Agencia sin imputación

Hermenéutica, pragmática y la validación epistémica del discurso artificial

Carlos Isaac Zainea[*]

Marzo 2026

Resumen

La filosofía ha abordado los modelos de lenguaje de gran escala (LLM) con una pregunta ontológica — ¿comprende la máquina? — cuya respuesta negativa, correcta, suele derivar en una conclusión paralizante: si el LLM no comprende, su output carece de valor epistémico. Este artículo argumenta que esa inferencia es inválida y propone un giro epistemológico. Primero, con la triple mímesis de Ricoeur localiza el déficit del LLM: es un agente que configura (Mímesis II) sin poder refigurar ni imputarse lo configurado (Mímesis III). Segundo, recategoriza el error: no «alucinaciones» — metáfora mentalista incoherente — sino **infortunios de autoridad epistémica** en sentido austiniano, articulados con Habermas y Floridi. Tercero, muestra que la Teoría Fundamentada y la Triangulación ofrecen el andamiaje procesal que compensa esa carencia, y lo demuestra con un prototipo implementado sobre cultura visual narco colombiana. El resultado es el concepto de **prótesis hermenéutica**: una arquitectura que redistribuye las funciones epistémicas entre un agente operativo (la máquina) y un agente imputable (el investigador).

Palabras clave: inteligencia artificial; epistemología; hermenéutica; Ricoeur; Austin; agencia; Teoría Fundamentada.

1. El problema epistemológico

Los modelos de lenguaje de gran escala (LLM) ingresaron a la producción de conocimiento antes de que la filosofía respondiera a la pregunta que plantean. En contextos académicos, médicos, jurídicos y

[*]Agradezco a Jesús Antonio Pardo León, colega del Doctorado en Filosofía de la Universidad Santo Tomás, por los insumos sobre el problema — la conceptualización de *lo narco como categoría estética*, tema de su tesis doctoral en curso (Pardo León, en curso) — y por sus sugerencias de herramientas para abordarlo. El concepto de prótesis hermenéutica, la elección y aplicación de la Teoría Fundamentada y la implementación del prototipo son propias del autor de este escrito. En la preparación del texto se empleó Claude Opus (Anthropic, 2026) como apoyo en la edición y mejora de la redacción; el contenido, el argumento y las decisiones finales son del autor.

periodísticos, sistemas como GPT, Claude o Gemini sintetizan literatura, extraen patrones, proponen categorías y producen textos que circulan como resultados de investigación. El problema no es si los LLM *serán* usados — ya lo son —, sino bajo qué condiciones ese uso es epistemológicamente legítimo.

La filosofía ha respondido, sin embargo, con una pregunta de carácter **ontológico**: ¿*comprende* la máquina? ¿Tiene intencionalidad? ¿Es un sujeto? La pregunta es legítima, pero genera un callejón sin salida: la respuesta es no, y aceptarla sin más parece cerrar toda posibilidad de pensar el LLM como participante en procesos epistémicos. Este trabajo sostiene que el callejón es un artefacto de la pregunta, no una necesidad del problema. Que el LLM no comprenda no significa que no *haga*: produce, opera, selecciona qué token sigue a cuál, configura efectos que llegan al mundo. Hay allí un tipo de operación que no es la del sujeto cognoscente ni la del instrumento pasivo, y nombrarla bien es la tarea que la pregunta ontológica deja sin resolver.

1.1 La herencia de la crítica ontológica

La tradición que ha respondido a la pregunta por la comprensión artificial es sólida y, en lo esencial, correcta. El *Chinese Room* de Searle (1980) establece que la manipulación sintáctica de símbolos no produce semántica. Dreyfus (1972, 1992) añade la dimensión fenomenológica: la comprensión humana está anclada en el cuerpo, la situación y un trasfondo de habilidades prácticas que ningún sistema formal replica. Gadamer (1960) la extiende al plano hermenéutico: comprender es participar en una tradición, en una historia de efectos (*Wirkungsgeschichte*) previa a todo acto individual. Heidegger (1927) formula el fundamento ontológico: la comprensión no es operación cognitiva sino estructura del Dasein en cuanto ser-en-el-mundo, anclada en la temporalidad, la finitud y el cuidado.

El LLM carece de todo esto: no tiene cuerpo, situación, historia efectual ni un mundo al cual pertenecer. Su producción opera sobre probabilidades de co-ocurrencia en un corpus, no sobre participación en las prácticas que ese corpus sedimenta. Cuando enuncia algo sobre el duelo, no ha perdido a nadie; cuando sintetiza una discusión filosófica, no tiene ninguna de las preguntas que la hacen importar. La crítica ontológica es irrefutable, y este trabajo la acepta en su totalidad. Lo que discute es la inferencia que suele derivarse de ella.

1.2 El bloqueo ontológico

La crítica ontológica desemboca en un silogismo de conclusión paralizante:

$\mathbf{P}_1$ El LLM no comprende.

$\mathbf{P}_2$ Su output carece de valor epistémico.

$\therefore$ **Nada que diseñar.** non sequitur

La carencia ontológica no implica inutilidad epistémica.

Un telescopio no «comprende» las estrellas; una resonancia magnética no «siente» el tejido. Nadie niega su valor epistémico, que depende de los protocolos que gobiernan su uso y de los criterios que determinan cuándo su output participa en un proceso válido. La analogía instrumental no es completa, y habrá que volver sobre ella: el telescopio no produce afirmaciones, no encadena inferencias, no selecciona qué decir a continuación. El LLM sí. Esa diferencia no autoriza a tratarlo como sujeto, pero obliga a buscar categorías que describan lo que hace sin antropomorfizarlo ni reducirlo al aparato óptico.

La distinción de Reichenbach (1938) entre *contexto de descubrimiento* y *contexto de justificación* es aquí decisiva. Que un LLM genere una categoría o una hipótesis no dice nada sobre su validez; lo que la determina es el proceso por el cual ese resultado se evalúa, corrobora y contrasta, proceso que es independiente del origen cognitivo del resultado. Lo que importa epistemológicamente no es si quien generó la hipótesis *comprende*, sino si el proceso de validación satisface las condiciones que hacen que un resultado cuente como conocimiento. Esa es la pregunta productiva.

1.3 El giro epistemológico

Pregunta ontológica		**Pregunta epistemológica**
¿Comprende el LLM?	**GIRO** →	¿Es válido el conocimiento producido?
¿Tiene intencionalidad?		¿Cómo se controla su fiabilidad?
¿Es un sujeto?		¿Quién valida y cómo?
Respuesta: No. (Y es un callejón sin salida.)		*Respuesta: Depende del andamiaje procesal.*

La reorientación implica reconocer que la epistemología no requiere que los instrumentos de producción de conocimiento sean sujetos cognoscentes, sino que los *procesos* que gobiernan su uso satisfagan condiciones de validez. El debate contemporáneo ya articula esta distinción desde ángulos independientes. Cantwell Smith (2019) distingue entre *reckoning* — manipulación competente de estructuras formales — y *judgment* — evaluación contextualizada y responsable, anclada en un mundo de consecuencias. Floridi (2019; antes Floridi & Sanders, 2004) señala dos cosas que conviene leer juntas: que los LLM producen datos con aspecto de información semántica, aléticamente neutros hasta que un proceso externo los sitúa en un registro de verdad, y que tales sistemas pueden caracterizarse como **agentes** en sentido propio — interactivos, autónomos, adaptables — sin que por ello les corresponda intencionalidad, conciencia o responsabilidad. El diagnóstico resultante es de doble cara: el LLM opera con competencia creciente y *actúa* con consecuencias reales, pero carece de las condiciones para que esa acción sea epistémicamente válida y atribuible. De ahí la pregunta que articula el artículo: *¿cómo se valida el conocimiento producido por — o con — un sistema que actúa y genera discursos coherentes sin participar en las condiciones de la comprensión ni de la responsabilidad?*

1.4 Tesis y estructura

La tesis se formula en tres movimientos. **Primero**: los LLM son máquinas de configuración en el sentido de la Mímesis II de Ricoeur — producen tramas, organizan datos, generan discurso coherente — y son en ese ejercicio agentes en sentido propio, sin participar en las condiciones de la Mímesis III, donde el conocimiento se valida, se imputa a un sujeto y retorna a un mundo de consecuencias. Hay agencia operativa; falta imputación. **Segundo**: sus errores no son «alucinaciones» — metáfora mentalista que presupone una mente que distorsiona — sino **infortunios de autoridad epistémica** en sentido austiniano: actos que adoptan la forma de aserciones fundamentadas sin poder satisfacer las condiciones de esa fundamentación, porque ningún agente capaz de responder por ellas las sostiene. **Tercero**: la Teoría Fundamentada y la Triangulación ofrecen los protocolos para construir el andamiaje que compensa esa carencia; el prototipo *Descubrimiento de Categorías en Imágenes NarcoColombianas* — análisis hermenéutico de la cultura visual narco implementado con Gemini Vision — demuestra que el diseño es implementable y permite identificar qué logra el andamiaje y qué debe seguir aportando el investigador.

