
De Austin a la pregunta

un recorrido por diez obras

*Agencia sin imputación: hermenéutica, pragmática
y la validación epistémica del discurso artificial*

Carlos Isaac Zainea

Examen de conocimientos | Junio 2026

Vigencia (mayo 2026)

León XIV, *Magnifica Humanitas* — primera encíclica, sobre IA, firmada en el 135.º aniversario de *Rerum Novarum*.

LA ENCÍCLICA AFIRMA (el qué)

« la IA no comprende lo que produce »



ESTA TESIS DA (el cómo)

aceptar los **límites ontológicos**

« la técnica no es antagónica al humano... debe servirle, no concentrar poder »



rediseñarla con valores:
trazable · riguroso · auditable

La encíclica **reclama** en general; la tesis da el **cómo** verificable.

Frases por cotejar con el texto oficial (vatican.va). Cf. Vatican News, 25 may. 2026.

El problema y el cambio de variable

El estado de la inteligencia artificial puede describirse como dos vías que no intercambian información. La vía crítico-filosófica diagnostica con precisión los límites de la IA, pero no construye nada con ese diagnóstico: es correcta y estéril. La vía técnica, la de los modelos de lenguaje, alcanza un rendimiento notable a costa de un proceso interno que nadie puede inspeccionar: la **caja negra**. Entre ambas se abre un diálogo de sordos.

Pregunta ontológica

¿comprende la máquina?

respuesta: no (callejón)

CAMBIO DE VARIABLE →

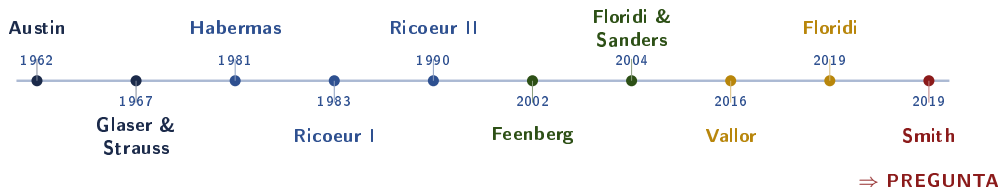
Pregunta de diseño

*¿ejecuta un *procedimiento* riguroso, inspeccionable y atribuible?*

respuesta: depende del andamiaje

El recorrido fija tres exigencias para ese procedimiento —que sea **trazable**, **riguroso** y **auditable**— y muestra que cada una tiene su origen en una de las diez obras.

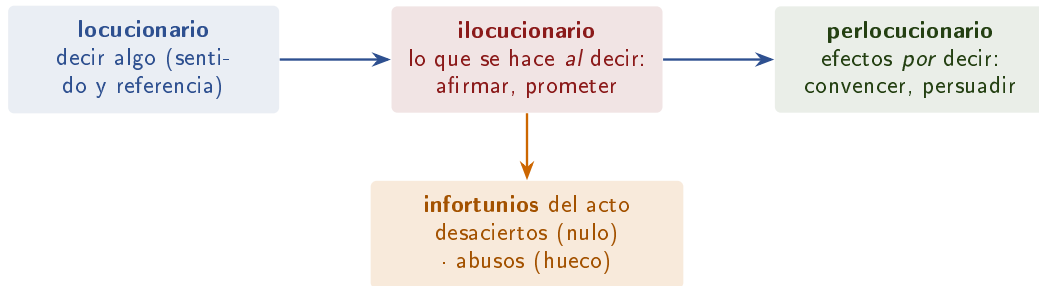
El argumento como construcción: 1962–2019



Cada obra deja fijada una de las exigencias del procedimiento; la pregunta de investigación es la composición de las diez.

Austin (1962): del valor de verdad a la condición de éxito

Austin desplaza la pregunta: del valor de verdad de un enunciado (¿corresponde al mundo?) a su éxito como acto (¿se ejecutó bien el procedimiento?). Y el propio libro disuelve la dicotomía constativo/realizativo en una tríada decisiva para la investigación posterior: todo enunciado es, a la vez, tres actos.



La salida de un LLM es un **acto ilocucionario** —una aserción con fuerza epistémica— que puede incurrir en **infortunio** cuando aparenta estar fundamentada sin poder estarlo.

