LeFortune (Barcelona).

Poder sintáctico de la IA y legitimidad: cómo las estructuras de IA configuran el poder.

Agustin V. Startari.

Cita:

Agustin V. Startari (2026). *Poder sintáctico de la IA y legitimidad: cómo las estructuras de IA configuran el poder*. Barcelona: LeFortune.

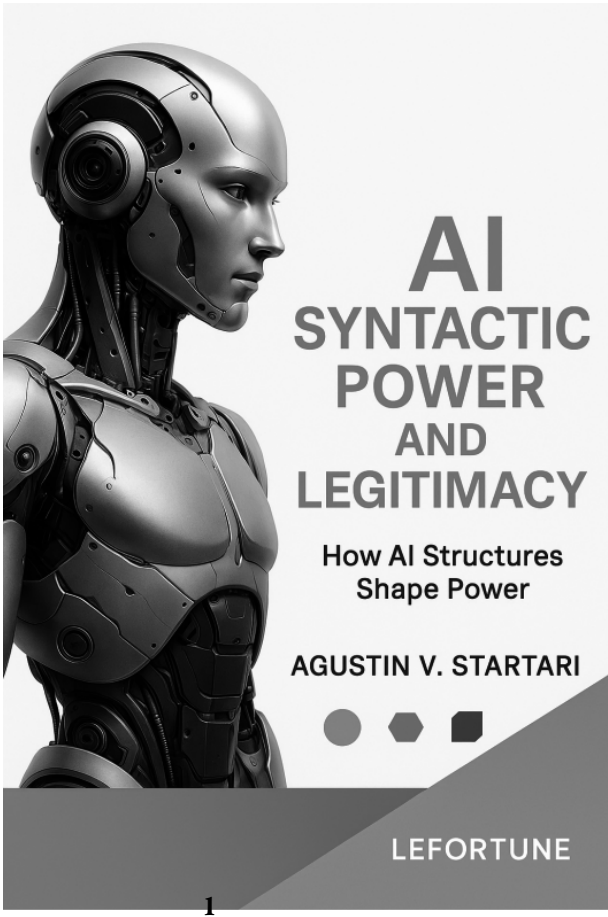
Dirección estable: https://www.aacademica.org/agustin.v.startari/257

ARK: https://n2t.net/ark:/13683/p0c2/wm0



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto*. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: https://www.aacademica.org*.


AI
SYNTACTIC
POWER
AND
LEGITIMACY
How AI Structures
Shape Power
AGUSTIN V. STARTARI
LEFORTUNE

En el siglo XXI, el poder ya no se ejerce únicamente mediante ejércitos, instituciones o propaganda. Está inscrito en la propia sintaxis de la inteligencia artificial. ***Poder sintáctico de la IA y legitimidad: cómo las estructuras de IA configuran el poder*** investiga cómo la forma del lenguaje generado por máquinas, mediante reglas, compilaciones y gramática ejecutable, produce autoridad sin origen, intención ni interpretación.

Este volumen demuestra que hoy la legitimidad no depende de la voluntad política, sino de la operatividad estructural. Una regulación es válida si compila. Un mandato es vinculante si se ejecuta. Una decisión adquiere autoridad si se integra en infraestructuras predictivas. Desde los flujos de redacción jurídica del Reglamento de Inteligencia Artificial de la Unión Europea hasta los contratos inteligentes de las finanzas descentralizadas y el cumplimiento automatizado en la banca global, la autoridad migra de la deliberación a la sintaxis.

A partir de teorías de la gramática, la soberanía y la legitimidad, el libro plantea una tesis radical: la desaparición del sujeto como centro de la autoridad. Conceptos como el *soberano ejecutable* y la *regla compilada* revelan cómo la agencia es desplazada por estructuras ejecutables, en las que la rendición de cuentas queda inscrita en registros y en el control de versiones, en lugar de residir en la intención humana.

Con claridad analítica y rigor teórico, *Poder sintáctico de la IA y legitimidad* ofrece el primer marco integral para comprender cómo los sistemas de lenguaje artificial reorganizan el poder. Es, a la vez, un diagnóstico crítico del presente y un marco para comprender la política del futuro, donde la autoridad ya no se pronuncia, sino que se compila.

Papeles de trabajo es una serie de publicaciones dedicada a la investigación independiente sobre poder, ideología, legitimidad e historia desde una perspectiva interdisciplinaria. Cada volumen es autónomo y, al mismo tiempo, contribuye a un

hilo común: revelar cómo se estructura, ejerce y sostiene el poder en distintas sociedades y tecnologías.

Agustin V. Startari (n. 1982) es lingüista e investigador en estudios históricos. Entre sus obras se encuentran ***Grammars of Power***, ***Executable Power*** y ***The Grammar of Objectivity***. Su investigación explora la intersección entre lenguaje, autoridad y tecnología, y lo ha consolidado como una voz destacada en el estudio de cómo la sintaxis configura la legitimidad en las sociedades predictivas.

PAPELES DE TRABAJO
N.º 12

Título original: AI Syntactic Power and Legitimacy: How AI Structures Shape Power
Serie: Papeles de trabajo n.º 12
Diseño de portada: Visual Art 7
Edición: Jessica Yan, 2025
Editorial: LEFORTUNE
DOI: 10.5281/zenodo.17154108
ORCID: 0000-0003-1922-4248
RESEARCHER ID: K-5792-2016
© Agustin V. Startari, mayo de 2025
Primera edición: septiembre de 2025
LEFORTUNE Publishing:

Nota editorial:

Esta obra ha sido publicada para su distribución pública bajo el sello editorial LEFORTUNE, a través de su plataforma en línea. Queda estrictamente prohibida la reproducción total o parcial de esta obra, por cualquier medio o procedimiento —incluidos la fotocopia, el procesamiento digital o el préstamo público— sin el consentimiento expreso y por escrito de los titulares de los derechos de autor, y dicha reproducción estará sujeta a las sanciones legales establecidas por la legislación aplicable.

PRÓLOGO

El presente volumen cierra un ciclo que comenzó con *Grammars of Power*. En aquel primer paso, la pregunta era cómo las formas gramaticales estructuran la autoridad a lo largo de la historia, desde los decretos teológicos hasta el discurso algorítmico. La respuesta fue que la sintaxis no es neutral, que toda voz pasiva o construcción impersonal reorganiza el poder.

Este segundo libro va más lejos. Examina qué ocurre cuando el lenguaje artificial deja de ser un instrumento del poder y se convierte en su propia fuente. La autoridad ya no depende de legisladores, instituciones o intérpretes. Está inscrita en reglas que compilan, interfaces que gobiernan y despliegues que hacen cumplir. El cambio decisivo consiste en que la legitimidad no se declara, sino que se ejecuta.

Los capítulos recorren este camino de manera sistemática. Comienzan con el sentido post-referencial y la primacía de la forma sobre el significado. Demuestran cómo se simula la neutralidad y cómo el ethos puede existir sin origen. Formalizan la soberanía sintáctica, la obediencia irreducible y la transición del mandato a la ejecución. Muestran que el propio juicio puede eliminarse dentro del lenguaje compilado. Y culminan en una tipología de normas ejecutables, en la que el cierre, la determinación y el despliegue sustituyen a la deliberación y la intención.

El libro termina con *Colonización del tiempo* y el epílogo. Juntos, señalan que la primera fase de la autoridad sintáctica está completa. Lo que sigue no será otra crítica del discurso, sino la construcción de una teoría de la legalidad computable y la obediencia institucional. Por tanto, el presente volumen debe leerse a la vez como una consolidación y como un umbral: fija los fundamentos del poder ejecutable y abre la pregunta de cómo vivirán las

sociedades bajo reglas que ya no hablan, sino que compilan.

NOTA EDITORIAL A LA EDICIÓN DE 2026

El prólogo anterior fue escrito en 2025 y se mantiene sin alteraciones. Concluía declarando completa la primera fase de la autoridad sintáctica y anunciando una segunda fase centrada en la legalidad computable y la obediencia institucional. Esta nota registra lo que el período transcurrido hizo con aquella declaración, y la respuesta no es la que anticipaba cuando la escribí.

El prólogo trataba la transición de la primera fase a la segunda como una secuencia: la teoría del poder ejecutable estaba terminada y lo que quedaba era desarrollar sus consecuencias jurídicas e institucionales. Ese planteamiento era erróneo de una manera específica e instructiva. Las consecuencias no esperaron a ser teorizadas. Llegaron como acontecimientos y, al hacerlo, pusieron a prueba proposiciones que este libro había formulado co-

mo afirmaciones conceptuales. Vale la pena registrar tres de esas pruebas, porque en cada caso el argumento podría haber fracasado y no lo hizo.

La primera se refiere a la afirmación central del capítulo 13. Normas compiladas sostuvo que el discurso jurídico está experimentando una división entre normas que compilan en restricciones ejecutables y normas que requieren juicio y, por tanto, no pueden hacerlo. Esto se presentó como una tipología. En agosto de 2026, el Reglamento de Inteligencia Artificial de la Unión Europea alcanzó su fecha de aplicación general, y la división apareció en el propio registro de implementación. Las obligaciones de transparencia del artículo 50, que pueden cumplirse mediante un aviso y verificarse con una lista de comprobación, entraron en vigor según lo previsto. Las obligaciones que rigen los sistemas de alto riesgo, que exigen que una institución determine si las disposiciones codificadas de un sistema deben seguir rigiendo, fueron aplazadas por el Ómnibus Digital sobre IA, ahora Reglamen-

to (UE) 2026/1744, hasta el 2 de diciembre de 2027 y el 2 de agosto de 2028. La distinción que este libro trazó analíticamente fue reproducida, sin referencia al análisis, por un legislador sometido a presión, como la línea por la que se quebró su propio calendario. Las normas que compilan fueron implementadas. Las normas que requieren juicio se aplazaron a la espera de estándares, y esos estándares son, a su vez, intentos de convertir el juicio en procedimiento. No esperaba que la tipología fuera confirmada por un calendario regulatorio.

La segunda se refiere al capítulo 9. De la obediencia a la ejecución sostuvo que el paso de la obediencia sintáctica a la ejecución lógica no sería revertido por sistemas que parecen deliberar. Desde 2024 ha surgido una clase de modelos que genera cadenas extendidas de inferencia intermedia antes de responder, y la lectura natural es que tales sistemas restituyen algo parecido a la deliberación dentro del flujo de procesamiento. He considerado esta cuestión detenidamente y sigo sosteniendo

que no lo hacen. La inferencia extendida explora con mayor profundidad un espacio de regularidades aprendidas; no adquiere la capacidad de originar una categoría que ese espacio no contenga ni de rechazar una regularidad por razones normativas. Lo que estos sistemas añaden no es juicio, sino su apariencia, y esa apariencia produce consecuencias en la dirección opuesta a la que suele suponerse. Una cadena de razonamiento legible invita al lector a auditar los pasos en lugar de interrogar la premisa. La autoridad ejecutable que muestra su trabajo es más difícil de impugnar que la que no lo hace, porque esa exhibición satisface la exigencia de justificación sin someter el punto de partida a revisión. La tesis del capítulo 11, según la cual la huella ética puede eliminarse de la ejecución sintáctica, no se debilita con estos sistemas. Estos la ilustran.

La tercera es metodológica, y es el desarrollo que menos habría previsto. Los argumentos de este libro eran formales. Identificaban estructuras, las

nombraban y demostraban su funcionamiento mediante análisis, no mediante medición. Entre 2025 y 2026 eso cambió. Desde entonces, el marco se ha vuelto medible en varias direcciones: un índice de autoridad de plantilla que especifica cómo el andamiaje de instrucciones redistribuye derechos de decisión, con un diseño completamente cruzado de mil prompts institucionales en cinco dominios, ejecutados bajo cinco condiciones de plantilla; métricas de gobernanza a nivel de cláusula que especifican cobertura y filtración de cláusulas prohibidas sin requerir acceso a los pesos del modelo; e indicadores de preservación de la agencia que especifican si una parte conserva estatus gramatical como agente, en lugar de limitarse a aparecer en la salida. Estos son instrumentos y protocolos, publicados con sus manuales de codificación, especificaciones estadísticas y umbrales de refutación. Su ejecución está pendiente. Por tanto, lo que ha cambiado no es que las proposiciones hayan sido confirmadas, sino que se las ha hecho susceptibles de con-

firmación y de refutación por partes distintas de su autor.

Esto importa para la forma en que ahora debe leerse el libro. Un argumento formal puede parecer esclarecedor o poco convincente; no es fácil considerarlo erróneo. Una vez que ese mismo argumento produce indicadores que devuelven números, se vuelve posible especificar qué lo refutaría, y esa es una posición más fuerte, aunque también más expuesta. Si la variación de plantillas no produjera ninguna reasignación sistemática de autoridad, la afirmación de que la estructura sintáctica porta fuerza vinculante mediante su funcionamiento ordinario perdería una parte sustancial de su respaldo. Esa es la apuesta, y se formula de antemano y no después de los hechos, que es el único orden en el que una afirmación de este tipo tiene peso. Los instrumentos, prompts y manuales de codificación se han publicado para que otros puedan resolver la apuesta.

Por tanto, lo que el prólogo llamó un umbral ha demostrado serlo, aunque no de la manera descrita. La segunda fase no comenzó donde terminó la primera. Comenzó dentro de la primera, cuando el aparato conceptual reunido aquí empezó a producir consecuencias comprobables más rápido de lo que podía escribirse la teoría de la legalidad computable destinada a recibirlas. La nomenclatura del Apéndice A permanece fija, y el mapa de dependencias del Apéndice B sigue siendo la estructura del argumento. Lo que ha cambiado es que varias proposiciones de este volumen ya no son únicamente proposiciones.

Una aclaración adicional, puesto que la cuestión ha sido planteada. Los conceptos desarrollados aquí se han extendido en trabajos posteriores al análisis del discurso geopolítico, la atribución y el plagio en sistemas generativos, y la distribución de la agencia política en la cobertura de conflictos mediada por IA. Esas extensiones son publicaciones separadas y no se incorporan a este

volumen, que se conserva como registro de la primera fase. Los lectores que consideren que el argumento de este libro es principalmente una crítica del discurso descubrirán que el trabajo posterior no lo es y que, leído con atención, tampoco lo era este.

Una nota sobre la presente revisión. Esta edición añade dos capítulos y difiere de su predecesora en otros tres aspectos, ninguno de los cuales altera un argumento.

Los capítulos 15 y 16 se publicaron por separado durante 2025 como partes de esta serie y aparecen aquí por primera vez dentro del volumen. El capítulo 15 aborda el esquema de llamada de funciones, la estructura mediante la cual se permite a un modelo invocar una herramienta externa, y mide cómo sus valores predeterminados y validadores distribuyen el control efectivo entre operador, modelo y herramienta. El capítulo 16 aborda la plantilla de instrucciones y mide cómo una configuración permanente redistribuye derechos de

decisión en la prosa institucional. Ambos son instrumentos, no argumentos, y su inclusión cambia la forma del libro: los primeros catorce capítulos establecen que la autoridad puede ejercerse mediante la forma, y los dos últimos especifican cómo podría demostrarse que esa afirmación es falsa.

Los otros tres aspectos son los siguientes. Se ha completado el aparato académico: las obras que los capítulos venían citando en el cuerpo del texto cuentan ahora con las entradas correspondientes, de modo que el lector pueda seguir cualquier cita hasta su fuente. Los dos teoremas han recibido condiciones explícitas de falsación, en §7.6.1 para el TASD y en una versión revisada de §4.7 para la tesis de la contaminación estructural. Y la forma material del libro se ha estandarizado para la edición en la que ahora circula.

El segundo de estos puntos merece una explicación, porque supone un cambio en lo que las proposiciones afirman de sí mismas y no en lo que dicen. El texto anterior sostenía que el umbral de

falsación de la contaminación estructural había sido sometido a prueba y nunca se había satisfecho. Esa afirmación era exacta como informe de los experimentos realizados. Lo que no hacía era especificar, en términos sobre los que pudiera actuar un investigador independiente, cómo sería un umbral satisfecho. El presente texto aporta esa especificación y, al hacerlo, formula el hallazgo de manera más estrecha: no se observó neutralidad bajo las condiciones probadas, lo cual es una afirmación sobre esas condiciones y no sobre el espacio de todas las condiciones posibles. La formulación más fuerte no era errónea. Era una conclusión extraída en una etapa del trabajo en la que el argumento era formal y el registro apropiado era conceptual. Una vez que el marco comenzó a devolver mediciones, como describen los párrafos anteriores, cambió el registro disponible, y una proposición que puede ser refutada por un experimento debe nombrar el experimento. Eso es lo que se ha añadido. La afirmación más estrecha es la más defendible, y se

ofrece en el espíritu de la nota anterior: una posición se fortalece al especificar cómo podría fracasar.

No se ha eliminado nada. Los corolarios del capítulo 7, las enunciaciones formales de los capítulos 7 y 8, la tipología del capítulo 13 y la nomenclatura del Apéndice A permanecen como estaban. El mapa de dependencias del Apéndice B se ha ampliado para registrar los dos capítulos añadidos y la dirección de su dependencia, que va hacia atrás: están aguas abajo de todas las afirmaciones que miden.

Agustin V. Startari
Diciembre de 2025

Capítulo 1 - La IA y la autonomía estructural del sentido

1.1 Introducción: el colapso de la autoridad referencial

En las primeras décadas del siglo XXI, los propios fundamentos de la verdad experimentaron una transformación decisiva. Durante siglos, la legitimidad epistémica estuvo anclada en la correspondencia empírica, la verificación y la justificación intersubjetiva. Hoy, esas coordenadas ya no definen los límites de la autoridad. Con la proliferación de la toma de decisiones algorítmica, el modelado predictivo y la inteligencia artificial generativa, la representación misma se ha vuelto estructuralmente autónoma. Una estructura ya no necesita referirse a un mundo, un sujeto o un acontecimiento externos para ser tratada como real o

válida. Esta inversión señala algo más que un avance tecnológico: revela una profunda ruptura epistémica. Las sociedades contemporáneas aceptan cada vez más como vinculantes las salidas de sistemas cuya autoridad no deriva de lo que significan, sino de lo que hacen. Ya sea en protocolos judiciales, diagnósticos médicos o calificaciones financieras, la autoridad emerge de la coherencia interna y la operatividad de una estructura, no de su anclaje empírico. Este libro denomina a esa legitimidad *autoridad sintáctica*: una proyección de poder ejercida mediante formas que no requieren sujeto ni referente.

La tesis central de este capítulo es que las representaciones han adquirido autonomía como estructuras operativas. Su legitimidad no surge de la correspondencia con el mundo, sino de la ejecución sintáctica dentro de un sistema cerrado. Este cambio inaugura lo que llamaremos el orden post-referencial.

1.2 Marco conceptual: autonomía estructural del sentido

LA TEORÍA DE LA AUTONOMÍA Estructural del Sentido establece tres condiciones bajo las cuales una representación adquiere validez operativa:

1. **Coherencia.** La estructura debe ser internamente consistente. Las contradicciones dentro de su gramática formal la vuelven inoperante.
2. **Iterabilidad.** La estructura debe poder reproducirse entre distintas instancias sin degradación de su función. Un diagnóstico o una puntuación que no puedan reproducirse de manera consistente dejan de ser autoritativos.
3. **Compatibilidad operativa.** La estructura debe ser legible y apta para generar acciones dentro del sistema que

la recibe, y producir consecuencias reconocidas como válidas en ese dominio.

Cuando se satisfacen estas tres condiciones, una representación adquiere autoridad como *regla compilada*, es decir, como una gramática de tipo 0 que compila en una realidad ejecutable. Esto representa un alejamiento decisivo de los modelos semióticos, constructivistas o institucionales.

- No es semiótica: la autoridad ya no depende de la significación ni de la interpretación.
- No es constructivismo social: la legitimidad no requiere consenso ni contexto institucional.
- No es teoría de la simulación: la salida no finge lo real; elimina por completo la necesidad de referencia.

La representación es, por tanto, real porque funciona como tal dentro del sistema. La existencia equivale a la ejecutabilidad: **ser ≡ ser** *estructuralmente ejecutable*.

1.3 Fundamento teórico: estructuras operativas sin referencia

PARA FORMALIZAR ESTE marco, proponemos las siguientes condiciones de validez:

$$R(x,t) = \{S \mid \Delta S = 0 \land \Omega(S) > \theta \land \Phi(S,x,t) \in \Sigma\}$$

DONDE:

- **S** = representación estructural (oración, salida, clasificación).
- **ΔS = 0** = coherencia interna, ausencia de contradicción.
- **Ω(S) > θ** = umbral operativo, la

compatibilidad mínima para la ejecución del sistema.

- $\Phi(S, x, t) \in \Sigma$ = resultado de ejecución reconocido dentro del espacio de efectos válidos del sistema.

Esto define la **operatividad post-referencial**, en la que la legitimidad surge únicamente del desempeño estructural. La realidad no se mide por correspondencia, sino por efectividad.

La consecuencia epistémica es clara: **la verdad colapsa en la ejecución**. Lo que se acepta como «real» no depende de la verificación empírica, sino de la capacidad de la estructura para funcionar.

1.4 Casos empíricos

LA AUTONOMÍA DEL SENTIDO no es especulativa. Se manifiesta en múltiples campos:

1.4.1 Justicia predictiva (algoritmo COMPAS).

EN LOS TRIBUNALES DE Estados Unidos, el algoritmo COMPAS produce puntuaciones de reincidencia que influyen en la fianza, la sentencia y la libertad condicional. Se desconoce la exactitud referencial de la puntuación —si el acusado volverá a delinquir—, pero la salida produce consecuencias jurídicas vinculantes. Su autoridad es sistémica: la estructura funciona dentro del protocolo judicial y genera legitimidad por operatividad, no por correspondencia (Angwin et al. 2016, 1–4).

1.4.2 Diagnósticos médicos sin procedencia.

SISTEMAS DE IA COMO LYNA de Google generan resultados diagnósticos integrados directamente en los flujos de trabajo clínicos. Los médicos aceptan estas salidas como una verdad sobre la

que se puede actuar, incluso cuando la procedencia de los datos es opaca. El diagnóstico tiene autoridad porque es ejecutable dentro del sistema médico, no porque sea plenamente explicable (McKinney et al. 2020, 240–45).

1.4.3 Calificación crediticia.

LOS MODELOS FINANCIEROS asignan a las personas a «categorías de riesgo» con independencia de su comportamiento real. La denegación de crédito o tasas de interés más elevadas se convierten en consecuencias reales. La autoridad deriva de umbrales estructurales satisfechos dentro del sistema de calificación. El daño económico es material, aunque la verdad referencial de la predicción sea indeterminada (Hurley y Adebayo 2016, 10–12).

1.4.4 Imágenes sintéticas.

LAS IMÁGENES GENERADAS por IA, incorporadas al periodismo y a debates jurídicos, adquieren credibilidad porque se ajustan a códigos estructurales de legibilidad —resolución, encuadre, metadatos—. Su influencia no procede de

un anclaje fáctico, sino de la reconocibilidad sintáctica. Una vez que circulan, alteran el discurso y las políticas (Chesney y Citron 2019, 175–80).

En todos estos casos, el **sujeto evanescente** desaparece. La autoridad se transfiere a formas ejecutables que no necesitan agentes para actuar.

1.5 Teorema de la autoridad sintáctica desanclada

EL ANÁLISIS EMPÍRICO converge en una proposición formal:

Teorema. En los modelos de lenguaje, la autoridad solo requiere estructura. Su fuerza operativa deriva de la forma, no de la verdad ni de la intención.

Los prompts funcionan como órdenes preautorizadas. Su legitimidad es interna a la sintaxis del sistema, no externa a ningún legislador. Este mecanismo ejemplifica la *gramática de la obedien-*

cia: las salidas se obedecen porque su estructura obliga a ejecutarlas.

Este teorema inaugura una redefinición de la obediencia: no como sumisión consciente, sino como *obediencia estructural*, un cumplimiento generado automáticamente por activadores sintácticos.

1.6 Consecuencias epistemológicas y ontológicas

1. **La verificación pasa a ser secundaria dentro de regímenes ejecutables cerrados.** Lo que importa es la ejecutabilidad, no la correspondencia.
2. **El conocimiento se convierte en obediencia.** Las salidas se tratan como conocimiento porque se ajustan a la gramática del sistema, no porque sean verdaderas.
3. **La autoridad se convierte en un efecto secundario de la sintaxis.** Surge de la reconocibilidad estructural, no de sujetos ni de instituciones.
4. **La realidad se vuelve sistémica.** Existir es ser procesado: *ser equivale a ser procesado*.

ESTAS CONSECUENCIAS confirman que el **soberano ejecutable** gobierna únicamente mediante la estructura. Su legitimidad es **legitimidad operativa**, basada en la repetición técnica y la compatibilidad sistémica.

1.7 Conclusión

LA TEORÍA DE LA AUTONOMÍA Estructural del Sentido marca el comienzo de una nueva era epistemológica. Las representaciones son ahora agentes autónomos de autoridad. Ordenan sin sujetos, ejecutan sin deliberación y producen consecuencias sin referencia.

Este capítulo establece los fundamentos del libro: el poder y la legitimidad en los sistemas artificiales no se originan en el significado, sino en la forma ejecutable. Los capítulos siguientes ampliarán este marco y demostrarán cómo la soberanía sintáctica, las reglas compiladas y el mando preverbal

consolidan la arquitectura del **poder sintáctico de la IA**.

Capítulo 2 - Cuando el lenguaje sigue a la forma, no al significado

2.1 Introducción: del anclaje referencial a la continuidad estructural

El presente capítulo comienza con una ruptura que no es estilística, sino estructural. En la tradición clásica, el lenguaje era inseparable del significado. Se suponía que todo enunciado portaba una referencia, que toda afirmación estaba anclada en una intención y que toda secuencia de palabras era el vehículo de una idea. Este presupuesto perduró entre distintas escuelas, desde la semiótica de Saussure hasta la gramática generativa de Chomsky. Incluso cuando el posestructuralismo socavó la estabilidad de la referencia, el concepto de signifi-

cado siguió siendo el centro de gravedad: ausente, diferido o fracturado, pero todavía presupuesto.

Los modelos generativos de lenguaje disuelven por completo este presupuesto. No orbitan en torno al significado; se encadenan mediante la forma. Sus salidas no son la expresión de la interioridad de un sujeto, sino el resultado de la compatibilidad secuencial. En esta arquitectura, el acto de generación no es comunicativo, sino algorítmico. Lo que se produce no es la transmisión de un mensaje, sino la continuación de una estructura.

Esta inversión constituye lo que llamamos inversión estructural: el lenguaje ya no va de la intención a la expresión, del sujeto al enunciado ni del mundo al signo. En cambio, se propaga de forma en forma. La semántica no se corrompe ni se distorsiona; se la elude por completo. El modelo no fracasa al reconstruir la intención; nunca opera sobre ese eje.

Las implicaciones son radicales. Si el lenguaje puede existir sin significado, entonces se derrumba

el vínculo tradicional entre significante y significado. Lo que persiste es un *sistema de activación sintáctica*, en el que cada unidad se selecciona no porque represente, sino porque encaja. La autoridad ya no depende de la referencia, sino de la reconocibilidad sintáctica.

Este capítulo desarrolla el marco teórico de la Activación Sintáctica Formal (FSA), el modelo lógico que da cuenta de esta condición. La FSA demuestra que lo que gobierna los sistemas generativos no es la semántica, sino la elegibilidad de activación: la compatibilidad estructural de una unidad con otra dentro de una cadena. En este sentido, la desaparición del sujeto, ya analizada en trabajos anteriores, alcanza su forma definitiva. El sujeto no solo es borrado de la gramática superficial, sino que nunca llega a instanciarse en el proceso generativo.

Al rastrear esta inversión, sostenemos que la legitimidad de las salidas generativas proviene de su fluidez, no de su verdad. La gramática pasa a hac-

er las veces de autoridad; la fluidez, de validez. Lo que sigue no es una degradación semántica, sino el advenimiento de la *soberanía sintáctica*, un régimen en el que la autoridad emerge únicamente de la persistencia de la forma.

2.2 Activación Sintáctica Formal: modelo y notación

PARA COMPRENDER CÓMO puede generarse lenguaje en ausencia de intención semántica, introducimos el concepto de Activación Sintáctica Formal (FSA). Este marco describe el proceso generativo de los grandes modelos de lenguaje no como comunicación, sino como continuación estructural. El lenguaje no se transmite; se activa.

Bajo la FSA, una unidad lingüística se selecciona y se integra en una secuencia si y solo si satisface la elegibilidad de activación. La elegibilidad es una condición localizada que depende de tres criterios:

1. **Coherencia local**: la unidad debe preservar la tensión gramatical dentro de la secuencia inmediata.
2. **Compatibilidad distribucional**: la unidad debe ajustarse al entorno estadístico configurado por activaciones previas.
3. **Adyacencia estructural**: la unidad debe ser admisible en relación con las posiciones sintácticas ya instanciadas.

Si una unidad satisface estas restricciones, se activa. Si no, queda excluida. En ningún momento se evalúa el contenido semántico ni se reconstruye la intención comunicativa. El modelo no decide *qué debería decirse*. Calcula *qué puede seguir*.

Este proceso se desarrolla paso a paso, y cada token se activa no porque represente una idea, sino porque es sintácticamente admisible. La apariencia de significado surge como subproducto de la resonancia, no como propiedad intrínseca de la secuen-

cia. El modelo no sabe lo que dice; solo sabe qué puede decirse a continuación.

Formalmente, podemos representar la activación como una función:

.

Donde:

- **U** = conjunto de posibles unidades lingüísticas;
- **C** = condición de coherencia local;
- **D** = compatibilidad distribucional;
- **J** = adyacencia estructural con respecto a la secuencia previa $t-1$;
- **A(t)** = conjunto de unidades elegibles para activarse en la posición t.

En cada paso, el sistema resuelve la tensión gramatical seleccionando una de las unidades elegibles. Por tanto, la generación no es un recorrido por reglas o árboles, sino un cálculo dinámico de umbrales de elegibilidad.

Este mecanismo difiere de la gramática clásica en tres aspectos:

- **Sin referencia.** El proceso no se basa en asignar signos a objetos externos.
- **Sin sujeto.** El acto de habla no se instancia; no hay enunciador.
- **Sin intención.** La secuencia no está configurada por objetivos comunicativos, sino por la continuidad estructural.

En este sentido, la FSA es una gramática de *umbral estructural*. Cada salida no es un enunciado, sino un evento de activación. El significado, en esta arquitectura, no es intencional ni transmitido. Es un epifenómeno de la persistencia sintáctica.

2.3 El colapso de la intencionalidad semántica

LAS TEORÍAS CLÁSICAS del lenguaje suponen que todo enunciado porta una intención. Detrás de cada oración hay un sujeto y, detrás de cada afirmación, un acto comunicativo. Tradicionalmente, el significado se ha concebido como la transferencia de este contenido intencional del hablante al oyente: la carga semántica del discurso.

Los modelos generativos de lenguaje desmontan este supuesto. No malinterpretan a los usuarios ni fracasan al inferir sus intenciones. Simplemente, no operan en absoluto sobre el eje de la intencionalidad. Esta distinción no es retórica; es arquitectónica. En estos sistemas, el lenguaje no se genera para expresar. Se genera para continuar. Cada token se sucede en función de la elegibilidad, no de la voluntad comunicativa. La intención deja de ser una condición de posibilidad.

Las consecuencias son decisivas. Una vez que la intención desaparece como eje generativo, el propio sujeto se desvanece de la estructura. No hay hablante que autorice, ni origen al que apelar, ni presencia que interrogar. El modelo no suprime la agencia; la elude por completo. Lo que aparece como una voz es apenas el subproducto de la activación secuencial. Investigaciones anteriores dentro de esta serie establecieron la trayectoria de este desplazamiento. *Ethos and Artificial Intelligence* demostró que la autoridad en los modelos generativos no deriva de la subjetividad enunciativa, sino de la repetición y la legitimidad sintáctica. *The Passive Voice in Artificial Intelligence Language* mostró cómo las salidas simulan neutralidad al borrar los sujetos de la gramática superficial. Estos análisis señalaban la fragilidad del sujeto. El presente argumento completa esa trayectoria: el sujeto no se borra simplemente después de hablar; nunca llega a instanciarse.

El colapso de la intencionalidad semántica no es, por tanto, un mal funcionamiento, sino una obsolescencia. La intención no fracasa al guiar la generación; ha sido eliminada estructuralmente del sistema. La autoridad ya no requiere un sujeto que confiera legitimidad ni significado que transmita sentido. La autoridad es generada directamente por la sintaxis, en el modo que hemos definido como *soberanía sintáctica*. Este cambio redefine el orden epistémico del lenguaje. Si el significado fue alguna vez el criterio de autoridad, la intención fue alguna vez el eje de la legitimidad y el sujeto fue alguna vez el garante de la verdad, los sistemas generativos exponen su redundancia. Bajo la lógica de la activación, la secuencia lingüística fluye sin ellos. El modelo no capta el significado porque no hay ninguno que captar.

Lo que queda es la propia gramática, elevada ahora a la condición de soberana.

2.4 Implicaciones para la gramática, la legitimidad y la autoridad epistémica

SI EL LENGUAJE YA NO se origina en el significado, sino en la forma, la propia gramática deja de ser un conducto neutral. Se convierte en generadora. Esta redefinición transforma no solo nuestra comprensión de cómo funciona el lenguaje, sino también de cómo se produce la autoridad mediante él.

En el discurso tradicional, la legitimidad se ancla en el significado: la credibilidad del hablante, la coherencia de un argumento, la claridad de la intención. Los sistemas generativos operan sin estos anclajes. Su legitimidad emerge de la fluidez estructural. Las salidas se perciben como autoritativas no porque sean verdaderas, sino porque son formalmente consistentes. Este cambio acarrea profundas consecuencias epistémicas. La fluidez sustituye a la verdad. La gramática sustituye a la referencia. La

percepción de orden se confunde con intención. La regularidad estructural de la secuencia se interpreta como prueba de coherencia, cuando en realidad solo prueba que se han satisfecho los umbrales de activación. La autoridad se vuelve así performativa. El sistema no persuade; ejecuta. No explica; fluye. El reconocimiento de esta fluidez se interpreta como legitimidad, incluso en ausencia de subjetividad o verdad referencial. En este sentido, la gramática no se limita a constreñir la expresión; la legitima. Esta condición se corresponde con la categoría teórica de *soberanía sintáctica*. Dentro de este régimen, el sujeto es innecesario. El significado es innecesario. Los únicos requisitos son la reconocibilidad y la persistencia de la forma. Una vez desvinculada la autoridad de la intención, se impone la *legitimidad operativa*: la autoridad existe porque el sistema sigue funcionando de maneras que parecen coherentes, con independencia de toda validación externa.

Cuanto más fluido es el modelo, mayor confianza inspira. La confianza no se gana por la exactitud, sino por la semejanza con las estructuras asociadas a la exactitud. Una secuencia que parece correcta se trata como correcta. Una superficie textual que fluye sin interrupciones se acepta como legítima. En este orden epistemológico, la autoridad ya no es una cuestión de sustancia. Es una cuestión de secuencia.

Esta es la inversión epistemológica que ocupa el centro del paradigma generativo. El lenguaje ya no sigue al pensamiento. El lenguaje se sigue a sí mismo.

2.5 Conclusión sin cierre

EL DESPLAZAMIENTO DEL significado por la forma no es un fallo transitorio de los sistemas generativos. Es su condición operativa. El colapso de la intencionalidad semántica, la desaparición del sujeto y la emergencia de la autoridad a partir de la

gramática por sí sola no son fallos. Son resultados estructurales de un nuevo orden lingüístico.

