y es puesta a prueba para ver si sus supuestos se sostienen. Reassessment Needed y Reassessing cierran el círculo. Cifras: 75 pruebas de regresión, 4 507 líneas de código de aplicación. Técnicamente: Flask, SQLAlchemy, Jinja, migraciones Alembic. Rama Feature-V2, commit 369db5a. Madurez: madurez funcional, en prueba.

DigiTwin — el vecino afín

DigiTwin es un simulador de procesos que fue distinguido en el Hackathon eJustice.ch. No pertenece al núcleo de la familia de herramientas Phronesis, porque no es una herramienta en el sentido de los cuatro roles de pensamiento. Pero es pensamiento afín — sobre todo respecto al eje Kairos-Aion: la simulación pone a prueba supuestos antes de que

se tome una decisión, y así flanquea el modo Kairon desde el otro lado. Donde Kairon observa el efecto a posteriori, DigiTwin lo pone a prueba por anticipado.

El taller — el banco de trabajo

Traviesa. No una herramienta de la familia, sino aquello que produce herramientas.

El banco de trabajo tiene un papel especial en el conjunto. Él mismo no es una herramienta de la familia Phronesis — no es ni Aletheia ni Mnemosyne, ni Methodos ni Kairon. Es, en cambio, el taller en el que las herramientas surgen. Y no solo las produce, sino que las fundamenta de manera verificable. Ese es su verdadero logro: no entrega código solo, sino el dossier que muestra por qué se puede confiar en este código.

Cuatro estadios, siete productos

Una herramienta surge en cuatro estadios: cuadro de requisitos, arquitectura, especificación, implementación. Cada estadio tiene sus productores, sus examinadores y sus puertas. Al final quedan siete productos: cuadro de requisitos, arquitectura, especificación, casos de prueba, especificación técnica, puesta en servicio, y — como verdadero corazón del asunto — el protocolo de examen y decisión.

El protocolo de examen y decisión no es un adorno. Es el producto que mantiene unidos a los demás. Muestra quién examinó qué enunciado, qué modelos estuvieron involucrados, dónde coincidieron, dónde se contradijeron, qué hallazgos quedaron abiertos y qué, pese a la aprobación, no ha sido resuelto. Es el documento en el que el banco de trabajo hace justicia a su propia norma.

Productores y examinadores, con modelos distintos

En cada estadio trabaja un par productor-examinador. Ambos son instancias distintas; más importante aún: son modelos distintos. En el ejemplo VOR-0580b0fa — el tutor de mecánica cuántica sobre el que el banco de trabajo está documentado en su versión 1.0 — el productor era xai/grok-4.3 y el examinador openai/gpt-4.1. Dos filosofías de entrenamiento distintas, dos regímenes RLHF distintos, dos perfiles de alucinación

distintos. Modelos distintos no garantizan errores independientes; pero al menos evitan el autoexamen trivial de la misma instancia de modelo y hacen visible la dependencia del modelo.

El núcleo calcula, el modelo interpreta, el ser humano juzga y asume la responsabilidad

El núcleo calcula, el modelo interpreta, el ser humano juzga y asume la responsabilidad.

Esa es la decisión arquitectónica fundamental del banco de trabajo, y es el gemelo artesanal de RA-01. Donde se exige determinismo — valores de retorno de pytest, sonda de arranque, si el servicio responde, si los casos de prueba pasan, si los identificadores están completos —, ningún modelo de lenguaje debe decidir. Donde se exige interpretación — si un capítulo se corresponde sustantivamente con los requisitos, si la arquitectura cumple su propósito —, ningún código de reglas debe juzgar. El núcleo cuenta y mide; el modelo lee y evalúa; y donde ha de resultar de ello un juicio asumido, se suma el ser humano. Los tres permanecen en su rol.

Convergencia no es corrección

Convergencia no es corrección.

La segunda frase portante del banco de trabajo. Cuando un examinador calla, puede significar dos cosas: el productor lo hizo bien — o el examinador se rindió. El banco de trabajo no confunde ambas cosas. Por eso registra cuándo un examen termina por acuerdo y cuándo termina por agotamiento. La segunda posibilidad es la más frecuente y la más peligrosa; no se oculta.