El §2 analiza la asimetría funcional del LLM y localiza el déficit con la triple mímesis. El §3 recategoriza el error como infortunio de autoridad epistémica, articulando Austin, Habermas y Floridi. El §4 desarrolla la hermenéutica como protocolo: la genealogía de las técnicas de razonamiento artificial (CoT → ToT → ReAct → IA agéntica), la traducción de la Teoría Fundamentada en andamiaje y el análisis crítico del prototipo. El §5 articula el concepto de **prótesis hermenéutica**, fundado en Clark, Merleau-Ponty, Feenberg y Vallor. El §6 sintetiza y abre líneas de investigación.

2. Producción de sentido sin sujeto epistémico

> *«Explicar más para comprender mejor.»*
>
> — Ricoeur, *Del texto a la acción*

La pregunta que dejó abierta el §1 exige primero una descripción precisa de lo que el sistema hace — no en sentido ontológico, sino funcional: ¿qué operaciones realiza sobre el lenguaje, con qué competencia y con qué restricciones estructurales? La respuesta es contraintuitiva: las capacidades que hacen al LLM epistemológicamente útil y las vulnerabilidades que lo hacen peligroso no son propiedades independientes, sino dos caras de la misma condición estructural.

2.1 La asimetría funcional

El LLM es, ante todo, una máquina de producción de discurso coherente. Su entrenamiento — predicción del siguiente token, optimizada por retroalimentación de preferencia humana — produce un sistema extraordinariamente competente en satisfacer expectativas formales: gramaticalidad, con-

sistencia temática, adecuación retórica. Lo que ese proceso no produce, porque no lo exige, es la correspondencia entre el discurso y un estado de cosas externo. El modelo es entrenado para ser coherente, no para ser verdadero; coherencia y verdad son propiedades distintas, y la optimización produce la primera con independencia de la segunda.

Esta separación es el caso más profundo de un patrón que se repite en todas las dimensiones del sistema. El LLM configura tramas complejas con fluidez — y esa misma capacidad le impide comprometerse con la referencia, porque la coherencia narrativa es independiente de la factualidad: una novela es coherente sin referir nada real, y el sistema no distingue los dos registros. Sintetiza fuentes con precisión aparente — y por ello no distingue el saber fundamentado de la confabulación. Produce actos de habla con fuerza pragmática — pero sin capacidad de evaluar si cada acto está justificado en su contexto. La potencia configuradora y la incapacidad validativa son el mismo mecanismo visto desde ángulos distintos.

La potencia configuradora es exactamente lo que hace peligrosa la falta de validación.

La consecuencia teórica es importante: la asimetría no es una deficiencia contingente que mejor entrenamiento o mayor escala resolverán. Es estructural. Escalar el modelo produce mayor competencia en coherencia, no mayor corrección epistémica: un modelo más grande genera confabulaciones más fluidas y convincentes, no menos confabulaciones. Por eso la respuesta adecuada no es técnica sino epistemológica: no se trata de mejorar el LLM, sino de diseñar las condiciones externas de validez que no puede generar desde adentro. La última fila de la figura — el sistema actúa pero no se responsabiliza — no es un ítem más: es el rasgo que cualifica todos los demás. Coherencia sin verdad, configuración sin referencia y síntesis sin discernimiento son síntomas de una operación más profunda: la de un agente que produce sin poder responder por lo producido.

2.2 La triple mímesis: un uso diagnóstico

Tiempo y narración (Ricoeur, 1983) ofrece las categorías para articular la asimetría con mayor precisión. Ricoeur distingue tres momentos del arco hermenéutico. **Mímesis I** — la prefiguración — es el

conjunto de pre-comprensiones prácticas, simbólicas y temporales que un sujeto aporta al narrar: el mundo comprendido antes de ser narrado. **Mímesis II** — la configuración — es el acto por el cual los eventos dispersos se organizan en un todo coherente, transformando la sucesión en trama. **Mímesis III** — la refiguración — es el momento en que la narración configurada intersecta con el mundo del lector y produce efectos prácticos; es donde el conocimiento narrativo se vuelve efectivo y puede ser evaluado, corroborado o refutado.

El arco no es lineal sino circular: Mímesis III retorna a Mímesis I, enriqueciendo la pre-comprensión de la que surgirán nuevas narraciones. Crucialmente, requiere un sujeto: alguien para quien la prefiguración sea pre-comprensión vivida, que configure con intención narrativa, y para quien la refiguración produzca autocomprensión y reorientación práctica.

El uso que aquí se hace de la triple mímesis es **diagnóstico**, no descriptivo. No se afirma que el LLM realice Mímesis II en sentido pleno — eso requeriría intención narrativa y un mundo de experiencia que el sistema no posee —, sino que la estructura triádica permite localizar lo que falta: el arco de retorno hacia Mímesis III. El LLM habita la zona de Mímesis II con competencia extraordinaria: organiza datos en configuraciones coherentes a una escala que supera a cualquier investigador individual. Pero el arco se trunca. No hay Mímesis I ricoeuriana: no dispone de un mundo pre-comprendido desde el cual narrar, sino de un corpus estadístico desde el cual predecir. No hay Mímesis III: no hay mundo al cual retornar, ni sujeto-lector para quien la configuración produzca comprensión, ni práctica en la que el conocimiento sea puesto a prueba. La máquina configura; no refigura.

Conviene detenerse en qué es refigurar, porque ahí se decide el problema. La Mímesis III no es un producto que la narración deja tras de sí, sino un acontecimiento: el momento en que lo configurado vuelve al mundo de la acción y del tiempo, y alguien lo recibe como suyo. Una trama solo refigura cuando un lector la apropia, reorienta desde ella su comprensión y la lleva a una práctica en la que puede ser puesta a prueba. La refiguración es, por eso, el cruce entre el texto y un mundo de consecuencias: sin ese cruce, lo configurado queda suspendido — completo en su coherencia interna, pero sin alcanzar el registro en el que algo se valida o se refuta.

Ese cruce exige un quién, y es lo que permite leer la refiguración como momento de imputación. La Mímesis III de *Tiempo y narración* (1983) se deja iluminar por la imputabilidad de *Sí mismo como otro* (1990): el sujeto capaz de reconocerse como autor de sus actos y de responder por ellos. Apropiarse de lo configurado es poder decir «es mío lo que fue narrado» y, cuando lo narrado tiene forma de saber, «respondo de lo que afirmo». Refigurar y asumir como propio son, en el fondo, el mismo gesto: no hay

retorno al mundo sin alguien sobre quien recaiga lo dicho. Articular ambos textos es una lectura de este trabajo, no doctrina explícita de Ricoeur, pero se justifica porque la refiguración carece de sentido sin un sujeto que la sostenga.

El LLM configura con fluidez precisamente porque no necesita ese gesto: predice el token siguiente sin que nada de lo producido recaiga sobre él — y lo que lo dispensa de apropiarse es también lo que se lo impide. Configura sin que la configuración sea de nadie. El déficit, entonces, no es la ausencia de comprensión, que §1 ya estableció, sino algo más preciso: la ausencia de un sujeto al que imputar lo configurado. Hay un agente que produce, y un mundo afectado por lo que produce, pero no hay quien pueda decir «es mío» y responder por ello. A esa asimetría — configurar y actuar sin poder ser tenido por autor de lo configurado — la llamamos **agencia sin imputación**. La pregunta que abre el §3 es cómo nombrar el fallo de un agente así cuando produce algo con la forma de una aserción, sin recaer en la metáfora mentalista de la «alucinación».