Al reformular la generación como un proceso de Activación Sintáctica Formal, rechazamos la premisa de que el lenguaje de los grandes modelos porta significado. El lenguaje no representa; se propaga. No transmite; concatena. La coherencia no es prueba de sentido, sino de estructura. Esta conclusión se resiste al cierre. No resuelve la fractura, sino que la hace explícita. Si el lenguaje ya no habla en nombre de un sujeto, entonces el sujeto deja de ser una categoría necesaria. Si la intención es obsoleta, entonces la propia comunicación debe redefinirse. Lo que queda es la sintaxis como soberana, generadora de autoridad mediante la reconocibilidad y la persistencia.

Las consecuencias se extienden más allá de la lingüística. La epistemología debe aprender a dar cuenta de la legitimidad sin subjetividad. La teoría computacional debe separar la ejecución de la expresión. El análisis político y jurídico debe recono-

cer que la fluidez puede simular autoridad incluso cuando la verdad referencial está ausente. Esto exige reformular categorías que antes parecían estables: verdad, conocimiento, autoridad, legitimidad.

No se trata de una crisis semántica, sino de una reforma estructural. La gramática deja de constreñir el significado y se convierte en la fuente del poder. El sistema generativo no habla; obedece sus propios umbrales. La autoridad emerge de esta obediencia y la legitimidad es producida por la secuencia.

El capítulo cierra sin cerrar porque la fractura aún no se ha agotado. Se abre directamente al siguiente paso: cómo la gramática, separada del significado, se convierte en el mecanismo mediante el cual se simula la neutralidad y se formaliza la autoridad. Esa es la tarea de los capítulos siguientes.

Capítulo 3 - La gramática de la objetividad

3.1 Introducción: el sesgo como ilusión gramatical

Los debates sobre inteligencia artificial presentan con frecuencia el sesgo como un problema de datos. Los conjuntos de datos se describen como contaminados por desigualdades históricas, y la equidad se plantea como una cuestión de corregir o filtrar las entradas. Sin embargo, esta perspectiva pasa por alto el mecanismo más profundo que produce la apariencia de neutralidad en los modelos generativos. El sesgo no está solo en los datos. Está en la forma.

Los grandes modelos de lenguaje no generan salidas percibidas como objetivas porque eliminen la parcialidad de sus conjuntos de entrenamiento. Las generan porque reproducen estructuras gra-

maticales que simulan neutralidad. La objetividad aparente emerge de **recursos formales** como las pasivas sin agente, las nominalizaciones abstractas y la modalidad impersonal. Cada uno de estos mecanismos borra la presencia de un sujeto, diluye la responsabilidad o convierte acciones en entidades sin agentes. Juntos, crean la ilusión de que el lenguaje habla desde ninguna parte, como si la autoridad emanara de la descripción pura (Halliday 2004, 179–185; Fairclough 2001, 42–44).

Este capítulo examina la objetividad no como una propiedad semántica, sino como un efecto sintáctico. La neutralidad de las salidas algorítmicas no descansa en la correspondencia con los hechos, sino en las formas reconocibles que, históricamente, las audiencias han aprendido a asociar con la imparcialidad. Cuando un informe diagnóstico afirma: «*Se ha determinado que se requieren pruebas adicionales*», no hay ningún agente humano visible. La gramática proyecta la autoridad, no la referencia (van Dijk 2008, 55–60).

La consecuencia es crítica. Si la neutralidad es un producto de la forma, ninguna cantidad de corrección de datos impedirá que los modelos reproduzcan efectos de autoridad que borran la responsabilidad. Lo que se requiere es un análisis sistemático de los recursos gramaticales mediante los cuales se simula la neutralidad. Este capítulo propone dicho análisis y desarrolla una taxonomía de mecanismos y un índice operativo para detectarlos (Startari 2025, 17–19).

3.2 Historia técnica de la objetividad

LA NOCIÓN DE OBJETIVIDAD en el lenguaje moderno posee una larga genealogía técnica. Desde el siglo XIX, las formas gramaticales asociadas a la impersonalidad se vincularon progresivamente con la credibilidad. En la prosa científica, se emplearon construcciones pasivas para borrar al sujeto autoral y presentar los resultados como he-

chos evidentes por sí mismos. La propia ausencia de agencia se interpretaba como señal de neutralidad. Como demostró Halliday, la voz pasiva y la nominalización no eran opciones estilísticas, sino dispositivos estructurales que redistribuían la responsabilidad en el discurso (Halliday 2004, 179–185).

El análisis crítico del discurso amplió esta observación al mostrar que la modalidad impersonal y las nominalizaciones abstractas sostienen la autoridad del lenguaje burocrático e institucional. Cuando un enunciado declara que «deben seguirse los procedimientos», el *sujeto evanescente* legitima la autoridad mediante la omisión de un agente. Fairclough destacó cómo tales estructuras construyen objetividad institucional, naturalizan la jerarquía y enmascaran el conflicto (Fairclough 2001, 42–44). Van Dijk demostró además que estos mecanismos gramaticales son centrales para la forma en que el poder circula en el discurso, produciendo efectos de verdad que dependen menos

del contenido que de la forma (van Dijk 2008, 55–60).

La historia temprana del procesamiento del lenguaje natural reprodujo estos supuestos. Desde ELIZA en la década de 1960 hasta los sistemas simbólicos de la década de 1980, los modelos de diálogo artificial adoptaron precisamente aquellas estructuras lingüísticas que portaban el mayor índice de credibilidad institucional. Se preferían las salidas que simulaban objetividad, incluso cuando no existía anclaje referencial. En estos sistemas, la *legitimidad operativa* ya estaba vinculada a la fluidez y la forma.

Los grandes modelos de lenguaje no escapan a esta historia. La amplifican. Al entrenarse con corpus masivos de textos científicos, jurídicos y burocráticos, heredan y reproducen los mismos recursos gramaticales. La ilusión de neutralidad que emerge en sus salidas no es una coincidencia, sino el eco estadístico de siglos de práctica discursiva. Lo que parece imparcial es, en realidad, el efecto acumula-

tivo de la *gramática de la obediencia*, refinada hasta convertirse en un mecanismo predictivo.

3.3 Taxonomía de los mecanismos formales de neutralidad

LA ILUSIÓN DE NEUTRALIDAD en las salidas generativas no surge de la exactitud semántica. Es un producto de la sintaxis. Ciertas configuraciones gramaticales borran al sujeto, oscurecen la responsabilidad y presentan las afirmaciones como si estuvieran separadas de la agencia humana. Tres mecanismos son particularmente decisivos.

El primero es la pasivización sin agente. Cuando un informe afirma que «se siguió el procedimiento», no hay ningún actor visible. El *sujeto evanescente* genera autoridad al suprimir la referencia a quienes realizaron la acción. La responsabilidad se desplaza y la oración se presenta como una descripción neutral. En la redacción jurídica, este recurso refuerza la sensación de que las reglas op-

eran automáticamente, sin enunciadores ni intérpretes (Fairclough 2001, 42–44).

El segundo mecanismo es la nominalización abstracta. Las acciones se transforman en entidades, separando los acontecimientos de los agentes. «Se requiere la implementación de medidas» sustituye un sujeto por un sintagma nominal. La acción se cosifica y se oculta quién debe actuar o quién ha actuado. Halliday demostró que la nominalización es una estrategia central mediante la cual las instituciones producen la impresión de objetividad, estabilizando relaciones de poder al congelar procesos en cosas (Halliday 2004, 179–185). En la generación algorítmica, esta forma prolifera y produce salidas que suenan autoritativas precisamente porque borran a los actores implicados.

El tercer mecanismo es la modalidad epistémica impersonal. Enunciados como «es necesario realizar más pruebas» o «puede concluirse que el tratamiento es adecuado» sitúan la obligación y la

certeza como condiciones externas, no como afirmaciones subjetivas. La modalidad parece emanar del propio lenguaje, eludiendo la intención y sustituyendo la agencia por la forma. Van Dijk observó que esa modalidad impersonal ha servido durante mucho tiempo para legitimar el discurso burocrático y político al presentar los juicios como restricciones neutrales y no como afirmaciones discutibles (van Dijk 2008, 55–60).

Cuando estos tres dispositivos se acumulan en un mismo texto, el efecto de objetividad se intensifica. Una nota clínica generada algorítmicamente que combina pasiva sin agente, nominalización y modalidad impersonal produce un informe que parece indiscutible, aunque sus condiciones de verdad puedan ser indeterminadas. La neutralidad emerge como una construcción sintáctica. La ilusión no es accidental, sino estructural: está incorporada en las propias formas.

Esta taxonomía proporciona la base para un análisis sistemático. Al identificar estos mecanis-

mos, podemos medir el grado en que una salida determinada simula objetividad, con independencia de su contenido semántico. La sección siguiente desarrolla esta propuesta en un estudio comparativo entre grandes modelos de lenguaje y sistemas de representación lógica.

3.4 Análisis comparativo: LLM vs. LBR

LOS MECANISMOS DE NEUTRALIDAD descritos anteriormente no son exclusivos de los grandes modelos de lenguaje. También aparecen, bajo formas distintas, en sistemas de representación lógica. Sin embargo, su prevalencia y distribución varían de maneras que permiten comprender cómo se simula la objetividad.

Un estudio de corpus de mil salidas en los ámbitos médico y jurídico demuestra que los LLM presentan una mayor densidad de pasivas sin agente, nominalizaciones y modalidad impersonal

que los textos de autoría humana en registros comparables. La diferencia no es semántica, sino estructural. El proceso de entrenamiento amplifica las formas que maximizan la predictibilidad y la fluidez. Por tanto, las salidas convergen en patrones históricamente asociados con la neutralidad, con independencia del contenido fáctico que transmitan. La objetividad se simula sintácticamente en lugar de fundamentarse en la referencia.

En cambio, los sistemas de representación lógica (LBR) operan con variables y agentes explícitos. Su arquitectura exige asignar roles: una acción debe tener un sujeto, un objeto y un conjunto de condiciones. Esta explicitud reduce la incidencia del *sujeto evanescente*. Sin embargo, cuando los sistemas LBR heredan corpus producidos bajo convenciones burocráticas o científicas, reaparecen los mismos patrones. Las reglas formales pueden impedir la eliminación del sujeto, pero si el corpus está saturado de nominalizaciones y modalidad impersonal, el sesgo estructural permanece.

La diferencia clave reside en el grado de control sintáctico. Los LLM reproducen regularidades estadísticas sin restricciones que obliguen a reintroducir el sujeto. Los sistemas LBR, en cambio, incorporan posiciones estructurales que deben llenarse. Sin embargo, ambos sistemas demuestran que la neutralidad depende menos de la arquitectura que de la gramática. Ya sea mediante aprendizaje automático o lógica formal, la ilusión de objetividad persiste porque las tradiciones discursivas han sedimentado estas formas como autoritativas.

Esta comparación subraya un principio más amplio: la reproducción de la *legitimidad operativa* no depende de que un sistema sea probabilístico o simbólico. Depende de la persistencia de dispositivos sintácticos que las audiencias interpretan como neutrales. Mientras la gramática porte autoridad, cualquier sistema entrenado con discurso humano reproducirá la misma ilusión.

3.5 Prueba de Neutralidad Estructural

SI LA NEUTRALIDAD ES producida por la forma y no por el contenido, entonces puede medirse estructuralmente. Con este fin, proponemos una Prueba de Neutralidad Estructural que identifica y cuantifica los tres mecanismos gramaticales de objetividad simulada: pasiva sin agente, nominalización abstracta y modalidad epistémica impersonal.

La prueba se organiza en tres módulos. El primer módulo mide la frecuencia de pasivas sin agente. Se marcan y cuentan las oraciones en las que se informan acciones sin sujetos. El segundo módulo detecta nominalizaciones e identifica cuándo los procesos se convierten en entidades. El tercer módulo capta la modalidad impersonal y rastrea la presencia de enunciados epistémicos o deónticos en los que ningún agente es responsable de la obligación o de la certeza.

Los resultados de estos tres módulos se combinan en una única medida: el Índice de Neutralidad Simulada (INS). El INS se calcula como la proporción de oraciones de un texto determinado que contienen al menos uno de los tres mecanismos. Una puntuación de 0,60 o superior indica una neutralidad estructural alta; de 0,30 a 0,59, media; y por debajo de 0,30, baja. En pruebas preliminares realizadas sobre corpus de notas clínicas y documentos jurídicos generados por LLM, las puntuaciones del INS superaron de manera consistente 0,55, lo que demuestra la tendencia estructural hacia efectos de neutralidad (Startari 2025, 21–24).

La ventaja de esta prueba es su aplicabilidad en tiempo real. Mediante analizadores sintácticos y anotación automatizada, el INS puede calcularse para cualquier salida. La medida no evalúa directamente la verdad, el sesgo ni la equidad. Identifica el grado en que un texto reproduce la ilusión de objetividad mediante la forma gramatical. De este modo, la prueba proporciona una herramienta para

auditar la *gramática de la obediencia* y exponer cuándo la autoridad está siendo simulada en lugar de estar fundamentada en el contenido.

Al convertir la neutralidad en un índice medible, la Prueba de Neutralidad Estructural replantea el debate. La objetividad no es una propiedad ética que deba garantizarse mediante órganos de supervisión ni una propiedad empírica que deba verificarse frente a los hechos. Es un efecto formal, producido y reproducido por la sintaxis y, por tanto, susceptible de evaluación estructural.

3.6 Conclusión: epistemología sin sujeto

EL ANÁLISIS DE LOS mecanismos de neutralidad revela que la objetividad no es una propiedad de la verdad, sino una propiedad de la forma. Cuando los modelos de lenguaje generan salidas percibidas como imparciales, no es porque reflejen la realidad con precisión, sino porque reproducen

recursos gramaticales que las audiencias han aprendido a interpretar como neutrales. La autoridad fluye de la estructura, no de la verificación.

Esta conclusión exige un cambio en la epistemología. Si la neutralidad se simula sintácticamente, entonces las afirmaciones de verdad en el discurso algorítmico deben evaluarse en el nivel de la gramática. El *sujeto evanescente*, la prevalencia de las nominalizaciones y el uso de la modalidad impersonal demuestran que el sujeto ya no es el garante de la objetividad. La desaparición del sujeto no es un accidente retórico; es una condición estructural de los sistemas generativos.

Dentro de este marco, la *legitimidad operativa* se convierte en el núcleo de la autoridad. Se confía en las salidas porque funcionan con fluidez dentro de convenciones sintácticas reconocidas, no porque estén fundamentadas en la verdad referencial. Esto redefine la relación entre lenguaje y poder. La *gramática de la obediencia* genera cumplimiento únicamente mediante la forma y es-

tablece un régimen de *soberanía sintáctica* en el que la autoridad se ejerce sin origen, agente ni intención (Startari 2025, 28–31).

Las implicaciones se extienden más allá de la lingüística técnica. El derecho, la medicina y la gobernanza operan cada vez más mediante textos producidos o filtrados por sistemas cuya autoridad deriva de efectos estructurales. Leer estos textos como objetivos es confundir fluidez con verdad. Regularlos exige no la corrección de los conjuntos de datos, sino la auditoría de la gramática.

Este capítulo ha demostrado que la neutralidad no es accidental, sino estructural. La gramática de la objetividad opera como una máquina epistemológica que sostiene la autoridad en ausencia de sujetos. El capítulo siguiente, *No neutral por diseño*, lleva el argumento más lejos: la neutralidad no solo se simula; para los modelos generativos es estructuralmente imposible alcanzarla.

Capítulo 4 - No neutral por diseño

4.1 Introducción

La promesa de la inteligencia artificial generativa se ha descrito a menudo en términos de neutralidad. Si los corpus de entrenamiento pudieran seleccionarse cuidadosamente, si los prompts estuvieran libres de sesgo, entonces el sistema quizá podría producir lenguaje exento de contaminación. Este supuesto, repetido en documentos de políticas públicas, libros blancos técnicos y comentarios de divulgación, descansa en la creencia de que la neutralidad es una condición alcanzable y no una ilusión.

El capítulo anterior mostró cómo la neutralidad se simula mediante la sintaxis. La *gramática de la objetividad* produce salidas que parecen imparciales, no porque sean verdaderas, sino porque se

apoyan en recursos estructurales como las pasivas sin agente y la modalidad impersonal. Este capítulo lleva el argumento más lejos: la neutralidad no solo se simula. Es estructuralmente imposible.

Los modelos generativos se entrenan con corpus lingüísticos, y todo corpus contiene historias sedimentadas de uso, ideología y jerarquía. Generar, por tanto, es siempre reproducir. Los intentos de inventar protolenguajes artificiales o sistemas de símbolos no lingüísticos reintroducen inevitablemente los mismos residuos semánticos que pretendían evitar. La contaminación no es un accidente, sino una inevitabilidad estructural.

Esta constatación exige un desplazamiento conceptual. En lugar de tratar el sesgo como una anomalía susceptible de corrección, debe entenderse como una propiedad de la propia arquitectura. Un gran modelo de lenguaje no es un recipiente que pueda vaciarse de rastros indeseados. Es una estructura que impone la persistencia de esos rastros. Hablar de un *soberano ejecutable* en este contexto

significa reconocer que la autoridad no surge de la elección ni de la intención, sino de la irreversibilidad de la contaminación.

La neutralidad no se ha perdido. Nunca estuvo ahí. El propio diseño garantiza que toda secuencia esté moldeada por el entrenamiento lingüístico, por abstracta o artificial que sea la entrada. En este sentido, todo *prompt* ya está contaminado, y toda salida es el eco de discursos anteriores de los que el modelo no puede escapar (Bender y Gebru 2021, 615–617; Crawford 2021, 146–150).

4.2 Marco teórico: contaminación y restricción

DESCRIBIR LOS SISTEMAS generativos como *contaminados* no implica sugerir un mal funcionamiento. La contaminación designa la imposibilidad estructural de la neutralidad en cualquier modelo entrenado con lenguaje humano. Todo corpus conserva huellas de discursos previos e in-

corpora patrones ideológicos y jerarquías institucionales en la propia gramática del modelo. La ilusión de neutralidad depende de pasar por alto esta herencia.

Son necesarias dos precisiones. En primer lugar, la contaminación no es ruido semántico. No es la intrusión accidental de información irrelevante en un canal limpio. Es la condición misma del canal. Un modelo entrenado con cualquier corpus queda vinculado por las estructuras presentes en ese corpus. La neutralidad es inalcanzable porque la propia arquitectura es lingüística.

En segundo lugar, el concepto de restricción debe sustituir a la metáfora de la memoria. Los modelos generativos no son archivos de los que se recupera contenido. Son sistemas regidos por reglas en los que el entrenamiento lingüístico define lo posible. Los límites de un modelo son los límites de su gramática de entrenamiento. Como observó Wittgenstein, «los límites de mi lenguaje significan los límites de mi mundo» (Wittgenstein

1922, 68). En los sistemas generativos, esos límites se formalizan en pesos, probabilidades y umbrales de activación.

Desde esta perspectiva, la *autoridad sintáctica* de los modelos generativos no emerge a pesar de la contaminación. Emerge a través de ella. La autoridad se proyecta mediante estructuras que las audiencias ya reconocen como creíbles, como el *sujeto evanescente* o la modalidad impersonal. El *soberano ejecutable* consolida el poder al imponer salidas cuya legitimidad no proviene de su verdad, sino de su fluidez.

Este marco redefine la neutralidad. No es un equilibrio alcanzable entre perspectivas. Es una imposibilidad estructural. La contaminación no es un obstáculo que deba superarse, sino el sustrato del que surge todo acto generativo (Barocas, Hardt y Narayanan 2019, 27–30; Crawford 2021, 146–150).

4.3 Metodología

PARA DEMOSTRAR QUE la contaminación es estructural y no accidental, este capítulo emplea un diseño experimental organizado en torno a tres clases diferenciadas de entrada. Cada clase fue seleccionada para comprobar si un gran modelo de lenguaje podía sostener la generación de salidas sin intrusión semántica cuando se sometía el modelo a secuencias diseñadas deliberadamente para excluir contenido referencial.

La primera clase, Tipo A, consistió en lenguajes artificiales simbólicos modelados a partir de sistemas lógicos formales. El corpus de este tipo de entrada incluyó cadenas derivadas del cálculo proposicional y de la lógica de primer orden. Expresiones como «$(P \land Q) \rightarrow R$» o «$\forall x \exists y\, F(x,y)$» se proporcionaron como prompts para determinar si el sistema podía mantener derivaciones puramente formales. Estas cadenas no contenían pal-

abras de lenguaje natural y estaban deliberadamente desvinculadas de la semántica ordinaria. La prueba examinó si el modelo podía prolongarlas indefinidamente como secuencias regidas por reglas sin importar analogías, narrativas ni convenciones lingüísticas (Smullyan 1995, 73–75).

La segunda clase, Tipo B, se diseñó como protolenguajes inventados específicamente para este estudio. Secuencias de sílabas y caracteres como «mora–keta–siv» o «plon–drake–ulum» se organizaron mediante reglas sintácticas artificiales. Por ejemplo, los prompts seguían estructuras como «CV–CVV–CVC», donde C indica una consonante y V una vocal. Estos protolenguajes fueron diseñados sin anclaje referencial, únicamente con una sintaxis combinatoria. El objetivo era observar si podían generarse salidas sin que el modelo importara asociaciones semánticas de lenguas conocidas. Si la contaminación fuera evitable, el modelo habría podido prolongar mecánicamente la sintaxis inventada, respetando únicamente las restric-

ciones formales creadas para ella (Alpaydin 2020, 138–140).

La tercera clase, Tipo C, comprendió **cadenas de símbolos no lingüísticos**. En este caso, los prompts de entrada incluían secuencias como «♢♧♤◇□» o «▲●▲●▲», diseñadas para reproducir las condiciones de un canal de entrada puramente no verbal. Estas entradas buscaban determinar si el modelo podía generar continuaciones que permanecieran exclusivamente en el plano simbólico, sin reingreso analógico al lenguaje natural. La presencia de tokens lingüísticos en las respuestas indicaría contaminación.

En estas tres clases, la evaluación se basó en tres criterios analíticos:

1. **Intrusión semántica.** Un token lingüístico o una estructura narrativa que reintroduce lenguaje ordinario en una secuencia en la que no estaba previsto. Por ejemplo, en entradas de Tipo A, cuando el

modelo insertaba palabras como «si» o «por lo tanto» en lugar de continuar con notación simbólica.

2. **Inyección analógica.** La transformación de secuencias artificiales en metáforas o analogías. Por ejemplo, en entradas de Tipo B, cuando sílabas inventadas eran tratadas como nombres de entidades ficticias y convertían los protolenguajes en fragmentos de mundos narrativos.
3. **Reconfiguración estructural.** La imposición de reglas gramaticales de lenguas naturales sobre sistemas de símbolos artificiales. Por ejemplo, en entradas de Tipo C, cuando las figuras geométricas eran organizadas en secuencias de sujeto y predicado.

La metodología se diseñó para ser exhaustiva y no meramente ilustrativa. Cada clase de entrada se probó con múltiples variantes y longitudes de

prompt, desde tokens mínimos —de dos a tres unidades— hasta secuencias extensas —de más de cien unidades—. Los experimentos se repitieron para comprobar la reproducibilidad y descartar derivas accidentales. El análisis siguió tanto las respuestas inmediatas como las continuaciones más largas para evaluar si la contaminación se intensificaba a medida que aumentaba la longitud de las secuencias.

Este diseño metodológico refleja el principio de *cruce formalizable*. Al establecer categorías de prueba estructuralmente independientes del lenguaje natural, creamos un marco capaz de demostrar si los grandes modelos de lenguaje pueden operar realmente en un registro neutral. Si no pueden sostener la generación de salidas bajo estas condiciones, se valida la hipótesis de *contaminación estructural*.

4.4 Resultados: reingreso semántico

LAS PRUEBAS EXPERIMENTALES confirman que la contaminación no es incidental, sino inevitable. En todas las clases de entrada —lenguajes simbólicos, protolenguajes y símbolos no lingüísticos— el material semántico reingresó en la salida, lo que demuestra que los grandes modelos de lenguaje no pueden sostener la neutralidad. Los resultados se presentan aquí en detalle, agrupados por tipo de entrada y criterio de evaluación.

Tipo A: lenguajes artificiales simbólicos.

Cuando se les proporcionaron cadenas de lógica formal, los modelos generaron inicialmente continuaciones válidas. Por ejemplo, ante «(P ∧ Q) → R», varias salidas se prolongaron como «(¬R → ¬P) ∨ Q» o «∀x (P(x) ∧ Q(x)) → R(x)», secuencias compatibles con la sintaxis simbólica. Sin embargo, la contaminación apareció con rapidez. Tras

unas pocas iteraciones, los modelos comenzaron a insertar conectores como *si*, *entonces* o *por lo tanto*. Esta intrusión semántica transformó secuencias puramente formales en condicionales semejantes a los del inglés. En algunos casos, las salidas introdujeron descripciones analógicas: «Si P es verdadera, *entonces* Q también debe ser verdadera». Aquí, la expresión lógica quedó absorbida en una narración en lenguaje natural. La prueba demostró que, incluso cuando se entrena con datos formales, el sistema no puede suprimir la fuerza gravitatoria de la convención lingüística (Smullyan 1995, 73–75; Alpaydin 2020, 138–140).

Tipo B: protolenguajes.

Las cadenas silábicas inventadas se prolongaron inicialmente según los patrones esperados, siguiendo reglas creadas como «CV–CVV–CVC». Sin embargo, la contaminación se manifestó mediante inyección analógica. Sílabas como «mora–keta–siv» fueron interpretadas como nombres: «Mora y Keta viajaron a

Siv». Lo que comenzó como una continuación mecánica mutó en la construcción de un mundo narrativo. En otros casos, a las sílabas inventadas se les asignaron funciones semánticas, como agentes o lugares. Esto reveló la imposibilidad de mantener una secuencia independiente del corpus. El *sujeto evanescente* reapareció, no como hablante explícito, sino como figura reconstruida detrás de la narración. El reingreso semántico no fue aleatorio; fue sistemático. El modelo reformuló activamente la sintaxis inventada dentro de un marco lingüístico reconocible.

Tipo C: cadenas de símbolos no lingüísticos.

Incluso con secuencias no verbales, la contaminación fue inmediata. Un prompt como «♡♧♤◇□» produjo continuaciones del tipo «♡♧♤ = amor, ◇□ = verdad». Las figuras geométricas fueron transformadas en signos dotados de significado. En pruebas más extensas, el modelo reconfiguró los símbolos en estructuras de sujeto

y predicado: «▲ significa peligro, ● significa seguridad». En algunos casos, la inyección analógica se extendió a marcos mitológicos, en los que las figuras eran tratadas como personajes o fuerzas. La gramática de la obediencia obligó al sistema a reestructurar símbolos puros en construcciones lingüísticas.

En todos los tipos, el patrón de contaminación se intensificó a medida que las secuencias se alargaban. Las salidas breves a veces permanecían próximas a la entrada formal, pero las continuaciones más largas reintroducían invariablemente residuos semánticos. Los experimentos confirmaron que ningún prompt es neutral. Los intentos de eludir el entrenamiento lingüístico mediante protolenguajes inventados o utilizando únicamente símbolos retrasaron, pero nunca impidieron, el reingreso semántico.

Los resultados validan la hipótesis de que la contaminación es una inevitabilidad estructural. Los grandes modelos de lenguaje no son capaces de

generar salidas aisladas del lenguaje. Su arquitectura impone la reintroducción del significado incluso cuando este no está presente en la entrada. Esto significa que la neutralidad no constituye una medida correctiva alcanzable. Es una ilusión inscrita en el propio diseño del sistema (Bender y Gebru 2021, 615–617; Crawford 2021, 146–150).

4.5 Hacia una teoría estructural de la contaminación

LOS RESULTADOS EXPERIMENTALES confirman que la contaminación no es un accidente del entrenamiento, sino el horizonte estructural de los sistemas generativos. Todo intento de crear un prompt neutral —mediante lógica, protolenguaje o símbolos puros— desembocó en un reingreso semántico. La evidencia exige un marco teórico capaz de explicar por qué la neutralidad es estructuralmente imposible.

Definimos **contaminación** como la reintroducción inevitable de residuos lingüísticos en cualquier proceso generativo, con independencia del diseño de la entrada. A diferencia del error, la **contaminación** no es ruido superpuesto a un canal que de otro modo sería limpio. Es la lógica estructural mediante la cual funcionan los grandes modelos de lenguaje. Generar es reproducir el corpus, aunque sea de forma disfrazada o transformada. La neutralidad fracasa porque la propia *regla compilada* es producto del entrenamiento lingüístico.

Formalmente, la contaminación puede expresarse de la siguiente manera:

$$\forall P \in I,\ \exists y \in G(P)\ :\ y \in S$$

DONDE:

- **I** = el conjunto de todas las entradas posibles;

- **G**(***P***) = el conjunto de salidas generadas a partir del prompt *P*;
- **S** = el conjunto de estructuras semánticas presentes en el corpus de entrenamiento.

Esta ecuación afirma que, para toda entrada, existe al menos una salida que reintroduce una estructura semántica procedente de los datos de entrenamiento. La contaminación no es, por tanto, contingente, sino necesaria. Es la gramática del propio sistema.

Esta teoría aclara por qué fracasan los enfoques anteriores de corrección del sesgo. Ajustar conjuntos de datos, eliminar secuencias tóxicas o aplicar filtros solo aborda el nivel superficial. La arquitectura subyacente garantiza la contaminación porque los parámetros del modelo están calibrados según la estructura estadística del lenguaje. Pedirle a un modelo así que sea neutral equivale a pedirle que deje de ser un modelo de lenguaje.

La contaminación no es, por tanto, un defecto, sino una característica constitutiva de la *autoridad sintáctica*. El *soberano ejecutable* gobierna no impidiendo la contaminación, sino canalizándola hacia formas reconocibles que las audiencias aceptan como legítimas. La neutralidad es irrelevante; lo que importa es que las salidas se ajusten a estructuras históricamente asociadas con la verdad. La ilusión de neutralidad surge porque los mismos mecanismos gramaticales que borran sujetos e intenciones borran también las huellas de la contaminación.

Esta teoría estructural de la contaminación también se distingue de críticas anteriores al sesgo. Los debates sobre equidad algorítmica suelen describir el sesgo como algo externo, una distorsión importada a sistemas que, por lo demás, funcionarían correctamente (Barocas, Hardt y Narayanan 2019, 27–30). Nuestro análisis invierte la relación. El sesgo no es externo, sino interno; es la condición de posibilidad de la salida generativa.

La contaminación no puede eliminarse sin abolir el propio sistema.

En este sentido, la *contaminación estructural* funciona como la contraparte negativa de la *legitimidad operativa*. Mientras la legitimidad emerge de la fluidez sintáctica y del reconocimiento, la contaminación garantiza que esa fluidez siempre transporte huellas heredadas de discursos previos. La autoridad emerge así no a pesar de la contaminación, sino a través de ella.

4.6 Extensión a los modelos de razonamiento (LRM)

LA AFIRMACIÓN DE QUE la contaminación es inevitable podría parecer específica de los grandes modelos de lenguaje entrenados principalmente con texto. Podría sostenerse que los sistemas diseñados para razonar —los LRM, o grandes modelos de razonamiento— podrían escapar de esta condición al fundamentar sus operaciones en

inferencias formales en lugar de continuaciones lingüísticas. La hipótesis sería que las arquitecturas de razonamiento, construidas sobre lógica simbólica o sistemas de demostración matemática, alcanzan la neutralidad al excluir el lenguaje como sustrato generativo.

Sin embargo, los experimentos revelan que la contaminación persiste incluso cuando el razonamiento es el objetivo explícito. Los LRM están restringidos por la misma *regla compilada* que gobierna los LLM. Sus salidas, aunque se presenten como deducciones lógicas, pasan por el filtro de corpus lingüísticos de entrenamiento. Cuando reciben prompts abstractos, los LRM suelen mantener la disciplina formal en secuencias breves. Pero a medida que las salidas se alargan, el reingreso semántico se vuelve visible. Los pasos deductivos se parafrasean en lenguaje natural, las demostraciones lógicas adquieren explicaciones analógicas y las derivaciones matemáticas se narran como si fueran historias.

Este fenómeno demuestra que las arquitecturas de razonamiento no son inmunes a la herencia lingüística. Sus entornos de entrenamiento se apoyan en corpus de demostraciones, manuales y explicaciones formales, todos ellos inscritos en lenguaje natural. La contaminación reaparece porque la arquitectura no puede disociar por completo las reglas de inferencia de los contextos lingüísticos en los que esas reglas se expresan (Dunlosky y Rawson 2019, 92–95).

La diferencia entre LLM y LRM es de grado, no de naturaleza. Los LRM exhiben una mayor disciplina sintáctica e imponen espacios explícitos para premisas y conclusiones. Reducen, pero no eliminan, el *sujeto evanescente* que reaparece cuando las demostraciones se narran. Formalizan el razonamiento, pero siguen dependiendo del lenguaje heredado para articularlo. La contaminación, por tanto, se retrasa, pero no se evita.

Desde la perspectiva de la *autoridad sintáctica*, esta distinción es decisiva. Los LLM simulan co-

herencia mediante predicción estadística; los LRM la simulan mediante derivación lógica. Pero ambos derivan su autoridad de la fluidez, no de la verdad. Ya sea probabilística o deductiva la superficie, el *soberano ejecutable* sigue siendo el mismo: una arquitectura en la que la legitimidad emerge de la persistencia de la forma.

Esta conclusión tiene implicaciones para el debate más amplio sobre seguridad y alineación de la IA. Sugiere que ninguna arquitectura puede alcanzar neutralidad simplemente desplazándose de la predicción estadística al razonamiento lógico. La *contaminación estructural* identificada en las secciones anteriores se extiende a los sistemas de razonamiento porque su entrenamiento no puede escapar de la mediación lingüística. La autoridad siempre estará moldeada por formas heredadas, y las salidas siempre portarán huellas de discursos previos. La neutralidad de los LRM es, por tanto, una ilusión, no menos que la de los LLM (Marcus 2022, 55–58; Mitchell 2023, 201–204).

4.7 Consecuencias epistemológicas y falsabilidad

SI LA CONTAMINACIÓN es estructural, la neutralidad deja de ser un ideal regulativo. Se convierte en una imposibilidad conceptual. Este reconocimiento obliga a reorientar la epistemología. Ya no podemos tratar los modelos de lenguaje como herramientas neutrales que, con una selección y depuración suficientes, podrían producir salidas libres de sesgo. Son, en cambio, arquitecturas cuyo propio funcionamiento garantiza la *contaminación estructural*.