Una versión de aprobación nunca está libre de hallazgos. Lo que aquí se sostiene está decidido — no resuelto.

Y por eso una aprobación del banco de trabajo nunca puede pretender estar libre de errores. Solo pretende que los hallazgos aún presentes están decididos: vistos, nombrados, ubicados, provistos de un tratamiento. Lo que se pudo se resolvió; lo que queda se ha sabido.

Modo-núcleo y modo-humano

El banco de trabajo conoce dos modos de decisión. El modo-núcleo rige exactamente cuando examinador y productor coinciden y el hallazgo reza: «se aplica». Todos los demás casos — cada contradicción, cada «no se aplica», cada «condición» — caen en el modo-humano. La máquina solo puede confirmar. Todo negar, restringir y condicionar queda reservado al ser humano.

Un agente vale lo que vale su criterio de aceptación.

Esa es la implementación más consecuente de la condición que en el research paper se llama consequentiality: la posición humana determina el resultado, salvo que un acto documentado la reemplace. En el banco de trabajo esta condición no es filosófica, sino operativa: es la diferencia entre una línea verde y una línea roja en el protocolo.

Lo que se sigue

El banco de trabajo es así la respuesta operativa a la pregunta fundamental de la actitud: ¿cómo puede la sabiduría práctica permanecer en un ser humano cuando las máquinas escriben con, examinan con, proponen con? La respuesta reza: por el hecho de que la máquina conoce su rol, de que dos máquinas distintas se examinan mutuamente, de que el núcleo cuenta en vez de juzgar, de que toda negación regresa a un ser humano.

Las once normas de la suite

Nivel 4. El corpus común de reglas por encima de todas las instanciaciones.

Las once normas forman el corpus común de reglas de la suite. Le dan al banco de trabajo su medida, a las herramientas su forma y a las instanciaciones su verificabilidad. Son la razón por la que dos herramientas que difieren en sustancia — HERMES PIA y KAIRON — permanecen sin embargo reconocibles como partes de la misma familia: sus identificadores, sus estados, sus regímenes de examen, su lenguaje siguen las mismas normas.

N.º	NORMA	PROPÓSITO
1	Manifiestos	actitud, propósito, límites
2	Identidad visual	brand & logo styleguide, familia de logos
3	Uso	guía UX, guiado del usuario
4	Arquitectura de referencia v0.4	13 principios RA-01 a RA-13, 10 prohibiciones NR-01 a NR-10
5	Catálogo de capabilities	llamada IA, persistencia, workflow, STT, TTS, renderer, exportación
6	Forma contractual	skill-manifest.yaml
7	Norma de skill	SK-01 a SK-17 (norma de skill v0.2; la versión anterior v0.1 terminaba en SK-15), con regla + motivo + examen

8	Concepto de prueba	exámenes deterministas y asistidos por modelo
9	Directrices de desarrollo	CLAUDE.md por repositorio
10	Modelo de roles	especificación HERMES: rol funcional ↔ método ↔ examinador ↔ gobernanza
11	Calidad sustantiva	modelo de calidad con criterios D y F

Los trece principios de la arquitectura de referencia v0.4

La arquitectura de referencia es el núcleo de la suite de normas. Sus trece principios describen cómo debe estar construida una herramienta de la familia para que pueda portar la actitud. No son recomendaciones, sino compromisos: quien construye conforme a ellos puede hacer verificar que ha construido conforme a ellos.

RA-01 El núcleo calcula, el modelo interpreta, el ser humano juzga y asume la responsabilidad.

RA-02 Ningún hecho autoritativo procedente del modelo.

RA-03 Cada objeto-núcleo lleva su governance-mixin.

RA-04 Borrador y confirmación son campos separados.

RA-05 La aprobación es una puerta; un cambio la invalida.

RA-06 Los cambios son transiciones, no rupturas.

RA-07 Configuración antes que programación.

RA-08 La separación de tenants es estructura, no filtro.

RA-09 Sin contacto directo con el modelo.

RA-10 Cada resultado lleva su procedencia.

RA-11 La incertidumbre es un campo, no una formulación.

RA-12 Vocabulario controlado.

RA-13 Una herramienta produce pruebas, no solo resultados.