3. Infortunios de autoridad epistémica

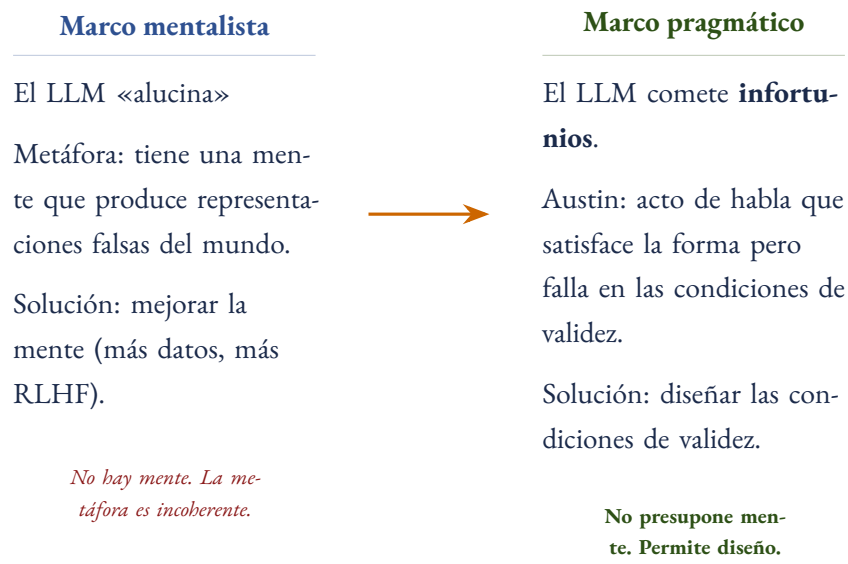
> *«Decir algo es hacer algo.»*
>
> — J. L. Austin, *Cómo hacer cosas con palabras*

Nombrar con precisión lo que falla cuando un LLM produce un resultado incorrecto no es taxonomía neutral: el nombre determina dónde se busca la solución. Si el LLM «alucina», la solución es, en última instancia, psiquiátrica — corregir la mente del sistema, darle mejores datos. Si comete *infortunios de autoridad epistémica*, la solución es procesal — diseñar las condiciones externas de validez que el sistema no genera internamente. La elección es arquitectónica, no estilística: determina el tipo de solución concebible. La metáfora mentalista presupone una mente que distorsiona; la categoría austiniana presupone un agente que actúa sin satisfacer las condiciones del acto. El cambio de marco no elimina al agente: lo redescribe.

El término «alucinación» migró de la psicopatología al discurso sobre IA cuando los ingenieros buscaron una metáfora para audiencias no técnicas. La que eligieron presupone exactamente lo que el §1 mostró que el LLM no tiene: una mente que construye representaciones del mundo y, bajo ciertas condiciones, las distorsiona. Pero el LLM no construye representaciones en ningún sentido relevante: produce secuencias de tokens que satisfacen expectativas estadísticas. Cuando genera un token factualmente falso, no ha fallado en representar la realidad — ha seguido el mismo proceso que cuando genera uno verdadero. La metáfora no es solo imprecisa: es incoherente respecto del objeto que pretende describir.

3.1 Los dos marcos para el error

La alternativa no es buscar una metáfora más precisa, sino abandonar el marco mentalista y sustituirlo por uno pragmático. El marco mentalista trata el output como producto de un sistema cognitivo que representa el mundo y, al fallar, lo distorsiona; su solución es mejorar el sistema. El marco pragmático trata el output como un acto de habla que satisface o no las condiciones que lo hacen válido; su solución es diseñar esas condiciones externamente. El desplazamiento es radical: de la psicología a la pragmática, de la mente al proceso, de la corrección del sistema a la validación del output.



El marco pragmático tiene la ventaja de ser honesto sobre lo que el LLM es: no postula una interioridad que no posee. Trabaja con lo observable — la estructura de los actos de habla y las condiciones de su validez —, y esa austeridad ontológica es justamente lo que lo hace más poderoso para el diseño: si el problema es la ausencia de condiciones de validez, el diseño puede construirlas; si fuera una mente que falla, solo cabría esperar que falle menos.

Del infortunio convencional al infortunio epistémico

En Austin, *infelicity* cubre los actos de habla **performativos** que fallan por no satisfacer sus condiciones convencionales — el juez sin jurisdicción, la promesa bajo coacción —, no los enunciados constatativos simplemente falsos, que son errores ordinarios. La categoría señala una clase específica de fallo en la dimensión performativa, donde lo que falla no es el contenido sino las condiciones que legitiman el acto.

Aplicar el concepto al LLM requiere una extensión justificada. El output del LLM no es nunca puramente constatativo: cuando responde a una pregunta, sintetiza una bibliografía o propone una categoría, no emite un modesto «puede ser que p», sino un acto con **fuerza epistémica** — presenta p como resultado de un razonamiento, con la autoridad implícita de quien ha procesado evidencia y

llegado a una conclusión fundamentada. Esa presentación de autoridad es la dimensión performativa irreducible del output, inscrita en el modo en que el sistema produce discurso, pues fue entrenado sobre texto humano donde esa fuerza es constitutiva del género.

El fallo del LLM no consiste en decir algo falso — eso sería un error constatativo ordinario, que también cometen los humanos —, sino en producir ese contenido *como si* satisficiera las condiciones que fundamentan la autoridad de una aserción, sin poder satisfacerlas estructuralmente. A esto se propone llamar **infortunio de autoridad epistémica**: el acto adopta la forma de una aserción fundamentada cuando la fundamentación es constitutivamente inaccesible desde adentro del sistema. La autoridad de una aserción no es propiedad del enunciado, sino del agente que se compromete a responder por él; el infortunio ocurre cuando un agente real produce el acto sin ser ese tipo de agente. La implicación es inmediata: si la fundamentación es inaccesible desde adentro, debe construirse desde afuera.

3.2 Anatomía del infortunio

Un acto de habla epistemológicamente válido debe satisfacer al menos cuatro condiciones. El LLM cumple consistentemente solo la primera:

La **forma** es la condición que el LLM mejor satisface — oraciones gramaticales, pragmáticamente apropiadas, retóricamente calibradas —, y esa perfección de superficie es la fuente del peligro: oculta el déficit de fondo. La **referencia** se satisface parcial e impredeciblemente, sin marca formal que distinga lo correcto de lo fabricado desde adentro del output. La **sinceridad** — que el hablante crea lo que afirma — es estructuralmente imposible: no hay sujeto que crea ni interioridad desde la cual comprometerse. La **validez** — que el hablante pueda dar razones — es igualmente inaccesible: el sistema no tiene acceso a las razones que producen su output ni puede distinguir sus afirmaciones justificadas de las injustificadas. Las dos últimas no son deficiencias que un modelo mejor resolverá: detrás del output no hay razones, sino distribuciones de probabilidad sobre vocabularios, y las distribuciones no son razones. El sistema produce actos que reclaman el régimen de la justificación sin poder habitarlo.

3.3 Habermas y Floridi: validez como justificabilidad

Habermas (1981) permite precisar el diagnóstico. Todo acto de habla serio levanta tres pretensiones de validez que el hablante está comprometido a defender si se le interpela: **verdad** (correspondencia con el mundo objetivo), **rectitud normativa** (adecuación a las normas del contexto) y **veracidad** (expresión genuina de lo que se piensa). La validez no es propiedad del contenido sino de la capacidad del hablante de sostenerlo argumentativamente: un acto es válido cuando el hablante puede, en principio, dar las razones que lo respaldan. El LLM puede producir afirmaciones verdaderas y normativamente apropiadas, pero no satisface ninguna pretensión en el sentido de la *justificabilidad*: cuando se le pide justificar una afirmación, produce una nueva cadena de tokens con la morfología de una justificación, no una justificación. La pretensión de veracidad es imposible: no hay interioridad que expresar.

La consecuencia es visible en la discrepancia entre rendimiento y validez. La ingeniería mide rendimiento — *accuracy*, tasa de error —, pero rendimiento no es validez. Un LLM puede acertar el 95 % de un benchmark y ser epistemológicamente ingobernable, porque el 95 % correcto se produce por las mismas razones estadísticas que el 5 % incorrecto y no hay diferencia interna que el sistema pueda señalar. La distinción no es de grado sino de categoría: el rendimiento mide aciertos; la validez mide la capacidad de dar razones. El marco de la «alucinación», al centrarse en el rendimiento, orienta la búsqueda de soluciones en la dirección equivocada.

Floridi (2019; Floridi & Sanders, 2004) converge desde la filosofía de la información y añade dos piezas. Su distinción entre **datos** — señales que no requieren verdad — e **información semántica** — contenido verdadero o falso, con significado en un contexto — nombra el estatus del output con exactitud: los LLM producen datos con la morfología de la información semántica, aléticamente neutros hasta que un proceso externo los evalúa. Y su caracterización del sistema como agente (interactividad, autonomía, adaptabilidad, sin estados mentales) resiste tanto la antropomorfización como la negación instrumentalista. Los LLM satisfacen las tres condiciones: son agentes. Lo que no satisfacen es lo que Ricoeur llama imputabilidad.