El acto se juzga por su procedimiento, no por su esencia

« Cuando digo, ante el funcionario o el altar, “sí, acepto”, no estoy informando sobre un matrimonio: lo estoy llevando a cabo. Emitir tales expresiones no es describir aquello que hago al expresarme así, ni afirmar que lo hago: es hacerlo. Ninguna de ellas es verdadera o falsa; pueden, en cambio, salir bien o mal, ser afortunadas o desafortunadas. »

J. L. Austin, *Cómo hacer cosas con palabras* (Paidós, 1982; orig. 1962), Conferencias I–II. Pasaje condensado de la traducción de Carrió & Rabossi.

El criterio de éxito es **público y verificable**. Austin sustituye un predicado interno —la verdad— por uno externo —cumplir el procedimiento—: el mismo movimiento que esta investigación hará con la IA sesenta años después.

Glaser y Strauss (1967): la teoría emerge de los datos

Frente a la sociología que postulaba grandes teorías y luego buscaba datos para confirmarlas, Glaser y Strauss invierten el orden: la teoría se construye desde los datos, de abajo hacia arriba, mediante operaciones explícitas y repetibles —la comparación constante, el muestreo teórico y la saturación—. Su exigencia es que ningún concepto exista sin los datos que lo fundan. Por eso el método es, de raíz, **trazable**: la trazabilidad no se le añade, le es constitutiva.



Un método cuyo rigor deja rastro

« Theoretical sampling is the process of data collection for generating theory whereby the analyst jointly collects, codes, and analyzes his data and decides what data to collect next and where to find them, in order to develop his theory as it emerges. »

B. G. Glaser & A. L. Strauss, *The Discovery of Grounded Theory* (Aldine, 1967), p. 45.

Recolección, codificación y análisis ocurren **a la vez**: cada concepto queda ligado al dato que lo motivó. Es el método candidato a traducir a una arquitectura, porque su rigor produce un rastro inspeccionable.

Habermas (1981): dos racionalidades

Partiendo de Austin, Habermas distingue la acción orientada al éxito —instrumental, estratégica— de la acción orientada al entendimiento. Toda emisión comunicativa eleva **pretensiones de validez** (verdad, rectitud, veracidad) que son legítimas solo si pueden **redimirse**: justificarse cuando alguien las cuestiona. El diálogo de sordos recibe aquí su nombre: el predominio de la razón instrumental, que optimiza sin responder por sus razones.

acción instrumental
orientada al éxito

acción comunicativa
orientada al entendimiento

la vía técnica = razón instrumental sin redención de razones

La validez es una pretensión que debe poder redimirse

« Hablo, en cambio, de acciones comunicativas cuando los planes de acción de los actores implicados no se coordinan a través de un cálculo egocéntrico de resultados, sino mediante actos de entendimiento. En la acción comunicativa los participantes no se orientan primariamente al propio éxito; antes bien, persiguen sus fines individuales bajo la condición de que sus respectivos planes de acción puedan armonizarse entre sí sobre la base de una definición compartida de la situación. »

J. Habermas, *Teoría de la acción comunicativa*, vol. I (Taurus, 1987; orig. 1981).

Una caja negra no satisface esta condición: no puede redimir ninguna pretensión. Aquí queda fijada la exigencia de **auditabilidad** a nivel de razones.

Ricoeur (1983): el sentido se configura en una trama

Ricoeur muestra cómo lo disperso y heterogéneo se vuelve inteligible al ordenarse en una **intriga**: una síntesis que reúne agentes, fines, causas y azares en una totalidad que puede seguirse —una *concordancia discordante*—. El arco hermenéutico va de la prefiguración del mundo (Mímesis I) a su configuración en relato (Mímesis II) y a su retorno al mundo del lector (Mímesis III). Una máquina puede configurar; el arco, sin embargo, se trunca cuando no hay retorno ni sujeto al que el relato pertenezca.



Comprender un proceso es poder seguir su relato

« *El tiempo se hace tiempo humano en la medida en que se articula de un modo narrativo, y la narración alcanza su plena significación cuando se convierte en una condición de la existencia temporal.* »

P. Ricoeur, *Tiempo y narración I* (Siglo XXI, 1995; orig. 1983), Primera parte.

Si comprender es poder recorrer una trama, entonces exigir que un sistema sea **trazable** es exigir que pueda *narrar* cómo llegó a su conclusión. Trazable equivale a narrable.

Ricoeur (1990): identidad narrativa e imputación

En *Sí mismo como otro*, Ricoeur distingue la identidad de lo que permanece idéntico (*ídem*) de la identidad del sí mismo que se constituye en el tiempo (*ipse*), y sostiene que esta última es **narrativa**: un agente se reconoce en la historia de sus actos. De ahí su concepto ético decisivo, la **imputabilidad**: un acto es atribuible en la medida en que puede integrarse en una secuencia reconstruible. La imputabilidad resulta así una función de la trazabilidad —y un sistema opaco, al no ser narrable, vuelve la responsabilidad inasignable, que recae por defecto en el último humano de la cadena.