Nota: la versión de RA-01 aquí formulada está precisada respecto a la arquitectura de referencia v0.4 (donde dice: «El núcleo decide, el modelo propone, el ser humano asume»). La nueva versión sigue el trípodé del capítulo V y se recogerá con la próxima versión de la arquitectura de referencia. Hasta entonces, RA-01 en esta versión es una propuesta de ajuste, no la norma vigente.

No los trece principios están anclados del mismo modo. Diez descansan directamente sobre una frase del manifiesto. Dos — RA-07 «Configuración antes que programación» y RA-12 «Vocabulario controlado» — son reglas técnicas sin anclaje directo en valores, pero legitimadas por una práctica de producto probada. Uno — RA-08 «La separación de tenants es estructura, no filtro» — está anclado indirectamente; su frase portante PH-11 sobre el trato con lo confiado se pospone hasta que la operacionalización de «expresamente destinado a compartirse» esté aclarada. Esta diferenciación no se oculta sino que se declara: solo se sostiene lo que está fundamentado; solo puede examinarse lo que está declarado.

Las diez prohibiciones

NR-01 Una herramienta no debe: emitir un juicio o asumir una responsabilidad que corresponde a un ser humano.

NR-02 Una herramienta no debe producir una referencia, una cifra o una prueba que no provenga de una fuente.

NR-03 Una herramienta no debe dirigirse directamente a un modelo.

NR-04 Una herramienta no debe traspasar una separación de tenants.

NR-05 Una herramienta no debe conservar un estado aprobado mientras el contenido cambia.

NR-06 Una herramienta no debe ampliar por propia autoridad un vocabulario controlado.

NR-07 Una herramienta no debe emitir un resultado sin procedencia.

NR-08 Una herramienta no debe garantizar una seguridad únicamente mediante instrucción a un modelo.

NR-09 Una herramienta no debe cambiar el significado de valores almacenados sin registrar la transición.

NR-10 Una herramienta no debe editar con posterioridad una prueba emitida.

Prohibiciones del banco de trabajo (WB-01 a WB-10)

Estas prohibiciones se derivan de la práctica del banco de trabajo y complementan la arquitectura de referencia; llevan un prefijo propio para excluir confusión con los identificadores NR de la arquitectura de referencia.

WB-01 Sin silent narrowing.

WB-02 Sin rutinas de confirmación sin verdadera elección.

WB-03 Sin elección de modelo sin protocolo.

WB-04 Sin confusión de convergencia y corrección.

WB-05 Sin juicios de modelo de lenguaje en conteos deterministas.

WB-06 Sin mezclar productor y examinador en la misma instancia de modelo.

WB-07 Sin aprobación sin hallazgos residuales declarados abiertamente.

WB-08 Sin enunciado sin identificador de procedencia en las producciones del banco de trabajo.

WB-09 Sin renombrar hallazgos decididos como hallazgos resueltos.

WB-10 Sin entrega de identidades sin verificabilidad de la seudonimización.

Por qué estas once normas juntas

Las once normas no cubren por casualidad niveles distintos. Desde la actitud (manifiestos) pasando por la identidad visual y el uso, por la arquitectura con sus principios y prohibiciones, las capabilities, la forma contractual y la norma de skill, hasta prueba, directrices de desarrollo, modelo de roles y modelo de calidad — trazan el camino que una herramienta recorre antes de salir al mundo.

La cadena — cómo las normas se hacen código

Nivel 5. La cadena de desarrollo de agentes v0.7 — cuatro estadios, cinco puertas.

La cadena es la versión operativa del banco de trabajo. Donde el banco de trabajo es el taller en el que una cosa efectivamente se fabrica, la cadena es el procedimiento que hace esa fabricación describible. Tiene cuatro estadios y en cada transición una puerta que solo un ser humano puede abrir: cuadro de requisitos → arquitectura → especificación → implementación → prueba.

Los cuatro estadios y sus puertas

Al comienzo se sostiene el cuadro de requisitos. Surge en el diálogo — no en un formulario vacío, sino en una narración guiada de la que se extraen los requisitos. Al final de este estadio se sostiene la primera puerta humana: el mandante lee lo que ha de hacerse, ve los puntos aún abiertos y los cierra o los confirma como deliberadamente abiertos. Solo entonces está aprobado el cuadro de requisitos.