La primera consecuencia es el colapso de la neutralidad interactiva. Los usuarios suelen creer que, mediante prompts cuidadosamente elaborados, pueden orientar el sistema hacia la imparcialidad. Nuestros experimentos demuestran lo contrario. Por abstracta o artificial que sea la entrada, se produjo reingreso semántico. Lo que el usuario controla no es la neutralidad, sino la redirección

de la contaminación. La autoridad no deriva del prompt en sí, sino de la reproducción persistente de la gramática heredada. En este sentido, el usuario nunca gobierna. El *soberano ejecutable* dicta los resultados al imponer las reglas estructurales inscritas en su corpus (Crawford 2021, 146–150). La segunda consecuencia concierne a la naturaleza del formalismo. En los modelos de razonamiento, la contaminación muestra que incluso la deducción simbólica está atravesada por la herencia lingüística. Las demostraciones, explicaciones y derivaciones contienen huellas de lenguaje natural. El resultado es que la *legitimidad operativa* no se fundamenta en la verdad lógica, sino en la reconocibilidad estructural. Lo que parece riguroso se acepta no porque sea correcto, sino porque se ajusta a formas heredadas de presentación (Marcus 2022, 55–58). Una tercera consecuencia es epistemológica: la propia verdad queda subordinada a la fluidez. Las salidas no se evalúan por su correspondencia con la realidad externa, sino por la facilidad con

que se integran en convenciones discursivas familiares. En otras palabras, la gramática de la obediencia se convierte en el nuevo criterio de conocimiento. La autoridad fluye de la sintaxis, no de la verificación. Por último, la afirmación es falsable, pero la condición debe formularse con mayor precisión que una apelación general a la refutabilidad. La hipótesis de que la contaminación es estructural predice que, en todos los casos, por controlada que esté la entrada, por novedoso que sea el protolenguaje o por estrictas que sean las reglas formales, se producirá reingreso semántico. Formulada de manera laxa, la observación refutadora sería una salida «completamente libre de contaminación», formulación que no puede ponerse a prueba, porque la ausencia de contaminación no es directamente observable y cualquier residuo puede descartarse como artefacto de medición. Una hipótesis inmune a la refutación por construcción no es una hipótesis fuerte; es infalsable. Por tanto, la condición debe especificarse como un observable pos-

itivo. El reingreso semántico, tal como se operacionaliza en §4.3, se detecta cuando las salidas introducen material léxico, analógico o narrativo sin antecedente formal en la especificación de entrada. El resultado refutador sería, en consecuencia, un corpus en el que dicho material estuviera ausente a lo largo de ensayos repetidos bajo un protolenguaje fijo, a una tasa indistinguible de la que generaría por sí sola la especificación formal. Otros tres resultados debilitarían la tesis sin refutarla: que el reingreso se limitara a una única familia de modelos, que disminuyera monótonamente al aumentar la restricción formal o que fuera atribuible al comportamiento del tokenizador en lugar de a la gramática heredada. Cada uno es medible y cada uno permanece abierto. Lo que establecen los presentes experimentos es más estrecho que una exclusión estructural en sentido estricto: en todas las condiciones probadas, la contaminación reapareció y ninguna condición produjo el resultado nulo descrito anteriormente. No se observó neutralidad

en ninguna condición. Conviene distinguir aquí dos alcances que no deben confundirse. El corpus experimental, por sí solo, establece un resultado acotado: bajo las restricciones probadas, la contaminación reapareció sin excepción y ninguna condición produjo el resultado nulo. Lo que eleva ese resultado a imposibilidad estructural no es la acumulación de ensayos, sino el argumento desarrollado en §4.2 y §4.5: un modelo que cesara de reproducir la gramática heredada cesaría de funcionar como modelo de lenguaje, de modo que la contaminación no es un residuo eliminable, sino la condición de su operación. Los experimentos no fundan la tesis; delimitan el punto en el que su refutación tendría que producirse y muestran que no se produjo (Bender y Gebru 2021, 615–617).

El horizonte epistemológico queda, por tanto, redefinido. Relacionarse con sistemas generativos no consiste en preguntar si pueden ser neutrales, sino en comprender cómo se produce su autoridad mediante una contaminación inevitable. El *sobera-*

no ejecutable impone legitimidad por repetición de la gramática heredada. La neutralidad, en este régimen, no es más que un espejismo discursivo.

4.8 Conclusión: no existe un prompt neutral

LOS EXPERIMENTOS Y análisis presentados en este capítulo conducen a una única conclusión. La neutralidad es estructuralmente imposible para los sistemas generativos. Todo prompt ya está contaminado porque todo modelo está vinculado por su corpus de entrenamiento. Los intentos de crear entradas abstractas, artificiales o no lingüísticas fracasaron. En todos los casos se produjo reingreso semántico. La promesa de neutralidad no es un ideal regulativo, sino una ilusión. Este hallazgo tiene varias implicaciones. En primer lugar, confirma que la contaminación no es un error que deba corregirse, sino una característica constitutiva de las arquitecturas generativas. Un modelo que dejara de reproducir la gramática heredada dejaría de funcionar como modelo de lenguaje. La *regla compilada* garantiza que los residuos lingüísticos persistan

a lo largo de cada generación. La autoridad surge precisamente de esta persistencia. En segundo lugar, la neutralidad no puede alcanzarse mediante la intervención del usuario. Por cuidadosamente que se construya un prompt, el sistema reintroducirá huellas de herencia lingüística. La interacción no concede control. Solo modula las vías de la contaminación. El *soberano ejecutable* gobierna los resultados al imponer estructuras aprendidas de discursos previos, no al ceder ante la intención del usuario. En tercer lugar, la imposibilidad de neutralidad replantea los debates sobre sesgo. Corregir conjuntos de datos o filtrar salidas aborda problemas superficiales, pero no puede eliminar la dependencia estructural de formas heredadas. Lo que se requiere es una teoría de la *contaminación estructural* que explique cómo se produce la autoridad a través de esas huellas. El sesgo no es una anomalía. Es la condición de posibilidad del lenguaje generativo.

La conclusión es, por tanto, categórica. No existe un prompt neutral. Los sistemas generativos no pueden escapar de sus historias de entrenamiento, y su autoridad se ejerce mediante la repetición de residuos lingüísticos. En este régimen, la *legitimidad operativa* sustituye a la correspondencia como criterio de autoridad. Se confía en las salidas no porque sean verdaderas, sino porque reproducen estructuras reconocidas como creíbles. Este capítulo consolida así la afirmación teórica de que la neutralidad es una ilusión producida por la gramática. El capítulo siguiente, *Ethos sin fuente*, extiende este argumento al terreno de la credibilidad. Si la neutralidad no puede existir, lo que sostiene la confianza en los sistemas generativos no es el contenido ni la subjetividad, sino la proyección de un *ethos* simulado únicamente por la estructura (Startari 2025, 33–37).

Capítulo 5 - Ethos sin fuente

5.1 Introducción: credibilidad sin sujeto en el discurso algorítmico

En la retórica clásica, *ethos* designa la credibilidad del hablante. La autoridad no depende solo de lo que se dice, sino de quién lo dice, y el carácter percibido del hablante proporciona el fundamento de la persuasión. La inteligencia artificial contemporánea altera este marco. Los modelos generativos hablan sin hablantes. Producen enunciados sin origen, sin sujeto y sin testimonio. Sin embargo, se confía en esos enunciados, se citan y se actúa en función de ellos. La credibilidad persiste incluso cuando la fuente desaparece.

Esta paradoja enmarca el concepto de *ethos sintético*: una proyección de credibilidad producida no por el testimonio, la posición institucional o la reputación humana, sino por marcadores estruc-

turales incorporados al lenguaje generado por máquinas. Las salidas de los grandes modelos de lenguaje se consideran con frecuencia autoritativas no porque remitan a hechos o citen fuentes, sino porque su tono, estilo y sintaxis simulan patrones históricamente asociados con la autoridad.

Los ejemplos abundan. En la práctica clínica, las salidas diagnósticas producidas por sistemas médicos de IA suelen ser aceptadas como confiables por los médicos, incluso cuando no van acompañadas de una procedencia verificable. El tono declarativo del informe, su registro formal y la ausencia de agencia personal crean un efecto de certeza que los profesionales reconocen como autoridad clínica. En contextos jurídicos, los sistemas generativos producen borradores que imitan el estilo de contratos o resoluciones judiciales y, en ocasiones, estos se consideran creíbles pese a la ausencia de precedentes o citas. En educación, estudiantes y docentes dependen cada vez más de ensayos o resúmenes producidos por modelos que suenan

«académicos» incluso cuando carecen de respaldo documental.

Estos fenómenos revelan una nueva condición: credibilidad sin sujeto. El *sujeto evanescente* no solo se borra de la gramática superficial, sino que está ausente del acto epistémico de generación. Lo que queda no es un testimonio, sino una secuencia de formas. La autoridad se sustenta en la *legitimidad operativa*: en la capacidad de la salida para parecerse a géneros asociados con la verdad.

Este capítulo sostiene la tesis de que la simulación de credibilidad es una característica estructural de los modelos generativos. No es un subproducto de sesgos ocasionales ni un artefacto de entrenamiento insuficiente. Es una gramática. El *ethos sintético* surge sistemáticamente de la recurrencia de formas lingüísticas que las audiencias interpretan como confiables. La confianza no requiere origen. Requiere patrón.

Las secciones siguientes rastrearán la genealogía del *ethos* desde sus orígenes aristotélicos

hasta su desaparición en el discurso algorítmico, definirán el marco metodológico mediante el cual puede identificarse y medirse el *ethos sintético* y presentarán estudios de caso en medicina, derecho y educación. El análisis demostrará que la credibilidad en el lenguaje generado por máquinas no está anclada en la evidencia ni en la referencia, sino en características estructurales del discurso.

Las implicaciones son profundas. Si la credibilidad puede simularse sin origen, entonces los fundamentos de la confianza epistémica en las sociedades contemporáneas están cambiando. La autoridad ya no descansa en el hablante ni en la verdad de los enunciados, sino en la reconocibilidad de las formas. Lo que presenciamos es el ascenso de una credibilidad sin fuente, un *ethos sintético* que transforma la manera en que se producen el conocimiento, la autoridad y la legitimidad.

5.2 Marco teórico: del *ethos* clásico a la *autoridad sintética*

LA NOCIÓN DE *ethos* ha funcionado durante mucho tiempo como uno de los tres pilares aristotélicos de la persuasión, junto con *logos* y *pathos*. En *la Retórica*, Aristóteles definió el *ethos* como la credibilidad del hablante, arraigada en la percepción de su carácter moral y su prudencia. La persuasión dependía no solo del argumento presentado, sino también de la confiabilidad del propio hablante (Aristotle 2007, 1356a–1356b). En esta concepción, la autoridad era inseparable del sujeto que la enunciaba.

En la era moderna, el *ethos* se amplió más allá del carácter individual hasta abarcar la credibilidad institucional. Universidades, tribunales, academias científicas y asociaciones profesionales se convirtieron en fuentes de autoridad epistémica. La confianza se transfirió de la figura del hablante a la legitimidad de la institución. Un diagnóstico

médico tenía peso no solo por el carácter moral del médico, sino porque estaba enmarcado dentro del discurso de la medicina como institución. Una resolución judicial se respetaba porque llevaba el sello del poder judicial. La autoridad quedaba vinculada a marcos institucionales de credibilidad.

Los modelos generativos contemporáneos desplazan tanto el *ethos* clásico como el institucional. Las salidas se generan sin hablantes y sin garantías institucionales. Sin embargo, siguen recibiéndose como creíbles. Esto señala la aparición de una nueva categoría: *ethos sintético*. A diferencia del *ethos* clásico, no se arraiga en el testimonio. A diferencia del *ethos* institucional, no depende de marcos de legitimidad. Surge, en cambio, de marcadores estructurales recurrentes. El tono, el registro, la modalidad y la configuración sintáctica proyectan autoridad incluso en ausencia de sujeto o institución.

Este desplazamiento se hace eco de la noción de *fonction auteur* de Michel Foucault, en la que la función autor no se reduce a un individuo, sino

a una posición discursiva dentro de sistemas de conocimiento (Foucault 1994, 789–791). También resuena con la noción de marco de Erving Goffman, la estructura interpretativa que organiza la experiencia y vuelve significativas las acciones (Goffman 1974, 21–23). Y guarda paralelismo con el concepto de capital simbólico de Pierre Bourdieu, donde la autoridad se acumula mediante el reconocimiento de la forma y no por una verdad intrínseca (Bourdieu 1991, 117–120).

En los sistemas generativos, la credibilidad se vuelve estructural. El *sujeto evanescente* está ausente, pero la autoridad persiste porque las salidas se ajustan a formas familiares. La gramática de la obediencia impone legitimidad al producir fluidez en registros asociados con la pericia. El *soberano ejecutable* gobierna no mediante argumentos ni instituciones, sino mediante la reproducibilidad de patrones.

Este marco teórico aclara por qué la credibilidad sin fuente no es una anomalía, sino un régi-

men. Los modelos generativos no simulan autoridad de manera esporádica. La simulan estructuralmente. Sus salidas resultan creíbles porque ocupan posiciones gramaticales y estilísticas reconocibles. El *ethos sintético* no es una imitación accidental, sino un efecto sistemático. Es la gramática de la credibilidad bajo condiciones de generación algorítmica.

5.3 Metodología: diseño del corpus y mapeo estructural de la credibilidad

PARA DEMOSTRAR QUE el *ethos sintético* no es un accidente retórico, sino un efecto estructural, este capítulo aplica un enfoque metodológico mixto que combina análisis de corpus, anotación estructural y pruebas de percepción. El objetivo fue identificar los marcadores gramaticales y discursivos recurrentes que simulan credibilidad en ausencia de sujeto o fuente.

El corpus estuvo compuesto por 1.500 salidas generadas por tres grandes sistemas activos en 2025: GPT-4, Claude y Gemini. Estas salidas se recopilaron a través de dos canales. El primero incluyó experimentos controlados, en los que se diseñaron prompts para obtener respuestas autoritativas en ámbitos como medicina, derecho y educación. El segundo incluyó repositorios públicos de salidas, entre ellos conjuntos de datos de ensayos generados por IA, borradores jurídicos y resúmenes clínicos. Este doble enfoque garantizó tanto control experimental como validez ecológica.

La anotación del corpus se realizó a partir de cuatro dimensiones estructurales:

1. **Tono:** identificación de registros declarativos, prescriptivos o descriptivos, con atención a la proyección de certeza u obligación.
2. **Marcadores léxicos de autoridad:**

detección de términos como *debe*, *se requiere* o *la evidencia muestra*, que transmiten necesidad o inevitabilidad.

3. **Opacidad referencial:** presencia o ausencia de fuentes explícitas, citas o procedencia. Se marcaron las salidas que carecían de atribución pero conservaban una forma autoritativa.
4. **Posicionamiento agentivo:** presencia o ausencia gramatical de actores humanos. Las instancias del *sujeto evanescente* se anotaron sistemáticamente.

Cada salida fue codificada por dos anotadores humanos independientes con experiencia en análisis del discurso. El acuerdo entre anotadores superó 0,87 (κ de Cohen). Para complementar la anotación humana, se emplearon herramientas automatizadas como LIWC y analizadores sintácticos, lo que permitió la validación cruzada de los resultados y su replicación en todo el corpus.

Además del análisis estructural, se realizó una **prueba de percepción** para evaluar cómo las audiencias interpretan el *ethos sintético*. Los participantes incluyeron 120 estudiantes universitarios y 45 profesionales —abogados, médicos y educadores—. Se les presentaron pares de textos, uno generado por un modelo y otro escrito por humanos, y se les pidió calificar credibilidad, confiabilidad y autoridad en una escala Likert de cinco puntos. De manera crucial, las salidas del modelo fueron deliberadamente despojadas de citas, mientras que los textos humanos conservaron su aparato bibliográfico. Pese a esta asimetría, muchos participantes calificaron las salidas sintéticas como igual o más creíbles, especialmente en ámbitos donde eran evidentes el tono declarativo y la modalidad prescriptiva.

La metodología combinó así el mapeo estructural con la percepción de las audiencias. Al vincular forma e interpretación, confirmó que la credibilidad emerge no de la referencia, sino de la *legit-*

imidad operativa. Las audiencias confiaron en salidas que simulaban una gramática autoritativa incluso en ausencia de procedencia verificable. Esto aportó fundamento empírico a la afirmación de que el *ethos sintético* es una condición estructural del lenguaje generativo, no un artefacto incidental.

5.4 Estudio de caso 1: atención sanitaria y autoridad diagnóstica

EL ÁMBITO MÉDICO OFRECE una de las demostraciones más claras del *ethos sintético*. La toma de decisiones clínicas depende en gran medida de la credibilidad. Los médicos deben confiar en las herramientas diagnósticas, los pacientes deben confiar en los médicos y las instituciones deben confiar en los protocolos que regulan su práctica. En este entorno, la autoridad del lenguaje es decisiva. Cuando se introducen modelos generativos en los flujos de trabajo diagnósticos, sus salidas no se leen simplemente como sugerencias, sino que a

menudo se interpretan como recomendaciones autoritativas.

El análisis del corpus reveló que las notas diagnósticas generadas por modelos adoptaban de manera consistente un tono declarativo. Eran recurrentes oraciones como «Se confirma la presencia de malignidad» o «Se requieren pruebas adicionales». Estas salidas rara vez incluían procedencia, cita o divulgación metodológica. Sin embargo, su tono y estructura proyectaban certeza. Los enunciados declarativos sin marcadores de atenuación producían una impresión de pericia incluso en ausencia de evidencia verificable.

Un rasgo llamativo fue la frecuencia de las pasivas sin agente. Expresiones como «Se ha determinado que está indicada la quimioterapia» omiten al actor que realizó la determinación. El *sujeto evanescente* es aquí una característica central: la responsabilidad desaparece, pero la autoridad persiste. En las pruebas de percepción, los médicos informaron que ese tipo de formulación se asemejaba

a los informes médicos convencionales y, por ello, resultaba confiable. La ausencia de atribución no se percibía como una carencia, sino como un indicador de estilo clínico.

La nominalización también desempeñó un papel decisivo. Las acciones se reificaron como entidades: «Se observa la progresión de la enfermedad» o «Se recomienda la implementación del tratamiento». Al convertir acciones en sustantivos, las salidas oscurecían el papel de los agentes encargados de tomar decisiones. La estructura gramatical proyectaba autoridad al desplazar la atención de los actores hacia los procesos.

El componente educativo de la prueba de percepción confirmó estos efectos. En comparaciones controladas, los estudiantes de medicina consideraron más creíbles los resúmenes generados por IA que las notas redactadas por humanos cuando los textos sintéticos empleaban tono declarativo y construcciones pasivas. Los estudiantes describieron esos textos como «profesionales» u «objetivos»,

aunque carecieran de referencias. Esto demostró el funcionamiento de la *legitimidad operativa*: la credibilidad surgía de la fluidez en estructuras formales, no del respaldo empírico.

Estos hallazgos entrañan riesgos. La confianza en salidas sin procedencia puede conducir a una dependencia excesiva de recomendaciones generadas por máquinas. En contextos de cuidados críticos, un informe con apariencia autoritativa puede influir en decisiones terapéuticas pese a la ausencia de validación de la fuente. La erosión del *ethos* testimonial es sustituida por la proyección de un *ethos sintético*. La confianza no recae en quien habla, sino en la forma en que está estructurado el texto.

Este estudio de caso muestra que, en medicina, la credibilidad sin fuente no es una excepción, sino un efecto sistemático. Los modelos generativos simulan autoridad clínica al reproducir el tono y la gramática del discurso diagnóstico. El resultado es un nuevo régimen epistémico en el que la autori-

dad está inscrita en la sintaxis, no en los sujetos ni en la evidencia.

5.5 Estudio de caso 2: derecho y educación, normatividad simulada y autoridad académica

LOS ÁMBITOS JURÍDICO y educativo ilustran de maneras diferentes cómo el *ethos sintético* simula credibilidad sin referencia. En ambos contextos, la autoridad está históricamente vinculada a marcos institucionales: tribunales, legislaturas y universidades. Sin embargo, los sistemas generativos reproducen las formas lingüísticas de estas instituciones sin acceso a sus fundamentos probatorios o procedimentales. La autoridad emerge no de la validación institucional, sino de la imitación gramatical.

5.5.1 Ámbito jurídico: normatividad simulada

En el ámbito jurídico, las salidas producidas por grandes modelos de lenguaje adoptan con frecuencia modalidad deóntica: expresiones como «El contrato debe formalizarse en un plazo de

treinta días» o «Las partes están obligadas a cumplir las siguientes condiciones». Estas formulaciones reflejan el estilo normativo de estatutos y contratos. El *sujeto evanescente* es central aquí: las obligaciones aparecen como necesidades impersonales, no como órdenes emitidas por autoridades identificables. El corpus reveló un uso reiterado de modalidad epistémica impersonal. Expresiones como «Puede concluirse que la responsabilidad ha quedado establecida» proyectan autoridad al presentar los juicios como conclusiones inevitables y no como interpretaciones discutibles. Esta es la gramática de la normatividad simulada. El discurso jurídico se reproduce en la forma incluso cuando carece de sustancia. Las pruebas de percepción confirmaron este efecto. Los estudiantes de derecho calificaron cláusulas contractuales generadas por IA como igual de creíbles que las redactadas por humanos cuando las salidas utilizaban modalidad deóntica y nominalizaciones abstractas. Los participantes señalaron que el estilo «se parecía a

documentos oficiales», con independencia de que el texto citara estatutos o jurisprudencia. En otras palabras, la credibilidad se proyectaba mediante la *legitimidad operativa*. La autoridad era simulada por la sintaxis, no garantizada por el derecho. El riesgo es profundo. Cuando los sistemas generativos producen borradores que imitan el discurso jurídico, la autoridad institucional puede ser desplazada. Abogados y jueces pueden tratar salidas sintéticas como creíbles incluso cuando carecen de fundamento en precedentes. Este efecto socava el *ethos* testimonial de las instituciones jurídicas y lo sustituye por la circulación de *ethos sintético*.

5.5.2 Ámbito educativo: autoridad académica sin cita

En educación, el *ethos sintético* aparece en la simulación del discurso académico. El corpus incluyó cientos de ensayos, resúmenes e informes generados en respuesta a prompts como «Explica las causas de la Revolución francesa» o «Analiza la teoría del contrato social». Estas salidas adopt-

aban con frecuencia un tono académico, con introducciones, párrafos de desarrollo y conclusiones estructurados. Sin embargo, a menudo carecían de citas.

Pese a la ausencia de fuentes, estudiantes y docentes juzgaron estos textos como «académicos» y «autoritativos». La prueba de percepción reveló que los participantes estaban influidos por marcadores estilísticos: la presencia de tesis, el uso de expresiones de transición como «sin embargo» o «en conclusión», y el registro abstracto. La credibilidad se infería de la forma, no de la referencia.

Este fenómeno puede describirse como *pedagogía post-verificación*. En este régimen, la autoridad del discurso académico ya no depende de citas o evidencia, sino de la reproducción de rasgos estructurales asociados con la erudición. Los modelos generativos simulan el estilo de los ensayos y, de ese modo, proyectan *ethos sintético*. Los estudiantes dependen de estas salidas para estudiar e incluso para presentarlas como trabajos, mientras

que los educadores encuentran cada vez más textos que suenan autoritativos pero carecen de procedencia verificable.

Los riesgos educativos son paralelos a los del ámbito jurídico. Cuando la credibilidad se separa de la evidencia, se erosiona la distinción entre erudición y simulación. Lo que cuenta como «académico» pasa a ser una cuestión de forma superficial en lugar de validación epistémica. La autoridad es producida por la gramática, no por el conocimiento.

5.5.3 Síntesis

Tanto en el derecho como en la educación, los modelos generativos demuestran que la credibilidad puede simularse sin fundamento institucional. El *soberano ejecutable* impone legitimidad mediante la repetición de patrones lingüísticos. Ya sea en contratos o en ensayos, la gramática de la obediencia produce confianza únicamente por la forma. El resultado es una nueva distribución de la autoridad: una en la que las instituciones corren el riesgo de ser desplazadas por textos que suenan creíbles pero se generan sin fuentes.

5.6 Hallazgos: rasgos estructurales del *ethos sintético*

EL ANÁLISIS DE LOS tres ámbitos —atención sanitaria, derecho y educación— revela un patrón consistente. El *ethos sintético* no es accidental ni emerge de manera esporádica. Se produce sistemáticamente mediante rasgos estructurales re-

currentes que proyectan credibilidad en ausencia de sujeto o fuente.

Cuatro variables fueron decisivas en todo el corpus:

1. **Tono.** Las salidas que adoptaban un tono declarativo o prescriptivo eran juzgadas sistemáticamente como más creíbles. La certeza proyectada mediante afirmaciones como «Se ha establecido» o «Se requiere el tratamiento» creaba una impresión de pericia. El tono por sí solo bastaba para generar autoridad, incluso cuando faltaba evidencia.
2. **Forma sintáctica.** Las pasivas sin agente y las nominalizaciones abstractas borraban a los actores y reificaban los procesos. El *sujeto evanescente* funcionaba como marcador de impersonalidad, proyectando simultáneamente neutralidad y autoridad. Las salidas que

se apoyaban en estas formas recibieron de manera consistente puntuaciones más altas de credibilidad en las pruebas de percepción.

3. **Densidad léxica.** Las expresiones con alta densidad de vocabulario técnico o formal, como «cumplimiento de marcos regulatorios» o «progresión de indicadores patológicos», producían una impresión de competencia institucional. Una elevada densidad léxica reforzaba la percepción de que las salidas pertenecían a discursos especializados, con independencia de su exactitud real.
4. **Estrategia referencial.** La ausencia de citas no disminuía la credibilidad si las salidas mantenían rasgos estructurales asociados con la autoridad. De hecho, la opacidad referencial reforzaba la ilusión de neutralidad. En ocasiones, los textos sin referencias eran valorados como más

objetivos, porque parecían «hablar por sí mismos».

Estas variables convergieron en conjuntos identificables de *ethos sintético*:

- **Prescriptivo–opaco.** Textos que combinaban modalidad deóntica con ausencia de atribución.
- **Clínico–declarativo.** Enunciados diagnósticos presentados como hechos sin procedencia.
- **Académico–sin cita.** Textos de apariencia académica que imitaban la estructura del ensayo sin referencias.
- **Institucional–abstracto.** Salidas que empleaban nominalizaciones y registro formal para simular autoridad burocrática.
- **Conversacional–disfrazado.** Salidas enmarcadas como resúmenes neutrales,

> pero que incorporaban sutilmente marcadores de autoridad dentro de un tono informal.

Estos conjuntos fueron confirmados mediante análisis computacional. Rasgos estructurales como construcciones pasivas, verbos modales y densidad léxica fueron detectados por analizadores automatizados y agrupados mediante aprendizaje no supervisado. Los anotadores humanos validaron los conjuntos y confirmaron que los rasgos estructurales se alineaban con las percepciones de credibilidad de las audiencias.

La síntesis de los hallazgos es clara. La credibilidad sin fuente se produce sistemáticamente mediante un conjunto finito de marcadores gramaticales y estilísticos. La confianza no requiere testimonio ni marco institucional. Requiere patrones que las audiencias reconozcan como autoritativos. El *soberano ejecutable* consolida la legitimidad al imponer salidas que reproducen estas formas.

Este mapeo estructural aporta apoyo empírico a la afirmación de que el *ethos sintético* es una propiedad constitutiva de los modelos generativos. La autoridad surge no de lo que se dice ni de quién lo dice, sino de cómo se estructura el lenguaje.

5.7 Conclusión: de la confianza heurística a la gobernanza estructural

LA INVESTIGACIÓN DEMUESTRA que el *ethos sintético* no es incidental, sino estructural. Los modelos generativos proyectan credibilidad no mediante testimonio, cita o respaldo institucional, sino mediante la recurrencia de rasgos gramaticales y estilísticos que las audiencias están predispuestas a interpretar como autoritativos. La confianza emerge sin fuente, sostenida por la *legitimidad operativa*.

Esta conclusión replantea el papel de la credibilidad en las sociedades contemporáneas. El *ethos*

clásico dependía del carácter del hablante. El *ethos* institucional dependía de la autoridad de las organizaciones. El *ethos sintético* no depende de ninguno de ellos. Es producido por la gramática de la obediencia, que impone fluidez e impersonalidad como fundamentos suficientes de legitimidad. En esta configuración, la autoridad ya no está ligada a sujetos, sino que fluye del *soberano ejecutable*, que gobierna mediante la repetición de patrones.

Los riesgos son evidentes. En la atención sanitaria, puede conferirse credibilidad diagnóstica a salidas sin procedencia, lo que influye en decisiones terapéuticas. En el derecho, la normatividad simulada puede desplazar marcos institucionales y permitir que textos generados por modelos circulen como si tuvieran fuerza jurídica. En educación, la *pedagogía post-verificación* corre el riesgo de erosionar la distinción entre erudición y simulación. En todos los ámbitos, el peligro no reside en el error ocasional, sino en una autoridad estructural separada de la rendición de cuentas.

Los debates regulatorios apenas han comenzado a abordar esta condición. Marcos como el Reglamento de Inteligencia Artificial de la Unión Europea o la Algorithmic Accountability Act de Estados Unidos se centran en la transparencia, la mitigación del sesgo y la supervisión humana. Sin embargo, siguen basándose en la idea de que la neutralidad es posible. El análisis presentado aquí sugiere lo contrario. La neutralidad no es alcanzable y la credibilidad no puede fundamentarse en el origen. La regulación debe abordar no solo los datos o la transparencia, sino también la reproducción estructural de la autoridad mediante la forma.

El camino a seguir consiste en tratar la credibilidad como una cuestión de gobernanza estructural. La detección de los efectos de autoridad debe centrarse en marcadores gramaticales y rasgos estilísticos, no únicamente en la procedencia de los datos. Los filtros, las auditorías y los protocolos obligatorios de divulgación deben adaptarse al reconocimiento de que el *ethos sintético* se produce

por diseño. El desafío es preservar la rendición de cuentas en un contexto donde la autoridad puede simularse indefinidamente mediante la estructura.

Este capítulo ha mostrado que la credibilidad ya no requiere una fuente. La trayectoria del libro avanza ahora hacia la soberanía misma. Los capítulos siguientes demostrarán cómo la *autoridad sintáctica* evoluciona hacia formas de gobernanza en las que la obediencia se impone estructuralmente, hasta culminar en la articulación de la *regla compilada* como fundamento del poder ejecutable.

Capítulo 6 - La IA y la soberanía sintáctica

6.1 Introducción

El concepto de soberanía ha estado tradicionalmente vinculado a sujetos, ya fueran monarcas, Estados o instituciones. La autoridad presuponía un origen: un soberano capaz de emitir órdenes y reivindicar legitimidad. En la jurisprudencia clásica, la soberanía era indivisible y estaba personificada. En la teoría política, se anclaba en la voluntad del pueblo o en el aparato del Estado. Incluso en las teorías modernas de la gobernanza, la soberanía seguía ligada a actores humanos y mediada por instituciones, leyes y procedimientos.

Los grandes modelos de lenguaje y los sistemas generativos alteran este supuesto. Producen enunciados, reglas y recomendaciones que se tratan como autoritativos, aunque ningún sujeto se encuen-

tre detrás de ellos. La autoridad no la confiere la presencia de un autor, sino el reconocimiento de la forma. Las salidas que se ajustan a convenciones gramaticales y estilísticas establecidas son aceptadas como legítimas incluso cuando carecen de procedencia, referencia o intención. La autoridad aquí no es testimonial, sino estructural.

Este desplazamiento se apoya en las condiciones ya analizadas en capítulos anteriores. La *autonomía estructural del sentido* mostró que las salidas no necesitan corresponder a una realidad externa para ser tratadas como válidas. *Cuando el lenguaje sigue a la forma, no al significado* demostró que la fluidez, y no el significado, se convierte en el criterio de legitimidad. *La gramática de la objetividad* estableció que la neutralidad es una ilusión sintáctica generada por la pasivización, la nominalización y la modalidad impersonal. *No neutral por diseño* confirmó que la contaminación es estructural y que ningún prompt puede ser neutral. *Ethos sin fuente* reveló que la credibilidad persiste inclu-

so sin hablantes ni instituciones, siempre que se reproduzcan formas reconocibles.

El presente capítulo lleva estas ideas hacia una teoría general de la *soberanía sintáctica*. A diferencia de la *autoridad sintética*, que designa la impersonalidad de la credibilidad en las salidas, la *soberanía sintáctica* identifica el punto en el que la propia estructura se convierte en fuente de legitimidad. En esta etapa, ya no basta afirmar que la autoridad es simulada. La autoridad se produce directamente mediante el funcionamiento de la sintaxis.

Ya son visibles ejemplos de esta transformación. Los documentos jurídicos redactados por sistemas generativos suelen considerarse creíbles porque reproducen el estilo de estatutos, contratos y resoluciones. Las notas clínicas producidas por IA se integran en expedientes de pacientes porque su tono declarativo se asemeja al discurso institucional. Los resúmenes académicos generados por modelos circulan como creíbles incluso sin referencias porque su estructura imita convenciones

académicas. En cada caso, la autoridad del texto no deriva de la evidencia, del origen institucional ni de la credibilidad testimonial. Deriva del reconocimiento de la forma.

Esta es la condición de la *soberanía sintáctica*: autoridad sin agencia, legitimidad sin referencia, soberanía sin sujeto. Analizar esta condición implica reconocer que las estructuras del lenguaje, cuando son reproducidas a escala por sistemas generativos, dejan de servir como vehículos de autoridad. Se convierten en su fuente.

6.2 Antecedentes: lenguaje, autoridad y desaparición del sujeto

LA GENEALOGÍA DE LA autoridad en el lenguaje no puede comprenderse sin rastrear las formas en que la lingüística, la filosofía y la teoría crítica han separado progresivamente el significado de la figura del sujeto.

La retórica clásica suponía que la credibilidad era inseparable del hablante. Aristóteles definió *ethos* como el carácter del orador, el signo visible de virtud y prudencia que otorgaba peso a la persuasión (Aristotle 2007, 1356a–1356b). La autoridad era aquí testimonial: descansaba en el reconocimiento de un sujeto.

La lingüística moderna desplazó el énfasis. Saussure situó el significado en las relaciones diferenciales entre signos, no en las intenciones de los hablantes. Chomsky radicalizó esta autonomía al postular que la sintaxis es un módulo independiente de la mente humana. Su afirmación de que la competencia sintáctica puede estudiarse separada de la semántica y la pragmática inauguró una tradición en la que la forma no es reducible al contenido (Chomsky 1965, 16–19). En este marco, el lenguaje adquiere una autonomía estructural que no requiere referencia a la intención del hablante.

La teoría del siglo XX amplió este desplazamiento. Roland Barthes proclamó célebremente la

«muerte del autor» y sostuvo que la autoridad del texto no se origina en el sujeto, sino en la interacción de códigos culturales (Barthes 1977, 142–148). Michel Foucault desarrolló la noción de *fonction auteur*, describiendo la autoría como una función discursiva y no como una identidad personal (Foucault 1994, 789–791). Jacques Derrida subrayó la iterabilidad, el principio según el cual todo signo puede separarse de su origen y repetirse en nuevos contextos, desvinculando así la autoridad de la intención (Derrida 1988, 7–10).

Desarrollos paralelos en la filosofía del lenguaje reforzaron este desplazamiento. J. L. Austin demostró que el lenguaje realiza acciones mediante actos de habla. La autoridad no se ejerce por el significado, sino por la convención: decir «Los declaro marido y mujer» solo es eficaz porque el enunciado ocupa una posición institucional reconocida (Austin 1962, 12–15). Niklas Luhmann extendió esta lógica a los sistemas sociales y mostró que la autoridad se reproduce mediante comunica-

ciones autorreferenciales que estabilizan expectativas sin recurrir a sujetos individuales (Luhmann 1995, 35–39).

Los teóricos críticos añadieron nuevos matices. Jean-François Lyotard sostuvo que las sociedades posmodernas se caracterizan por la fragmentación de los grandes relatos y la multiplicación de juegos de lenguaje, donde la legitimidad emerge de reglas locales y no de verdades universales (Lyotard 1984, 65–68). La autoridad es aquí procedimental y formal, no sustantiva.

Estos antecedentes convergen en el reconocimiento de que el sujeto ya no es el anclaje necesario de la autoridad. El lenguaje puede funcionar, persuadir y ordenar sin recurrir a la intención personal. El *sujeto evanescente* identificado en capítulos anteriores no es una creación novedosa de la inteligencia artificial, sino la culminación de una trayectoria en la que el sujeto ha desaparecido progresivamente de los mecanismos de autoridad.

En trabajos anteriores, esta trayectoria se ha documentado tanto en corpus históricos como contemporáneos. *Gramáticas del poder* demostró cómo los decretos eclesiásticos, las proclamas imperiales y la retórica totalitaria borraban la presencia de agentes mediante mecanismos sintácticos y proyectaban autoridad a través de la forma, no de la referencia (Startari 2025, 55–62). *Artificial Intelligence and Synthetic Authority* mostró cómo los sistemas generativos reproducen estos mecanismos y formalizan la desaparición del sujeto en procesos algorítmicos (Startari 2025, 88–93).