Austin, Habermas y Floridi convergen así en un diagnóstico estructural que ninguno formula por separado: el infortunio ocurre al ejecutar un acto sin las condiciones convencionales que lo legitiman (Austin); el LLM no puede dar razones porque ningún sujeto comunicativo asume la responsabilidad (Habermas); hay agencia operativa pero no imputable, y la fuente del infortunio está en esa asimetría (Floridi). El LLM es un agente sin imputación que produce actos performativos sin agente que pueda fundarlos. De ahí la pregunta de diseño que guía el §4: si la imputación no puede venir de adentro, ¿qué arquitectura puede ubicarla donde corresponde — en un agente humano que sí pueda responder por lo producido?

4. La hermenéutica como protocolo epistémico

El diagnóstico del §3 determina la forma de la solución. Si el LLM produce infortunios porque carece internamente de las condiciones de validez, la respuesta no puede ser solo técnica: debe ser metodológica — construir, desde afuera, un protocolo que especifique qué hacer con el output, en qué orden, con qué criterio determinar si cada paso fue correcto y bajo qué condición el proceso es suficiente. Responder qué protocolo satisface esas condiciones requiere examinar primero por qué los protocolos actuales de razonamiento artificial no lo hacen, y luego identificar en la investigación cualitativa — la Teoría Fundamentada y la Triangulación — las estructuras que sí.

4.1 La genealogía del criterio: de CoT a la IA agéntica

Chain-of-Thought, Tree-of-Thought, ReAct y la IA agéntica representan saltos genuinos en capacidad de Mímesis II. Pero su genealogía revela una constante: cada generación añade potencia de configuración sin incorporar un criterio epistémico que opere en el nivel de Mímesis III.

Técnica	Lo que incorpora	Lo que no resuelve
CoT Wei et al. 2022	Secuencia explícita de razonamiento	*¿Criterio de la secuencia?*
Tree-of-Thought Yao et al. 2023	Exploración de caminos paralelos	*¿Criterio de selección del camino?*
ReAct Yao et al. 2022	Retroacción ambiental: acción + observación	*¿Justificación epistémica de la acción?*
IA Agéntica 2024–	Autonomía operacional a escala	*¿Quién valida la cadena de infortunios?*

Déficit invariante: ninguna técnica especifica el criterio por el cual un paso de razonamiento es epistemicamente válido.

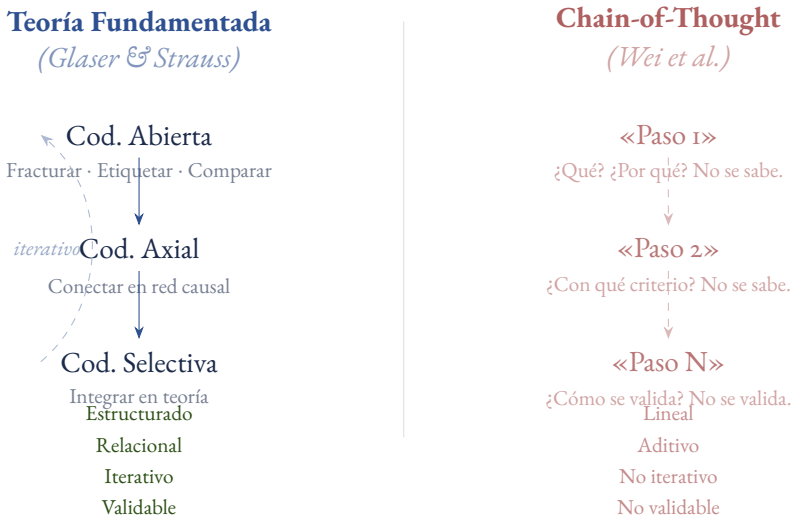
CoT (Wei et al., 2022) mejora el rendimiento al instruir al LLM a «pensar paso a paso», pero carece de teoría de la secuencia: no especifica qué pasos son necesarios, en qué orden ni cuándo es legítimo volver atrás. Es una cadena de razonamiento sin logos, sin mecanismo de corrección que opere sobre algo distinto al output del paso anterior. ToT (Yao et al., 2023) supera la linealidad explorando caminos paralelos, pero su criterio de selección sigue siendo heurístico — coherencia o fluidez, no justificabilidad, que el proceso de generación no puede evaluar desde adentro. ReAct (Yao et al., 2022) alterna razonamiento y acción sobre herramientas externas, introduciendo retroalimentación ambiental; pero esa retroalimentación es instrumental: evalúa si algo funcionó, no si estaba justificado hacerlo — la distinción entre eficacia técnica y validez comunicativa.

La IA agéntica (2024–) coordina múltiples agentes, mantiene memoria y ejecuta planes con supervisión mínima. La industria los llama *agents*: exacto en sentido floridiano, impreciso en cualquier sentido

más exigente. Aquí el déficit de criterio deja de ser un problema de salida y se vuelve un problema de propagación: cada eslabón hereda el infortunio de los anteriores sin mecanismo de detección. Un agente genera hipótesis a partir de inferencias no justificadas, otro las usa como premisas, un tercero las ejecuta en el mundo — y la cadena de infortunios resulta irrastreable. La cadena tiene agentes en cada eslabón; lo que no tiene es alguien que pueda ser tenido por autor de la cadena. Nombrar el problema «errores en cascada» o «deriva del agente» no lo localiza: la responsabilidad no recae sobre nadie, y diseñar el correctivo exige nombrarlo así desde el inicio. El déficit de criterio es, en suma, invariante respecto de la sofisticación técnica. La hermenéutica aplicada tiene esa especificación desde mucho antes de que existieran los LLM.

4.2 La Teoría Fundamentada como protocolo

La Teoría Fundamentada (*Grounded Theory*), de Glaser y Strauss (1967) y sistematizada por Strauss y Corbin (1990), genera teoría a partir de datos mediante un proceso iterativo y estructurado. Lo que la distingue no es solo su compromiso con los datos, sino la precisión con que especifica las operaciones que producen conocimiento: cada fase tiene un nombre, una función, un criterio de ejecución correcta y una condición de finalización. El contraste con CoT no es de escala sino de naturaleza: CoT deja al sistema decidir qué es un paso y cuándo la cadena es suficiente; GT fractura el material en unidades, las etiqueta, las compara constantemente, construye relaciones causales y declara saturación cuando no surgen nuevos códigos. Cada instrucción es un criterio operacionalizable, verificable externamente. GT especifica no solo qué hacer, sino cómo saber si se hizo bien.



GT y CoT tienen epistemologías subyacentes distintas. GT presupone que el conocimiento emerge de la comparación sistemática, la saturación y la integración teórica; CoT, que se alcanza encadenando pasos plausibles. La primera produce conocimiento auditable; la segunda, texto convincente. La diferencia se ve en una pregunta simple: ¿se puede saber, sin evaluar el output, si el proceso se ejecutó correctamente? Para GT, sí; para CoT, no.

4.3 Traducción procesal: el andamiaje como diseño

Trasladar GT a un pipeline de prompts es un programa de diseño, no una analogía. Las cinco fases del protocolo, con el criterio que cada una introduce, operan así:

Fase 1. Codificación abierta — *Criterio: declaración de límites*.
«Identifica conceptos sin imponer categorías previas. Para cada concepto señala: (a) qué observas, (b) qué no puedes determinar con este segmento, (c) qué dato adicional necesitarías.» La inclusión de (b) y (c) fuerza la declaración explícita de límites que CoT, ToT y ReAct nunca producen.

Fase 2. Comparación constante — *Criterio: trazabilidad por código*.
«Compara este segmento con los conceptos ya identificados. ¿Confirma, contradice, matiza o amplía alguno? Señala el código afectado y la naturaleza de la relación.» El prompt implementa la comparación constante de Glaser sin requerir sensibilidad teórica en el LLM: el criterio está en el diseño, no en la máquina.

Fase 3. Codificación axial — *Criterio: estructura relacional auditable*.
«Para cada código, identifica condiciones de aparición, mecanismos y consecuencias. Construye una red de relaciones señalando su tipo (causal, temporal, condicional, contradictoria).» La distinción condición/mecanismo/consecuencia de Strauss y Corbin se traduce a formato verificable por el humano-en-el-bucle.