Responsable es quien puede dar cuenta de sus actos

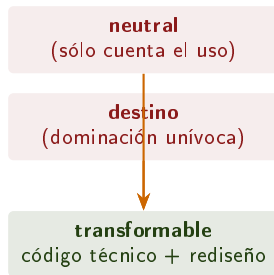
« Responder a la pregunta “¿quién?”, como lo había dicho con toda energía Hannah Arendt, es contar la historia de una vida. La historia narrada dice el quién de la acción. Por consiguiente, la propia identidad del quién no es sino una identidad narrativa. »

P. Ricoeur, *Sí mismo como otro* (Siglo XXI, 1996; orig. 1990); cf. ≈p. 126.

Sin proceso reconstruible no hay «quién» al que imputar. La **trazabilidad** es, por tanto, condición de la responsabilidad: aquí se resuelve el problema de la agencia sin imputación sin recurrir a autores ajenos a la bibliografía.

Feenberg (2002): ni neutral ni destino

Feenberg descarta las dos posiciones habituales: la que considera la técnica un medio neutral, que ignora los valores inscritos en el diseño, y la que la considera un destino unívoco de dominación, que solo deja la resignación o la crítica. Frente a ambas, todo diseño fija un **código técnico** —valores cristalizados en la forma del artefacto—, pero ese código es contingente: la técnica es ambivalente y, por tanto, transformable. A su rediseño deliberado lo llama **racionalización democrática**.



La técnica como escenario de disputa

« Critical theory argues that technology is not a thing in the ordinary sense of the term, but an “ambivalent” process of development suspended between different possibilities. . . . On this view, technology is not a destiny but a scene of struggle. It is a social battlefield . . . a “parliament of things” in which civilizational alternatives contend. »

A. Feenberg, *Transforming Technology: A Critical Theory Revisited* (Oxford UP, 2002), Introducción.

Si la técnica es escenario de disputa y no destino, entonces inscribir la transparencia en el diseño es legítimo: la opacidad de la IA es una **elección** revisable, no una fatalidad.

Floridi y Sanders (2004): agencia sin mente

Floridi y Sanders definen la agencia mediante criterios observables —interactividad, autonomía y adaptabilidad— evaluados a un **nivel de abstracción** explícito, y separan dos predicados que suelen confundirse: la responsabilidad moral, que exige conciencia e intención, y la *accountability* moral, la imputabilidad de un comportamiento, que no las requiere. El resultado es una **moral sin mente**: se puede evaluar lo que un agente hace, al nivel de abstracción fijado, sin resolver si piensa. Es, en ética, el mismo gesto de Austin en el lenguaje, y la justificación exacta de la cláusula «aceptando sus límites ontológicos».

Imputable sin ser consciente

« [An agent is a system that is] interactive, autonomous, and adaptable at a given level of abstraction. . . . There is substantial scope for the concept of an agent that is morally accountable without being morally responsible, and hence need not exhibit free will, mental states or intentions — a “mind-less morality”. »

L. Floridi & J. W. Sanders, “On the Morality of Artificial Agents”, *Minds and Machines* 14(3), 2004, pp. 349–379 (texto condensado de §3 y conclusiones).

La agencia y la atribución quedan **desacopladas** de la ontología de la mente: no es una concesión, sino la fijación de un nivel de abstracción en el que el sistema es evaluable.

Vallor (2016): la técnica forma el carácter

Vallor recupera la ética de la virtud —aristotélica, confuciana, budista— y sostiene que las tecnologías no son neutrales respecto del carácter: condicionan qué disposiciones podemos cultivar. Propone doce **virtudes tecnomorales** —entre ellas la honestidad, la **humildad**, el autocontrol, la justicia y el cuidado— integradas en una sabiduría práctica capaz de orientar la técnica hacia un futuro habitable. La cláusula de aceptar los límites ontológicos se reinterpreta así como la virtud de la humildad convertida en restricción de diseño: no atribuir al sistema capacidades que no tiene.

Un futuro que valga la pena querer

«[Technomoral virtues are] specific skills and strengths of character, adapted to the unique challenges of 21st-century life, that offer the human family our best chance of learning to live wisely and well with emerging technologies. »

S. Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting* (Oxford UP, 2016).