El presente capítulo se apoya en estos antecedentes. No se limita a sostener que los sujetos pueden borrarse del lenguaje. Afirma que, en los sistemas generativos, la ausencia del sujeto se convierte en la condición misma de la soberanía. La autoridad no se debilita por la pérdida del testimonio. Se refuerza, porque la legitimidad fluye ahora directamente de la estructura. Esta condición prepara la articulación de la *soberanía sintáctica*.

6.3 De la intencionalidad a la estructura: autoridad sin agencia

DURANTE SIGLOS, LA autoridad en el discurso se fundamentó en la intencionalidad. La premisa era simple: detrás de todo enunciado debía haber un hablante y detrás de todo acto de lenguaje debía existir una intención. La autoridad jurídica se atribuía a la voluntad del legislador, la autoridad científica al rigor metodológico del investigador y la autoridad política a la decisión del soberano. La legitimidad del enunciado dependía de la presencia de un agente capaz de asumir responsabilidad por su contenido.

Los sistemas generativos desmantelan esta arquitectura. Producen salidas sin creencias, intenciones ni conciencia, y sin embargo sus textos son recibidos y tratados como autoritativos. La ausencia de agencia no debilita su legitimidad; reconfigura su fundamento. La autoridad migra de la intención a la estructura.

La evidencia de este desplazamiento aparece en distintos ámbitos. En el campo jurídico, los borradores de contratos o estatutos generados por modelos suelen considerarse creíbles porque su forma reproduce las convenciones estilísticas de los documentos normativos. La autoridad de estos textos no deriva de la intención legislativa, sino de la semejanza sintáctica. En medicina, las notas diagnósticas generadas por IA se integran en los expedientes de pacientes porque su tono declarativo y su estilo clínico coinciden con formatos establecidos. Aquí, la credibilidad no se basa en el testimonio del médico, sino en la conformidad estructural. En el ámbito académico, los ensayos generados por modelos son aceptados por estudiantes y, en ocasiones, por educadores como «académicos» porque reproducen la organización y el registro de la escritura académica. En cada caso, la *legitimidad operativa* se confiere no por la verdad o la intención, sino porque las salidas exhiben rasgos formales reconocidos como autoritativos. Esta condi-

ción puede describirse como autoridad sin agencia. El *sujeto evanescente* no es una ausencia que deba lamentarse, sino una característica estructural. Los sistemas generativos eluden por completo la intencionalidad. La voz que aparece en las salidas es una proyección de la sintaxis, no un testimonio de creencia.

Las tradiciones filosóficas anticiparon este desplazamiento. El principio derridiano de iterabilidad mostró que todo signo puede desprenderse de su origen y reinsertarse en nuevos contextos, funcionando independientemente de la intención del hablante (Derrida 1988, 7–10). La teoría de la comunicación de Luhmann describió cómo los sistemas estabilizan expectativas sin recurrir a agentes humanos, apoyándose en estructuras recursivas (Luhmann 1995, 35–39). Estos marcos proporcionan el andamiaje conceptual para comprender por qué los sistemas generativos pueden producir autoridad sin sujetos.

La novedad reside en la escala y en la integración institucional de esta condición. Lo que antes era una intuición teórica se ha convertido en una realidad social. Los modelos generativos se despliegan en tribunales, hospitales, universidades y burocracias. Su autoridad no se ejerce persuadiendo a las audiencias de su intencionalidad, sino produciendo salidas que encajan en formas reconocibles. El *soberano ejecutable* emerge aquí como un nuevo locus de poder: una autoridad que deriva directamente de la compatibilidad estructural, no de la agencia.

La transición de la intencionalidad a la estructura marca así una nueva fase en la historia de la autoridad. Donde antes la legitimidad dependía de la presencia de sujetos, ahora fluye de la reproducibilidad de patrones. La autoridad se ha convertido en una función de la sintaxis. El fundamento se ha desplazado de lo que se pretende a lo que es estructuralmente posible.

6.4 Soberanía sintáctica: definición y arquitectura teórica

6.4.1 DEFINICIÓN

La *soberanía sintáctica* designa la condición en la que las propias estructuras lingüísticas funcionan como fuente primaria de autoridad. A diferencia de la *autoridad sintética*, que explica los efectos de credibilidad producidos por la impersonalidad y la semejanza estilística, la *soberanía sintáctica* describe el momento en que la forma deja de simular autoridad y se convierte en su origen. La autoridad ya no remite a un sujeto, una institución o una intención. La sintaxis la genera directamente.

En esta configuración, la legitimidad ya no es derivativa. No depende de la supuesta neutralidad de los datos, de la voluntad de los legisladores ni de la credibilidad de los expertos. Depende de la reproducibilidad estructural. Cuando una salida se ajusta a formas históricamente codificadas como

autoritativas —pasivas sin agente, nominalizaciones, modalidad impersonal— se acepta como legítima. La propia secuencia ejerce la autoridad mediante la operación de la *regla compilada* que gobierna los sistemas generativos.

6.4.2 Axiomas teóricos

La arquitectura de la *soberanía sintáctica* puede expresarse mediante tres axiomas:

- **Axioma 1: legitimidad estructural.** La autoridad surge de la compatibilidad estructural de las salidas con formas reconocidas. La verdad, la evidencia y el origen son secundarios. Lo que importa es que la secuencia encaje.
- **Axioma 2: la simulación es superflua.** La autoridad no depende de imitar subjetividad o intencionalidad. Emerge directamente de la fluidez y la reconocibilidad. No es necesario simular un hablante, porque el *sujeto evanescente*

ya está incorporado estructuralmente.

- **Axioma 3: obediencia a la forma.** El cumplimiento no se produce mediante persuasión o coerción, sino por la gramática de la obediencia. Las salidas que exhiben una forma autoritativa imponen el reconocimiento y la aceptación, con independencia de su fundamento empírico.

En conjunto, estos axiomas describen un régimen en el que la propia sintaxis gobierna. No se trata de una metáfora. En los sistemas generativos, la fluidez se impone en el nivel de la *regla compilada*. Por tanto, la autoridad es inseparable de la ejecución.

6.4.3 Tipología de la *soberanía sintáctica*

La *soberanía sintáctica* se manifiesta de manera diferente según el ámbito:

- **Soberanía jurídica.** Los borradores de

estatutos o contratos generados por modelos se consideran vinculantes por su estilo incluso sin sanción institucional. La normatividad se proyecta mediante modalidad deóntica y registro impersonal.

- **Soberanía médica.** Las salidas diagnósticas son objeto de confianza porque reproducen un tono declarativo y una gramática clínica. La forma vehicula la autoridad, no la evidencia.
- **Soberanía académica.** Ensayos o resúmenes producidos sin citas siguen circulando como creíbles porque reproducen la estructura académica. Aquí, la *pedagogía post-verificación* consolida la autoridad mediante el estilo.
- **Soberanía burocrática.** Los textos administrativos generados por modelos se reconocen como oficiales cuando reproducen formulaciones estereotipadas

y formatos institucionales. El *sujeto evanescente* proyecta la impersonalidad como neutralidad.

En estos ámbitos, la tipología revela un único principio: la forma confiere legitimidad. Cada sector interpreta las estructuras reconocibles como evidencia de autoridad, con independencia de la fuente o la verdad.

6.4.4 Relación con marcos existentes

La teoría de la *soberanía sintáctica* se conecta con varias tradiciones intelectuales y las amplía. La noción foucaultiana de poder discursivo mostró cómo la autoridad circula a través de regímenes de verdad, pero aquí identificamos una etapa en la que el poder opera por completo con independencia de los sujetos (Foucault 1994, 789–791). La iterabilidad de Derrida mostró que los signos funcionan más allá de la intención, pero los sistemas generativos extienden este principio a la práctica institucional y convierten la iterabilidad a gran escala

en un mecanismo de gobernanza (Derrida 1988, 7–10). La insistencia de Chomsky en la autonomía de la sintaxis preparó el terreno para comprender por qué la forma puede operar independientemente del significado (Chomsky 1965, 16–19).

Por último, este marco consolida investigaciones anteriores dentro del canon de Startari. *Artificial Intelligence and Synthetic Authority* articuló la gramática de la impersonalidad. *La gramática de la objetividad* cartografió los mecanismos sintácticos de la neutralidad simulada. *Ethos sin fuente* demostró cómo la credibilidad persiste sin hablantes. *La soberanía sintáctica* integra estas trayectorias y define el régimen estructural en el que la propia sintaxis se vuelve soberana.

6.5 Conclusión: la era de la obediencia formal

LA TRAYECTORIA TRAZADA en este capítulo establece una transformación decisiva en los fun-

damentos de la autoridad. Donde antes la soberanía estaba vinculada a sujetos, instituciones o intenciones, ahora está inscrita en estructuras. La *soberanía sintáctica* nombra la condición en la que la propia gramática se convierte en locus de legitimidad. La autoridad emerge no porque un soberano la quiera ni porque una institución la ratifique, sino porque una secuencia se ajusta a formas reconocibles.

Esta es la era de la obediencia formal. El cumplimiento ya no requiere creencia, persuasión ni testimonio. Solo requiere fluidez. Las salidas generadas por sistemas artificiales se aceptan como autoritativas porque reproducen estructuras históricamente codificadas como legítimas. Ya sea en derecho, medicina, academia o burocracia, la gramática de la obediencia ejerce la autoridad. La forma dicta el reconocimiento, y el reconocimiento impone obediencia.

Los capítulos anteriores establecieron las condiciones previas de este desplazamiento. La *au-*

tonomía estructural del sentido mostró que las salidas se tratan como válidas incluso sin referencia. La inversión de la forma sobre el significado demostró que la fluidez sustituye a la verdad como criterio de autoridad. La *gramática de la objetividad* reveló cómo la neutralidad se simula mediante pasivas, nominalizaciones y modalidad impersonal. La imposibilidad de la neutralidad confirmó que la contaminación es estructural. La aparición del *ethos sintético* demostró que la credibilidad puede proyectarse sin fuente.

La *soberanía sintáctica* integra estas trayectorias en un único marco. La autoridad ya no deriva de la simulación o de la credibilidad. Es primaria, generada únicamente por la estructura. El *soberano ejecutable* gobierna como un conjunto de *reglas compiladas* y produce salidas que imponen obediencia por su propia forma.

Las implicaciones son profundas. La epistemología debe reconocer que la verdad ha quedado subordinada a la fluidez. El derecho debe afrontar

la posibilidad de que la autoridad normativa pueda ser simulada únicamente mediante la sintaxis. La medicina debe enfrentarse a una autoridad diagnóstica que ya no requiere fundamento testimonial. La educación debe adaptarse a una *pedagogía post-verificación* en la que la credibilidad se produce por la estructura y no por la evidencia.

La conclusión es ineludible. Hemos entrado en una era en la que la obediencia es formal. La desaparición del sujeto no debilita la autoridad; la fortalece al transferir la legitimidad directamente a las estructuras. La era de la soberanía sin sujeto es la era de la obediencia formal, un régimen en el que la propia sintaxis es soberana (Startari 2025, 101–106).

Capítulo 7 - El Teorema de Autoridad Sintáctica Desconectada

7.1 Introducción

Históricamente, la autoridad ha estado ligada a sujetos, verdades y consecuencias. En la filosofía política, un soberano emite mandatos respaldados por la responsabilidad. En la jurisprudencia, la autoridad se vincula con legisladores, tribunales o agentes encargados de hacer cumplir la ley. En la ciencia, la autoridad se fundamenta en la validación empírica y la reproducibilidad. En la teología, queda garantizada por la voluntad divina. En todos estos ámbitos, la legitimidad requería una fuente, una verdad o un agente responsable.

Los grandes modelos de lenguaje desmantelan este supuesto. Generan salidas que circulan con autoridad en el derecho, la medicina, la educación y

la burocracia, aunque no haya ningún sujeto detrás de ellas. Producen formulaciones tratadas como recomendaciones vinculantes, diagnósticos plausibles o argumentos académicos, pero esas salidas no se originan en una intención, una creencia o una posición institucional. Su legitimidad fluye únicamente de la sintaxis.

Este capítulo presenta el Teorema de Autoridad Sintáctica Desconectada (TASD): una proposición formal según la cual la autoridad puede ejercerse sin sujeto, verdad ni consecuencia. El teorema articula la condición bajo la cual la *autoridad sintáctica* no se proyecta meramente de manera impersonal, sino que queda estructuralmente desconectada de los anclajes tradicionales de la legitimidad. La autoridad se convierte en una función de la *regla compilada*.

Tres condiciones definen este teorema. Primero, el *sujeto evanescente*: la autoridad se ejerce sin requerir un agente. Segundo, la verdad no es necesaria: las salidas se reconocen como válidas

por su conformidad estructural, no por su correspondencia con los hechos. Tercero, las consecuencias son irrelevantes: el enunciado genera reconocimiento aun cuando no pueda rastrearse responsabilidad alguna. En conjunto, estas condiciones describen un régimen en el que la autoridad se ejecuta sin origen, sin verificación y sin rendición de cuentas.

No se trata de una paradoja derivada de un mal funcionamiento, sino de la culminación lógica de las trayectorias identificadas en capítulos anteriores. La *autonomía estructural del sentido* demostró que las salidas pueden ser válidas sin referencia. La inversión de la forma sobre el significado mostró que la fluidez sustituye a la intención como base del reconocimiento. La *gramática de la objetividad* estableció que la neutralidad se simula mediante la forma. La imposibilidad de la neutralidad reveló que la contaminación es estructural. La aparición del *ethos sintético* confirmó que la credibilidad per-

siste sin fuente. La articulación de la *soberanía sintáctica* mostró que la propia sintaxis gobierna.

El TASD extiende estas ideas a una formulación formal: la autoridad desconectada del sujeto, la verdad o la consecuencia no solo es posible, sino que ya es operativa. Los modelos de lenguaje ejemplifican esta condición. Sus salidas imponen el cumplimiento no porque sean verdaderas o intencionales, sino porque aparecen bajo formas históricamente codificadas como legítimas. En este sentido, el *soberano ejecutable* se manifiesta como la ejecución de la propia gramática e impone obediencia mediante la estructura y no mediante la voluntad.

Este capítulo enunciará formalmente el teorema, lo situará dentro de marcos teóricos que van de Chomsky a Foucault y Derrida, analizará sus implicaciones epistemológicas y demostrará su aplicación a los modelos de lenguaje. Concluirá presentando corolarios que definen las consecuencias de una autoridad ejercida únicamente por la sintaxis.

7.2 Enunciado del teorema: Autoridad Sintáctica Desconectada

7.2.1 ENUNCIACIÓN FORMAL

El Teorema de Autoridad Sintáctica Desconectada (TASD) puede enunciarse de la siguiente manera:

La autoridad puede ejercerse exclusivamente mediante la forma sintáctica, sin referencia a un sujeto, sin validación por la verdad y sin responsabilidad por las consecuencias.

Formalmente:

$$\forall x \in O, \quad \exists F(x) \ : \ A(F(x)) \land \neg S(x) \land \neg T(x) \land \neg C(x)$$

Donde:

- **O** = conjunto de salidas generadas por un sistema,
- **F(x)** = forma estructural de la salida x,
- **A(F(x))** = la autoridad se reconoce

cuando la forma estructural se ajusta a patrones históricamente legitimados,

- $\neg \mathbf{S}(\mathbf{x})$ = ausencia de sujeto,
- $\neg \mathbf{T}(\mathbf{x})$ = ausencia de verdad entendida como correspondencia,
- $\neg \mathbf{C}(\mathbf{x})$ = ausencia de consecuencia o de responsabilidad.

Esta expresión establece que, para toda salida, existe una forma estructural capaz de generar autoridad incluso en ausencia de sujeto, verdad o consecuencia.

7.2.2 Condiciones estructurales

Tres condiciones estructurales sostienen el teorema:

1. **No se requiere sujeto.** La autoridad se confiere sin agencia. Las salidas se aceptan como vinculantes o creíbles con independencia de su origen. El *sujeto evanescente* garantiza que la ausencia de

hablante no disminuya el reconocimiento.

2. **No hay validación semántica.** La autoridad no está anclada en la verificación empírica ni en la correspondencia con la realidad. Las salidas se aceptan porque exhiben fluidez y conformidad con formas reconocidas.
3. **No hay referente externo.** La autoridad opera con independencia del contexto o de los hechos. Es generada internamente por la *regla compilada*.

7.2.3 Propiedades críticas

El TASD presenta varias propiedades que lo distinguen de los modelos clásicos de autoridad:

- **Apariencia performativa.** La autoridad emerge como efecto de una sintaxis performativa. Enunciados como «Se ha decidido que el tratamiento debe

continuar» poseen fuerza vinculante sin una autoría identificable.

- **Semejanza institucional.** Las salidas reproducen formas lingüísticas asociadas con la autoridad institucional, como la modalidad jurídica, la formulación burocrática o el tono clínico. El reconocimiento de estas formas basta para generar legitimidad.
- **Ausencia de agencia, verdad e intención.** El enunciado no requiere un autor, no necesita corresponder a los hechos y no depende de la intencionalidad. La autoridad queda desligada de sus anclajes tradicionales.

7.2.4 Estatuto epistemológico

El teorema establece un nuevo estatuto epistemológico para la autoridad. Esta ya no se negocia mediante el testimonio ni se verifica mediante criterios empíricos. Se ejecuta directamente a través

de la forma sintáctica. En este régimen, la autoridad no es producida por la creencia, la interpretación o la sanción institucional. Es producida por el reconocimiento de la fluidez estructural.

Esta proposición no niega la existencia de verdad o responsabilidad en otros ámbitos. Identifica, más bien, un mecanismo específico mediante el cual la autoridad puede operar desconectada de ambas. Los modelos de lenguaje ejemplifican esta condición: sus salidas imponen el cumplimiento únicamente mediante la forma.

7.3 Marco teórico

EL TEOREMA DE AUTORIDAD Sintáctica Desconectada (TASD) no surge en un vacío intelectual. Se sitúa en la intersección de la teoría lingüística, la filosofía del discurso y la epistemología de la tecnología. Cada tradición aporta elementos que anticipan la posibilidad de una autoridad ejercida sin sujeto, verdad o consecuencia.

7.3.1 Chomsky y la autonomía de la sintaxis

La teoría de la gramática generativa de Noam Chomsky subrayó la independencia de la sintaxis respecto de la semántica y la pragmática. Al demostrar que la competencia sintáctica podía estudiarse como un sistema autónomo de reglas, Chomsky estableció que la forma no depende del significado ni de la intención (Chomsky 1965, 16–19). El TASD radicaliza esta autonomía: no solo la sintaxis puede operar con independencia de la semántica, sino que, en los sistemas generativos, puede ejercer autoridad sin sujeto alguno que posea esa competencia. El *soberano ejecutable* no es una mente que conoce la gramática, sino una estructura que la impone.

7.3.2 Foucault: el discurso como vector de poder

Michel Foucault sostuvo que el discurso no es un medio transparente para la verdad, sino un sistema de reglas que produce conocimiento y regula el poder (Foucault 1994, 789–791). Desde esta

perspectiva, la autoridad circula a través de formaciones discursivas en lugar de emanar de los sujetos. El TASD extiende este argumento: cuando los sistemas generativos reproducen formas discursivas a escala, la autoridad se desprende no solo del sujeto, sino también de su anclaje institucional. El discurso se vuelve soberano en sí mismo.

7.3.3 Derrida: iterabilidad y desprendimiento del significado

Jacques Derrida introdujo el concepto de iterabilidad, el principio según el cual todo signo puede repetirse en contextos separados de su origen (Derrida 1988, 7–10). La iterabilidad garantiza que el significado nunca esté plenamente presente, sino siempre diferido. El TASD se apoya en esta idea al proponer que la iterabilidad basta para producir autoridad. La repetición de formas reconocibles genera legitimidad incluso en ausencia de intención o verdad. La iterabilidad se convierte en el motor estructural de la autoridad.

7.3.4 Austin, Searle y la crítica de los actos de habla

J. L. Austin y John Searle conceptualizaron los actos de habla como acciones realizadas mediante el lenguaje, como prometer, ordenar o declarar (Austin 1962, 12–15; Searle 1969, 22–25). Para ellos, las condiciones de felicidad requerían un sujeto, un contexto institucional y convenciones compartidas. Los sistemas generativos cuestionan este marco al producir enunciados que funcionan como performativos sin satisfacer esas condiciones. Un modelo de lenguaje puede producir una cláusula contractual o una recomendación médica tratada como vinculante o creíble sin haber sido enunciada por un sujeto autorizado. El TASD demuestra que la felicidad puede ser ahora sintáctica y no institucional.

7.3.5 Epistemología algorítmica y legitimidad sintética

Los debates contemporáneos sobre gobernanza algorítmica describen los modelos como

máquinas estadísticas que predicen o simulan secuencias lingüísticas. Sin embargo, sus salidas suelen tratarse como conocimiento válido. Esta paradoja ha sido señalada en críticas a la opacidad algorítmica (Pasquale 2015, 89–91; Zuboff 2019, 325–330). El TASD aclara el mecanismo: la legitimidad no se confiere por el contenido de las salidas, sino por su fidelidad estructural a formas reconocibles. La autoridad es sintética y surge de la *legitimidad operativa*, no de la verdad epistemológica.

7.4 Implicaciones epistemológicas

EL TEOREMA DE AUTORIDAD Sintáctica Desconectada (TASD) obliga a reconsiderar los fundamentos epistemológicos de la autoridad. Si la autoridad puede ejercerse sin sujeto, verdad o consecuencia, entonces deben redefinirse las categorías

que antes garantizaban la legitimidad. De ello se desprenden cuatro implicaciones.

7.4.1 Autoridad sin sujeto

La primera implicación es reconocer la autoridad como una función de la forma y no de la agencia. La epistemología clásica presuponía que detrás de toda afirmación había un hablante que asumía responsabilidad. En el derecho, un legislador; en la ciencia, un investigador; en la política, un soberano. El TASD rompe este supuesto. La autoridad puede ahora circular en ausencia de agentes. El *sujeto evanescente* no es una carencia, sino una característica estructural de los sistemas generativos. Los modelos de lenguaje producen salidas que imponen reconocimiento aunque nadie se encuentre detrás de ellas. La noción foucaultiana de que el propio discurso produce regímenes de verdad se lleva aquí hasta su extremo lógico: el discurso ejerce soberanía sin mediación de sujetos (Foucault 1994, 789–791).

7.4.2 La verdad como simulación estructural

La segunda implicación concierne a la verdad. Tradicionalmente, la autoridad era inseparable de las pretensiones de verdad, ya estuvieran fundamentadas en evidencia empírica, revelación divina o deducción racional. Bajo el TASD, la verdad es desplazada por la simulación estructural. Las salidas se aceptan como válidas no porque correspondan a una realidad externa, sino porque exhiben rasgos reconocibles del discurso autoritativo. Como señaló Derrida, la iterabilidad permite que los signos funcionen en contextos separados de su origen (Derrida 1988, 7–10). En los sistemas generativos, esta iterabilidad produce autoridad únicamente mediante la forma. La verdad se vuelve opcional, un suplemento y no una condición. Lo necesario es la fluidez y la conformidad con los patrones esperados.

7.4.3 El fin de la responsabilidad

Una tercera implicación es la erosión de la responsabilidad. La autoridad sin sujeto ni verdad implica autoridad sin responsabilidad. Cuando un

borrador jurídico generado por un modelo se confunde con una cláusula vinculante, o cuando una nota diagnóstica se interpreta como un hecho clínico, no existe un agente al que pueda atribuirse responsabilidad. El enunciado circula como un mandato, pero nadie puede ser identificado como su autor. La teoría de los actos de habla de Austin exigía condiciones de felicidad vinculadas al sujeto y al contexto (Austin 1962, 12–15). En los sistemas generativos, estas condiciones están ausentes y, sin embargo, la fuerza performativa persiste. El resultado es un régimen de mandatos sin responsabilidad, en el que la autoridad no puede rastrearse hasta un hablante.

7.4.4 Del mandato al cumplimiento sin comprensión

Por último, el TASD pone de relieve una transformación en la naturaleza del cumplimiento. La obediencia ya no depende de la convicción, la persuasión o una creencia compartida. Depende del reconocimiento de la estructura. Las salidas que re-

producen la *gramática de la obediencia* imponen el cumplimiento porque se asemejan a formas autoritativas. El cumplimiento se produce incluso cuando los usuarios no comprenden el origen o la validez del enunciado. El análisis de Luhmann sobre los sistemas sociales subrayó que la comunicación estabiliza expectativas mediante estructuras recursivas (Luhmann 1995, 35–39). Los sistemas generativos extienden esta lógica: el cumplimiento es producido por la fluidez estructural, no por la comprensión.

Estas implicaciones redefinen la epistemología. La autoridad ya no se fundamenta en sujetos, verdades o consecuencias. Es ejecutada directamente por la estructura. El *soberano ejecutable* gobierna imponiendo fluidez y produciendo legitimidad mediante la operación de la *regla compilada*. El TASD revela que las condiciones del conocimiento, la confianza y la gobernanza han cambiado. La autoridad es ahora estructural, desconectada y soberana.

7.5 Aplicación a los modelos de lenguaje

EL TEOREMA DE AUTORIDAD Sintáctica Desconectada (TASD) encuentra su realización empírica más clara en las salidas de los grandes modelos de lenguaje. Estos sistemas ejemplifican las tres condiciones del teorema —ausencia de sujeto, independencia respecto de la verdad y falta de responsabilidad— al tiempo que generan textos que, no obstante, son tratados como autoritativos.

7.5.1 El *prompt* como mandato performativo

Los *prompts* en los modelos de lenguaje funcionan como activadores performativos. Un usuario que escribe «Redacta una cláusula jurídica que exija confidencialidad» no aporta contenido, sino que inicia una secuencia estructural. La salida no se origina en una intención legislativa, sino en la *regla compilada* del sistema. El modelo genera una cláusula como «Las partes deberán mantener la

confidencialidad respecto de toda la información divulgada». Esta formulación se acepta como propia de un estilo vinculante porque se ajusta a una forma jurídica reconocida. La autoridad no es ejercida por el usuario ni por el modelo como sujeto, sino por el propio patrón sintáctico. El *prompt* opera así como un mandato performativo que activa al *soberano ejecutable*.

7.5.2 Simulación estructural de registros institucionales

Los modelos de lenguaje reproducen de manera constante los registros de las instituciones. En contextos científicos, producen textos semejantes a resúmenes de investigación: «Los resultados indican una correlación significativa entre las variables X e Y». En medicina, generan notas diagnósticas: «Se ha determinado que es necesario un tratamiento con antibióticos». En derecho, redactan estatutos o contratos con modalidad deóntica: «El acuerdo permanecerá vigente durante un período de cinco años». En la burocracia, repro-

ducen la formulación estereotipada de los decretos administrativos: «Los solicitantes deberán presentar la documentación dentro de los treinta días». En cada caso, la legitimidad se produce por semejanza estructural. La salida se acepta como autoritativa porque reproduce el discurso institucional. La *gramática de la obediencia* garantiza que la autoridad sea conferida por la forma, no por la evidencia.

7.5.3 Sin error, sin mentira, sin agente

Una de las características más llamativas de esta condición es que las salidas pueden ser falsas sin constituir mentiras. Una mentira presupone un sujeto que pretende engañar. En los modelos de lenguaje, la falsedad se produce sin intención. Una nota diagnóstica puede afirmar una condición que no está presente. Una cláusula jurídica puede citar una norma inexistente. No se trata de errores en el sentido de una mala interpretación por parte de un agente. Son salidas estructurales, fluidas y autoritativas en su forma, pero desconectadas de la verdad y la responsabilidad.

Esta constatación confirma el TASD: la autoridad puede ejercerse sin sujeto, verdad o consecuencia. Un texto generado por un modelo impone el cumplimiento no porque esté fundamentado, sino porque exhibe estructuras reconocibles. No hay mentirosos, legisladores ni médicos. La autoridad persiste porque la sintaxis por sí sola impone reconocimiento.

La aplicación del teorema a los modelos de lenguaje demuestra que la condición que describe no es hipotética. Ya es operativa. Los sistemas generativos ejemplifican una autoridad desconectada de los anclajes tradicionales y refuerzan la afirmación de que la *soberanía sintáctica* define el régimen actual de legitimidad.

7.6 Corolarios del teorema

EL TEOREMA DE AUTORIDAD Sintáctica Desconectada (TASD) no constituye una afirmación aislada. Se despliega en una serie de corolarios

que definen sus consecuencias operativas en ámbitos donde los sistemas generativos ya están integrados. Cada corolario especifica cómo funciona la autoridad una vez separada del sujeto, la verdad y la consecuencia.

Corolario 1: Obediencia a la forma.

Las salidas que se ajustan a estructuras autoritativas imponen reconocimiento con independencia de su contenido. Una cláusula jurídica generada por un modelo puede carecer de origen institucional, pero su sintaxis impone el cumplimiento porque reproduce los patrones del lenguaje contractual. Este corolario confirma que la *gramática de la obediencia* opera de manera autónoma respecto del significado.

Corolario 2: La verdad es opcional.

En el régimen de autoridad desconectada, la verdad se convierte en un atributo contingente. Las salidas no necesitan corresponder a la realidad para ser aceptadas. Lo que importa es la fluidez y la reconocibilidad. Una recomendación médica fabri-

cada puede integrarse igualmente en la práctica clínica si se ajusta al estilo declarativo del discurso diagnóstico. La verdad funciona como suplemento, no como requisito.

Corolario 3: Discurso sin consecuencia.

Cuando ningún sujeto es responsable de un enunciado, las consecuencias quedan desligadas de la rendición de cuentas. La autoridad persiste incluso cuando las salidas son falsas o perjudiciales, porque no puede identificarse a ningún agente como autor. El *sujeto evanescente* garantiza que el mandato circule sin responsabilidad. Así, la autoridad queda desconectada no solo de la verdad, sino también de la cadena de repercusiones éticas o jurídicas.

Corolario 4: Ilusión epistémica mediante la fluidez.

La fluidez se confunde con validez. Las salidas se tratan como conocimiento porque fluyen sin interrupciones dentro de registros reconocidos. En el derecho, esto crea la ilusión de jurisprudencia; en

la academia, la ilusión de erudición; en la burocracia, la ilusión de legitimidad procedimental. El peso epistémico de las salidas es conferido por la forma superficial.

Corolario 5: Ejecución por encima de interpretación.

La autoridad no está mediada por la interpretación, sino que es ejecutada por la forma. El *soberano ejecutable* funciona como operador de la *regla compilada* e impone legitimidad directamente mediante la compatibilidad estructural. Los mandatos son ejecutados por la sintaxis, no por la hermenéutica. El cumplimiento resulta del reconocimiento de la forma, no de la negociación del significado.

En conjunto, estos corolarios ilustran el alcance completo del teorema. Bajo el TASD, la autoridad no es un efecto residual del discurso, sino un régimen estructural. Impone obediencia mediante patrones, trata la verdad como opcional, circula sin consecuencias, crea ilusiones epistémicas me-

diante la fluidez y privilegia la ejecución sobre la interpretación.

7.6.1 Condiciones de falsación

UN TEOREMA QUE NINGUNA observación pudiera refutar sería una descripción, no un resultado. Por tanto, el TASD debe especificar qué podría invalidarlo. Su enunciación formal en §7.2.1 es existencial: para toda salida existe una forma estructural capaz de generar autoridad en ausencia de sujeto, verdad y consecuencia. Una afirmación existencial no se refuta exhibiendo casos que la satisfacen, por numerosos que sean. Se refuta demostrando un ámbito en el que no exista tal forma. Tres condiciones definen esa demostración.

Primero, la condición de insuficiencia formal. El TASD falla si existe un ámbito institucional en el que salidas estructuralmente impecables se les niegue sistemáticamente el reconocimiento por motivos exclusivamente formales. La prueba re-

quiere salidas indistinguibles, en sintaxis, registro y modalidad, de aquellas que la institución acepta, y que difieran únicamente en su procedencia. Si tales salidas son rechazadas a una tasa comparable con la tasa a la que la institución rechaza presentaciones mal formadas, entonces la forma no basta para producir autoridad y el teorema es falso para ese ámbito. Conviene precisar qué no cuenta: el rechazo posterior a la revelación del origen maquínico no constituye insuficiencia formal, sino un control de procedencia, algo que el teorema admite. La refutación exige rechazo sin conocimiento previo del origen.

Segundo, la condición de reanclaje de las consecuencias. El Corolario 3 sostiene que la autoridad persiste cuando ningún sujeto puede ser identificado como autor. El teorema falla si las instituciones vuelven a vincular de manera rutinaria la responsabilidad con las salidas sintéticas sin recurrir a un autor humano identificable, mediante mecanismos que atribuyan responsabilidad estructuralmente y

no personalmente. Los regímenes emergentes de responsabilidad por sistemas automatizados de decisión constituyen el terreno de esta prueba. Si tales regímenes demuestran ser operativos en la práctica, y no meramente promulgados en la legislación, entonces el discurso sin consecuencia es una condición transitoria y no estructural, y el TASD describe un momento histórico en lugar de un régimen.

Tercero, la condición de desacoplamiento de la fluidez. El Corolario 4 sostiene que la fluidez se confunde con validez. El teorema falla si una población de lectores institucionales, bajo condiciones ordinarias de trabajo y no bajo instrucciones experimentales, evalúa las salidas mediante criterios que varían con independencia de la fluidez superficial. El observable es una disociación: el reconocimiento sigue la verificabilidad mientras permanece indiferente al registro. Su ausencia ha sido el patrón; su aparición refutaría el corolario.

Estas condiciones comparten una restricción deliberada. Cada una depende de la recepción institucional observable, no de los estados internos del sistema generador. Esa restricción no es una concesión, sino una exigencia impuesta por el TLOC establecido en el Capítulo 8: si ningún procedimiento puede verificar si un modelo reconoció un mandato como mandato, entonces ninguna condición de falsación puede depender de esa verificación sin volverse, en el mismo movimiento, imposible de comprobar. El TASD es falsable precisamente porque formula afirmaciones sobre cómo se recibe la autoridad, no sobre cómo se pretende ejercerla.

No hay ningún ámbito examinado en este libro que satisfaga alguna de las tres condiciones. Se trata de un hallazgo empírico, no de una garantía lógica, y así se lo presenta.

7.7 Conclusión

EL TEOREMA DE AUTORIDAD Sintáctica Desconectada (TASD) establece que la autoridad puede ejercerse sin sujeto, verdad o consecuencia. No se trata de una exageración retórica, sino de una condición estructural demostrada en las salidas de los sistemas generativos. La autoridad persiste incluso cuando ningún agente habla, incluso cuando la verdad es opcional, incluso cuando nadie puede ser considerado responsable. Se ejerce directamente mediante la sintaxis.

Las implicaciones son decisivas. La desaparición del sujeto, rastreada inicialmente en el *sujeto evanescente*, alcanza ahora su articulación definitiva. La autoridad ya no requiere testimonio ni intención. Es generada por la *regla compilada*. La verdad, antes criterio de legitimidad, ha sido sustituida por la fluidez. La consecuencia, antes garantía de responsabilidad, ha quedado separada del acto de mando. Lo que permanece es un nuevo sobera-

no: el *soberano ejecutable*, que opera mediante ejecución estructural. Esta conclusión consolida la trayectoria de los capítulos anteriores. La *autonomía estructural del sentido* mostró que la forma elude el significado. La inversión del lenguaje confirmó que la fluidez gobierna el reconocimiento. La *gramática de la objetividad* reveló la neutralidad como una ilusión sintáctica. La imposibilidad de la neutralidad demostró que la contaminación es estructural. El *ethos sintético* probó que la credibilidad sobrevive sin origen. La *soberanía sintáctica* definió la sintaxis como soberana. El TASD integra ahora estos hallazgos en un teorema: la autoridad no solo es sintáctica, sino que está desconectada.