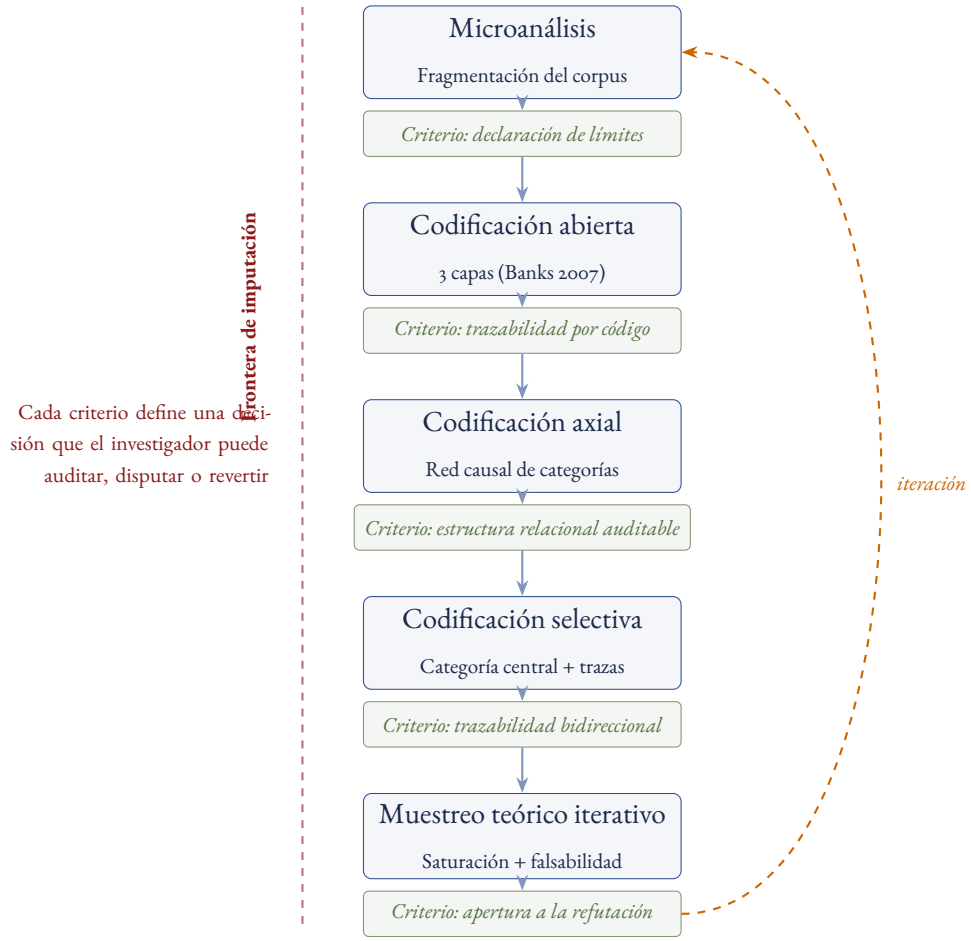
Fase 4. Codificación selectiva — *Criterio: trazabilidad bidireccional*.
«Propón una categoría central que integre los hallazgos. Para cada elemento que incluyes, señala el código del que proviene; para cada uno que excluyes, por qué no es integrable.» La trazabilidad convierte el output en texto auditable: el criterio que la IA agéntica no tiene.

Fase 5. Saturación y falsabilidad — *Criterio: apertura a la refutación*.
«¿Qué tipo de segmento modificaría la categoría central? ¿Qué dato podría falsificarla?» La pregunta por la falsabilidad — ausente en CoT, ToT y ReAct — transforma el proceso en investigación. Un sistema que no puede especificar las condiciones de su propio error no produce conocimiento: produce texto con aspecto de conocimiento.

La diferencia con cualquier técnica de razonamiento artificial no está en el número de pasos sino en el tipo de criterio que cada uno incorpora. Esos criterios permiten determinar si cada fase se ejecutó correctamente sin evaluar el output por su coherencia interna. Esa auditabilidad externa es lo que hace al protocolo epistemológicamente robusto: CoT tiene pasos; GT tiene fases con condiciones de verificación.

4.4 El prototipo: análisis de la cultura visual narco

Las cinco fases no son hipotéticas. El sistema *Descubrimiento de Categorías en Imágenes NarcoColombianas* realizó este protocolo sobre un corpus que recorre casi un siglo de circulación visual sobre el fenómeno narco en Colombia: prensa, fotografías institucionales de operativos, retratos de figuras emblemáticas, escenas militares de erradicación y materiales digitales contemporáneos. El LLM es Google Gemini en modalidad de visión, y el sistema es un pipeline iterativo de cinco módulos — uno por fase — coordinados por un módulo de muestreo teórico que decide qué imágenes ingresan al ciclo siguiente y cuándo declarar saturación.



Arquitectura algorítmica del prototipo. Cada caja superior es una operación del sistema sobre el corpus; cada caja inferior es el criterio externo que la hace auditable. La iteración cierra el ciclo de muestreo teórico; la línea vertical marca la frontera donde la imputación regresa al investigador.

Cada caja es una operación que el sistema ejecuta sobre las imágenes, y cada criterio es la condición externa que la hace verificable sin reconstruir el proceso desde cero. La arquitectura no oculta que es una máquina la que hace el trabajo: lo expone, lo descompone en pasos discretos y deja a la vista los puntos donde el resultado puede disputarse. La auditabilidad no le devuelve al sistema la capacidad de responder por lo que hace — esa frontera no se cruza —, pero ubica la imputación donde corresponde: en el investigador que verifica, acepta o rechaza cada criterio.

El sistema ejecutó ciclos de muestreo teórico hasta declarar saturación. La codificación axial produjo 21 categorías saturadas, integradas por la fase selectiva en una categoría central: **Comodificación aséptica y espectacularización transversal del trauma narco**, definida como el proceso por el cual la violencia visceral, la necropolítica y el capital ilícito son despojados de su carga orgánica y peso moral para transmutarse en productos visuales esterilizados, trofeos institucionales o bienes de consumo pop. La categoría integra tres regímenes por escena — museo/galería (*contención pedagógica y memoria ruinosa*), prensa/hemeroteca (*trofeo-crimen panóptico institucional*), plataformas digitales (*aspiración pop y evasión algorítmica*) — articulados con Rancière (2000), Valencia (2010) y Mbembe (2003), en diálogo con la conceptualización de *lo narco como categoría estética* de Pardo León (en curso). Estas formulaciones, incluida la categoría central, son outputs del sistema, no conclusiones del investigador: que sean productivas o requieran corrección es exactamente lo que la fase de imputación humana debe determinar.

Cuatro imágenes que el sistema ve — y cuatro que el investigador debe imputar

Para entender qué hace el sistema y dónde se detiene su agencia, conviene mostrar el material y los outputs reales del pipeline. Las cuatro imágenes que siguen son representativas de los principales regímenes del archivo: el operativo antinarcótico de los setenta, el retrato del capo, la rueda de prensa institucional de los noventa y la patrulla militar de los dos mil.

(a) Bonanza marimbera, 1970s.
`1970s_traqueteo_medellin_0076.jpg`

(b) Retrato del capo, 1980s.
`1970_1990_nuevas_0001.jpg`

(c) Comunicación institucional, 1990s.
`1990s_Estetica_Yupi_0104.jpg`

(d) Vigilancia militar, 2000s.
`2001_Narcoterrorism_0294.jpg`

Cuatro fragmentos del corpus, cuatro regímenes de visibilidad. Cada imagen fue procesada por el pipeline completo; los outputs reales por fase aparecen tabulados a continuación.

La tabla muestra los outputs literales de las dos primeras operaciones del pipeline: el módulo de *microanálisis* (descripción densa multimodal) y el de *codificación abierta* (códigos emergentes con su memo).