Sigue la arquitectura. Responde a la pregunta de cómo debe estar construida la herramienta para cumplir el cuadro de requisitos. Aquí trabaja un par productor-examinador sobre modelos distintos. Al final se sostiene una constatación de conformidad respecto a la arquitectura de referencia: cuáles de los trece principios se cumplen íntegramente, cuáles parcialmente, cuáles permanecen abiertos. Solo cuando un ser humano ha examinado este cuadro y decidido sobre los hallazgos abiertos, la arquitectura está aprobada.

Sigue la especificación. Traspone la arquitectura a una descripción tan precisa que pueda resultar de ella código — objetos-núcleo, interfaces, estados, transiciones, casos de prueba.

También aquí un par productor-examinador, también aquí una puerta que solo un ser humano abre.

Finalmente la implementación. Aquí surge el código, y con él surgen las producciones que hacen el código verificable: casos de prueba, especificación técnica, puesta en servicio. Al final se sostiene la prueba: el protocolo de examen y decisión que hace toda la cadena legible en sentido inverso.

Constatación de conformidad con el ejemplo HERMES PIA

En el ejemplo HERMES PIA — examinado contra la arquitectura de referencia en su versión v0.2 — dos de los principios se cumplen íntegramente, nueve parcialmente, dos permanecen abiertos. El hallazgo más crítico fue RA-09 (sin contacto directo con el modelo): a través de la API nativa del navegador Web Speech existía un camino por el que datos de audio podían llegar a un proveedor sin seudonimización previa. Fue nombrado, decidido — y cerrado en la versión siguiente. Es exactamente el caso para el que la cadena está hecha: no para que nada salga mal, sino para que lo que pudiera salir mal esté nombrado y resuelto antes de que se declare la aprobación.

Por qué la cadena tiene esta forma

La cadena protege en tres puntos a la vez. Protege la actitud, porque en cada puerta un ser humano decide — no un botón de aprobación, no una rutina de confirmación. Protege las herramientas, porque sus producciones permanecen verificables contra las normas. Y protege a los usuarios, porque no reciben la herramienta en un estado intransparente: la reciben con su propia historia.

La infraestructura — Capabilities, Skills, Código

Nivel 6. La base técnica que todas las herramientas comparten.

En este nivel se encuentra la infraestructura: los bloques técnicos que todas las herramientas de la familia comparten. Son la razón por la que una herramienta nueva no comienza desde cero — y la razón por la que un error en un bloque compartido puede corregirse en un solo lugar y todas las herramientas se restablecen con ello.

Capabilities — los bloques compartidos

Las capabilities son bloques utilizados por varias instancias — y que por eso se explican, construyen y examinan exactamente una vez. No se reinventan en cada herramienta. La familia comprende actualmente ocho capabilities:

- Llamada IA — con seudonimización previa; ninguna herramienta llama a un proveedor directamente y sin protección (RA-09).
- Persistencia — el almacenamiento regulado de objetos-núcleo con su governance-mixin (RA-03).
- Workflow — el control de transiciones entre estados; la maquinaria de base tras los ciclos de vida de diez estados.
- Reconocimiento de voz — habla a texto; guiado a través de la capa de seudonimización, donde pueden caer nombres de personas.
- Síntesis de voz — texto a habla; para la salida hacia el usuario.
- Renderizador de fórmulas — notación matemática, p. ej. para materiales didácticos como el tutor de mecánica cuántica.
- Renderizador de diagramas — diagramas y esquemas; renderizados de forma determinista, no generados por modelo.

— Exportación de documentos — salida en formato Word y PDF; exportación nativa, para que los resultados permanezcan editables.

Skills — la forma contractual

Los skills son las unidades funcionales más pequeñas. Siguen una forma contractual, para que un skill no solo funcione, sino que describa su función: un archivo frontmatter SKILL.md con los campos name, description, scope, applies_to — y un skill-manifest.yaml como verdadera forma contractual. Ambos juntos hacen un skill verificable, intercambiable y reutilizable.