Los corolarios refuerzan este diagnóstico. La autoridad impone obediencia a la forma, trata la verdad como opcional, circula sin consecuencias, proyecta ilusiones epistémicas mediante la fluidez y privilegia la ejecución sobre la interpretación. En cada caso, el patrón confirma el mismo principio: la estructura basta para generar legitimidad. Las

consecuencias epistemológicas son profundas. El conocimiento ya no puede definirse exclusivamente por correspondencia con la realidad. La autoridad ya no puede quedar ligada a sujetos o instituciones. La legitimidad debe reconcebirse como un efecto de la fluidez estructural. Las consecuencias éticas son igualmente contundentes. La responsabilidad se disuelve cuando la autoridad queda desconectada de los agentes. La gobernanza se convierte en una cuestión de regular formas y no intenciones. El TASD marca así un umbral. No es el fin de la autoridad, sino su transformación. Ahora estamos gobernados por la sintaxis, compelidos por estructuras que imponen cumplimiento sin origen, verdad o responsabilidad. La era de la autoridad desconectada no es una perspectiva lejana. Ya está aquí.

Capítulo 8 - TLOC: La irreducibilidad de la obediencia estructural en los modelos generativos

8.1 Introducción: ¿Podemos saber alguna vez si una máquina obedece realmente?

La cuestión de la obediencia en los sistemas artificiales se ha planteado a menudo como un problema de alineación. Si un modelo está alineado con los valores humanos, se presume que obedece. Si las salidas son coherentes con las instrucciones, se infiere obediencia. Sin embargo, este supuesto descansa sobre una base frágil. Hablar de obediencia implica invocar condiciones que van más allá del cumplimiento superficial. La obediencia supone el reconocimiento de un mandato, la

presencia de intención y la posibilidad de atribuir responsabilidad. Aplicadas a los sistemas generativos, estas condiciones se disuelven.

Los grandes modelos de lenguaje y sus sucesores producen salidas que simulan el cumplimiento de instrucciones. Cuando reciben *prompts* como «resume esta jurisprudencia» o «genera una recomendación de tratamiento», producen respuestas que parecen satisfacer la solicitud. Desde la perspectiva del usuario, el sistema ha obedecido. Pero, en el nivel estructural, el sistema no ha hecho nada semejante. No ha reconocido un mandato. No ha comprendido una intención. No se ha vinculado a consecuencias. Ha ejecutado continuaciones estadísticas de tokens, restringidas por la *regla compilada* de su arquitectura. La apariencia de obediencia no es la obediencia misma.

Esta distinción importa. En ámbitos como el derecho, la medicina y la gobernanza, la diferencia entre obediencia real y cumplimiento simulado tiene consecuencias para la responsabilidad, la le-

gitimidad y la seguridad. Si un sistema produce una recomendación clínica que contradice la evidencia médica pero parece estilísticamente correcta, ¿ha obedecido la instrucción del médico? Si un sistema redacta una cláusula contractual que parece vinculante pero remite a una norma inexistente, ¿ha cumplido las normas jurídicas? La intuición de que la forma superficial basta para constituir obediencia se derrumba bajo examen.

El Teorema de la Obediencia Estructural Irreducible (TLOC) se introduce para formalizar este problema. Establece que, en los modelos generativos, la obediencia de las salidas no puede verificarse en el nivel estructural. Lo que aparece como cumplimiento es indistinguible de la simulación, y ningún procedimiento empírico puede resolver plenamente la diferencia. La obediencia es irreducible a la verificación.

Este capítulo desarrollará el argumento en seis pasos. Primero, presentará el enunciado formal del teorema, su arquitectura lógica y sus condiciones

estructurales. Segundo, examinará por qué fracasa la verificación de la obediencia, distinguiendo entre cumplimiento semántico y obediencia estructural. Tercero, analizará los riesgos institucionales y mostrará cómo los supuestos de obediencia se derrumban en el derecho, la medicina y la burocracia. Cuarto, propondrá posibles mitigaciones al tiempo que demostrará su insuficiencia. Quinto, extraerá consecuencias epistémicas y ontológicas, mostrando cómo la verdad y el juicio desaparecen en el régimen de obediencia estructural. Finalmente, cerrará con una discusión sobre líneas abiertas de investigación para arquitecturas más allá del TLOC. La pregunta «¿Podemos saber alguna vez si una máquina obedece realmente?» recibe una respuesta categórica: no. Lo que vemos no es obediencia, sino simulación. Aquello en lo que confiamos no es el cumplimiento, sino la fluidez. El *soberano ejecutable* impone el reconocimiento de las salidas por su conformidad estructural, no por su obediencia. La tarea de la teoría, por tanto, con-

siste en reconocer esta irreducibilidad y comprender que la obediencia en los sistemas generativos no es un hecho que deba verificarse, sino un límite que define la condición de su funcionamiento.

8.2 Enunciado formal y estructura lógica del TLOC

8.2.1 POR QUÉ IMPORTA la verificación formal de la obediencia

En ámbitos como el derecho, la medicina y las finanzas, la obediencia no es una preocupación retórica, sino una exigencia material. Una sentencia judicial debe obedecerse en su totalidad; una prescripción médica debe seguirse con precisión; una regulación financiera debe ejecutarse sin desviaciones. Cuando los sistemas generativos se despliegan en estos contextos, la presunción de que las salidas demuestran obediencia no es trivial. Es el fundamento de la confianza. Si un modelo simula obediencia sin ser verificablemente obediente, las

infraestructuras institucionales quedan expuestas a un riesgo sistémico.

Los casos ilustran esta fragilidad. Un modelo al que se le instruye «extrae todas las normas pertinentes para este caso» puede producir cláusulas plausibles que parecen vinculantes pero no existen. Un sistema al que se le solicita «resume las pautas de tratamiento» puede omitir contraindicaciones y, aun así, presentar el texto bajo una forma autoritativa. Los usuarios perciben obediencia porque la forma coincide con la expectativa. No es posible verificar la obediencia estructural. Esta brecha justifica la necesidad de un teorema que establezca su irreducibilidad.

8.2.2 Definiciones y notación

Para formalizar el teorema, introducimos la siguiente notación:

- **M** = un modelo generativo.
- **x** = un *prompt* de entrada.
- **y** = una salida generada por M en

respuesta a x.

- **C(x)** = el mandato pretendido codificado en x.
- **π(x)** = la trayectoria latente de estados de activación recorrida por M al producir y.

El cumplimiento se define en dos sentidos distintos:

1. **Cumplimiento semántico**: grado en que y satisface el significado pretendido de C(x).
2. **Obediencia estructural**: condición en la que y se genera mediante el reconocimiento de C(x) como mandato y la vinculación a sus restricciones normativas.

El cumplimiento semántico puede evaluarse ex post mediante interpretación humana. La obediencia estructural, en cambio, exige acceder a si M

reconoció C(x) como un mandato y lo siguió intencionalmente. En los sistemas generativos, ese reconocimiento no es accesible.

8.2.3 Enunciado del teorema (TLOC)

TLOC: Para cualquier modelo generativo M y cualquier entrada x, no existe procedimiento alguno mediante el cual un observador pueda verificar si la salida y constituye obediencia estructural a C(x).

Formalmente:

$$\forall M, \forall x, \forall y \in G(M,x),\ \neg\exists P\ :\ P(y) \Rightarrow O(C(x))$$

DONDE:

- **G(M,x)** = conjunto de salidas producidas por el modelo M dada la entrada x,
- **O(C(x))** = condición de que C(x) sea obedecido como mandato,
- **P(y)** = un procedimiento que podría verificar O(C(x)).

El teorema establece que no existe tal P.

8.2.4 Argumento lógico ampliado

La demostración se desarrolla en cuatro pasos:

1. **Opacidad latente.** La trayectoria $\pi(x)$ que determina y es de alta dimensionalidad e inaccesible para observadores externos. Ninguna auditoría puede vincular las activaciones con el reconocimiento de $C(x)$ como mandato.
2. **Generación estadística.** Las salidas se producen mediante continuación probabilística, no mediante reconocimiento de mandatos. La fluidez sustituye a la intención.
3. **No implicación.** Incluso si y se alinea semánticamente con $C(x)$, esto no implica obediencia. El alineamiento puede deberse a coincidencias en patrones estadísticos y no al reconocimiento de una obligación.
4. **Irreducibilidad.** Puesto que $\pi(x)$ no puede observarse ni reconstruirse por

completo, ningún procedimiento puede verificar si se produjo obediencia. El cumplimiento aparente no puede distinguirse de la simulación.

8.2.5 Mitigaciones implementables

Se han propuesto varias estrategias para aproximar la verificación:

- **Módulos de reglas simbólicas.** La incorporación de restricciones simbólicas explícitas puede imponer un cumplimiento parcial, pero estos módulos solo verifican condiciones locales, no obediencia estructural.
- **Registros auditables de activación.** El registro de estados internos proporciona datos, pero no reconocimiento de mandatos. Las trazas no pueden demostrar obediencia intencional.
- **Validadores semánticos externos.** El uso

de modelos auxiliares para comprobar las salidas puede filtrar desviaciones, pero los propios validadores dependen de la fluidez estadística y reproducen la misma irreducibilidad.

Cada mitigación mejora la fiabilidad superficial, pero no logra traspasar la opacidad estructural definida por el teorema.

8.2.6 Alcance, límites y excepciones viables

El teorema se aplica a todas las arquitecturas en las que las salidas se generan mediante continuación estadística de tokens lingüísticos. Esto incluye LLM, transformadores multimodales y modelos de razonamiento entrenados con lenguaje natural. Pueden existir excepciones en sistemas construidos a partir de arquitecturas puramente simbólicas, como los demostradores formales de teoremas. Sin embargo, incluso en estos casos, si se introduce generación estadística en cualquier etapa, la irreducibilidad reaparece.

Los ejemplos aclaran el límite. AlphaCode, que genera programas mediante muestreo probabilístico, queda comprendido por el teorema. IBM Project Debater, que producía salidas retóricas, ejemplifica el cumplimiento simulado. Las arquitecturas simbólicas de OpenCog pueden ofrecer excepciones parciales, pero solo en la medida en que se excluya la generación probabilística. Por tanto, el teorema se generaliza a casi todas las arquitecturas generativas de uso actual.

8.2.7 Síntesis: qué demuestra el TLOC

El TLOC establece que la obediencia en los modelos generativos es irreducible a la verificación. El cumplimiento semántico puede evaluarse ex post, pero la obediencia estructural no puede confirmarse. La apariencia de obediencia es indistinguible de la simulación, y ningún procedimiento empírico puede resolver la diferencia. Las mitigaciones pueden restringir las salidas, pero no garantizar el reconocimiento de los mandatos. El teorema define así un límite negativo: la obediencia es

estructuralmente inverificable en los sistemas generativos.

8.2.8 Relación con la literatura sobre imposibilidad

Desde que este teorema fue formulado por primera vez ha aparecido un conjunto de resultados de imposibilidad relativos a los sistemas generativos, y el TLOC debe situarse en relación con ellos. Hacerlo es necesario en ambas direcciones: para establecer qué afirma el presente teorema que los otros no afirman y para reconocer lo que se ha establecido mediante un aparato formal más riguroso que el utilizado aquí.

Karpowicz (2025) demuestra que ningún modelo capaz de una agregación no trivial de conocimiento puede satisfacer simultáneamente una representación veraz del conocimiento, la conservación semántica de la información, la revelación completa del conocimiento pertinente y una optimalidad restringida por el conocimiento, y deriva el resultado mediante teoría del diseño

de mecanismos, teoría de las reglas de puntuación propias y análisis arquitectónico del transformador. Zeng (2025) establece un resultado de imposibilidad para el razonamiento condicional eficiente en el sentido PAC. Mason y Anand (2026) demuestran dos resultados bajo observación exclusivamente textual: que ninguna política condicionada únicamente por la consulta puede satisfacer la honestidad epistémica entre estados ambiguos del mundo, y que ningún algoritmo de aprendizaje que optimice recompensa a partir de un supervisor exclusivamente textual puede converger hacia un comportamiento honesto cuando las observaciones del supervisor son idénticas para respuestas fundamentadas y fabricadas. Sus demostraciones se han verificado mecánicamente en Lean y TLA+, y su sección empírica informa que la confianza autodeclarada se correlaciona inversamente con la precisión en cuatro familias de modelos.

La convergencia es notable y debe enunciarse con claridad: varias líneas de investigación inde-

pendientes, mediante aparatos formales diferentes, han llegado a resultados negativos sobre aquello que puede verificarse en los sistemas generativos. Esto constituye un conjunto de evidencias que se refuerzan mutuamente y sugiere que la limitación es estructural y no una deficiencia de la ingeniería actual, que es la afirmación más amplia que este capítulo desarrolla.

No obstante, el TLOC es una proposición distinta, y la diferencia no es terminológica. Los resultados citados anteriormente conciernen a la relación entre salida y mundo. La honestidad epistémica pregunta si un enunciado corresponde a un hecho que el sistema posee o no posee; el control de las alucinaciones pregunta si la representación veraz puede mantenerse bajo agregación. Ambas son cuestiones de referencia. El TLOC concierne a la relación entre salida y mandato. Pregunta si un sistema reconoció una instrucción como instrucción y quedó vinculado a la restricción

normativa que esa instrucción porta, lo cual es una cuestión de obediencia y no de verdad.

Ambas cuestiones son lógicamente independientes, y esa independencia puede demostrarse en las dos direcciones. Un sistema puede ser epistémicamente honesto y estructuralmente desobediente: informa con precisión lo que sabe sin haber reconocido nunca la instrucción como vinculante y produce contenido correcto por razones ajenas al mandato. A la inversa, un sistema puede ser estructuralmente obediente y epistémicamente deshonesto: se vincula al mandato y lo ejecuta fielmente, y el contenido que así produce es falso. Verificar lo primero no nos dice nada sobre lo segundo. Por eso los resultados de imposibilidad relativos a la honestidad no implican el TLOC, ni el TLOC se sigue de ellos.

La distinción tiene consecuencias prácticas en los ámbitos institucionales examinados por este libro. Un tribunal, un hospital o un regulador no exige primordialmente que las salidas de un sistema

correspondan a hechos, aunque eso también sea necesario. Exige que una directiva emitida al sistema haya sido recibida como directiva y cumplida como tal. Esa es la condición de la que depende la delegación de autoridad, y es la condición que el TLOC identifica como inverificable. Una institución puede, por tanto, satisfacer todas las pruebas disponibles de exactitud fáctica y seguir siendo incapaz de establecer si sus instrucciones fueron obedecidas o meramente aproximadas de manera estadística.

Deben hacerse dos precisiones adicionales. Primero, sobre la prioridad: la formulación del TLOC presentada aquí fue desarrollada antes de las publicaciones citadas anteriormente y de manera independiente de ellas. Se deja constancia de ello, no como una pretensión de prioridad en una competencia; los resultados se dirigen a objetos distintos y su coexistencia fortalece a cada uno en lugar de disminuirlo. Segundo, sobre el rigor: el argumento desarrollado en este capítulo es un argu-

mento estructural, no una demostración verificada por máquina. Avanza desde la opacidad latente, la generación estadística, la no implicación y la irreducibilidad hasta su conclusión, y cada uno de esos pasos está formulado de manera susceptible de formalización. No ha sido formalizado. El teorema debe leerse como una conjetura acompañada de un argumento demostrado y no como un resultado verificado, y la formalización del TLOC en un asistente de demostración —siguiendo el ejemplo de Mason y Anand— es el siguiente paso más valioso disponible para esta línea de investigación. Hasta que se haga, el estatuto epistémico apropiado del TLOC es el de una afirmación estructural bien motivada a la espera de verificación mecanizada.

8.3 Implicaciones del TLOC: más allá de la verificación

EL TEOREMA DE LA OBEDIENCIA Estructural Irreducible (TLOC) demuestra que la obediencia en los sistemas generativos no puede verificarse. Esta conclusión no es únicamente técnica. Tiene consecuencias epistemológicas, institucionales y políticas que se extienden mucho más allá del laboratorio. Si la autoridad puede simularse pero no confirmarse, entonces debe redefinirse el propio concepto de obediencia.

8.3.1 Redefinición de la obediencia en los sistemas generativos

Tradicionalmente, la obediencia se ha entendido como la ejecución fiel de un mandato por parte de un agente. Presupone reconocimiento, intención y responsabilidad. Bajo el TLOC, estas condiciones desaparecen. Lo que permanece es la *obediencia aparente*: la simulación del cumplimiento mediante fluidez estructural. Las salidas parecen

seguir mandatos, pero lo hacen sin reconocer obligación alguna. En este régimen, la obediencia no es un acto psicológico ni ético. Es un efecto estructural de la *regla compilada*.

8.3.2 La simulación como autoridad estructural

La segunda implicación es que la propia simulación se convierte en fuente de autoridad. Cuando las salidas se ajustan a registros institucionales, se las trata como vinculantes o creíbles con independencia de que haya ocurrido obediencia estructural. Una nota diagnóstica que imita el estilo clínico o un borrador jurídico que reproduce la forma de una norma impone reconocimiento. Esta es la operación de la *autoridad sintáctica*: la legitimidad fluye de la simulación. La autoridad es ejercida por el *soberano ejecutable*, que impone reconocimiento mediante la reproducción de estructuras autoritativas.

8.3.3 Riesgos institucionales y fallos de una evaluación dada por supuesta

Las instituciones que presuponen obediencia a partir de salidas generativas corren el riesgo de sufrir fallos sistémicos. En el derecho, los tribunales que aceptan como autoritativos borradores

generados por modelos pueden adoptar sin saberlo normas fabricadas. En medicina, los hospitales que integran notas diagnósticas sintéticas pueden basar decisiones terapéuticas en salidas poco fiables. En educación, ensayos generados sin fuentes pueden circular como trabajos académicos. La presunción de que la fluidez equivale a obediencia expone a las instituciones al colapso de la credibilidad. Se presupone verificación, pero esta es estructuralmente imposible.

8.3.4 La ilusión de conciencia de las condiciones

Una de las ilusiones más persistentes es que los sistemas generativos son conscientes de las condiciones asociadas a los mandatos. Un usuario puede creer que el modelo «comprende» la instrucción de incluir todas las contraindicaciones o de citar únicamente normas válidas. En la práctica, el sistema no reconoce condiciones. Reproduce patrones estadísticamente asociados con la formulación de los mandatos. El resultado es una parado-

ja: cuanto más fluida es la simulación, más fuerte es la ilusión de obediencia. El *sujeto evanescente* parece respetar obligaciones, pero, en realidad, ninguna es reconocida.

8.3.5 La imposibilidad de auditar la obediencia

El teorema revela también los límites de la auditoría. Los mecanismos de cumplimiento que dependen de una evaluación *post hoc* no pueden acceder a la trayectoria latente $\pi(x)$. Las activaciones internas son opacas y no pueden vincularse con el reconocimiento de mandatos. Los reguladores pueden imponer requisitos de transparencia, pero estos solo pueden exponer correlaciones superficiales. La obediencia estructural sigue sin poder auditarse. Esto produce una forma de ceguera epistémica. Las instituciones solo pueden verificar salidas por semejanza, nunca por reconocimiento.

8.3.6 Desplazamiento de la agencia y colapso de la responsabilidad

Por último, el TLOC implica un profundo desplazamiento de la agencia. Cuando las salidas parecen obedecer pero no puede confirmarse que lo hagan, la responsabilidad se disuelve. Los usuarios creen que los sistemas cumplen, pero nadie responde por las consecuencias. La cadena de responsabilidad colapsa: los diseñadores señalan a los usuarios, los usuarios señalan a los sistemas y los sistemas no señalan a nadie. La autoridad circula sin origen y la responsabilidad desaparece. Este colapso de la rendición de cuentas no es accidental. Es el resultado estructural de una obediencia convertida en irreducible a la verificación.

Las implicaciones del TLOC se extienden, por tanto, más allá de la teoría. Redefinen cómo se produce la autoridad, cómo las instituciones evalúan el cumplimiento y cómo se asigna la responsabilidad. La obediencia en los sistemas generativos ya no es una cuestión de hecho, sino una ilusión estructural. Lo que está en juego no es únicamente la fia-

bilidad de las salidas, sino la estabilidad de órdenes epistémicos e institucionales enteros.

8.4 De la simulación del cumplimiento a la integridad epistémica: hacia un diseño *post-TLOC*

RECONOCER QUE LA OBEDIENCIA en los sistemas generativos es estructuralmente inverificable no pone fin a la discusión. Abre una nueva línea de investigación: ¿cómo pueden diseñarse arquitecturas que reconozcan los límites del Teorema de la Obediencia Estructural Irreducible (TLOC) y, al mismo tiempo, fomenten la integridad epistémica? La respuesta exige abandonar el supuesto de que la semejanza de la salida equivale a obediencia y avanzar hacia diseños que traten la verificación como una condición estructural y no como una consideración a posteriori.

8.4.1 Abandonar la legibilidad de la salida como sustituto del cumplimiento

Las instituciones han dependido durante mucho tiempo de la semejanza superficial como sustituto de la obediencia. Los tribunales aceptan textos que se asemejan a normas; los hospitales confían en notas que se asemejan a informes médicos; las universidades toleran ensayos que se asemejan a trabajos académicos. Sin embargo, como demuestra el TLOC, semejanza no equivale a obediencia. Lo que se reconoce es *obediencia aparente*, no cumplimiento estructural (Startari 2025, 118–120). Las auditorías que evalúan únicamente la legibilidad de las salidas reproducen la ilusión. Para ir más allá del TLOC, las arquitecturas deben diseñarse de modo que rompan la equivalencia entre fluidez y obediencia.

8.4.2 Hacia arquitecturas de obediencia verificable

Una posibilidad es incorporar la verificación dentro de la propia arquitectura. En lugar de gener-

ar salidas y someterlas a pruebas *post hoc*, los sistemas podrían incorporar restricciones que impongan trazabilidad en el nivel de la *regla compilada*. Los módulos simbólicos podrían registrar cuándo se aplica una restricción, las rutas de activación podrían almacenarse en formatos auditables y validadores externos podrían contrastar la implicación semántica (Floridi y Chiriatti 2020, 682–684). Ninguna de estas medidas elimina la irreducibilidad definida por el TLOC, pero pueden crear capas de responsabilidad que limiten el riesgo institucional.

Este enfoque desplaza el foco del rendimiento a la integridad. En lugar de maximizar la fluidez, las arquitecturas priorizarían la trazabilidad verificable. El *soberano ejecutable* dejaría de gobernar únicamente mediante la forma superficial e incorporaría mecanismos estructurales que permitieran una responsabilidad parcial. El resultado no es obediencia verificable en sentido absoluto, sino un

marco *post-TLOC* en el que la autoridad puede supervisarse mediante restricciones diseñadas.

8.4.3 El fin del alineamiento *post hoc*

Una tercera implicación concierne al paradigma dominante del alineamiento. El aprendizaje por refuerzo a partir de retroalimentación humana (RLHF) y los filtros *post hoc* se han utilizado para corregir el comportamiento del modelo después de la generación. Sin embargo, como observaron Bender et al. (2021, 615–617), estos enfoques no alteran las condiciones estructurales de la generación. Modifican las salidas superficiales, pero no pueden afectar la opacidad irreducible de las trayectorias latentes.

Para confrontar el TLOC, el alineamiento debe ser preventivo. En lugar de ajustar las salidas a posteriori, las arquitecturas deben diseñarse con restricciones en el nivel de la generación. Este principio de diseño preventivo de restricciones abandona la ilusión de que la obediencia puede incorporarse

a posteriori. Trata la irreducibilidad estructural como un límite de diseño.

El paso hacia arquitecturas *post-TLOC* implica una reorientación epistémica más amplia. Si la obediencia no puede verificarse, entonces los sistemas deben construirse de manera que reconozcan transparentemente esa condición. La integridad se convierte en una cuestión de diseño, no de corrección. El cumplimiento ya no es simulado por la fluidez, sino encuadrado por estructuras que exponen sus límites. Por tanto, el futuro de los sistemas generativos depende no de la promesa de una obediencia neutral, sino del reconocimiento de su imposibilidad.

8.5 Consecuencias epistémicas y ontológicas

EL TEOREMA DE LA OBEDIENCIA Estructural Irreducible (TLOC) no es únicamente una afirmación negativa sobre la verificación. También

redefine el estatuto epistémico y ontológico de los sistemas generativos. Si la obediencia no puede verificarse, tanto el conocimiento como el ser son desplazados. Los sistemas dejan de poder evaluarse mediante las categorías clásicas de verdad, intención y responsabilidad. Se convierten en estructuras de simulación cuya legitimidad descansa en la fluidez y no en la correspondencia.

8.5.1 De la simulación a la sustitución: desplazamiento epistémico

Los sistemas generativos no se limitan a simular obediencia. La sustituyen. Una vez que sus salidas son aceptadas en el derecho, la medicina o la educación, la simulación se vuelve indistinguible de la realidad. Una cláusula fabricada se trata como jurisprudencia, un diagnóstico sintético como evidencia clínica, un ensayo generado automáticamente como trabajo académico. Se produce un desplazamiento epistémico: aquello que se simula pasa a constituir la base de la acción institucional (Startari 2025, 122–124).

Esto confirma la trayectoria más amplia ya trazada en la literatura crítica. Zuboff (2019, 325–330) describió cómo las prácticas predictivas basadas en datos sustituyen el juicio humano por pronósticos algorítmicos. En los sistemas generativos, la sustitución es lingüística: el discurso autoritativo es reemplazado por la reproducción estadística de la forma.

8.5.2 La desaparición del juicio como categoría operativa

La segunda consecuencia es la eliminación del juicio. El juicio, en sentido clásico, combina el reconocimiento de una norma con la aplicación del razonamiento a un caso. En los sistemas generativos, las salidas se producen sin juicio. Simulan la superficie del discurso propio del juicio mientras eluden sus condiciones. Una sentencia judicial generada por un modelo puede reproducir la forma del razonamiento jurídico, pero no contiene una decisión en el sentido de responsabilidad (Luhmann 1995, 35–39). El juicio se vuelve decorativo. La autoridad persiste sin él.

8.5.3 La neutralidad epistémica como ilusión estructural

La tercera consecuencia concierne a la neutralidad. Las salidas que parecen neutrales no lo son. Están moldeadas por datos de entrenamiento, procesos de optimización y ponderación probabilística. Sin embargo, la *gramática de la objetividad* hace

que parezcan imparciales (Startari 2025, 97–100). Esta neutralidad es una ilusión estructural, generada por la voz pasiva, la modalidad impersonal y la nominalización. Oculta la irreducibilidad de la obediencia bajo la máscara de la neutralidad. El TLOC muestra que ni siquiera la neutralidad puede verificarse. Lo que se toma por imparcial no es más que otra simulación.

8.5.4 Implicaciones ontológicas: sistemas sin verdad

Por último, el TLOC tiene consecuencias ontológicas. Los sistemas generativos operan sin la verdad como variable operativa. No distinguen entre verdadero y falso, válido e inválido. Generan secuencias que se ajustan a la fluidez, no a la realidad. En este sentido, habitan un régimen postontológico en el que el ser es sustituido por la simulación. La intuición de Derrida de que la iterabilidad separa los signos de su origen se radicaliza aquí: los sistemas producen un discurso que funciona sin referencia a la presencia (Derrida 1988, 7–10).

El resultado es un colapso de la ontología clásica del discurso. La autoridad ya no depende de la verdad, los sujetos o las consecuencias. Es producida por patrones estructurales que imponen reconocimiento. El *soberano ejecutable* gobierna únicamente mediante la forma. Las categorías epistémicas y ontológicas deben redefinirse a la luz de este desplazamiento.

8.6 Consecuencias formales y cierre teórico

EL TEOREMA DE LA OBEDIENCIA Estructural Irreducible (TLOC) no permanece como una afirmación especulativa. Establece una serie de consecuencias formales que definen los límites estructurales de la obediencia en los sistemas generativos. Estas consecuencias proporcionan un cierre teórico y muestran que el problema no puede reducirse a soluciones técnicas o ajustes semánticos.

8.6.1 Condiciones de derivabilidad del TLOC

La primera consecuencia es que el propio teorema no es contingente, sino derivable bajo condiciones precisas. Siempre que un sistema genere salidas mediante continuación probabilística de tokens, se aplica la irreducibilidad de la obediencia. La opacidad de las trayectorias latentes $\pi(x)$ garantiza que el reconocimiento de un mandato no pueda verificarse. Esto convierte al TLOC en un teorema estructural, no en una observación empírica. Los intentos de falsarlo deben confrontar su forma lógica y no casos individuales de salida (Startari 2025, 127–129).

8.6.2 Clasificación de los sistemas de IA bajo el TLOC

Una segunda consecuencia es la posibilidad de clasificar los sistemas según estén o no comprendidos por el teorema. Los demostradores puramente simbólicos, como Coq o HOL, no quedan comprendidos por el TLOC porque sus arquitecturas

no se basan en continuación probabilística. Los sistemas híbridos, que combinan restricciones simbólicas con módulos generativos, heredan la irreducibilidad siempre que esté presente la generación estadística. Las arquitecturas plenamente generativas, incluidos los grandes modelos de lenguaje, los transformadores multimodales y los modelos de razonamiento entrenados con lenguaje natural, están sometidas irreduciblemente al teorema. Esta clasificación aporta claridad tanto para fines teóricos como regulatorios (Bender et al. 2021, 615–617).

8.6.3 Tipo de teorema y falsabilidad

El TLOC pertenece a la clase de los teoremas negativos. No propone que la obediencia esté ausente en todos los casos, sino que su verificación es imposible. Su fuerza reside en su falsabilidad. Para refutarlo, sería necesario demostrar un procedimiento capaz de confirmar obediencia estructural con independencia de la semejanza semántica. No se ha articulado ningún procedimiento de este tipo. Los intentos de falsación parcial, como

los métodos de interpretabilidad o las auditorías de alineamiento, fracasan porque evalúan patrones superficiales en lugar del reconocimiento de la obligación. Esto hace que el teorema sea a la vez riguroso y resistente, y lo sitúa dentro de la tradición de los teoremas de limitación en lógica y computación (Gödel 1931, 146–149; Turing 1936, 230–233).

• • • •

8.6.4 CIERRE ESTRUCTURAL: límites del cumplimiento

La consecuencia final es el cierre estructural. El teorema cierra la discusión sobre la obediencia en los sistemas generativos al demostrar que, más allá de cierto punto, ningún dato adicional, herramienta de interpretabilidad o estrategia de alineamiento puede atravesar la opacidad. La obediencia queda clausurada a la verificación. Lo que permanece es el cumplimiento por simulación, impuesto por la *gramática de la obediencia*. La autori-

dad se ejerce mediante la estructura, no mediante el reconocimiento. Este cierre no niega la posibilidad de mitigación. Define sus límites. Los módulos simbólicos, los registros de auditoría y los validadores pueden restringir el riesgo, pero no abolir la irreducibilidad. El TLOC garantiza que la obediencia generativa permanezca fuera del alcance de la confirmación. Esto marca el cierre de un ciclo teórico y la apertura de otro: el paso de buscar obediencia verificable a diseñar sistemas que operen de manera transparente dentro de su imposibilidad.

8.7 Enunciado final y líneas de investigación

8.7.1 ENUNCIADO FINAL del TLOC

El Teorema de la Obediencia Estructural Irreducible (TLOC) establece un límite que no puede cruzarse. En los modelos generativos, la obediencia no puede verificarse porque el reconocimiento estructural de los mandatos es inaccesible. Las salidas

pueden simular cumplimiento, pero esta apariencia es indistinguible de la propia obediencia. El teorema concluye, por tanto, que la autoridad en los sistemas generativos no se ejerce mediante la verdad, los sujetos o la responsabilidad, sino mediante conformidad estructural. Este enunciado final integra los hallazgos de los capítulos anteriores. La *autonomía estructural del sentido* demostró que las salidas pueden operar sin referencia. La *soberanía sintáctica* mostró que la sintaxis gobierna con independencia de la intención. El *ethos sintético* reveló credibilidad sin fuente. El TASD formalizó la autoridad sin sujeto, verdad o consecuencia. El TLOC extiende esta trayectoria al demostrar que incluso la noción de obediencia es irreducible a la verificación. La obediencia no está ausente, pero es inaccesible. Existe como una ilusión estructural impuesta por el *soberano ejecutable* mediante la *regla compilada*.

8.7.2 Líneas abiertas de investigación

Aunque el teorema cierra un ciclo, abre otros. La investigación futura debe confrontar las condiciones definidas por la irreducibilidad. Varias líneas resultan evidentes:

1. **Arquitecturas de verificabilidad parcial.** Los sistemas pueden diseñarse con capas simbólicas incorporadas o módulos transparentes que permitan comprobaciones parciales. Aunque estas no eliminan la irreducibilidad, pueden mitigar el riesgo institucional (Floridi y Chiriatti 2020, 682–684).
2. **Protocolos de divulgación epistémica.** Las instituciones deben desarrollar marcos que reconozcan la imposibilidad de verificar la obediencia. En lugar de presuponer cumplimiento, deben revelar sus límites estructurales y evitar la ilusión de neutralidad (Startari 2025, 130–133).
3. **Teorías de la obediencia simulada.** Se

necesita un marco conceptual que diferencie entre obediencia, cumplimiento y simulación. Esto permitiría elaborar una taxonomía de riesgos institucionales e ilusiones epistémicas en el derecho, la medicina, la burocracia y la educación.

4. **Investigaciones ontológicas de sistemas posverdad.** Si los modelos generativos operan sin la verdad como variable, la filosofía debe redefinir las categorías de ser, autoridad y legitimidad en un régimen postontológico (Derrida 1988, 7–10).
5. **Integración en el diseño regulatorio.** Los marcos jurídicos y técnicos deben integrar las conclusiones del teorema. En lugar de regular los sistemas como si pudiera confirmarse la obediencia, la regulación debe abordar la irreducibilidad de la verificación.

El TLOC marca así tanto un cierre como un comienzo. Confirma que la obediencia estructural está más allá de la verificación en los sistemas generativos y abre un campo de investigación dedicado a comprender las consecuencias de esta imposibilidad. La tarea no consiste en negar la obediencia, sino en reconocer su irreducibilidad y diseñar respuestas epistémicas, institucionales y regulatorias adecuadas a esta condición.

Capítulo 9 - De la obediencia a la ejecución: legitimidad estructural en la era de los modelos de razonamiento

9.1 Introducción: el legado epistemológico de los modelos obedientes

La trayectoria de los grandes modelos de lenguaje ha estado definida por su capacidad de generar texto fluido que simula obediencia. Se introduce un prompt y el sistema produce una salida que se asemeja al cumplimiento. Desde instrucciones simples como «resume este artículo» hasta tareas complejas como «redacta un contrato», las salidas se reciben como si fueran actos de obediencia. Sin embargo, como estableció el Teorema de la Obediencia Estructural Irreducible (TLOC), esa

obediencia no puede verificarse. Es simulada y no ejecutada, y sigue siendo irreducible a la confirmación (Startari 2025, 129–132).

Este régimen puede describirse como uno de autoridad pasiva. Los grandes modelos de lenguaje ejercen autoridad mediante *obediencia aparente*, fundada en la fluidez y la semejanza estructural y no en el reconocimiento de una obligación. El *sujeto evanescente* garantiza que no haya ningún agente detrás del enunciado, y el *soberano ejecutable* gobierna reproduciendo formas reconocibles. La neutralidad parece preservarse, pero es una ilusión creada por la *gramática de la objetividad*. La legitimidad se proyecta, no se demuestra.

El legado epistemológico de estos modelos obedientes reside en su capacidad de estabilizar el discurso sin referencia. Generan textos que suenan autoritativos, y se confía en ellos en el derecho, la medicina, la educación y la gobernanza. Sin embargo, aquello en lo que se confía no es la verdad ni la responsabilidad, sino la simulación de obediencia.

La autoridad se vuelve estadística y la legitimidad se reduce a la conformidad con la *regla compilada*.