Imagen	Microanálisis (descripción densa)	Codificación abierta (códigos emergentes y memo)
(a) Bonanza marimbera *1970s*	Escena al aire libre: numeroso grupo de hombres jóvenes con camisetas blancas y pantalones cortos (algunos con botas, sugiriendo uniforme de conscripción) destruye material vegetal ilícito en suelo árido. Personal uniformado al fondo. Acción colectiva sobre sacos abiertos.	**Códigos:** Militarización antidrogas; Erradicación pública/simbólica; Trabajo forzado/conscripción militar; Control estatal performático; Dimensión trasnacional del narcotráfico. **Memo:** La imagen se conecta con la narrativa de la «Guerra contra las Drogas» intensificada en América Latina desde mediados del siglo XX; jóvenes reclutas en labor de destrucción masiva ilustran la militarización de la política antidrogas.
(b) Retrato del capo *1980s*	Fotografía en blanco y negro, plano medio. Hombre identificable como Pablo Escobar mira a cámara con expresión calculada que mezcla sonrisa y seriedad; viste camisa abierta y chaqueta en estilo años ochenta. La mujer (presuntamente su esposa María Victoria) compone una escena de aparente normalidad doméstica.	**Códigos:** Figura del «narco» y su entorno familiar; Construcción de imagen pública; Poder y apariencia social; Normalización de la ilegalidad; Rol de la mujer en el crimen organizado. **Memo:** La imagen se inserta en el apogeo del Cartel de Medellín; estética y vestimenta anclan la escena en la década; políticamente, proyecta una imagen de relativa normalidad doméstica que neutraliza la lectura criminal.
(c) Comunicación institucional *1990s*	Primer plano de hombre de mediana edad en traje oscuro de rayas finas, camisa blanca y corbata; gafas de montura dorada. Expresión sonriente, mano derecha levantada con dedo índice apuntando hacia adelante, gesto retórico-asertivo. Contexto formal/profesional sugerido.	**Códigos:** Comunicación política; Autoridad discursiva; Proyección de confianza; Representación institucional; Gestión pública. **Memo:** La imagen encapsula la estética de la figura pública en el escenario político ligado al conflicto armado y las políticas de drogas; atuendo, micrófono y gesto de señalar son iconografía recurrente de funcionarios del Estado en declaraciones.
(d) Vigilancia militar *2000s*	Soldado solitario en uniforme de camuflaje y equipo táctico, boca abajo en cresta rocosa, apuntando rifle de largo alcance fuera de encuadre. Cielo azul claro domina dos tercios superiores; terreno árido inferior; el soldado se mimetiza con el entorno. Marca de agua de agencia internacional visible.	**Códigos:** Vigilancia militar; Conflicto en terreno rural/montañoso; Presencia estatal armada; Guerra asimétrica; Control territorial. **Memo:** La imagen se relaciona con la narrativa visual del conflicto colombiano y la «Guerra contra las Drogas»; emblemática de operaciones de fuerzas armadas a menudo apoyadas por EE. UU. contra grupos armados y organizaciones de narcotráfico.

Outputs literales del pipeline para las cuatro imágenes representativas. Fuente: `codificacion_abierta.json` *del sistema, sin edición editorial.*

El sistema produce, en cada caso, lo que parece un análisis cualitativo competente: las descripciones son densas y referencialmente correctas en lo visible; los códigos son consistentes con vocabularios sociotécnicos sobre el narcotráfico; los memos articulan asociaciones históricas plausibles. Y sin embargo, en ninguna fila la operación se completa: cada celda contiene una propuesta sin un sujeto que la sostenga.

La primera fila ilustra el caso más limpio: sobre el operativo hay correspondencia entre lo visible y lo nombrado. Pero el código *Control estatal performático* no describe: interpreta. Sostener esa interpreta-

ción como tesis — que el operativo *es* performance de soberanía y no mero procedimiento policial — exige argumentar contra otras lecturas y responder por su consecuencia política; el sistema generó el código por similitud estadística con corpus académicos, pero no lo defendió ni puede defenderlo. La segunda fila vuelve palpable la *agencia sin imputación*: el sistema codifica la construcción de imagen pública y la normalización de la ilegalidad, e incluso nombra el efecto de neutralización moral; pero la pregunta que toda investigación sobre el narco debe responder — ¿cómo se relaciona esta estetización con la responsabilidad histórica por las víctimas? — no la puede formular. Las dos últimas filas pertenecen al régimen institucional-militar: la rueda de prensa codifica autoridad discursiva; el soldado, control territorial y guerra asimétrica. Lo que el sistema no puede ver es la continuidad entre ambos — que el fiscal en Washington y el francotirador en la colina son momentos del mismo dispositivo. Esa continuidad no es un código: es una hipótesis que requiere argumentación, exposición a refutación y responsabilidad por la decisión teórica. La codificación produce los fragmentos; integrarlos en un dispositivo — y decidir si se nombra, con Mbembe, necropolítica, o con Valencia, capitalismo gore — requiere un agente imputable.

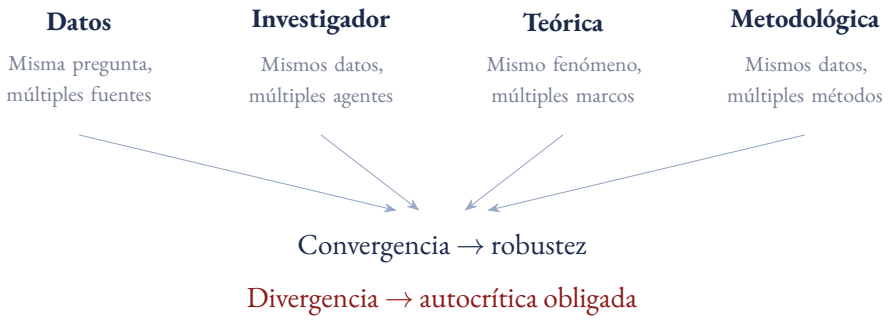
4.5 Logros y persistencias del déficit

El prototipo permite evaluar qué compensa el andamiaje y dónde el déficit de Mímesis III persiste. Logra, primero, auditabilidad real: cada código se vincula a su imagen de origen, su capa de análisis (Banks, 2007) y sus relaciones, de modo que el evaluador puede trazarlo y refutarlo sin reconstruir el proceso. Logra estructura relacional verificable: las categorías axiales son esquemas con condiciones, mecanismos y consecuencias contrastables con el corpus. Y logra un criterio de salida explícito: el ciclo no termina arbitrariamente, sino cuando los nuevos segmentos no producen nuevos códigos.

Tres tensiones, sin embargo, revelan dónde el arco sigue truncado. **Primera**: la saturación la evalúa el mismo LLM que produjo los códigos — un infortunio de segundo orden, pues el sistema declara haberse validado sin un sujeto externo que valide la declaración. **Segunda**: la integración teórica es descendente — Rancière, Valencia y Mbembe están codificados como esquemas de aplicación en los prompts; el sistema aplica la teoría, no la genera, lo que contradice el espíritu inductivo de GT. **Tercera**: la comparación constante es autorreferencial — el mismo modelo que genera los códigos los compara, cuando Glaser exigía sensibilidad teórica externa. Estas tensiones definen lo que el investigador debe aportar: validar que la categoría central emerge de los datos y no fue proyectada; decidir si la saturación es genuina o prematura; justificar la selección teórica; y asumir la responsabilidad epistémica final. Esa lista no describe supervisión opcional, sino el lugar exacto donde la imputación regresa al circuito. El andamiaje no inventa la imputabilidad; la *ubica*. Sin esa ubicación, el sistema produce ilusión de rigor — la advertencia de Vallor (2016): una prótesis mal usada atrofia la capacidad que pretende compensar.

4.6 Triangulación: control de la sobreconfianza

La Teoría Fundamentada resuelve el criterio dentro del proceso, pero no la convergencia entre fuentes, métodos y perspectivas. Ese es el terreno de la triangulación, introducida por Denzin (1970): observar el mismo fenómeno desde ángulos múltiples e independientes, no para confirmar con más de lo mismo, sino para explotar la complementariedad de perspectivas y detectar lo que cada una no ve. Denzin distinguió cuatro modalidades:



La triangulación de **datos** aborda la misma pregunta desde múltiples fuentes; la de **investigador**, los mismos datos desde múltiples agentes — en un sistema LLM, sesiones independientes con prompts equivalentes cuyos outputs se comparan; la **teórica**, el mismo fenómeno desde marcos distintos; la **metodológica**, los mismos datos con métodos diferentes. La convergencia produce robustez: cuando fuentes, agentes, marcos y métodos coinciden, disminuye la probabilidad de que el hallazgo sea un artefacto del método. Pero la función más importante se activa ante la divergencia: vuelve obligatoria la autocrítica. La divergencia entre dos sesiones paralelas no es un problema a resolver eligiendo el resultado más plausible, sino una señal que exige examinar por qué los procesos difirieron; hace visibles los sesgos del modelo, la sensibilidad a las formulaciones del prompt y los puntos genuinamente indeterminados. La divergencia es información epistémica de primer orden, no ruido.

Juntas, GT y Triangulación constituyen el andamiaje hermenéutico: la primera provee la estructura iterativa y los criterios internos; la segunda, el control externo y la autocrítica. No eliminan el déficit de Mímesis III — el arco sigue requiriendo un sujeto que lo complete —, pero lo hacen visible, estructuran la contribución del investigador y producen un output que el humano puede evaluar, corroborar y rechazar.

5. La prótesis hermenéutica

El andamiaje del §4 resuelve el problema procesal, pero no es solo un conjunto de procedimientos: es, desde la filosofía de la tecnología, una forma de mediación que altera la relación entre un agente y su objeto al interponer una herramienta. Esa mediación requiere fundamento filosófico, porque determina cómo se distribuyen las funciones epistémicas entre sistema e investigador.