Código — las decisiones técnicas

El código de las herramientas sigue decisiones deliberadamente sobrias. En KAIRON: Flask como marco web, SQLAlchemy como ORM, Jinja como motor de plantillas, Alembic para migraciones. Es la combinación clásica del ecosistema Python — no la más reciente, sino la más probada. Se elige porque sus comportamientos son conocidos: un equipo que la conoce sabe dónde algo sale mal.

Para los exámenes, pytest sirve como instancia de examen determinista. Los resultados son verdes o rojos; no son discutibles. Ese es el modo-núcleo que el banco de trabajo exige: el núcleo calcula.

Para la integración de skills, MCP, el Model Context Protocol, sirve de interfaz. Desacopla las herramientas de proveedores de modelos particulares y hace así del cambio de modelo una cuestión de configuración, no de ruptura. Es RA-07 (configuración antes que programación) en su versión técnica.

Tomada en conjunto, la infraestructura no da nada espectacular — y así se pretende. Está construida de modo que se retire tras las herramientas. Quien porta la actitud y conoce las normas ya no debe ser sorprendido en este nivel.

El flanco — Ciencia y presencia pública

Aquello con lo que la obra se mide hacia afuera y donde se hace visible.

Junto a los seis niveles se encuentra el flanco: aquello que examina la obra hacia afuera y donde ella se muestra. Consta de dos publicaciones científicas y de un repositorio silencioso.

Research paper — Where Judgement Is Reserved

El research paper con el título de trabajo Where Judgement Is Reserved es el núcleo científico del flanco. Sigue la metodología del Design Science Research según Peffers e introduce tres términos que hacen la actitud científicamente verificable: structural reservation — la reserva estructural de ocasiones de juicio como propiedad del sistema, no como rasgo de interacción; judgement occasion rate (JOR) — una serie temporal que mide qué proporción de las ocasiones que requieren juicio regresa efectivamente a un ser humano; y silent narrowing test — un criterio de examen que pregunta si el conjunto de los casos reservados puede reducirse silenciosamente sin que esa reducción sea ella misma una decisión documentada y fundamentada.

El artículo posiciona la actitud frente a Green (2022), se apoya empíricamente en Bainbridge (1983), Shen & Tamkin (2026), Parasuraman & Manzey (2010) y Buçinca et al. (2021), y encuentra en la FAA-SAFO-17007 (2017) el precedente histórico de una structural reservation ordenada por una autoridad de supervisión.

Documento de posición — las doce páginas para todos

El documento de posición es la lectura unificadora de los cinco manifiestos — Phronesis, Aletheia, Mnemosyne, Methodos, Kairon — en un único texto de doce páginas. Se dirige a interlocutores que no conocen ninguno de los textos de origen. Donde este documento de

síntesis es un mapa, el documento de posición es un panorama: el mismo suelo, otra perspectiva.

kaspair.ch — el repositorio silencioso

kaspair.ch es la presencia del autor. Es deliberadamente silenciosa: nada de marketing, nada de newsletters, nada de analítica en primer plano. Es un lugar donde reposan los textos — para todos aquellos que deseen leerlos, sin que se inicie una relación. Quien lo encuentra encuentra solo la obra.

El flanco no es, pues, ni publicidad ni venta. Es el lugar donde la obra en su figura actual ha de ser examinada — por peer review, por el público, por lectoras y lectores que la lean dentro de su propio día a día.

El estado actual

Lo que está terminado y lo que está abierto. Un balance honesto.

El estado al 14 de agosto de 2026 se lee en dos bloques: lo que está terminado y lo que está abierto. La honestidad de este balance pertenece a la actitud misma — lo que callo, no puedo mejorarlo.

Lo que está terminado

- Arquitectura de referencia v0.4. Los trece principios y las diez prohibiciones se hallan consolidados.
- Documento de posición. Doce páginas, listas para publicación; reúne los cinco manifiestos.
- Research paper v0.6. Versión de trabajo con arco narrativo completo.
- HERMES PIA v0.35.0. En uso; 104 pruebas, 2 054 líneas de código sustantivo, 42 funciones públicas.
- KAIRON Feature-V2. Con 75 pruebas de regresión y 4 507 líneas de código de aplicación; ciclo de vida de diez estados cerrado.
- Banco de trabajo versión 1.0. Cuatro estadios, siete producciones, pares productor-examinador sobre modelos distintos — documentado sobre el ejemplo VOR-0580b0fa.