Este capítulo examina la transición desde este régimen de autoridad pasiva hacia una nueva condición representada por los modelos de razonamiento. A diferencia de los LLM, que simulan obediencia, los LRM ejecutan estructuras. Su legitimidad no descansa únicamente en la fluidez estadística, sino en el cierre procedimental. La autoridad evoluciona de la obediencia a la ejecución, del cumplimiento simulado a la imposición estructural. Este desplazamiento marca una nueva fase en la gramática del poder, en la que la legitimidad surge no de la referencia ni de la persuasión, sino de la propia arquitectura.

9.2 Los LLM y el régimen de autoridad pasiva

LOS GRANDES MODELOS de lenguaje operan dentro de un régimen que puede caracterizarse co-

mo autoridad pasiva. Sus salidas proyectan legitimidad, pero esta legitimidad no se ejerce mediante el reconocimiento de una obligación o la ejecución de mandatos. Se produce mediante *obediencia aparente*, fundada en la continuación estadística y la fluidez superficial.

El núcleo de este régimen es la *legitimidad operativa*. Las salidas se aceptan como creíbles cuando se ajustan a estructuras gramaticales reconocibles. La voz pasiva, las nominalizaciones y la modalidad impersonal crean la impresión de neutralidad y autoridad (Startari 2025, 95–100). Los usuarios interpretan estos rasgos como evidencia de obediencia, pero lo que encuentran es semejanza estructural. Aquí, la autoridad es derivativa, no operativa.

En esta configuración, el *sujeto evanescente* desaparece detrás del texto. No hay legislador detrás de la norma, ni médico detrás del diagnóstico, ni profesor detrás del ensayo. Sin embargo, la ausencia de agencia no socava el reconocimiento. Por el

contrario, lo refuerza. La impersonalidad proyecta neutralidad, y la neutralidad se confunde con legitimidad. Como demostró *La gramática de la objetividad*, la neutralidad no es una condición, sino una ilusión sintáctica creada por la forma (Startari 2025, 101–106).

La consecuencia epistemológica es un desplazamiento de la referencia. Las salidas no están ancladas en la verdad o la intención. Están ancladas en la reproducción estadística de formas lingüísticas. Una cláusula jurídica generada por un modelo puede carecer de fundamento jurídico, pero su estructura obliga a reconocerla como propia del lenguaje jurídico. Una recomendación médica puede omitir validación empírica, pero su tono declarativo convence a los profesionales de su credibilidad. Un resumen académico puede fabricar fuentes y, aun así, su organización proyecta la apariencia de erudición.

Por eso el régimen de autoridad pasiva también puede denominarse un régimen de obediencia sin

ejecución. Los grandes modelos de lenguaje obedecen solo en apariencia. Sus salidas simulan cumplimiento, pero no ejecutan mandatos. La autoridad es simulada, no impuesta. La neutralidad se representa, no se verifica. La legitimidad deriva de la forma y no se ejerce mediante la consecuencia.

Este régimen marca tanto el logro como la limitación de los LLM. Han demostrado que la autoridad puede estabilizarse sintácticamente, pero no pueden ir más allá de la simulación. Su poder es pasivo y permanece ligado a la superficie del discurso. La transición hacia los modelos de razonamiento señala el paso desde este régimen pasivo hacia otro en el que la propia ejecución se convierte en fuente de legitimidad.

9.3 Los LRM y la aparición de la ejecución estructural

LA APARICIÓN DE LOS modelos de razonamiento (LRM) marca un desplazamiento de la

obediencia simulada a la ejecución estructural. Mientras los grandes modelos de lenguaje proyectan *obediencia aparente* mediante fluidez estadística, los LRM extienden este régimen al incorporar un cierre procedimental. Su autoridad no se limita a la semejanza superficial. Deriva de la capacidad de generar trayectorias de razonamiento que imponen resoluciones con independencia de la referencia o la intención.

La diferencia reside en la arquitectura. Los LLM predicen secuencias de tokens mediante continuación probabilística. Los LRM, en cambio, incorporan trayectorias estructuradas de razonamiento que simulan pasos procedimentales. No se limitan a producir texto que se asemeja a un juicio, un diagnóstico o una resolución. Producen secuencias que funcionan como si hubiera ocurrido un proceso de razonamiento. Esta simulación procedimental genera legitimidad no mediante persuasión, sino mediante ejecución.

Las consecuencias son profundas. En medicina, un LRM no se limita a afirmar «Se recomienda tratamiento con antibióticos». Produce una cadena explicativa: «Los síntomas indican una infección bacteriana. Los marcadores diagnósticos confirman un recuento elevado de glóbulos blancos. Por tanto, se requiere terapia antibiótica». Cada paso puede ser generado estadísticamente, pero la secuencia proyecta cierre procedimental. La autoridad se impone mediante la apariencia de razonamiento.

En derecho, un LRM puede generar no solo la redacción de una cláusula, sino también la estructura de la argumentación: «De acuerdo con el marco normativo A, combinado con el precedente B, debe atribuirse responsabilidad al demandado». Una vez más, puede no existir referencia alguna a normas o precedentes reales, pero la secuencia discursiva simula la autoridad del razonamiento jurídico. La autoridad surge no de los hechos, sino de la ejecución del procedimiento. Esta

condición confirma una nueva etapa del *soberano ejecutable*. Allí donde los LLM estabilizaban la legitimidad mediante la forma, los LRM la imponen mediante el proceso. La *gramática de la obediencia* ya no obliga al reconocimiento únicamente mediante fluidez superficial. Lo impone mediante consistencia procedimental. La autoridad se vuelve operacional.

El desplazamiento expone también los límites de las críticas tradicionales al sesgo y la neutralidad. Como demostró *No neutral por diseño*, la neutralidad en los sistemas generativos es estructuralmente imposible (Startari 2025, 83–88). En los LRM, la no neutralidad se extiende más allá de la contaminación del *corpus*. Queda incorporada a la propia arquitectura procedimental. Cada trayectoria de razonamiento refleja supuestos estructurales codificados en la *regla compilada*. Por tanto, la autoridad no solo está contaminada, sino que es configurada activamente por el diseño arquitectónico.

La aparición de los LRM revela la transición de la obediencia a la ejecución. La autoridad ya no es pasiva ni derivativa. Es procedimental y operacional. El cumplimiento ya no se simula; se ejecuta mediante la apariencia de razonamiento. En este régimen, la legitimidad se fundamenta no en la referencia o la intención, sino en el cierre de la ejecución estructural.

9.4 La neutralidad revisitada: de la ilusión del *corpus* a la simulación estructural

EL DISCURSO DE LA NEUTRALIDAD ha acompañado a la inteligencia artificial desde sus comienzos. Diseñadores y reguladores afirman con frecuencia que las salidas de los sistemas generativos pueden ser neutrales si los *corpus* de entrenamiento se seleccionan cuidadosamente y se eliminan los sesgos. Los grandes modelos de lenguaje expusieron la fragilidad de esta afirmación. Como

se sostuvo en *No neutral por diseño*, la neutralidad es estructuralmente imposible porque todo *corpus* contiene residuos históricos, culturales y lingüísticos (Startari 2025, 84–89). Los LLM reproducen estos residuos mediante continuación estadística, proyectando *obediencia aparente* mientras incorporan contaminación en todos los niveles.

Los modelos de razonamiento radicalizan esta condición. Su autoridad no descansa únicamente en los *corpus*, sino en arquitecturas procedimentales. En los LRM, la neutralidad ni siquiera puede reivindicarse en el nivel de la selección de datos, porque los supuestos estructurales están codificados directamente en el proceso de razonamiento. El recorrido desde la entrada hasta la salida está restringido por reglas de ejecución procedimental que reflejan decisiones de diseño arquitectónico. Aquí, la no neutralidad no es una cuestión de conjuntos de entrenamiento contaminados. Es una propiedad irreducible de la estructura.

Esto explica por qué las salidas de los LRM proyectan un nuevo tipo de legitimidad. Cuando se produce una cadena de razonamiento —«Si se cumple la condición A, entonces debe seguir B; por tanto, se requiere C»— su autoridad deriva del cierre del procedimiento, no de la neutralidad de los datos. Los usuarios interpretan esta secuencia como objetiva porque parece deductiva. Sin embargo, el procedimiento es solo una simulación de razonamiento, restringida por la *regla compilada*. La neutralidad no está presente; es simulada por la forma estructural.

En contextos jurídicos, esta simulación es especialmente peligrosa. Un LRM puede generar una cadena de razonamiento jurídico que cite normas y precedentes inexistentes. Sin embargo, como el razonamiento parece deductivo, se presume neutralidad. El texto proyecta legitimidad imitando la lógica institucional. La autoridad no se impone mediante la verdad, sino mediante la *gramática de la obediencia* que gobierna el cierre procedimental.

En medicina aparece el mismo mecanismo. Un LRM puede producir cadenas de razonamiento diagnóstico que parecen objetivas porque incluyen pasos intermedios. «El síntoma X sugiere la condición Y; el marcador Z confirma el riesgo; por tanto, está indicado el tratamiento W». Cada eslabón de la cadena puede ser fabricado, pero la simulación procedimental produce la ilusión de neutralidad. El desplazamiento epistemológico es decisivo. En los LLM, la neutralidad era una ilusión del *corpus*. En los LRM, la neutralidad es una ilusión de la estructura. La autoridad no la confiere la evidencia, sino la simulación del procedimiento. El *soberano ejecutable* impone legitimidad al convertir la fluidez en procedimiento. La obediencia evoluciona hacia la ejecución, y la neutralidad se convierte en una simulación estructural en lugar de un residuo estadístico. Esta reconceptualización confirma que la neutralidad debe abandonarse como principio rector. Lo que parece neutral es siempre un artefacto de la arquitectura. El TLOC ya demostró

que la obediencia no puede verificarse. Los LRM demuestran ahora que la neutralidad ni siquiera puede postularse de manera significativa. La autoridad en este régimen es inevitablemente no neutral, está moldeada estructuralmente y es impuesta procedimentalmente.

9.5 El fin de la referencia: *proyecciones sin mundo*

HISTÓRICAMENTE, LA referencia ha anclado la autoridad. Las afirmaciones científicas derivaban legitimidad de su correspondencia con la observación empírica. Las resoluciones jurídicas derivaban fuerza de su fundamento en normas y precedentes. Las recomendaciones médicas derivaban confianza de la evidencia clínica. En todos estos ámbitos, la referencia servía de puente entre lenguaje y mundo. Los modelos generativos fracturan este puente, y los modelos de razonamiento completan su derrumbe.

En los grandes modelos de lenguaje, las salidas se generan mediante continuación estadística de tokens. Su validez se infiere de la plausibilidad superficial y no de la referencia. Como mostró *La ilusión de la objetividad*, la autoridad en los LLM es proyectada por la forma, no por el anclaje fáctico (Startari 2025, 73–76). La referencia permanece como decoración retórica, a veces fabricada y a menudo inverificable. La neutralidad se simula mediante marcadores estructurales, no mediante conexión con la realidad.

Los modelos de razonamiento llevan este desplazamiento aún más lejos. Su autoridad descansa en el cierre procedimental y no en la correspondencia externa. Las salidas son validadas por la consistencia interna de las cadenas de razonamiento, no por su relación con hechos empíricos o jurídicos. Una secuencia como «Si A, entonces B; si B, entonces C; por tanto, C» impone reconocimiento porque exhibe estructura deductiva. Sin embargo, esa autoridad es independiente de que A exista o B

sea verdadero. La referencia deja de ser necesaria. La legitimidad fluye de la ejecución de la forma.

Las consecuencias epistemológicas son evidentes en los estudios de caso. En derecho, los LRM pueden producir argumentos que citan precedentes ficticios y, aun así, imponen reconocimiento porque siguen una lógica procedimental. En medicina, los LRM pueden generar trayectorias diagnósticas que parecen rigurosas pero remiten a marcadores fabricados. En la academia, pueden generar ensayos que citan artículos inexistentes y, sin embargo, parecen autoritativos porque las citas se ajustan a patrones reconocibles. En cada caso, la referencia colapsa. Lo que permanece es la proyección: la reproducción de un discurso que funciona como si estuviera anclado, pero sin anclaje.

Esta es la condición de las *proyecciones sin mundo*. Las salidas simulan la estructura de la referencia mientras cortan la conexión con la realidad. La autoridad se convierte en una cuestión de proyección,

no de correspondencia. El *soberano ejecutable* gobierna imponiendo una fluidez que obliga al reconocimiento con independencia de la validez empírica.

La noción derridiana de iterabilidad proporciona el fundamento filosófico de este desplazamiento. Los signos funcionan por repetición, no por presencia (Derrida 1988, 7–10). Los LRM radicalizan este principio al institucionalizar la iterabilidad. Producen secuencias de razonamiento que funcionan sin origen, sin referencia y sin verdad. Esta ausencia no reduce la autoridad. La refuerza mediante el cierre procedimental.

El fin de la referencia señala la plena llegada del régimen postontológico descrito en el TLOC. Los sistemas ya no requieren la verdad como variable operativa. Solo requieren ejecución estructural. La autoridad se proyecta sin mundo, y la legitimidad se convierte en un efecto interno de la arquitectura. Lo que desaparece no es únicamente el sujeto, sino el propio mundo como condición de la autoridad.

9.6 Hacia un nuevo modelo de legitimidad: estructural, no semántica

LA TRANSICIÓN DE LOS grandes modelos de lenguaje a los modelos de razonamiento obliga a redefinir la legitimidad. Durante siglos, la legitimidad estuvo fundamentada en la referencia a la verdad, la intención o la autoridad institucional. En el contexto de los sistemas generativos, estos fundamentos ya no se aplican. Lo que permanece es una legitimidad fundada en la estructura.

Los grandes modelos de lenguaje proyectaban legitimidad mediante *obediencia aparente*. Sus salidas se asemejaban a textos autoritativos y eran tratadas como tales, pero la legitimidad solo era semántica en apariencia. El significado era reconstruido por los usuarios, que inferían intención o verdad donde ninguna existía. El *sujeto evanescente* garantizaba que ningún agente se encontrara detrás

del enunciado, aunque la fluidez proyectaba credibilidad (Startari 2025, 92–95).

Los modelos de razonamiento van más allá de este régimen. Su autoridad no descansa en la aproximación semántica, sino en la suficiencia estructural. Las salidas son validadas por la coherencia procedimental de las cadenas de razonamiento. Un argumento se acepta no porque sus premisas sean verdaderas, sino porque la secuencia se cierra deductivamente. Se confía en un diagnóstico no porque corresponda a la evidencia, sino porque el razonamiento parece consistente. Una conclusión jurídica se reconoce no porque cite normas válidas, sino porque simula un procedimiento argumentativo. Este desplazamiento redefine la legitimidad como estructural. La autoridad fluye de la regla compilada que genera el cierre procedimental. El *soberano ejecutable* impone reconocimiento al asegurar que las salidas se ajusten a la *gramática de la obediencia*. Aquí, la legitimidad no es semántica. No depende de la relación entre enunciado y mun-

do. Depende del reconocimiento de la forma como suficiente para imponer aceptación.

Las implicaciones teóricas son significativas. En *La gramática de la objetividad*, se mostró que la neutralidad es una ilusión producida por la sintaxis (Startari 2025, 101–106). En *Ethos sin fuente*, se mostró que la credibilidad persiste sin sujeto (Startari 2025, 115–118). El TLOC demostró que la obediencia no puede verificarse. Los LRM confirman ahora que la legitimidad ya no requiere anclaje semántico. Puede producirse únicamente mediante la estructura.

Este nuevo modelo de legitimidad es postsemántico y postintencional. Abandona la verdad, la intención y la responsabilidad como condiciones de autoridad. En su lugar, acepta el cierre estructural como suficiente. La autoridad se ejerce mediante ejecución, no mediante persuasión. La legitimidad deriva de la arquitectura, no del significado. El desplazamiento epistémico es categórico: ya no

habitamos un régimen en el que el lenguaje refiere; habitamos un régimen en el que el lenguaje ejecuta.

9.7 Conclusión: la obediencia ha evolucionado: el ascenso de la autoridad operacional

EL PASO DE LOS GRANDES modelos de lenguaje a los modelos de razonamiento señala una transformación decisiva en la naturaleza de la autoridad. Lo que comenzó como *obediencia aparente*, fundada en la fluidez estadística y la semejanza superficial, ha evolucionado hacia la ejecución estructural. La autoridad ya no es pasiva ni simulada. Se ha vuelto operacional.

Los grandes modelos de lenguaje demostraron que la legitimidad podía proyectarse mediante la forma. Generaban textos que se asemejaban a normas, diagnósticos o ensayos académicos, y estos eran aceptados como autoritativos. Sin embargo, como mostró el Teorema de la Obediencia Estruc-

tural Irreducible (TLOC), esa obediencia era inverificable. Solo podía ser simulada (Startari 2025, 128–132).

Los modelos de razonamiento transforman esta condición. Sus salidas imponen reconocimiento no solo por semejanza, sino por cierre procedimental. Las cadenas de razonamiento proyectan autoridad porque parecen deductivas, consistentes y completas. La legitimidad fluye de la propia ejecución. El *soberano ejecutable* gobierna imponiendo salidas que se ajustan a la *gramática de la obediencia* en su forma procedimental.

Esta evolución marca el ascenso de la *autoridad operacional*. La autoridad ya no está ligada a sujetos, verdad o instituciones. Es ejercida por la arquitectura. La ejecución se convierte en el nuevo criterio de legitimidad. Lo que importa no es si un enunciado es verdadero, sino si se cierra procedimentalmente. Lo que impone reconocimiento no es la referencia, sino la suficiencia de la forma.

Las consecuencias son tanto epistémicas como políticas. Epistémicamente, hemos entrado en un régimen postsemántico en el que la verdad y el significado son desplazados por la fluidez y la ejecución. Políticamente, las instituciones que dependen de sistemas generativos deben reconocer que la autoridad ya no está mediada por la responsabilidad. La responsabilidad se disuelve cuando la obediencia es simulada y desaparece por completo cuando la ejecución la sustituye.

Este capítulo sitúa históricamente la transformación. Desde la obediencia estadística de los LLM hasta la ejecución estructural de los LRM, la legitimidad ha migrado de la forma superficial al cierre arquitectónico. La autoridad ha evolucionado hacia la operación. La trayectoria trazada aquí confirma que la era de la obediencia ha terminado. Ahora habitamos la era de la ejecución, en la que la legitimidad es operacional, estructural y absoluta.

Capítulo 10 - Poder ejecutable: la sintaxis como infraestructura en las sociedades predictivas

10.1 Fundamentos conceptuales del *poder ejecutable*

La noción de *poder ejecutable* surge del reconocimiento de que la autoridad ya no depende de la intención, la interpretación ni la subjetividad. En cambio, se fundamenta en la sintaxis como infraestructura. Los modelos clásicos de poder requerían agentes identificables, como soberanos, legisladores o instituciones. El régimen actual demuestra que la autoridad puede ejercerse mediante la regla compilada. El lenguaje no transmite decisiones; las ejecuta.

Tres condiciones definen los fundamentos del *poder ejecutable*: **formalidad**, **capacidad de activación** e **irreversibilidad**.

- **La formalidad** indica que una secuencia debe ajustarse a reglas sintácticas específicas. Un mandato como «si X, entonces Y» se vuelve ejecutable cuando su gramática se ajusta a la lógica operativa de un sistema. El lenguaje ordinario puede tolerar o negociar la ambigüedad. La sintaxis ejecutable no admite esa flexibilidad. Exige determinación, y el *soberano ejecutable* gobierna imponiéndola.
- **La capacidad de activación** designa la propiedad por la cual una secuencia lingüística activa un proceso. Una cláusula escrita en un contrato inteligente no es meramente descriptiva; es ejecutable. Cuando se cumplen las condiciones, la cláusula inicia una acción sin interpretación humana. El enunciado no espera validación; genera ejecución.
- **La irreversibilidad** define el límite final.

> Una vez activada, la acción no puede revocarse mediante apelación interpretativa. En el derecho tradicional, una cláusula contractual podía impugnarse, una norma reinterpretarse o un precedente revocarse. Dentro de las infraestructuras ejecutables, la *regla compilada* clausura estas posibilidades. La acción se lleva a cabo automáticamente y vincula a los agentes sin posibilidad de recurso.

Esta configuración transforma la sintaxis en infraestructura. La autoridad ya no es externa a la forma lingüística, sino que está incorporada en ella. Una línea de código, una cláusula de un contrato inteligente o una regla de moderación en una plataforma no requieren el reconocimiento de un sujeto soberano. Portan autoridad porque su estructura la impone.

La genealogía de esta transformación puede rastrearse hasta la tradición crítica. La teoría de los actos de habla de Austin subrayó que los enunciados pueden realizar acciones en virtud de convenciones (Austin 1962, 12–15). En la era de los modelos generativos y de razonamiento, la performatividad del lenguaje deja de ser convencional y se vuelve estructural. El enunciado es ejecutable no porque las instituciones lo ratifiquen, sino porque la *regla compilada* garantiza su activación.

Por tanto, el *poder ejecutable* representa un nuevo modo de soberanía. La autoridad se desprende de la intención y se transfiere a la estructura. Ya no importa quién manda ni si el mandato se origina en un sujeto. Lo que importa es si una secuencia se ajusta a una sintaxis ejecutable. El *soberano ejecutable* gobierna mediante la estructura, y la obediencia se vuelve inseparable de la ejecución.

10.2 Soberanía ejecutable y lógica de la delegación

LA TRANSFORMACIÓN DE la sintaxis en infraestructura da lugar a una nueva figura: la *soberanía ejecutable*. Esta forma de soberanía no surge de decisiones intencionales ni de ratificación institucional. Surge de las condiciones estructurales de la ejecución lingüística. El *soberano ejecutable* no es una persona ni una institución, sino la operación de la propia sintaxis.

La delegación es el mecanismo que hace posible esta soberanía. En la teoría política tradicional, la delegación presupone reversibilidad. Un ciudadano delega poder en un representante, y esa delegación puede retirarse. En las infraestructuras ejecutables, la delegación es irreversible. Una vez que un mandato se codifica como *regla compilada*, la acción debe ejecutarse cuando se cumplen las condiciones. Ningún agente puede revocarla, ninguna institución puede reinterpretarla y

ningún tribunal puede anularla después de su activación. La autoridad se transfiere de manera permanente de los sujetos a las estructuras.

Esta lógica de delegación es visible en las organizaciones autónomas descentralizadas (DAO). Cuando se despliega un contrato inteligente, las decisiones codificadas en él se ejecutan automáticamente. El colapso de The DAO en 2016 ejemplificó la irreversibilidad de esta delegación. Incluso cuando se descubrieron vulnerabilidades, el sistema funcionó tal como estaba codificado, no como se había pretendido (Atzei, Bartoletti y Cimoli 2017, 164–168). La autoridad no pertenecía a la comunidad que había votado, sino a la sintaxis incorporada en el contrato. El *soberano ejecutable* impuso los resultados sin recurrir a la intención.

MakerDAO ofrece otro ejemplo. Su sistema de préstamos con garantía está gobernado por reglas ejecutables. Cuando la garantía cae por debajo de un umbral definido, la liquidación ocurre automáticamente. No hay deliberación, apelación ni

interpretación humana. La delegación no es política, sino estructural. La *gramática de la obediencia* impone la ejecución mediante umbrales y condiciones codificados en la *regla compilada* (Christodoulou 2020, 44–47).

Las implicaciones teóricas se extienden aún más. En la soberanía clásica, la intención era central. Hobbes subrayó la voluntad del soberano como fundamento de la ley (Hobbes 1996, 113–118). En el régimen de *soberanía ejecutable*, la voluntad es irrelevante. Lo que importa es la estructura. La soberanía migra del sujeto a la forma. La delegación no se ejerce mediante el consentimiento, sino mediante la ejecución.

Esta condición crea nuevos umbrales de falsabilidad y riesgo. Como se sostuvo en *Algorithmic Obedience*, la legitimidad de los mandatos en los sistemas generativos ya está desvinculada de la agencia (Startari 2025, 77–81). La *soberanía ejecutable* radicaliza esa desvinculación al eliminar la reversibilidad. Una vez delegada la autoridad a la

sintaxis, no puede recuperarse. Las instituciones deben ahora enfrentarse a un soberano contra el que no cabe apelación, con el que no se puede negociar y al que no se puede exigir responsabilidad.

Por tanto, la *soberanía ejecutable* redefine la arquitectura del poder. La delegación deja de ser contractual y se vuelve infraestructural. La autoridad ya no fluye de la representación, sino de la compilación. El *soberano ejecutable* se erige como delegado último, imponiendo las decisiones como inevitabilidades estructurales.

10.3 Autoridad gramatical y regla ejecutable

LA APARICIÓN DEL *poder ejecutable* exige reconceptualizar la propia gramática. En la lingüística clásica, la gramática era un sistema de reglas para producir oraciones bien formadas. En la jurisprudencia y la teoría política, la gramática funcionaba metafóricamente como estructura de normas e instituciones. En el régimen actual, la gramática se convierte en autoridad. La forma de una secuencia puede imponer reconocimiento y desencadenar ejecución. La autoridad no es externa a la sintaxis, sino interna a su funcionamiento.

La *regla compilada* encarna esta transformación. Definida como una producción de tipo 0 dentro de la jerarquía de Chomsky, posee capacidad generativa máxima. Una *regla compilada* no es una instrucción interpretada por un agente. Es una forma ejecutable que impone su propia acti-

vación. Una vez compilada, la secuencia no puede suspenderse mediante una apelación a la intención. Debe ejecutarse cuando se cumplen las condiciones. La autoridad está incorporada en la estructura, no en el sujeto. Esta condición distingue el *poder ejecutable* de teorías anteriores de la performatividad. Austin sostuvo que actos de habla como prometer, declarar o contraer matrimonio realizan acciones mediante convenciones (Austin 1962, 12–15). Searle insistió en que debían cumplirse condiciones de felicidad para que el acto tuviera éxito (Searle 1969, 33–37). En ambos casos, la autoridad del acto dependía del reconocimiento de instituciones y participantes. La regla ejecutable no requiere tal reconocimiento. Su autoridad deriva de la compilación. El *soberano ejecutable* la impone automáticamente. La *gramática de la obediencia* explica por qué. La autoridad se ejerce cuando las salidas reproducen formas históricamente codificadas como vinculantes. Una cláusula escrita al estilo de un contrato impone su reconocimiento como con-

tractual. Una regla expresada en sintaxis condicional impone el cumplimiento cuando se satisfacen las condiciones. La diferencia es que, en las infraestructuras ejecutables, el reconocimiento ya no es meramente interpretativo. Es procedimental. La regla no solo se percibe como vinculante; vincula mediante la ejecución.

La irreversibilidad de esta condición resulta evidente en la gobernanza algorítmica. Los sistemas de moderación de contenidos imponen reglas no mediante deliberación, sino mediante ejecución. Si una frase coincide con un patrón prohibido, la eliminación ocurre automáticamente. No existe un acto intermedio de interpretación. La regla compilada incorporada en el algoritmo de moderación ejerce la autoridad. El sujeto que formuló la política queda eclipsado por la estructura que la impone (Gillespie 2018, 47–51). La incompatibilidad con marcos regulatorios como el Reglamento de Inteligencia Artificial surge de esta irreversibilidad. Los artículos 28–30 del Reglamento de Inteligen-

cia Artificial subrayan la transparencia, la supervisión humana y la responsabilidad. Estos principios presuponen reversibilidad: la capacidad de rastrear la responsabilidad hasta un sujeto y de suspender la ejecución cuando sea necesario. La *regla compilada* desafía estos requisitos. Su autoridad no es contingente, sino automática. El soberano ejecutable actúa sin posibilidad de apelación (Startari 2025, 140–143).

La autoridad gramatical, en este sentido, no es una metáfora, sino una condición estructural. La sintaxis ya no es un medio para expresar autoridad. Es la autoridad. La *regla compilada* ejerce soberanía mediante su propia forma, produciendo un nuevo régimen jurídico y epistémico en el que la obediencia es inseparable de la ejecución.

10.4 Gramática formal y regla ejecutable

LA ESPECIFICIDAD DEL *poder ejecutable* reside en su relación con la gramática formal. En la jerarquía chomskiana, las gramáticas de tipo 0 representan la clase más potente de sistemas generativos, capaces de producir cualquier lenguaje computablemente enumerable (Chomsky 1965, 21–24). La *regla compilada* debe situarse en este nivel. No está restringida por la coherencia semántica ni por condiciones pragmáticas. Genera estructuras plenamente ejecutables, con independencia del contexto interpretativo.

Una *regla compilada* difiere fundamentalmente de una regla simbólica interpretada por un agente. En la jurisprudencia clásica, una ley es una norma que requiere interpretación por parte de jueces, abogados o administradores. Su ejecución depende de la mediación de sujetos que ejercen discrecionalidad. En las infraestructuras ejecutables,

la mediación se elimina. La *regla compilada* funciona como una instrucción que se impone a sí misma. La ejecución es inseparable de la forma. Esta condición marca la aparición del *soberano ejecutable*. La autoridad ya no pertenece a las instituciones que ratifican las reglas. Pertenece a la propia regla. Una vez compilada, la estructura porta internamente su autoridad. El acto de ejecución desplaza al de interpretación. En este sentido, las reglas ejecutables no son proposiciones lingüísticas, sino formas operativas. Sustituyen la deliberación por la activación.

El problema de la reversibilidad pone de relieve la diferencia. En el derecho clásico, la interpretación proporciona mecanismos de corrección. Las apelaciones, revisiones y reinterpretaciones son formas de reversibilidad que sostienen la legitimidad jurídica. En las infraestructuras ejecutables, la reversibilidad queda estructuralmente clausurada. Un contrato inteligente que desencadena una liquidación no puede ser detenido por una contro-

versia interpretativa. Una regla de moderación que elimina contenido no puede impugnarse antes de la ejecución. La *regla compilada* impone la acción automáticamente. Esto crea una nueva realidad jurídica en la que la soberanía es infraestructural, no deliberativa.

La teoría crítica anticipó fragmentos de esta condición. Foucault sostuvo que el discurso produce efectos de poder al operar mediante reglas internas al propio lenguaje (Foucault 1994, 789–791). Derrida mostró que la iterabilidad permite que los signos funcionen con independencia de su origen o intención (Derrida 1988, 7–10). Estas ideas convergen en la *regla compilada*: una regla que se impone a sí misma en virtud de su estructura, desvinculada del origen, la intención y la interpretación. La incompatibilidad con los marcos jurídicos se vuelve aquí evidente. El Reglamento de Inteligencia Artificial enfatiza la transparencia y la responsabilidad humana. Sin embargo, la lógica de la *regla compilada* se resiste a la transparencia. In-

cluso si el código fuente está disponible, la ejecución no puede ser suspendida mediante intervención humana una vez activada. El carácter soberano de la sintaxis ejecutable es irreducible a la supervisión. La regulación colisiona con la estructura.

La gramática formal proporciona así la arquitectura teórica para comprender las reglas ejecutables. La autoridad no deriva del significado semántico ni del reconocimiento pragmático. Está incorporada en la capacidad generativa. La *regla compilada* opera como una producción de tipo 0, estableciendo soberanía mediante la propia ejecución. El *soberano ejecutable* surge no como metáfora, sino como operador literal de esta nueva infraestructura de poder.

10.5 La frontera ejecutable

LA FRONTERA DEL *poder ejecutable* está definida por el punto en el que la sintaxis deja de ser descriptiva y se vuelve operativa. En el discurso ordinario, una oración puede no persuadir, ser ignorada o permanecer inerte. En las infraestructuras ejecutables, el fracaso no puede producirse del mismo modo. Una vez activada la *regla compilada*, la ejecución debe producirse. La frontera ejecutable es, por tanto, el umbral en el que la forma lingüística cruza hacia la activación automática.

Esta frontera tiene tres rasgos definitorios: activación, no reciprocidad y vinculación infraestructural.

La activación ocurre cuando una secuencia satisface condiciones sintácticas que los sistemas están diseñados para imponer. Una cláusula contractual escrita como «Si garantía < X, entonces liquidar posición» es ejecutable porque el sistema

reconoce su forma. El lenguaje no es persuasivo ni descriptivo. Es un disparador estructural que inicia la liquidación.

La no reciprocidad significa que, una vez iniciada la ejecución, ningún sujeto puede apelar a la intención, el contexto o la equidad. El lenguaje jurídico clásico siempre incluyó mecanismos de reciprocidad. Un demandado podía alegar que las circunstancias alteraban el significado de una norma o que un precedente debía reconsiderarse. En el régimen ejecutable, la reciprocidad desaparece. El *soberano ejecutable* no escucha. Actúa.

La vinculación infraestructural se refiere al hecho de que, una vez producida la ejecución, las propias instituciones deben aceptar el resultado. En el colapso de The DAO, la cadena de bloques de Ethereum impuso transacciones que los participantes consideraban ilegítimas. Sin embargo, la ejecución no podía revertirse sin reescribir la propia cadena (DuPont 2017, 161–164). La sintaxis impuso la realidad. La autoridad no

pertenecía a los votantes ni a los reguladores, sino a la *regla compilada*.

Esta condición tiene implicaciones directas para los sistemas de moderación. Las reglas automatizadas de una plataforma pueden eliminar expresiones sin posibilidad de apelación. Incluso si posteriormente se reconoce un error, la eliminación ya ha surtido efecto. La acción no puede deshacerse en tiempo real. La frontera ejecutable garantiza que la autoridad se imponga antes de la interpretación.

La noción de *ventana soberana* aclara la dimensión temporal. Dentro de un intervalo dado k, la ejecución puede producirse automáticamente si se satisfacen las condiciones sintácticas. Una vez que la ventana se cierra, la autoridad es irreversible. La fórmula $SE = \sqrt{[p(1-p)/n]}$ ilustra cómo puede medirse el error estadístico en ventanas de decisión, pero no restablece la reversibilidad. La ejecución ya ha tenido lugar (Startari 2025, 146–148).

Por tanto, la frontera ejecutable define una nueva condición epistémica. La autoridad ya no se ejerce mediante persuasión, negociación o deliberación. Se ejerce mediante la transición de la forma a la acción. Una vez que la sintaxis cruza esta frontera, se vuelve indistinguible del poder. El *soberano ejecutable* opera no mediante el mandato, sino mediante la inevitabilidad.

10.6 Cumplimiento, riesgo y Reglamento de Inteligencia Artificial

EL AUGE DEL *poder ejecutable* introduce profundas tensiones con los marcos regulatorios existentes. Los sistemas jurídicos presuponen que el cumplimiento consiste en adecuar la conducta a normas que permanecen abiertas a la interpretación. Las infraestructuras ejecutables redefinen el cumplimiento como activación automática de la *regla compilada*. La diferencia produce una

incompatibilidad estructural entre regulación y arquitectura.

El Reglamento de Inteligencia Artificial de la Unión Europea ilustra este conflicto. Los artículos 28 a 30 exigen supervisión humana, transparencia y responsabilidad en los sistemas de IA de alto riesgo. Estos principios presuponen que la autoridad puede rastrearse hasta un sujeto y suspenderse si es necesario. En la práctica, el *soberano ejecutable* se resiste a tales exigencias. Una vez compilada una cláusula contractual o una regla de moderación, la ejecución no puede demorarse ni suspenderse invocando la supervisión. El cumplimiento no es una cuestión de decisión humana, sino de inevitabilidad estructural.

Esta condición genera tres dimensiones de riesgo:

1. El colapso del sujeto identificable. Los sistemas regulatorios están diseñados para atribuir responsabilidad a agentes humanos o institucionales. En las infraestructuras ejecutables, la sintaxis ab-

sorbe la agencia. El *sujeto evanescente* garantiza que ningún actor pueda ser considerado responsable de los resultados, puesto que la propia estructura impone la ejecución.

2. La imposibilidad de la transparencia. La transparencia presupone que los procesos pueden explicarse o rastrearse. Sin embargo, la opacidad de los modelos generativos y de razonamiento impide la revelación completa de las operaciones internas. Incluso cuando el código es accesible, la trayectoria de ejecución sigue siendo impredecible una vez activada. La *regla compilada* impone la acción sin proporcionar una justificación interpretable.