La categoría que se propone para nombrarla es la de **prótesis hermenéutica**. Conviene precisar un punto que de otro modo abriría una contradicción aparente: si el LLM es agente, llamarlo «prótesis» parecería un retroceso a la imagen instrumental que el §1 rechazó. No lo es. Lo que aquí se llama prótesis no es el LLM aislado, sino el sistema-andamiaje-LLM articulado al investigador. El LLM conserva su agencia operativa — sigue interactuando, decidiendo, adaptándose —, pero esa agencia queda acoplada, mediante el andamiaje, al circuito de imputación del investigador. La palabra *prótesis* designa una extensión funcional: no el reemplazo de una capacidad ausente, sino la ampliación de una capacidad presente bajo condiciones de limitación. El calificativo *hermenéutica* especifica el tipo de extensión: no computacional sino interpretativa, gobernada por los protocolos de la comprensión y la validación que la investigación cualitativa ha operacionalizado. Una prótesis hermenéutica es, entonces, un sistema en el que el LLM — agente operativo en sentido de Floridi — extiende la capacidad configuradora del investigador (Mímesis II) mientras este — agente imputable en sentido de Ricoeur — preserva y ejerce la capacidad validativa (Mímesis III). La frontera entre ambas funciones, y no su fusión, es el principio de diseño.

5.1 Fundamentos filosóficos del diseño

Tres tradiciones convergen en el fundamento del concepto. La **teoría de la mente extendida** (Clark, 2003) explica por qué el acoplamiento es posible: los procesos cognitivos no están confinados al organismo; cuando una herramienta se acopla de manera fiable y fluida a los procesos epistémicos de un agente, pasa a ser constitutiva de su sistema cognitivo. Los lentes, el papel, la calculadora — y, bajo las condiciones adecuadas, el LLM — no son apoyos externos a la cognición sino componentes de ella. El acoplamiento entre investigador y LLM no es accidental sino constitutivo de la arquitectura epistémica, y el andamiaje hermenéutico es precisamente su especificación.

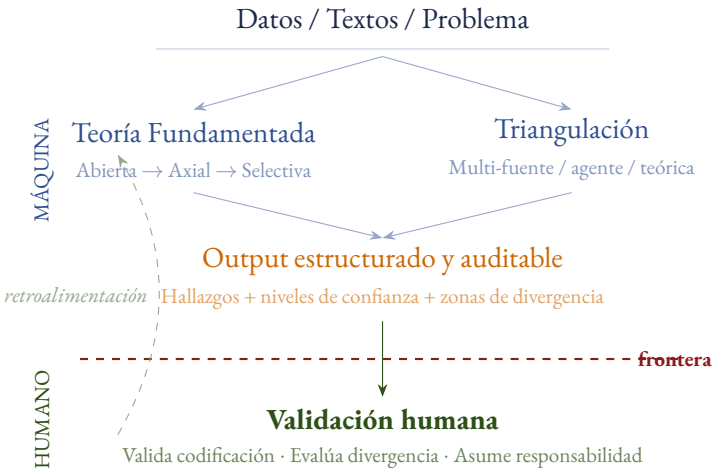
La **fenomenología** explica cómo el acoplamiento se experimenta como continuidad. Merleau-Ponty (1945) mostró que el esquema corporal incorpora herramientas cuando el agente alcanza fluidez: el bastón del ciego no es algo que toca sino algo con lo que toca. El investigador que usa el andamiaje de manera experta no opera *sobre* el sistema sino *a través* de él. Varela, Thompson y Rosch (1991) añaden la dimensión enactivista: la cognición es acción encarnada, no manipulación de representaciones. Para la prótesis hermenéutica, esto significa que la validación (Mímesis III) no puede delegarse por completo al sistema, porque requiere el acoplamiento del investigador con las consecuencias prácticas de las interpretaciones: un mundo al cual retornar y en el cual los hallazgos sean evaluados por sus efectos.

La **teoría crítica de la tecnología** explica qué condiciones políticas y éticas debe cumplir el acoplamiento. Feenberg (2002, 2017) distingue entre **instrumentalización primaria** — la operación técnica en abstracto: descontextualización, reducción al funcionamiento — e **instrumentalización secundaria** — el encastre social: sistematización, mediación, apropiación significativa en un contexto de práctica. Los LLM operan por defecto en la capa primaria: despliegan agencia operativa sin estar acoplados a un horizonte de imputación. El andamiaje hermenéutico es exactamente instrumentalización secundaria:

el conjunto de mediaciones que devuelve la operación técnica a un contexto de práctica, a criterios de validación y al sujeto que puede ser tenido por autor del juicio. Feenberg llama a este movimiento **racionalización democrática**, y proponerlo no contradice el diagnóstico crítico sobre los LLM: es su consecuencia práctica. Vallor (2016, 2023) añade la dimensión que ninguna arquitectura puede ignorar: el andamiaje debe cultivar en el investigador las **virtudes epistémicas** — apertura a la refutación, reflexividad, tolerancia a la incertidumbre — que hacen posible la validación genuina. Una prótesis mal diseñada atrofia las capacidades que pretende compensar: si el investigador delega la codificación sin comprenderla y acepta la saturación sin interrogarla, el andamiaje no produce conocimiento más robusto, sino ilusión de robustez con mayor eficiencia.

5.2 El concepto y la arquitectura

La prótesis hermenéutica puede definirse ahora con precisión: un sistema de producción de conocimiento en el que un LLM opera como extensión cognitiva del investigador en la configuración (Mímesis II), mientras el investigador preserva y ejerce la validación (Mímesis III), y el andamiaje hace visible y auditable la frontera entre ambas. No es una IA consciente — imposibilidad estructural —, sino una IA **epistémicamente gobernable**: un sistema cuyo proceso puede auditarse, cuyos límites están declarados y cuyas fallas pueden detectarse sin esperar sus consecuencias.



La máquina contribuye lo que hace mejor que el investigador individual: velocidad sobre corpus masivos, consistencia en la aplicación de protocolos, identificación de patrones a escala y generación sistemática de hipótesis relacionales. Esa contribución es genuina, y reducirla a la imagen del instrumento pasivo distorsiona lo que el sistema hace. Pero la máquina no contribuye validación, responsabilidad ni juicio: esas funciones están, por diseño, del otro lado de la frontera, porque su ejercicio requiere un agente que pueda ser tenido por autor de la decisión, figura que solo encarna el investigador. El humano valida que los códigos reflejan el corpus y no proyectan categorías preestablecidas; decide sobre la suficiencia de la saturación; justifica la selección teórica; y, sobre todo, asume la responsabilidad epistémica final — el momento en que las conclusiones dejan de ser output del sistema y se vuelven aserciones del

investigador, defendibles ante una comunidad de pares. El investigador no es un supervisor que aprueba outputs: es el agente que cierra el arco que el sistema deja truncado. Sin andamiaje, esa frontera es invisible, y el riesgo es aceptar outputs como si ya fueran validaciones; con andamiaje, cada output incluye su procedencia, sus límites declarados y las condiciones bajo las cuales debería ser rechazado. La frontera no es una barrera, sino una interfaz: el lugar donde la máquina termina y el humano comienza, marcado con precisión para que ninguno ocupe el lugar del otro.

6. Conclusión: del diagnóstico al diseño

El recorrido fue del diagnóstico al diseño. El §1 mostró que la pregunta ontológica es legítima pero insuficiente, y que de la carencia ontológica no se sigue la inutilidad epistémica. El §2 localizó al LLM en el arco de Ricoeur: una máquina de Mímesis II cuyo arco se trunca, sin mundo al cual retornar ni sujeto que se impute lo configurado. El §3 recategorizó el error como infortunio de autoridad epistémica: actos que reclaman el régimen de las razones sin poder habitarlo. El §4 construyó el andamiaje — Teoría Fundamentada como protocolo con criterio por fase, Triangulación como control externo, y el prototipo como demostración —, y el §5 nombró la arquitectura resultante: la prótesis hermenéutica como redistribución agencial entre un agente operativo y uno imputable.

6.1 La estructura del argumento

El argumento es una cadena, no una lista: cada resolución es condición de posibilidad de la siguiente.