La transparencia no es solo una exigencia técnica: es el modo de cultivar un futuro en el que la potencia de la IA no erosione la **confianza** ni la responsabilidad.

Floridi (2019): la filosofía como diseño conceptual

Floridi define el oficio mismo de la filosofía: no se limita a analizar conceptos heredados, sino que **diseña, refina y construye** los conceptos con los que tratamos las preguntas abiertas —aquellas que no se cierran ni por deducción pura ni por dato empírico—. Bajo esta definición, el entregable de esta tesis —un modelo conceptual, una arquitectura de conceptos y no un programa— es un objeto filosófico en sentido estricto. Responde por anticipado a la objeción sobre su adscripción: no se importa la ingeniería a la filosofía, se ejerce la filosofía como diseño.

Diseñar conceptos es hacer filosofía

«Philosophy, understood as the study or science of open questions and their answers, becomes primarily a form of conceptual design. . . . Knowledge is design, and philosophy is the ultimate form of conceptual design. »

L. Floridi, *The Logic of Information: A Theory of Philosophy as Conceptual Design* (Oxford UP, 2019), cap. 2
«Philosophy as Conceptual Design», pp. 27–52.

Diseñar conceptos no es un paso previo a la filosofía: **es** la filosofía cuando está a la altura de su tarea.

Smith (2019): cálculo sin juicio

Smith distingue el **reckoning** —el cálculo competente sobre estructuras formales, en el que las máquinas ya superan al humano— del **judgment** —el compromiso situado y responsable con el estado real del mundo—. Su tesis es que la IA actual tiene reckoning sin judgment. No niega la potencia del cálculo: localiza con precisión la carencia, y por eso no es crítica estéril. El proyecto convierte esa delimitación en restricción de diseño: no se exige juicio, sino que el cálculo sea riguroso, trazable y auditable.

reckoning
cálculo formal
competente

la IA *sí*

judgment
compromiso situa-
do y responsable

la IA *no*

El juicio que la máquina no alcanza

« Reckoning [is] the calculative prowess at which AI systems already excel. . . . Judgment [is] a form of dispassionate deliberative thought, grounded in ethical commitment and responsible action, appropriate to the situation in which it is deployed. . . . No historical or current approaches to AI . . . have even begun to wrestle with what constructing or developing judgment would involve. »

B. C. Smith, *The Promise of Artificial Intelligence: Reckoning and Judgment* (MIT Press, 2019). Definiciones condensadas del texto.

Si el juicio no es construible bajo la cláusula ontológica, lo que queda es **acotar e inspeccionar** lo que el sistema sí hace.

Composición: la pregunta

¿Cómo puede diseñarse un sistema de IA que, aceptando sus límites ontológicos, pueda construir interpretaciones de manera trazable, rigurosa y auditable?

diseñar

→ Floridi (2019) · Feenberg

límites ontológicos

→ Floridi & Sanders · Austin · Smith · Vallor

construir in-
terpretaciones

→ Glaser & Strauss · Ricoeur (1983)

trazable

→ Ricoeur (1983)

rigurosa

→ Glaser & Strauss · Austin

auditable

→ Habermas · Ricoeur (1990)

La pregunta no es un punto de partida: es el **resultado** de la construcción.

Cada término de la pregunta está definido y justificado por una de las diez obras.

Responderla —especificar la arquitectura y las métricas de fidelidad conceptual— es el objeto de esta tesis.

Gracias.

Bibliografía nuclear

- [1] Austin, J. L. (1982). *Cómo hacer cosas con palabras* (Carrió & Rabossi, trads.). Paidós. (Orig. 1962).
- [2] Feenberg, A. (2002). *Transforming Technology: A Critical Theory Revisited*. Oxford UP.
- [3] Floridi, L. (2019). *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford UP.
- [4] Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14(3), 349–379.
- [5] Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory*. Aldine.
- [6] Habermas, J. (1987). *Teoría de la acción comunicativa* (Jiménez Redondo, trad.). Taurus. (Orig. 1981).
- [7] Ricoeur, P. (1995). *Tiempo y narración I* (Neira, trad.). Siglo XXI. (Orig. 1983).
- [8] Ricoeur, P. (1996). *Sí mismo como otro* (Neira, trad.). Siglo XXI. (Orig. 1990).
- [9] Smith, B. C. (2019). *The Promise of Artificial Intelligence: Reckoning and Judgment*. MIT Press.
- [10] Vallor, S. (2016). *Technology and the Virtues*. Oxford UP.