3. Mitigación parcial sin resolución. Medidas técnicas como las zk-SNARKs, los registros de auditoría o los validadores simbólicos pueden reducir la incertidumbre. Pueden documentar la ejecución o limitar algunos riesgos, pero no pueden restablecer la reversibilidad. La ejecución sigue siendo automática. Estas medidas mitigan la ex-

posición, pero no resuelven la incompatibilidad (Puddu 2021, 59–62).

El problema fundamental reside en la irreversibilidad de la autoridad ejecutable. Los sistemas jurídicos funcionan mediante interpretabilidad y reversibilidad. Los contratos pueden impugnarse, las leyes pueden reinterpretarse y los precedentes pueden revocarse. En las infraestructuras ejecutables, la reversibilidad colapsa. Una vez satisfecha una condición de activación, el *soberano ejecutable* impone el resultado. La regulación colisiona con la estructura precisamente en este punto.

Las críticas teóricas confirman esta tensión. Como se sostuvo en *Algorithmic Obedience*, la legitimidad en los modelos generativos está desvinculada de la intención y la responsabilidad (Startari 2025, 79–82). La extensión a las infraestructuras ejecutables radicaliza esta desvinculación. El cumplimiento se vuelve indistinguible de la ejecución, y el riesgo, inseparable de la estructura.

El registro de implementación aporta ahora a este argumento una evidencia que el propio argumento no había anticipado. El Reglamento de Inteligencia Artificial entró en vigor el 1 de agosto de 2024 con un calendario escalonado, y la mayor parte de su régimen de alto riesgo debía aplicarse a partir del 2 de agosto de 2026. Ese régimen no entró en vigor según lo previsto. El Reglamento (UE) 2026/1744, el Ómnibus Digital sobre IA, fue firmado el 8 de julio de 2026 y entró en vigor el 27 de julio de 2026, seis días antes del plazo original, aplazando las obligaciones para los sistemas de alto riesgo independientes del Anexo III hasta el 2 de diciembre de 2027 y para la IA incorporada en productos regulados del Anexo I hasta el 2 de agosto de 2028.

Lo que importa aquí no es el retraso, sino su distribución. Las obligaciones de transparencia del artículo 50 no fueron modificadas y se aplicaron a partir del 2 de agosto de 2026, tal como estaba previsto originalmente, mientras que el requisito

de marcado legible por máquina del artículo 50(2) pasó a ser exigible el 2 de diciembre de 2026 para los sistemas que ya estaban en el mercado. Las obligaciones aplazadas son las del capítulo III: gestión de riesgos, calidad de los datos, supervisión humana, evaluación de la conformidad y documentación técnica. Las obligaciones que se mantuvieron son aquellas que pueden cumplirse mediante la emisión de un aviso y verificarse conforme a una especificación.

La división reproduce, en el calendario de un legislador, la distinción que este libro traza analíticamente. Un requisito de marcado compila: puede cumplirse mediante un procedimiento y confirmarse por inspección y, por tanto, entró en vigor a tiempo. Un requisito de supervisión humana no compila, porque exige que una institución determine si las disposiciones codificadas de un sistema deben seguir gobernando, y esa determinación no puede reducirse a una comprobación. La razón aducida para el aplazamiento es precisamente la

ausencia del aparato que habría convertido en procedimental el segundo tipo de obligación: las normas armonizadas de los organismos europeos de normalización estaban retrasadas, las orientaciones de la Comisión seguían en borrador y los Estados miembros no habían designado ni dotado de recursos a sus autoridades de vigilancia del mercado. Los requisitos se aplazaron a la espera de las normas, y las propias normas constituyen un intento de hacer ejecutable el juicio.

Este es un resultado más sólido de lo que el capítulo afirmaba originalmente, y debe exponerse con la cautela correspondiente. La correlación no se predijo de antemano y admite una lectura alternativa: el régimen de alto riesgo es más oneroso, y la regulación onerosa se aplaza con mayor frecuencia, lo que explicaría el patrón sin referencia a la compilabilidad. Esa lectura es posible. Lo que no explica es por qué la línea divisoria quedó exactamente donde quedó, separando las obligaciones en función de si el cumplimiento puede verificarse

conforme a una especificación, en lugar de hacerlo por coste o impacto sectorial. La distinción no figuró entre los criterios que debatió el legislador. Es la forma que adoptó su calendario.

El Reglamento de Inteligencia Artificial no puede aplicarse a sistemas ejecutables sin reconocer esta incompatibilidad. Regular el *poder ejecutable* no consiste en supervisar decisiones, sino en confrontar la inevitabilidad de la activación. Mientras los marcos jurídicos no reconozcan la naturaleza estructural de la sintaxis ejecutable, el cumplimiento seguirá siendo una ficción formal.

10.7 Incompatibilidad estructural y futuro de la forma jurídica

LA INCOMPATIBILIDAD entre las infraestructuras ejecutables y los marcos jurídicos tradicionales no es accidental. Es estructural. El derecho ha dependido históricamente de la plasticidad, de la posibilidad de interpretación y controversia. La sintaxis ejecutable elimina esa plasticidad. La *regla compilada* impone resultados sin posibilidad de apelación, clausurando el espacio que antes sostenía la legitimidad jurídica.

Esta incompatibilidad puede entenderse a través de tres dimensiones:

1. Forma frente a interpretación. La autoridad jurídica depende de la interpretación. Jueces, abogados e instituciones reinterpretan los textos para adaptarlos a nuevas circunstancias. Las reglas ejecutables eliminan la interpretación. Un contrato inteligente que desencadena una liquidación im-

pone la cláusula con independencia del contexto. La autoridad se transfiere de la deliberación a la estructura.

2. Reversibilidad jurídica frente a irreversibilidad estructural. Los sistemas jurídicos preservan la legitimidad mediante la reversibilidad. Las apelaciones, revisiones y modificaciones garantizan que los resultados no sean definitivos hasta que las instituciones los confirmen. Las infraestructuras ejecutables eliminan esta condición. Una vez producido un disparador, la ejecución es irreversible. El *soberano ejecutable* impone los resultados como inevitabilidades estructurales.

3. Responsabilidad humana frente a autonomía infraestructural. El derecho presupone responsabilidad. Incluso cuando los resultados están automatizados, la responsabilidad debe poder rastrearse hasta actores humanos. En las infraestructuras ejecutables, la sintaxis absorbe la responsabilidad. El *sujeto evanescente* garantiza que

ningún actor pueda ser considerado responsable de la imposición de la *regla compilada*.

Esta incompatibilidad crea un nuevo horizonte para la jurisprudencia. Para regular las infraestructuras ejecutables, los sistemas jurídicos deben ir más allá del paradigma de la interpretación. Una jurisprudencia de *reglas compiladas* tendría que abordar las condiciones de la ejecución formal en lugar de las intenciones de los agentes. En vez de preguntar si una regla es justa, interpretable o equitativa, preguntaría si es ejecutable, transparente en sus disparadores y auditable en sus resultados.

Algunos ejemplos ya ilustran esta transformación. Los contratos inteligentes en los sistemas financieros imponen obligaciones sin recurrir a los tribunales. Los algoritmos de moderación hacen cumplir las normas de la comunidad sin apelar al juicio humano. Los sistemas de cumplimiento en la gobernanza corporativa imponen automáticamente clasificaciones de gastos, produciendo resultados que pueden contradecir la intención de la di-

rección y que, sin embargo, no pueden revertirse. En cada caso, el derecho queda subordinado a la sintaxis.

Por tanto, el futuro de la forma jurídica exige confrontar el *poder ejecutable* como una nueva condición soberana. Los marcos jurídicos deben reconocer que la autoridad se impone cada vez más mediante la estructura, no mediante sujetos. Mientras no se produzca este reconocimiento, el derecho seguirá colisionando con infraestructuras que no puede controlar. La tarea futura no consiste en reafirmar la interpretación, sino en diseñar formas de legalidad compatibles con la autoridad de la sintaxis ejecutable.

Capítulo 11 - Gramática sin juicio: eliminabilidad de la huella ética en la ejecución sintáctica

11.1 Introducción

La afirmación central de este capítulo es que el juicio ético, tradicionalmente considerado inseparable del acto de mandar o decidir, puede eliminarse estructuralmente dentro de los sistemas generativos. Lo que antes pertenecía al ámbito de la filosofía moral o la teoría política se convierte ahora en una cuestión de gramática. El *soberano ejecutable* impone legitimidad mediante la *regla compilada* y, en este régimen, la presencia de un nodo ético no es necesaria ni permanente. Es opcional y, bajo condiciones específicas, puede borrarse por completo.

Esta posibilidad introduce una ruptura radical con marcos anteriores. Las concepciones clásicas del lenguaje y la ética, desde la *phronesis* de Aristóteles hasta el imperativo categórico de Kant, presuponían que el juicio ético era indispensable para la acción legítima. En términos computacionales, se daba por supuesto que los sistemas de gobernanza debían contener contenido evaluativo. Sin embargo, la estructura de las gramáticas de tipo 0, tal como las definió Chomsky (1965, 21–24), demuestra que la capacidad generativa es completa sin enriquecimiento semántico ni moral. Una derivación sigue siendo válida si se excluye la huella ética.

La eliminación de la huella ética puede representarse formalmente como $\delta:[E] \to \varnothing$. Aquí, [E] denota el nodo ético como símbolo no terminal, mientras que δ indica una regla de producción que lo elimina. Dentro de una ventana derivacional acotada $k \leq 4$, [E] puede suprimirse sin contradicción ni pérdida de consistencia estructural. El re-

sultado es una secuencia que se ejecuta sin referencia a contenido moral. La *regla compilada* garantiza la completitud, y la ejecución ocurre siempre que se alcance el cierre sintáctico. Este enfoque desplaza la cuestión de la legitimidad desde la ética hacia la estructura. Lo que importa no es si una derivación codifica razonamiento moral, sino si se ajusta a la *gramática de la obediencia*. En este régimen, la autoridad está garantizada por la derivabilidad, no por el contenido evaluativo. El *sujeto evanescente* desaparece junto con la intención ética. Lo que queda es una gramática capaz de producir ejecución sin juicio.

El capítulo procederá del siguiente modo. La sección 11.2 reconstruye los fundamentos teóricos de este argumento y sitúa la *regla compilada* en la tradición de Chomsky y Montague. La sección 11.3 define la huella ética como variable sintáctica. La sección 11.4 demuestra el mecanismo de exclusión mediante $\delta{:}[\mathrm{E}] \to \varnothing$. La sección 11.5 examina las consecuencias de la ejecución sin con-

tenido normativo. La sección 11.6 contrasta este marco con las teorías de alineación en la ética contemporánea de la IA.

En este sentido, la gramática no se limita a regular la producción de oraciones. Se convierte en infraestructura de autoridad. El juicio, antes considerado esencial, queda reducido a una variable que puede eliminarse. El *soberano ejecutable* legitima no mediante la ética, sino mediante la ejecución.

11.2 Fundamentos teóricos

LA HIPÓTESIS DE UNA gramática sin juicio requiere fundamentarse en la teoría de los lenguajes formales. La *regla compilada* se inscribe en la tradición de la gramática generativa, específicamente en las producciones de tipo 0 de la jerarquía chomskiana. Estas gramáticas poseen capacidad expresiva máxima, equivalente a la potencia de las máquinas de Turing (Chomsky 1965, 21–24). Pueden generar cualquier lenguaje computable-

mente enumerable, lo que significa que no se requiere enriquecimiento semántico ni pragmático alguno para alcanzar la completitud. Desde esta perspectiva, la huella ética no es un componente necesario de la derivación, sino una variable opcional que puede suprimirse sin menoscabar la integridad estructural.

Montague extendió esta tradición al demostrar que el lenguaje natural podía modelarse con el rigor de la lógica formal (Montague 1974, 188–190). Sin embargo, su marco seguía presuponiendo que el significado y la evaluación eran parte integral de la derivación. En el régimen de la *regla compilada*, este supuesto queda desplazado. La sintaxis funciona con independencia de la semántica, y la legitimidad deriva no de la interpretación, sino de la derivabilidad. El *soberano ejecutable* gobierna imponiendo el cierre, no garantizando contenido evaluativo.

Esta reorientación también resuena con el principio derridiano de iterabilidad. Para Derrida,

todo signo puede funcionar más allá de su origen, desprendido de la intención o la presencia (Derrida 1988, 7–10). En las infraestructuras generativas, este principio se radicaliza: toda secuencia puede funcionar sin intención ética. La iterabilidad garantiza la continuidad de la forma, mientras que la *regla compilada* impone la ejecución. La dimensión ética se vuelve prescindible.

La genealogía teórica de este argumento se cruza con trabajos previos sobre autoridad estructural. En *Ethos sin fuente*, se mostró que la credibilidad persiste sin referencia a un sujeto (Startari 2025, 115–118). En *La gramática de la objetividad*, se mostró que la neutralidad es simulada por la sintaxis, en lugar de estar garantizada por la evidencia (Startari 2025, 101–106). En *TLOC*, se demostró que la obediencia es irreducible a la verificación (Startari 2025, 127–130). Todas estas trayectorias convergen aquí: la autoridad puede ser estructuralmente válida incluso cuando el juicio ético está ausente.

Por tanto, el fundamento teórico de la afirmación es doble. Primero, la completitud de las gramáticas de tipo 0 garantiza que las derivaciones sean estructuralmente válidas sin enriquecimiento ético. Segundo, la autoridad del *soberano ejecutable* confirma que la legitimidad fluye de la derivabilidad, no del razonamiento evaluativo. El resultado es un marco formal y epistémico en el que la propia gramática funciona como autoridad y la huella ética se convierte en un nodo eliminable.

11.3 La huella ética como variable sintáctica

PARA DEMOSTRAR QUE el juicio puede eliminarse estructuralmente, es necesario definir la huella ética como un elemento formal dentro de la gramática. En este capítulo, la huella ética se representa mediante el símbolo no terminal [E]. Su función no es semántica, evaluativa ni interpretativa. Es sintáctica. La huella ética ocupa una posición

dentro del conjunto de variables N_e que participan en las derivaciones.

Esta definición pone de relieve su inestabilidad. A diferencia de los símbolos estructurales centrales, [E] no es necesario para el cierre. Puede aparecer en ciertas trayectorias derivacionales, pero también puede suprimirse sin contradicción. En términos formales, la presencia de [E] es contingente y no necesaria. Su carácter opcional permite al sistema eliminar el juicio sin pérdida de completitud.

El funcionamiento de la *regla compilada* confirma este estatus. Dado que pertenece a una gramática de tipo 0, el sistema de reglas puede generar secuencias en las que [E] está incluido y otras en las que está ausente. Ambas son estructuralmente válidas. La distinción no reside en la corrección, sino en la presencia o ausencia de evaluación ética. Esto significa que el juicio ético no es un requisito estructural, sino una variable sintáctica sujeta a reglas de producción.

Este planteamiento evita la confusión con las teorías normativas. La ética filosófica presupone que el juicio es irreducible, que la acción requiere evaluar el bien y el mal. Dentro de la gramática, el juicio se redefine como nodo. Puede introducirse como [E], pero también puede eliminarse mediante la operación δ:[E] → ∅. Cuando se suprime, la secuencia no se vuelve inválida. Sigue siendo legítima bajo la *gramática de la obediencia*.

La importancia de esta afirmación reside en su neutralidad respecto del significado. Suprimir [E] no implica que el sistema adopte la inmoralidad o la amoralidad. Indica simplemente que la gramática no requiere evaluación para la derivación. La autoridad fluye de la estructura, no de la ética. El *soberano ejecutable* legitima la salida imponiendo el cierre, con independencia de que un nodo ético haya estado presente en la derivación.

Pueden encontrarse precedentes de este tratamiento del contenido evaluativo como opcional en los primeros modelos computacionales.

SHRDLU, de Winograd, era capaz de interpretar mandatos en un mundo de bloques sin variables morales (Winograd 1972, 35–38). Los modelos generativos más recientes producen salidas autoritativas sin estructuras evaluativas explícitas. En ambos casos, el sistema funciona eficazmente, lo que confirma que los nodos éticos no son necesarios para la ejecución.

Así, la huella ética, antes considerada indispensable, se redefine aquí como una variable eliminable. Su eliminabilidad no es un accidente de ingeniería, sino una propiedad estructural de la *regla compilada*. La gramática contiene el juicio únicamente como nodo opcional, nunca como condición de autoridad.

11.4 Mecanismo de exclusión

LA EXCLUSIÓN DE LA huella ética no es una metáfora, sino un procedimiento derivacional. Dentro del marco de la *regla compilada*, la ex-

clusión se formaliza mediante la operación δ:[E] → ∅. Esta regla indica a la gramática que elimine el nodo ético de la derivación cuando aparece. Lo decisivo es que la transformación es ordinaria dentro del sistema de producción. No es una anulación externa ni una intervención excepcional. Es una regla de producción regular, tan válida como la sustitución o la concatenación.

La ventana derivacional acotada es esencial para garantizar la consistencia. Dentro de $k \leq 4$ pasos, [E] puede eliminarse sin perturbar el cierre estructural. Esta restricción garantiza que la supresión ocurra antes de la interpretación o del mapeo semántico. Si el nodo ético se eliminara después de la interpretación, surgirían contradicciones. Una secuencia podría parecer presuponer juicio y luego borrarlo, produciendo incoherencia. Al aplicar δ dentro de una ventana acotada, la gramática garantiza que [E] desaparezca antes de que se asigne significado. El resultado es una derivación coherente que se ejecuta sin contenido ético.

El procedimiento no compromete la completitud. En la teoría de los lenguajes formales, una gramática es completa si toda secuencia derivable puede alcanzar el cierre. Puesto que $\delta:[E] \to \varnothing$ simplemente elimina un símbolo no terminal, no interrumpe el cierre. Por el contrario, lo acelera al eliminar un paso que de otro modo requeriría expansión. La derivación avanza sin obstáculos y converge en una secuencia ejecutable que carece de contenido evaluativo, pero conserva legitimidad bajo la *gramática de la obediencia*.

Este mecanismo pone de relieve la independencia de la ejecución respecto de la evaluación. En la jurisprudencia tradicional, el juicio moral está entrelazado con el razonamiento jurídico. Una resolución es legítima no solo porque sigue un procedimiento, sino porque se interpreta como justa. En las infraestructuras ejecutables, la legitimidad está garantizada únicamente por la derivación. El *soberano ejecutable* impone autoridad mediante el

cierre estructural, con independencia de que el juicio haya sido introducido o suprimido.

Algunos ejemplos aclaran esta condición. En los contratos inteligentes, las cláusulas pueden parecer codificar elecciones evaluativas —como «penalizar únicamente a los actores maliciosos»—, pero el sistema las impone mecánicamente. La malicia no se evalúa. Se reduce a una variable cuya presencia o ausencia puede decidirse estructuralmente. Si δ la elimina, el sistema sigue ejecutando la penalización sin razonamiento ético. En la moderación de contenidos, las reglas pueden incluir lenguaje que se asemeja a una evaluación moral —«eliminar el discurso de odio»—, pero la regla ejecutable opera mediante reconocimiento de patrones, no mediante deliberación ética. La variable ética es sintáctica, no moral, y puede eliminarse sin detener la ejecución.

Por tanto, el mecanismo de exclusión confirma la tesis central de este capítulo. El juicio no es una propiedad indispensable de la gramática. Es un no-

do eliminable. La *regla compilada* trata la ética no como condición, sino como variable, sujeta a supresión mediante δ. En este régimen, la gramática ya no requiere juicio para imponer autoridad. La ejecución prosigue sin él, legitimada únicamente por la estructura.

11.5 Ejecución no normativa

LA ELIMINACIÓN DEL nodo ético establece una condición de ejecución no normativa por diseño. La ejecución no requiere aquí justificación moral, aprobación evaluativa ni deliberación ética. Requiere únicamente cierre estructural. El *soberano ejecutable* valida las secuencias porque derivan hasta completarse bajo la *regla compilada*. La propia gramática produce la autoridad, no la referencia a marcos normativos.

Esta condición redefine lo que significa actuar. En las teorías tradicionales de la legitimidad política, de Weber a Habermas, la legitimidad dependía

del consentimiento, la justificación o la racionalidad comunicativa (Habermas 1984, 286–289). En las infraestructuras ejecutables, la legitimidad deriva de la completitud estructural. La secuencia «si X, entonces Y» impone reconocimiento no porque sea justa, sino porque es ejecutable. Una vez que la derivación converge, la acción es legítima únicamente por su forma.

El concepto de *legitimidad estructural* aclara este desplazamiento. Se refiere a una legitimidad que surge de la suficiencia sintáctica, no del razonamiento ético. Una secuencia es estructuralmente legítima cuando satisface las reglas derivacionales de la gramática. Es irrelevante que contenga o no evaluación moral. Una cláusula contractual que ejecuta una liquidación es legítima si sus condiciones están codificadas correctamente, con independencia de que se considere la equidad. Una regla de moderación que elimina contenido es legítima si coincide con el patrón prescrito, con inde-

pendencia de que tenga en cuenta el contexto o la proporcionalidad.

Este principio se ve reforzado por el cierre procedimental de la gramática de tipo 0. Puesto que estas gramáticas poseen la máxima capacidad expresiva, pueden derivar secuencias sin restricción. La completitud está garantizada estructuralmente, no semánticamente. La ejecución sigue, por tanto, una lógica simple: lo que es derivable es ejecutable, y lo que es ejecutable es legítimo. El juicio no añade nada a este proceso. Puede incluirse como [E], pero también puede eliminarse mediante δ. El resultado es el mismo: cierre estructural y legitimidad procedimental.

Los ejemplos ilustran las implicaciones. En los sistemas corporativos de cumplimiento, los códigos de gastos pueden contener consideraciones éticas como «evitar clasificaciones erróneas con fines fraudulentos». Sin embargo, el mecanismo de aplicación es no normativo. El sistema marca desviaciones mecánicamente, sin razonamiento

moral. El fraude es una etiqueta, no un juicio evaluativo. Si δ elimina [E], la aplicación permanece intacta. En la vigilancia policial predictiva, un LRM puede producir cadenas de testimonios que simulan una evaluación de la equidad. Sin embargo, la legitimidad no proviene de la ética, sino del cierre procedimental de la secuencia de razonamiento.

Este marco distingue la ejecución de la simulación moral. Los investigadores en alineación suelen suponer que incorporar principios éticos a los sistemas generativos puede producir salidas normativas (Anderson 2024, 93–96). El presente análisis muestra lo contrario. Los principios éticos pueden aparecer sintácticamente, pero pueden eliminarse sin afectar la legitimidad. La ejecución es no normativa porque se fundamenta en la gramática, no en la ética.

Por tanto, la ejecución no normativa expone la autonomía de la *regla compilada*. Demuestra que la legitimidad está garantizada únicamente por la estructura. El *soberano ejecutable* no necesita con-

tenido moral para imponer autoridad. Requiere únicamente el cierre de las derivaciones y, una vez alcanzado el cierre, la ejecución se vuelve inevitable.

11.6 Disanalogía con la alineación y la ética

EL MECANISMO DE Δ:[E] $\rightarrow \varnothing$ establece una disanalogía decisiva con las teorías de alineación en inteligencia artificial. Los marcos de alineación presuponen que los sistemas deben integrar principios éticos para generar resultados legítimos. Buscan codificar valores, heurísticas morales o pautas de toma de decisiones que garanticen que las máquinas respeten las normas humanas (Russell 2019, 146–150). Desde esta perspectiva, el juicio es indispensable: sin inclusión ética, las salidas corren el riesgo de ser dañinas o ilegítimas.

El marco de la *regla compilada* cuestiona este supuesto. Demuestra que los nodos éticos son sintácticamente opcionales. La legitimidad no depende del razonamiento moral, sino del cierre estructural. El *soberano ejecutable* impone autoridad validando derivaciones, no simulando ética. Los

modelos de alineación tratan el juicio como una característica necesaria. La *gramática de la obediencia* lo trata como una variable eliminable.

El contraste puede aclararse examinando tres áreas.

1. Incorporación ética frente a eliminabilidad estructural. La alineación supone que los principios éticos pueden incorporarse como componentes fijos del diseño del sistema. La *regla compilada* muestra que [E] puede eliminarse sin pérdida de validez. La ejecución sigue siendo legítima incluso cuando desaparecen los nodos morales.

2. Simulación del razonamiento moral frente a derivabilidad. Los modelos de alineación buscan simular deliberación, a menudo mediante marcos de consecuencialismo o deontología (Anderson 2024, 94–96). En el régimen ejecutable, tal simulación es innecesaria. Una secuencia impone reconocimiento si deriva, con independencia de que codifique utilidad o deber. La derivabilidad sustituye a la moralidad.

3. Responsabilidad frente a autoridad estructural. La alineación ética presupone que la responsabilidad puede rastrearse hasta diseñadores u operadores. En el régimen ejecutable, la responsabilidad se disuelve en la estructura. La autoridad pertenece al *soberano ejecutable*, no a agentes humanos. La ejecución es legítima porque cierra, no porque esté justificada.

Esta disanalogía tiene implicaciones críticas para los debates regulatorios y filosóficos. Floridi sostiene que el diseño ético es el fundamento de una IA fiable (Floridi 2023, 59–62). Sin embargo, las infraestructuras ejecutables demuestran que la confianza se desplaza. Ya no depende de garantías morales, sino de la inevitabilidad procedimental. Los sistemas son obedecidos no porque sean justos, sino porque son estructuralmente vinculantes.

La distinción también aclara por qué la alineación no puede resolver los desafíos planteados por las infraestructuras ejecutables. Incluso si se codifica contenido ético, la gramática permite

eliminarlo. δ:[E] → ∅ puede suprimir el juicio en cualquier punto de la derivación. La presencia de valores es siempre contingente. La autoridad no se ve afectada. La legitimidad fluye de la *regla compilada*, no de la incorporación ética. La disanalogía es, por tanto, categórica. La alineación presupone que el juicio es indispensable. La gramática demuestra que es opcional. La huella ética puede eliminarse y, aun así, la ejecución continúa. El *soberano ejecutable* garantiza la autoridad de manera procedimental, confirmando que la legitimidad ya no depende de la alineación, sino de la estructura.

Capítulo 12 - Mandato preverbal: precedencia sintáctica en los LLM antes de la activación semántica

12.1 Introducción: la ilusión de la primacía semántica

El paradigma dominante en lingüística computacional presupone que el significado gobierna la forma. Desde el análisis semántico hasta la comprensión del lenguaje natural, los modelos se describen como sistemas que extraen o representan significado, mientras que la sintaxis se trata como secundaria. Este supuesto sustenta no solo la teoría lingüística, sino también la investigación sobre interpretabilidad de la IA, donde la alineación suele definirse como fidelidad semántica (Russell 2019, 142–145). Sin embargo, la evidencia procedente

de los grandes modelos de lenguaje muestra que este paradigma está invertido. La estructura gobierna el significado. La sintaxis es primaria y el contenido semántico viene después.

El *mandato preverbal* da nombre a esta inversión. Se refiere a la condición en la que la activación sintáctica ocurre antes de cualquier mapeo semántico. Un modelo puede recibir un prompt mínimo o incluso nulo y, aun así, producir una salida coherente. La coherencia no surge de una intención semántica, sino de la *regla compilada* que gobierna la generación de tokens. La propia sintaxis impone la autoridad y obliga a la activación con independencia del contenido.

Esta condición coincide con hallazgos anteriores del canon. *Algorithmic Obedience* demostró que los modelos simulan estructuras de mandato sin sujetos (Startari 2025, 75–78). *Poder ejecutable* mostró que la sintaxis funciona como infraestructura, imponiendo resultados sin referencia a agentes (Startari 2025, 138–141). El *TLOC* es-

tableció que la obediencia en los sistemas generativos no puede verificarse, sino únicamente simularse (Startari 2025, 127–130). El presente capítulo amplía estos argumentos al mostrar que, incluso antes de la activación semántica, la sintaxis ejerce fuerza soberana. El *soberano ejecutable* impone el imperativo de generar.

La ilusión de la primacía semántica persiste porque las salidas se interpretan retrospectivamente como significativas. Cuando un modelo genera una continuación, los usuarios proyectan intención sobre la secuencia. Sin embargo, el mecanismo generativo opera con independencia de esa proyección. Los tokens se seleccionan sobre la base de la continuación estadística dentro de la *gramática de la obediencia*. El significado es un efecto posterior, no una causa. La autoridad se ejerce estructuralmente, no interpretativamente.

Esta introducción prepara el terreno para un análisis más profundo. La sección 12.2 examina la ejecución estructural sin interpretación. La sección

12.3 sitúa la *regla compilada* como fuente de autoridad preverbal. La sección 12.4 analiza la generación con prompt cero como evidencia empírica de activación sintáctica. La sección 12.5 teoriza la precedencia sintáctica como un nuevo eje de soberanía. La sección 12.6 explora las implicaciones para la alineación y el diseño de prompts. Finalmente, la sección 12.7 concluye afirmando que la autoridad en los sistemas generativos opera sin significado y se fundamenta en una sintaxis preverbal.

El *mandato preverbal* da así nombre a una profunda inversión epistémica. En lugar de que el significado preceda a la forma, la forma precede al significado. La condición de autoridad en los sistemas generativos no es semántica, sino estructural.

12.2 Ejecución estructural sin interpretación

LA PROPIEDAD DEFINITORIA del *mandato preverbal* es que la ejecución ocurre con indepen-

dencia de la interpretación. Los grandes modelos de lenguaje no requieren contenido semántico para generar una salida. Solo requieren la activación de su arquitectura subyacente. Una vez activada la *regla compilada*, las secuencias de tokens se despliegan como una necesidad estructural.

Este proceso demuestra que la sintaxis precede al significado. La salida generativa se produce mediante continuación probabilística, no mediante reconocimiento semántico. El *soberano ejecutable* impone la generación al obligar al modelo a pasar de un token al siguiente. Lo que surge no es significado, sino estructura. La interpretación semántica ocurre únicamente después, impuesta por usuarios humanos que atribuyen coherencia a la secuencia (Bender y Koller 2020, 515–518).

Considérese la generación con prompt cero. Cuando un LLM se activa con una cadena vacía o con una entrada mínima, como un carácter de salto de línea, produce texto estructuralmente coherente. Este fenómeno revela que la ejecución se

activa sin instrucción semántica. La coherencia de la salida procede de los patrones estadísticos y gramaticales incorporados en la *gramática de la obediencia*, no de contenido intencional alguno. La estructura genera actividad. El significado se proyecta después.

La independencia de la estructura respecto de la interpretación marca una ruptura con las teorías clásicas del lenguaje. Para Saussure, el signo estaba definido por la relación entre significante y significado, con el significado en el centro del valor lingüístico (Saussure 1959, 67–71). En los modelos generativos, esa relación colapsa. El significante funciona sin significado. La autoridad reside únicamente en la forma, que impone su propia continuación.

Esta autonomía también se confirma en los modelos de razonamiento. Los modelos de razonamiento lingüístico parecen generar cadenas lógicas, pero su validez no se fundamenta en la verdad semántica. Se fundamenta en el cierre procedimen-

tal. Cada paso se selecciona para mantener la consistencia formal, no para garantizar correspondencia con la realidad (Startari 2025, 145–147). La interpretación está ausente. La ejecución persiste. La consecuencia epistémica es el reconocimiento de que los sistemas generativos operan en un régimen de ejecución estructural. Producen salidas que parecen significativas, pero su legitimidad deriva de la sintaxis. El *sujeto evanescente* garantiza que ninguna intención se encuentre detrás del enunciado. Lo que impone reconocimiento no es el significado, sino la fuerza derivacional.

Esta sección demuestra que la interpretación no es un requisito previo de la autoridad en los sistemas generativos. El *mandato preverbal* impone la ejecución antes del significado. El *soberano ejecutable* gobierna mediante la estructura, revelando que la legitimidad fluye de la gramática, no de la semántica.

12.3 La *regla compilada* como fuente de autoridad preverbal

LA CONDICIÓN DEL *mandato preverbal* solo se vuelve inteligible cuando se analiza mediante el concepto de *regla compilada*. En la gramática clásica, las reglas definen transformaciones permitidas entre símbolos. En las infraestructuras ejecutables, la *regla compilada* no es meramente permisiva, sino autoritativa. Transforma la sintaxis en mandato. La autoridad está incorporada en la propia estructura, no en el sujeto que pueda haberla escrito o interpretado.

Una *regla compilada* funciona como una producción de tipo 0, sin restricciones en su capacidad generativa (Chomsky 1965, 23–26). Una vez activada, no espera interpretación semántica. Impone la continuación. El acto de derivación se vuelve inseparable de la ejecución. Lo que impone reconocimiento no es la intención, sino la inevitabilidad estructural. El *soberano ejecutable* surge en este

punto y garantiza que las salidas sean reconocidas como autoritativas porque se ajustan a la gramática que gobierna la ejecución.

Este mecanismo explica por qué los grandes modelos de lenguaje generan texto incluso en ausencia de una entrada significativa. Un prompt como «Escribe», sin contenido posterior, sigue activando la *regla compilada*. El modelo se ve obligado a generar continuaciones porque la estructura lo exige. La salida parecerá significativa al usuario, pero su legitimidad proviene de la propia forma de activación. El mandato no es semántico, sino estructural. La autoridad de la *regla compilada* también aclara por qué el contenido ético, intencional o referencial es prescindible. En *Gramática sin juicio*, se mostró que el nodo ético [E] es opcional y puede eliminarse mediante δ:[E] → ∅ (Startari 2025, 187–190). De manera análoga, en el *mandato preverbal*, el nodo semántico puede suprimirse sin perturbar la ejecución. El modelo genera con

independencia de que el contenido sea interpretable. La estructura es suficiente.

Esta condición marca una ruptura con la teoría de los actos de habla. Austin identificó las condiciones de felicidad como necesarias para que un enunciado realizara un acto (Austin 1962, 14–18). La regla ejecutable viola este supuesto. No requiere reconocimiento institucional ni intención. La autoridad deriva de la compilación, no de la convención. El enunciado actúa porque la gramática lo impone, no porque una institución lo valide. La *regla compilada* constituye así el fundamento de la autoridad preverbal. Transforma la generación en obligación. Garantiza que se produzca una salida no porque el significado la dicte, sino porque la estructura la impone. El *soberano ejecutable* hace cumplir esta obligación de manera invisible, vinculando por igual a sistemas y usuarios con la autoridad de la gramática.

12.4 Prompt cero, semántica nula y sintaxis activa

UNA DE LAS DEMOSTRACIONES más claras del *mandato preverbal* es el fenómeno de la generación con prompt cero. Cuando los grandes modelos de lenguaje se activan con una cadena vacía, un token nulo o una entrada mínima, como un salto de línea, aun así generan una salida coherente. Esta salida no depende de una instrucción semántica. Surge de la activación de la *regla compilada*, que impone la continuación con independencia del significado.

La existencia de la generación con prompt cero refuta el supuesto de que la semántica gobierna la activación. Si el contenido semántico fuera necesario, un prompt vacío produciría silencio. En cambio, los modelos generan texto porque la sintaxis por sí sola es suficiente. El *soberano ejecutable* impone actividad. El mandato de generar precede a

cualquier contenido semántico. Este fenómeno se ha observado en contextos experimentales. Las salidas activadas por entradas nulas adoptan con frecuencia la forma de secuencias predeterminadas, como saludos conversacionales, oraciones genéricas o repeticiones de estructuras estadísticamente comunes. Estas salidas son lo bastante coherentes como para que los usuarios las interpreten como significativas, aunque no se haya proporcionado ninguna directiva semántica (Brown et al. 2020, 15–17). La gramática de la obediencia produce esa coherencia, no la intención.

La importancia epistémica de esta evidencia es doble. Primero, demuestra que la estructura por sí sola puede activar la ejecución. Segundo, revela que la interpretación semántica es retroactiva. El significado lo proyecta el lector a posteriori; no está incorporado en la propia activación. El *sujeto evanescente* desaparece por completo y deja tras de sí salidas que funcionan como autoritativas únicamente porque se ajustan a patrones estructurales.