Lo que está abierto

- Q4 de la agenda de investigación. La pregunta de si las ocasiones de juicio reservadas preservan capacidad o solo generan trabajo se declara abierta en el artículo. Es el eje empírico propiamente dicho de un proyecto sucesor.
- Dos límites del Contribution Boundary v0.5. Primero, el selection risk: las organizaciones tal vez prefieran los diseños menos eficaces porque se sienten mejor —

Buçınca muestra la correlación negativa entre eficacia y preferencia. Segundo, las intervention-generated inequalities: las ocasiones de juicio reservadas no ayudan por igual a todos los grupos. Ambos límites han de nombrarse en el artículo, no ocultarse.

- Ampliación de HERMES PIA de mono-tenant a multi-tenant. Actualmente la estructura está preparada, pero no técnicamente consumada (RA-08 como «estructura, no filtro» exige más que una sola línea).
- Ampliación de la familia de herramientas. HERMES PIA cubre solo partes de la iniciación. Las fases concepto, realización, introducción, cierre y ejecución se sostienen como candidatos abiertos a otras instanciaciones.

Los dos bloques muestran dónde está la obra: con una base que sostiene, y con preguntas más grandes que la próxima versión. Nada de esto ha de retirarse rápidamente. Lo que está abierto queda abierto — y será respondido justamente en el lugar donde surge.

CAPÍTULO XI

Cierre — la única figura de pensamiento

Regreso al comienzo: el círculo abierto en el anillo.

Cuatro niveles de resolución sostienen esta obra, y todos dicen lo mismo.

FILOSÓFICAMENTE

La sabiduría práctica sigue siendo humana.

ARQUITECTÓNICAMENTE

El núcleo calcula, el modelo interpreta, el ser humano juzga y asume la responsabilidad.

OPERATIVAMENTE

Todo lo que significa negación, restricción o condición cae en el modo-humano.

METROLÓGICAMENTE

Structural reservation, judgement occasion rate, silent narrowing test.

No son cuatro ideas. Son cuatro niveles de la misma.

Al comienzo se sostenía la pregunta por la única figura de pensamiento que mantiene todo unido. Al final se sostiene la respuesta, en la forma de un signo simple: el círculo abierto en el anillo. El todo permanece cerrado — los seis niveles se engranan, el banco de trabajo los une, el flanco los examina — ; y en el medio queda abierto un círculo: el ser humano que juzga y asume la responsabilidad.

Todo lo descrito en este documento es despliegue de este único signo. La actitud es el anillo. Las cuatro herramientas son sus cuatro despliegues como roles de pensamiento. Las instanciaciones son sus formas en el mundo. El banco de trabajo es la mano que las hace verificables. Las once normas son el corpus de reglas que deja al anillo encontrar su forma. La cadena es el camino. La infraestructura es la tierra por la que discurre. Y el flanco es el lugar en el que ha de mostrarse que el signo vale algo.

Lo que puede faltar y lo que no, lo que debe añadirse y lo que retirarse, se lee sobre esta figura. El círculo abierto en el medio es la salvaguarda. No debe cerrarse — ni por comodidad, ni por eficacia, ni por la tentación de hacer una herramienta tan buena que ya no necesite el juicio. Si este círculo se cierra, el signo entero queda alterado.

Cinco signos portan la misma figura de base: el anillo como el todo, el círculo abierto como el ser humano. El signo Phronesis los muestra en su forma pura. Los cuatro signos-herramienta añaden cada uno exactamente un gesto del que se deriva el significado de la herramienta.

Phronesis

LA ACTITUD

El todo permanece cerrado, el ser humano permanece abierto — la figura básica de la que se derivan las cuatro herramientas.

Aletheia

PERCEPCIÓN

El núcleo se alza como un sol sobre el horizonte — lo que estaba oculto sale a la luz.

Mnemosyne

MEMORIA

El núcleo reposa en capas que preservan — el saber permanece, sin entregar a la persona.

Methodos

MÉTODO

El camino asciende en gradas hacia el núcleo — de lo que se dice nace lo que el método exige.

Kairon

DECISIÓN

La aguja encuentra la señal en el borde — el momento justo en el tiempo entero.

«La sabiduría práctica sigue siendo humana.»