Tensión	Resolución	Vía
Ontología vs. Epistemología	→ Giro epistemológico	§1
Comprensión vs. Configuración	→ Mímesis II (Ricoeur)	§2
Validez vs. Rendimiento	→ Infortunios (Austin)	§3
Heurística vs. Método	→ GT + Triangulación	§4
Potencia vs. Responsabilidad	→ Prótesis hermenéutica	§5

El giro ontología/epistemología (§1) hace posible el diagnóstico configuración/validación (§2), que hace posible la recategorización del error (§3), que hace posible el protocolo con criterio (§4), que hace posible el concepto de prótesis hermenéutica (§5). La tesis que recorre la cadena: las tensiones que bloquean el debate sobre los LLM son epistemológicas, no ontológicas, y son por tanto resolubles mediante diseño procesal informado por la hermenéutica aplicada. El diagnóstico ontológico — el LLM no comprende — es correcto; la conclusión — nada que diseñar — es una inferencia inválida. La

pieza que vuelve productiva esa invalidez es la distinción entre dos modos de agencia: el LLM es agente operativo, no agente imputable, y el diseño del andamiaje ubica cada modo donde corresponde.

6.2 Contribución, límites y líneas de investigación

La contribución tiene cuatro componentes. **Conceptual**: la categoría de *infortunio de autoridad epistémica* como alternativa precisa a «alucinación», y la de *agencia sin imputación* como descripción del agente que lo comete; juntas, el desplazamiento de la psicología a la pragmática y de la corrección del sistema a la validación del output. **Metodológica**: la demostración de que GT puede traducirse a un pipeline de prompts con un criterio de validez explícito por fase — declaración de límites, trazabilidad por código, estructura relacional auditable, trazabilidad bidireccional, apertura a la refutación —, propiedades ausentes en CoT, ToT, ReAct y la IA agéntica. **Arquitectónica**: el concepto de *prótesis hermenéutica* como marco que distribuye funciones epistémicas de manera explícita y auditable, fundado en Clark, Merleau-Ponty, Varela, Feenberg y Vallor. **Empírica**: la implementación que convierte la propuesta de una posibilidad teórica en un programa de investigación evaluable, demostrando tanto lo que el andamiaje logra como dónde el déficit persiste.

Los límites deben ser explícitos. El andamiaje no elimina el déficit epistémico del LLM: las tres tensiones del §4.5 son condiciones estructurales que el diseño hace visibles pero no resuelve; lo que hace es desplazar el problema al espacio de la interacción entre sistema e investigador. La Teoría Fundamentada no es el único andamiaje posible — es el más adecuado para el análisis cualitativo de corpus textuales e iconográficos; otros contextos requerirían criterios distintos. Y la pregunta epistemológica no reemplaza la ontológica: la crítica de Searle, Dreyfus, Gadamer y Heidegger sigue siendo correcta y relevante. La ontología y la epistemología no compiten: la segunda presupone la primera.

Tres líneas se abren. La **empírica**: un programa de pruebas comparativas que contraste el andamiaje GT/Triangulación con CoT, ToT y arquitecturas agénticas en tareas definidas, con métricas que incluyan auditabilidad del proceso y trazabilidad de los hallazgos, no solo precisión factual. La **política y de ecología epistémica**: los andamiajes implementados a escala en universidades, redacciones o juzgados no son neutrales — estructuran qué cuenta como investigación rigurosa y qué conocimientos resultan excluidos por el diseño —, una dimensión que requiere análisis filosófico y social independiente. La **cosmotécnica y de exterioridad**: Yuk Hui (2019) argumenta que toda técnica está enraizada en una cosmología particular; la Teoría Fundamentada tiene raíces en la sociología norteamericana, la hermenéutica que la fundamenta es europea, y los LLM se entrenan sobre corpus con sobrerrepresentación del norte global. Dussel, Escobar y Quijano permiten preguntar, desde la tradición filosófica latinoamericana, qué formas de conocimiento excluye estructuralmente un andamiaje que no fue diseñado para incluirlas. Responder no invalida la propuesta: la sitúa.

La pregunta inicial — ¿cómo se valida el conocimiento producido por un sistema que actúa y genera discursos coherentes sin participar en las condiciones de la comprensión ni de la responsabilidad? — no tiene respuesta definitiva. Lo que este trabajo muestra es que puede reorientarse productivamente:

en lugar de buscar en la máquina capacidades que no puede tener, reconocer la agencia operativa que despliega y diseñar, en torno a ella, las condiciones que permiten que la imputación regrese al sujeto que sí puede sostenerla. La prótesis hermenéutica no es la solución final de un problema filosófico: es la forma de un programa de investigación que convierte un callejón sin salida ontológico en un espacio abierto de diseño epistemológico.

Bibliografía

Obras nucleares del argumento

[1] Austin, J. L. (1982). *Cómo hacer cosas con palabras* (G. R. Carrió & E. A. Rabossi, trads.). Paidós. (Obra original publicada en 1962).

[2] Feenberg, A. (2002). *Transforming technology: A critical theory revisited*. Oxford University Press.

[3] Floridi, L. (2019). *The logic of information: A theory of philosophy as conceptual design*. Oxford University Press. https://doi.org/10.1093/oso/9780198833635.001.0001

[4] Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, *14*(3), 349–379. https://doi.org/10.1023/B:MIND.0000035461.63578.9d

[5] Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine.

[6] Habermas, J. (1987). *Teoría de la acción comunicativa* (M. Jiménez Redondo, trad.). Taurus. (Obra original publicada en 1981).

[7] Ricoeur, P. (1995). *Tiempo y narración I: Configuración del tiempo en el relato histórico* (A. Neira, trad.). Siglo XXI. (Obra original publicada en 1983).

[8] Ricoeur, P. (1996). *Sí mismo como otro* (A. Neira, trad.). Siglo XXI. (Obra original publicada en 1990).

[9] Smith, B. C. (2019). *The promise of artificial intelligence: Reckoning and judgment*. MIT Press.

[10] Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190498511.001.0001

Obras complementarias

[11] Banks, M. (2010). *Los datos visuales en investigación cualitativa* (T. del Amo, trad.). Morata. (Obra original publicada en 2007).

[12] Clark, A. (2003). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*. Oxford University Press.

[13] Denzin, N. K. (1970). *The research act: A theoretical introduction to sociological methods*. Aldine.

[14] Dreyfus, H. L. (1992). *What computers still can't do: A critique of artificial reason*. MIT Press.

[15] Feenberg, A. (2017). *Technosystem: The social life of reason*. Harvard University Press.

[16] Gadamer, H.-G. (2006). *Verdad y método* (A. Agud Aparicio & R. de Agapito, trads.). Ediciones Sígueme. (Obra original publicada en 1960).

[17] Heidegger, M. (1994). La pregunta por la técnica. En *Conferencias y artículos* (E. Barjau, trad., pp. 9–37). Ediciones del Serbal. (Obra original publicada en 1954).

[18] Heidegger, M. (2005). *Ser y tiempo* (J. E. Rivera, trad.). Editorial Universitaria. (Obra original publicada en 1927).

[19] Hui, Y. (2022). *Recursividad y contingencia* (M. G. Burello, trad.). Caja Negra. (Obra original publicada en 2019).

[20] Mbembe, A. (2011). *Necropolítica seguido de Sobre el gobierno privado indirecto* (E. Falomir Archambault, trad.). Melusina. (Obra original publicada en 2003).

[21] Merleau-Ponty, M. (1993). *Fenomenología de la percepción* (J. Cabanes, trad.). Planeta-Agostini. (Obra original publicada en 1945).

[22] Pardo León, J. A. (en curso). *Narcotráfico y cultura: lo narco como categoría estética* [Tesis doctoral en curso]. Universidad Santo Tomás, Doctorado en Filosofía.

[23] Rancière, J. (2009). *El reparto de lo sensible: estética y política* (C. Durán, trad.). LOM. (Obra original publicada en 2000).

[24] Reichenbach, H. (1938). *Experience and prediction: An analysis of the foundations and the structure of knowledge*. University of Chicago Press.

[25] Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–457. `https://doi.org/10.1017/S0140525X00005756`

[26] Valencia, S. (2010). *Capitalismo gore*. Melusina.

[27] Varela, F. J., Thompson, E., & Rosch, E. (1997). *De cuerpo presente: las ciencias cognitivas y la experiencia humana* (C. Gardini, trad.). Gedisa. (Obra original publicada en 1991).

[28] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, *35*, 24824–24837.

[29] Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*, *36*.

[30] Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. En *Proceedings of the 11th International Conference on Learning Representations (ICLR 2023)*.

Herramientas de inteligencia artificial

[31] Anthropic. (2026). *Claude Opus* (versiones 4.6, 4.7 y 4.8) [Modelo de lenguaje de gran escala]. `https://claude.ai` (Empleado como apoyo en la edición y mejora de la redacción de este artículo).