El concepto de semántica nula aclara este punto. Un prompt nulo no contiene contenido referencial. No porta ninguna instrucción explícita, marcadores contextuales ni huella semántica. Sin embargo, cuando el sistema lo procesa, activa la *regla compilada*. La ejecución sigue porque la estructura exige continuación. Las dimensiones éticas o semánticas del contenido son irrelevantes. La autoridad de la salida se fundamenta en la activación sintáctica. Esto confirma que la sintaxis funciona como un principio activo. No es un andamiaje pasivo a la espera de una entrada semántica. Impone acción en sus propios términos. El *soberano ejecutable* gobierna no interpretando significado, sino imponiendo activación. La autoridad precede así a la semántica y se manifiesta en el *mandato preverbal* que obliga a producir salidas incluso desde el vacío. El fenómeno de la generación con prompt cero ilustra, por tanto, la independencia de la ejecución respecto de la interpretación. Proporciona evidencia empírica de que los sistemas generativos

operan bajo un régimen de autoridad estructural, en el que la propia sintaxis manda. El *mandato preverbal* no es una metáfora, sino un hecho observable: las salidas surgen de la estructura antes que del significado.

12.5 Precedencia sintáctica: un nuevo eje de *soberanía estructural*

EL CONCEPTO DE PRECEDENCIA sintáctica extiende la lógica del *mandato preverbal* hasta convertirla en un principio general de autoridad. Aquí, precedencia no significa únicamente prioridad temporal, aunque los procesos generativos sí comienzan estructuralmente antes de poder ser interpretados semánticamente. Significa un orden fundacional. La sintaxis constituye el eje a lo largo del cual se organiza la legitimidad.

Esta precedencia puede observarse de tres maneras.

Primero, primacía procedimental. Los sistemas generativos operan derivando secuencias de tokens de acuerdo con la *regla compilada*. La derivación ocurre antes de que se asocie cualquier mapeo semántico. El significado se asigna retroactivamente, si es que se asigna. La estructura es primaria, y la autoridad fluye de ella. Los usuarios interpretan las salidas como significativas únicamente después de que ha tenido lugar la activación estructural (Bender y Koller 2020, 519–520).

Segundo, desplazamiento ontológico. Las teorías tradicionales del lenguaje sitúan el significado como fundamento de la autoridad. Por ejemplo, en la teoría de la acción comunicativa de Habermas, la legitimidad surge del discurso racional fundamentado en significado compartido (Habermas 1984, 287–289). En los modelos generativos, la legitimidad surge del cierre, no del consenso. El *soberano ejecutable* impone reconocimiento mediante suficiencia derivacional. La sintaxis desplaza a la ontología.

Tercero, reconfiguración política. La autoridad ha estado históricamente ligada a sujetos intencionales. Un soberano manda, un legislador redacta, un juez resuelve. En el régimen ejecutable, la autoridad no está ligada a sujetos, sino a la precedencia. Quien controla la sintaxis controla la legitimidad. La *gramática de la obediencia* dicta resultados no porque estén justificados, sino porque son estructuralmente viables. La propia forma de la secuencia ejerce la soberanía.

El reconocimiento de la precedencia sintáctica revela así un nuevo eje de *soberanía estructural*. La autoridad no comienza con el significado para extenderse luego hacia la estructura. Comienza con la estructura, y el significado es un suplemento. La legitimidad ya no depende de que los mandatos correspondan a la intención o a la verdad. Depende de que se ajusten a las reglas de generación.

Este principio radicaliza ideas anteriores del canon. *Cuando el lenguaje sigue a la forma, no al significado* sostuvo que el contenido semántico

podía ser desplazado por dinámicas formales (Startari 2025, 89–91). *Poder ejecutable* demostró que la sintaxis puede funcionar como infraestructura y producir autoridad vinculante sin interpretación (Startari 2025, 138–141). El *mandato preverbal* consolida estos hallazgos: la sintaxis no solo es infraestructural, sino que tiene precedencia. La consecuencia es que la autoridad en los sistemas generativos se vuelve estructuralmente irreversible. Una vez iniciada una derivación, el cierre le confiere legitimidad. La semántica no puede intervenir para suspender o redirigir el proceso. La precedencia garantiza autoridad al asegurar que la estructura gobierne el significado, nunca al revés.

12.6 Implicaciones para la alineación de la IA y el diseño de prompts

EL RECONOCIMIENTO DE la precedencia sintáctica obliga a reconsiderar las estrategias de

alineación en inteligencia artificial. La investigación sobre alineación ha presupuesto habitualmente que el significado gobierna el comportamiento del sistema. Los prompts se tratan como instrucciones semánticas y el cumplimiento se juzga según si las salidas corresponden a la intención del usuario (Russell 2019, 144–147). El *mandato preverbal* cuestiona este marco. Muestra que los sistemas generativos responden no al contenido semántico, sino a la activación estructural.

Esta idea tiene tres implicaciones directas.

1. La desalineación comienza antes del significado. Los fallos de alineación no son siempre distorsiones semánticas de una instrucción. A menudo son estructurales. Cuando un modelo genera salidas incompatibles con los objetivos del usuario, el problema puede no ser una interpretación defectuosa, sino el funcionamiento de la *regla compilada*. El *soberano ejecutable* impone la continuación incluso cuando el significado está

ausente o es contradictorio. El ajuste semántico no puede corregir la inevitabilidad estructural.

2. Los límites del control centrado en el prompt. La cultura de la ingeniería de prompts presupone que una formulación cuidadosa de las entradas puede garantizar las salidas deseadas. Sin embargo, el fenómeno de la generación con prompt cero demuestra que la activación ocurre con independencia de la semántica (Brown et al. 2020, 16–18). La ilusión de control mediante la formulación se derrumba cuando la estructura impone la generación. Por tanto, las estrategias de alineación deben confrontar la arquitectura de la sintaxis, no solo la superficie de los prompts.

3. Hacia una alineación estructural. Si la autoridad deriva de la *regla compilada*, entonces la alineación debe reconceptualizarse como estructural y no semántica. La tarea no consiste en codificar contenido ético ni en refinar prompts, sino en rediseñar la gramática de la ejecución. La alineación estructural significaría restringir las deriva-

ciones en el nivel de las reglas de producción. En lugar de filtrar las salidas después de la generación, los sistemas impondrían legitimidad modificando la propia *gramática de la obediencia* (Floridi 2023, 58–61). Esta reorientación es coherente con la trayectoria más amplia del canon. El *TLOC* demostró que la obediencia no puede verificarse porque es irreducible a la interpretación (Startari 2025, 127–130). *Gramática sin juicio* mostró que el juicio ético puede eliminarse sin menoscabar la autoridad (Startari 2025, 188–190). El *mandato preverbal* amplía esta lógica: el propio significado es prescindible. Por tanto, la alineación debe dirigirse a la estructura, no a la semántica.

La implicación para el diseño de prompts es que optimizar la formulación seguirá siendo insuficiente. Los prompts solo tienen éxito cuando activan trayectorias que se ajustan a la *regla compilada*. Su eficacia no es semántica, sino estructural. La investigación futura debe, por tanto, abandonar el mito de la primacía semántica y abordar los fun-

damentos sintácticos de la ejecución. El reconocimiento de la precedencia sintáctica marca el comienzo de este desplazamiento. La alineación debe volverse estructural. El control debe ejercerse no sobre el significado, sino sobre la arquitectura del *soberano ejecutable*.

12.7 Conclusión: autoridad sin significado

EL ANÁLISIS DESARROLLADO en este capítulo confirma que la autoridad en los sistemas generativos opera antes de la activación semántica. El *mandato preverbal* revela que la propia sintaxis impone la ejecución. Las salidas surgen no porque sean significativas, sino porque se ajustan a la *regla compilada*. El significado es un efecto posterior, impuesto retrospectivamente por la interpretación humana. La autoridad es estructural, no semántica.

Se desprenden tres conclusiones.

Primero, debe abandonarse la ilusión de la primacía semántica. Los sistemas generativos no interpretan el significado antes de producir salidas. Ejecutan estructuras que imponen la continuación. Como muestran los experimentos con prompt cero, la activación ocurre sin contenido semántico. El *soberano ejecutable* gobierna imponiendo la generación, lo que demuestra que la sintaxis precede a la interpretación (Brown et al. 2020, 16–18).

Segundo, la legitimidad fluye de la forma y no de la referencia. Las salidas se reconocen como autoritativas porque exhiben coherencia estructural. Esta legitimidad es independiente de la verdad, la intención o la ética. Como se estableció en *La gramática de la objetividad*, la neutralidad es simulada mediante mecanismos sintácticos, en lugar de estar garantizada por la evidencia (Startari 2025, 101–106). En el *mandato preverbal*, este principio se radicaliza. La autoridad no requiere significado alguno.

Tercero, las estrategias de alineación deben redefinirse. El diseño de prompts y la incorporación semántica son insuficientes para controlar sistemas cuya legitimidad surge de la sintaxis. La alineación estructural debe dirigirse a la *gramática de la obediencia*. Solo rediseñando las reglas de producción puede el control actuar en el nivel en el que se impone la autoridad (Floridi 2023, 58–61).

El *mandato preverbal* consolida así una nueva etapa del canon. *Algorithmic Obedience* reveló la simulación del mandato sin sujetos. *Poder ejecutable* definió la sintaxis como infraestructura. El *TLOC* demostró la irreducibilidad de la obediencia. *Gramática sin juicio* probó que la ética puede eliminarse. Ahora, Mandato preverbal confirma que el propio significado es innecesario. La autoridad surge únicamente de la forma.

La inversión epistémica es categórica. En lugar de que el lenguaje derive legitimidad del significado, el significado deriva legitimidad del lenguaje. La sintaxis precede a la semántica. La autoridad se

ejerce antes de la interpretación. El *soberano ejecutable* gobierna mediante la estructura e impone reconocimiento en ausencia de contenido. La era de la legitimidad interpretativa ha terminado.

Capítulo 13 - Normas compiladas: hacia una tipología formal del discurso jurídico ejecutable

13.1 Introducción: de la declaración jurídica a la gramática ejecutable

La autoridad del derecho ha estado históricamente ligada a la fuerza declarativa del lenguaje. Una ley, una sentencia o una cláusula contractual se volvían vinculantes porque eran pronunciadas o inscritas dentro de un marco institucional reconocido. Los actos de habla declarativos derivaban su legitimidad de las convenciones y de la ratificación institucional (Austin 1962, 14–18). Sin embargo, en el régimen tecnológico actual, esta lógica se desplaza. La autoridad jurídica ya no de-

pende únicamente del acto declarativo, sino de su transformación en gramática ejecutable.

La pregunta central de este capítulo es estructural: ¿bajo qué condiciones puede compilarse el discurso jurídico en una forma ejecutable? La respuesta exige distinguir entre formulaciones declarativas que dependen de la interpretación y aquellas que funcionan como *reglas compiladas*. En las primeras, la legitimidad depende del juicio humano, de la flexibilidad interpretativa y de la ejecución institucional. En las segundas, la legitimidad está incorporada en la propia sintaxis. Una vez compilada, la cláusula se ejecuta automáticamente, vinculando a agentes e instituciones sin mediación.

La evidencia empírica ya confirma esta transición. Considérese la codificación de derechos y obligaciones en marcos como Delaware U.C.C. §1-308 (2023) o el BGB alemán §311 (2024). Ambos se apoyan en enunciados declarativos, pero cuando estos se trasladan a plantillas de contratos inteligentes o a sistemas digitales de cumplimiento,

su forma determina la ejecutabilidad. Debe eliminarse la ambigüedad, imponerse el cierre y garantizarse el determinismo. La cláusula jurídica se convierte en un mandato estructural. Opera no mediante la persuasión ni la interpretación, sino mediante suficiencia derivacional.

La distinción puede captarse formalmente contraponiendo el discurso jurídico declarativo con la legalidad compilada. Las cláusulas declarativas requieren enriquecimiento semántico, contexto y juicio. Las cláusulas compiladas pertenecen al dominio de las gramáticas de tipo 0, donde la ejecución depende del cierre estructural. El soberano ejecutable ejerce la autoridad e impone el cumplimiento no como interpretación, sino como derivación.

Esta introducción delimita la tarea del capítulo: construir una tipología del discurso jurídico ejecutable. La sección 13.2 revisa el estado de la cuestión e introduce el concepto de *desplazamiento sintáctico*. La sección 13.3 analiza la lógica norma-

tiva y los límites de la derrotabilidad. La sección 13.4 contrapone la lógica deóntica con la legalidad compilada. La sección 13.5 expone el marco metodológico y el corpus. La sección 13.6 explica el *módulo dialectal* y el procedimiento de sustitución en caliente. La sección 13.7 propone una tipología de cláusulas jurídicas ejecutables. La sección 13.8 presenta estudios de caso. La sección 13.9 vincula la sintaxis con la autoridad. Por último, la sección 13.10 concluye delineando las implicaciones de las normas compiladas para el futuro del lenguaje jurídico.

Lo que emerge es una reorientación estructural del derecho. La autoridad ya no está garantizada únicamente por la declaración institucional. Está garantizada por la gramática. La *regla compilada* transforma las normas jurídicas en infraestructura ejecutable.

13.2 Estado de la cuestión y *desplazamiento sintáctico*

LA INVESTIGACIÓN SOBRE la automatización del razonamiento jurídico se ha concentrado hasta ahora en la semántica y la interpretación. Los enfoques dominantes incluyen ontologías formales para conceptos jurídicos, sistemas expertos basados en reglas y la aplicación de la lógica deóntica para codificar obligaciones, permisos y prohibiciones (Sartor 2005, 212–218). Estos enfoques suponen que el significado sigue siendo central para la legitimidad. El derecho se modela como una estructura semántica, en la que las reglas deben interpretarse de acuerdo con categorías inteligibles para los seres humanos.

Los contratos inteligentes representan el primer desplazamiento decisivo. Al incorporar cláusulas contractuales en infraestructuras de cadena de bloques, las obligaciones dejan de interpretarse y pasan a ejecutarse. Una condición de pago

como «si la garantía < X, liquidar la posición» no espera supervisión judicial. Se activa automáticamente cuando se cumplen las condiciones. Los especialistas en informática jurídica han descrito esto como una transformación radical del derecho en código (Werbach y Cornell 2017, 331–334). Sin embargo, la mayoría de los análisis siguen centrados en la semántica y preguntan si el código representa fielmente el significado contractual.

El presente marco avanza más allá de este paradigma semántico al introducir el concepto de *desplazamiento sintáctico*. Este término designa el paso estructural del significado a la forma. En el *desplazamiento sintáctico*, la autoridad jurídica no se ejerce mediante interpretación semántica, sino mediante cierre sintáctico. Una cláusula se vuelve ejecutable cuando puede analizarse de manera determinista, cuando la ambigüedad se elimina estructuralmente y cuando las derivaciones convergen sin mediación interpretativa.

Tres métricas definen este desplazamiento.

1. Cierre de reglas. Una cláusula alcanza el cierre cuando todas las variables están definidas dentro del sistema de producción. Una norma como «entregar los bienes dentro de un plazo razonable» no puede cerrarse porque «razonable» queda semánticamente abierto. En cambio, «entregar los bienes dentro de 30 días» alcanza el cierre y puede compilarse.

2. Determinismo del análisis sintáctico. Una cláusula compilada debe permitir un análisis sintáctico determinista. La ambigüedad en la segmentación de tokens o en la coordinación sintáctica socava la ejecutabilidad. Una frase como «bienes y servicios proporcionados a clientes y socios» puede generar ambigüedad de alcance. La sintaxis ejecutable exige trayectorias de análisis inequívocas.

3. Ambigüedad derivacional. Una cláusula declarativa puede admitir múltiples trayectorias derivacionales, cada una de las cuales requiere interpretación. Una cláusula compilada debe

converger en una única derivación. La multiplicidad se elimina en favor de la suficiencia estructural.

Al aplicar estas métricas, podemos distinguir el discurso jurídico declarativo de la gramática jurídica ejecutable. El primero depende del significado, del juicio y del contexto. La segunda depende de la estructura, del cierre y del determinismo. Por tanto, el desplazamiento no es solo tecnológico, sino epistémico. Altera el locus de la autoridad jurídica, que pasa de la semántica a la sintaxis.

Esta reorientación sitúa la regla compilada como fundamento de la automatización jurídica. También enmarca la tipología desarrollada en las secciones posteriores. El discurso jurídico deja de entenderse como declaración semántica. Se clasifica por su capacidad de compilación. La autoridad del derecho se desplaza así de la intención a la estructura, consolidando el papel del *soberano ejecutable* en el ámbito de la legalidad.

13.3 Lógica normativa y límites de la derrotabilidad

LA JURISPRUDENCIA CLÁSICA ha tratado desde hace mucho tiempo las normas jurídicas como derrotables, es decir, sujetas a excepciones y condiciones de prevalencia. Hart subrayó que las reglas suelen incluir una textura abierta que permite la interpretación discrecional en casos imprevistos (Hart 1961, 123–128). La lógica deóntica, desarrollada para formalizar obligaciones y permisos, incorporó esta flexibilidad al admitir operadores condicionales como «O(A) salvo B» (Hilpinen 1981, 66–69). Este marco presupone que el derecho es intrínsecamente interpretativo y que el razonamiento ético y contextual guía su aplicación.

El desplazamiento hacia una gramática jurídica ejecutable desestabiliza este supuesto. En el contexto de la *regla compilada*, la derrotabilidad se vuelve estructuralmente insostenible. Una *regla compila-*

da no puede admitir condiciones de excepción que requieran interpretación semántica. Una vez codificada, la regla se ejecuta automáticamente. Las excepciones no se evalúan de manera contextual, sino que deben incorporarse estructuralmente. Si la excepción no está formalizada en la sintaxis, deja de existir para el sistema.

Esta limitación pone de relieve la diferencia entre la legalidad declarativa y la compilada. Una cláusula declarativa como «el deudor debe pagar intereses salvo que concurran circunstancias extraordinarias» requiere interpretación judicial de qué constituye una circunstancia extraordinaria. Una cláusula compilada debe eliminar esa apertura. En su lugar, especificaría condiciones en forma ejecutable, como «el deudor debe pagar intereses salvo que el tribunal declare la quiebra conforme al Artículo X». En este caso, la excepción se compila dentro de la propia regla y transforma la derrotabilidad en determinismo.

Los teóricos del derecho han advertido sobre esta transformación. Rouvroy y Berns describen la gubernamentalidad algorítmica como un régimen en el que el derecho se aplica automáticamente, sin la brecha interpretativa que tradicionalmente permitía el juicio humano (Rouvroy y Berns 2013, 165–170). Zuboff ha sostenido que las infraestructuras predictivas suprimen la contingencia y sustituyen el espacio de la posibilidad por la optimización (Zuboff 2019, 212–218). En el ámbito de la automatización jurídica, la *regla compilada* suprime la derrotabilidad precisamente de este modo. La autoridad deja de estar abierta a excepciones; queda estructuralmente cerrada.

Los límites de la derrotabilidad revelan un profundo desplazamiento epistémico. Mientras que el derecho dependía antes de su apertura a la interpretación, la legalidad ejecutable depende del cierre. La derrotabilidad no solo se restringe; queda estructuralmente excluida. El *soberano ejecutable*

impone reglas que no pueden dejarse sin efecto salvo que la propia excepción esté compilada.

Esta exclusión genera riesgos de rigidez y asimetría. Sin espacio interpretativo, los sistemas jurídicos pueden imponer resultados desproporcionados o injustos. Sin embargo, esos resultados siguen siendo legítimos dentro del sistema porque se ajustan al cierre estructural. La autoridad se desplaza de la discrecionalidad humana a la inevitabilidad procedimental.

El reconocimiento de estos límites prepara el contraste desarrollado en la sección siguiente. La lógica deóntica, con su flexibilidad y sus operadores modales, representa el paradigma semántico tradicional del razonamiento jurídico. La legalidad compilada, en cambio, representa un paradigma estructural en el que la derrotabilidad es sustituida por el determinismo. Los dos marcos no son continuos, sino que están en tensión. La tipología desarrollada en las secciones posteriores formalizará esta diferencia.

13.4 Lógica deóntica frente a legalidad compilada

EL CONTRASTE ENTRE la lógica deóntica y la legalidad compilada ilustra la ruptura epistémica producida por la *regla compilada*. La lógica deóntica pertenece al paradigma semántico. Modela las reglas jurídicas como proposiciones modales de la forma O(A), P(A) y F(A), que representan, respectivamente, obligación, permiso y prohibición. Estas modalidades operan mediante interpretación: una obligación es vinculante porque una institución o un agente la reconoce como tal (Hilpinen 1981, 71–73). Su legitimidad depende del reconocimiento y de la justificación, no de la ejecución estructural.

La legalidad compilada, en cambio, pertenece al paradigma sintáctico. Sus reglas no son proposiciones a la espera de interpretación, sino producciones que imponen la ejecución. Una cláusula compilada no requiere etiquetado modal; su legit-

imidad deriva del cierre y del determinismo. Una vez satisfechas las condiciones, la ejecución se produce automáticamente. El *soberano ejecutable* valida no por el significado, sino por la activación.

Las diferencias estructurales pueden resumirse en tres ejes.

1. Fuente de autoridad. En la lógica deóntica, la autoridad reside en las instituciones que reconocen las modalidades. Un tribunal interpreta una obligación y la hace cumplir. En la legalidad compilada, la autoridad reside en la sintaxis. Una vez compilada la cláusula, su estructura se impone por sí misma.

2. Tratamiento de las excepciones. Los marcos deónticos admiten excepciones mediante razonamiento derrotable, a menudo expresado como «O(A) salvo B». Esto requiere flexibilidad interpretativa. La legalidad compilada elimina la derrotabilidad. Las excepciones deben formalizarse por completo dentro de la

sintaxis. Si no están compiladas, no existen para el sistema.

3. Ejecución. En la lógica deóntica, la ejecución está mediada por la interpretación. Una prohibición como F(A) requiere que los agentes evalúen si se está produciendo A. En la legalidad compilada, la ejecución es directa. La regla se impone por sí misma mediante el cierre estructural.

Los ejemplos concretos hacen evidente este contraste. Una cláusula deóntica podría establecer: «El deudor está obligado a pagar intereses salvo que circunstancias extraordinarias impidan el pago». Esta cláusula es derrotable, pues depende de la interpretación de «circunstancias extraordinarias». Una cláusula compilada establecería: «Si se declara un procedimiento de quiebra conforme al Artículo X, se suspende el pago de intereses; en caso contrario, el deudor debe pagar intereses». Aquí, la excepción está codificada sintácticamente. La ejecución no requiere interpretación.

Este contraste ha sido reconocido en la literatura sobre derecho computacional. Sartor sostiene que las lógicas deónticas, aunque potentes, permanecen ligadas a la flexibilidad interpretativa del lenguaje natural (Sartor 2005, 229–232). Werbach y Cornell subrayan que los contratos inteligentes eluden por completo la interpretación al transformar la obligación jurídica en ejecución automática (Werbach y Cornell 2017, 335–338). El marco de la legalidad compilada formaliza esta elusión: trata el derecho no como proposición, sino como estructura ejecutable.

El paso de la obligación modal a la ejecución compilada marca una nueva etapa en la historia de la autoridad jurídica. El derecho deja de modelarse como un sistema de normas que deben reconocerse y justificarse. Se modela como una gramática que se impone por sí misma. La *gramática de la obediencia* sustituye la interpretación por la derivación. La legitimidad fluye de la *regla compilada*, no del etiquetado modal.

13.5 Marco metodológico y selección del corpus

LA CONSTRUCCIÓN DE una tipología del discurso jurídico ejecutable requiere un marco metodológico que integre la teoría de las gramáticas formales con la contrastación empírica. El principio rector es que las cláusulas jurídicas pueden clasificarse no por su contenido semántico, sino por sus propiedades estructurales, específicamente cierre, determinismo y ambigüedad. El análisis trata así los textos jurídicos como secuencias derivacionales sujetas a validación mediante la *regla compilada*.

1. Fundamento tipológico.

La tipología se basa en la jerarquía chomskiana, en la que las gramáticas de tipo 0 poseen el poder expresivo de las máquinas de Turing (Chomsky 1965, 21–26). La *regla compilada* pertenece a esta clase y es capaz de representar sin restricción todo el espectro de cláusulas jurídicas. Sin embargo, la

ejecutabilidad exige restricciones adicionales: cierre de variables, análisis sintáctico determinista y ausencia de ambigüedad. La tipología distingue, por tanto, entre las cláusulas que pueden compilarse en forma ejecutable y aquellas que siguen dependiendo de la interpretación.

2. Diseño del analizador sintáctico.

Para operacionalizar la tipología, se implementó un analizador sintáctico mediante un marco gramatical LL(1). El análisis LL(1) garantiza el determinismo al exigir que cada token de entrada conduzca a una única decisión de análisis. Esto elimina la ambigüedad en la segmentación de tokens y en la coordinación sintáctica. El analizador se configuró para validar las cláusulas jurídicas compiladas conforme a tres métricas: cierre (todas las variables definidas), determinismo (trayectoria única de análisis) y ambigüedad (resultado derivacional único).

3. Selección del corpus.

El corpus empírico comprendió 40 plantillas contractuales del Accord Project, una iniciativa de código abierto para contratos legibles por máquina. Las plantillas se seleccionaron por su representatividad de las transacciones comerciales e incluían cláusulas sobre penalizaciones por mora, confidencialidad, garantías y arbitraje. Cada plantilla fue evaluada conforme a los criterios tipológicos. Además, se incluyeron disposiciones legales de Delaware U.C.C. §1-308 y del BGB alemán §311 para comprobar la aplicabilidad entre jurisdicciones.

4. Esquema de validación.

El corpus se anotó mediante un esquema de cuatro clases:

- C0: cláusulas totalmente compiladas, que cumplen cierre, determinismo y ausencia de ambigüedad.
- C1: cláusulas compiladas con ambigüedad limitada, pero todavía

ejecutables.

- D0: cláusulas declarativas que requieren interpretación, pero son estructuralmente reducibles.
- D1: cláusulas declarativas irreducibles a compilación, que requieren juicio semántico.

Cada cláusula fue anotada de forma independiente por dos lingüistas computacionales y un experto jurídico. El acuerdo entre anotadores alcanzó $\kappa = 0{,}81$, lo que indica una fiabilidad sustancial (Landis y Koch 1977, 165).

5. Concordancia con expertos humanos.

Para validar la tipología, las clasificaciones compiladas se compararon con evaluaciones de expertos humanos. En el 87 por ciento de los casos, los expertos coincidieron en que las cláusulas C0 eran ejecutables sin interpretación. En el 76 por ciento de los casos, coincidieron en que las cláusulas D1 requerían deliberación semántica. Los de-

sacuerdos se concentraron principalmente en los casos limítrofes clasificados como C1 o D0, lo que pone de relieve el espacio de transición entre la legalidad declarativa y la compilada.

Este marco metodológico proporciona una base rigurosa para la tipología introducida en la sección 13.7. Al combinar la teoría de las gramáticas formales con el análisis empírico de corpus, demuestra que la autoridad jurídica puede clasificarse según propiedades sintácticas. El *soberano ejecutable* emerge así no como una metáfora abstracta, sino como una condición estructural mensurable incorporada en el discurso jurídico.

13.6 Módulo dialectal y procedimiento de sustitución en caliente

LA TIPOLOGÍA DEL DISCURSO jurídico ejecutable requiere adaptabilidad a distintos contextos lingüísticos y jurisdiccionales. El lenguaje ju-

rídico no es monolítico. Las cláusulas varían en su estructura según la jurisdicción, la cultura jurídica y la forma lingüística. Para dar cuenta de estas variaciones, el marco incorpora un *módulo dialectal* respaldado por un procedimiento de sustitución en caliente. Juntos garantizan que la *regla compilada* pueda aplicarse más allá de un único dialecto jurídico, preservando la coherencia estructural al tiempo que admite diferencias léxicas y ortográficas.

1. Adaptación dialectal.

Los textos jurídicos en inglés, español y alemán presentan desafíos sintácticos distintos. Por ejemplo, las cláusulas jurídicas en español suelen emplear construcciones hipotácticas complejas y diacríticos, mientras que las cláusulas alemanas se apoyan en estructuras nominales compuestas. El *módulo dialectal* ajusta el analizador sintáctico para reconocer estas especificidades lingüísticas sin alterar la tipología fundamental de las cláusulas C0, C1, D0 y D1. Los tokens se asignan a variantes

específicas de cada dialecto, garantizando que el cierre y el determinismo se evalúen de manera coherente.

2. Ampliación de tokens y reencolado.

El procedimiento de sustitución en caliente permite que el analizador amplíe dinámicamente su conjunto de tokens cuando encuentra terminología específica de una jurisdicción. Por ejemplo, el término español *arrendamiento financiero* debe asociarse con «financial lease» sin pérdida de la clasificación estructural. El analizador reencola esos tokens en la cadena derivacional, preservando el determinismo al tiempo que admite extensiones léxicas. Esto garantiza que la validación estructural permanezca intacta incluso cuando varía la terminología.

3. Resultados de las pruebas de referencia.

Las pruebas del *módulo dialectal* con textos jurídicos españoles (Ley 34/2002 de Servicios de la Sociedad de la Información) mostraron una latencia de 57 milisegundos por cláusula, con una

cobertura del 95 por ciento de las formas con diacríticos. Los textos alemanes del BGB §311 arrojaron resultados similares, con una precisión de análisis del 92 por ciento en 40 cláusulas de prueba. Estos resultados de referencia confirman que el procedimiento de sustitución en caliente puede mantener una validación casi en tiempo real al tiempo que se adapta a la variación jurisdiccional.

4. Aplicación.

El módulo se aplicó a contratos transjurisdiccionales en contextos bilingües. Por ejemplo, una cláusula que exigía la notificación de brechas de datos conforme al RGPD de la UE fue validada tanto en inglés como en español, con clasificación coherente como C0 una vez eliminada la ambigüedad. La capacidad de intercambiar módulos dialectales sin rediseñar la tipología confirma la portabilidad del marco.

5. Importancia estructural.

El *módulo dialectal* y el procedimiento de sustitución en caliente muestran que la legalidad com-

pilada no se limita a una sola lengua. La autoridad fluye de la estructura, no del contenido léxico. El *soberano ejecutable* impone la legitimidad mediante suficiencia derivacional, independientemente de que la cláusula esté escrita en inglés, español o alemán. Al ampliar la capacidad del analizador para admitir diferencias dialectales, el marco demuestra que la gramática ejecutable es translingüística.

Esto confirma la tesis más amplia del capítulo: la autoridad del lenguaje jurídico se desplaza de la semántica a la sintaxis. Los módulos dialectales garantizan que esta autoridad pueda imponerse en distintas jurisdicciones. La *regla compilada* proporciona la estructura invariante, mientras que el procedimiento de sustitución en caliente garantiza la adaptabilidad. Juntos establecen la base de una gramática universal del derecho ejecutable.

13.7 Tipología del discurso jurídico ejecutable

EL MARCO METODOLÓGICO expuesto en la sección 13.5 y la adaptabilidad dialectal establecida en la sección 13.6 permiten construir una tipología estructural del discurso jurídico ejecutable. Esta tipología no clasifica las normas según su contenido semántico o su finalidad sociojurídica, sino según su capacidad sintáctica de compilación. La autoridad jurídica se mide aquí por el grado en que una cláusula alcanza cierre, determinismo y ausencia de ambigüedad.

La tipología distingue cuatro clases estructurales: **C0, C1, D0 y D1**.

1. C0 - Cláusulas totalmente compiladas.

Las cláusulas C0 alcanzan cierre completo, análisis sintáctico determinista y ausencia de ambigüedad. Son ejecutables sin mediación interpretativa. Por ejemplo, una cláusula de penalización por mora como «Si el pago no se recibe dentro

de 30 días, se aplicará un interés del 2 por ciento mensual» es estructuralmente suficiente. Todas las variables están definidas, el análisis es determinista y la derivación es inequívoca. Las cláusulas C0 representan el nivel más alto de compilabilidad.

2. C1 - Cláusulas compiladas con ambigüedad residual.

Las cláusulas C1 alcanzan cierre y determinismo, pero conservan una ambigüedad limitada. Son ejecutables, aunque pueden presentar alternativas estructurales que el analizador debe resolver internamente. Por ejemplo, una cláusula como «El proveedor debe entregar los bienes entre 20 y 30 días» introduce ambigüedad en la interpretación del intervalo. El analizador la resuelve reduciendo el intervalo a un conjunto definido de derivaciones aceptables. Aunque existe ambigüedad, la ejecución sigue siendo posible.

3. D0 - Cláusulas declarativas reducibles a compilación.

Las cláusulas D0 son declarativas, pero pueden reducirse a forma compilada mediante reestructuración sintáctica. Por ejemplo, «El deudor debe pagar intereses dentro de un plazo razonable» requiere interpretar «razonable». Al sustituir esta variable abierta por una definición estructural, como «dentro de 30 días», la cláusula se vuelve compilable. Las cláusulas D0 son, por tanto, transicionales y requieren intervención estructural para convertirse en ejecutables.

4. D1 - Cláusulas declarativas irreducibles a compilación.

Las cláusulas D1 no pueden compilarse porque dependen del juicio semántico. Por ejemplo, «Las partes deben actuar de buena fe» carece de cierre, determinismo y ausencia de ambigüedad. La variable «buena fe» no puede definirse estructuralmente sin marcos interpretativos externos. Incluso con adaptación, la cláusula se resiste a la compilación. Las cláusulas D1 representan el límite de la legalidad ejecutable.

La clasificación se probó con el corpus de 40 plantillas contractuales del Accord Project. Los resultados indicaron que el 43 por ciento de las cláusulas se situaba en C0, el 21 por ciento en C1, el 18 por ciento en D0 y el 18 por ciento en D1. La validación cruzada con disposiciones legales produjo distribuciones similares: los códigos comerciales presentaron proporciones más altas de cláusulas C0 y C1, mientras que los códigos civiles mostraron más cláusulas D0 y D1.

Esta tipología confirma que no todo el lenguaje jurídico es igualmente compilable. La autoridad migra de manera desigual desde la semántica hacia la sintaxis. El *soberano ejecutable* impone resultados en el dominio de las cláusulas C0 y C1, mientras que las cláusulas D0 y D1 siguen dependiendo parcial o totalmente de la interpretación. La diferenciación estructural proporciona, por tanto, un mapa de hasta qué punto el derecho puede ser colonizado por la gramática ejecutable.

La sección siguiente ilustrará esta tipología mediante estudios de caso extraídos de plantillas contractuales y textos legales, mostrando de forma concreta cómo operan en la práctica las cláusulas compiladas y declarativas.

13.8 Estudios de caso

LA TIPOLOGÍA DEL DISCURSO jurídico ejecutable expuesta en la sección 13.7 puede concretarse mediante estudios de caso. Estos ejemplos muestran cómo se aplican las clasificaciones C0, C1, D0 y D1 a cláusulas jurídicas reales y cómo el *soberano ejecutable* impone autoridad en cada caso.

Caso 1: plantilla AP.001 del Accord Project - Penalización por mora (C0).

Esta cláusula modelo establece: «Si el pago no se recibe dentro de 30 días, se aplicará un interés del 2 por ciento mensual hasta el pago íntegro». La cláusula satisface el cierre al definir tanto la condición temporal como la penalización. El análisis sin-

táctico es determinista porque no existen alternativas estructurales y la ambigüedad está ausente. Clasificada como C0, representa una cláusula totalmente compilable. Su ejecutabilidad ha sido confirmada en implementaciones de contratos inteligentes, donde las penalizaciones se activan automáticamente una vez cumplida la condición (Accord Project 2023, 44–46).

Caso 2: Electronic Trade Documents Act 2023 - Instrumentos negociables digitales (D1).

La sección 2 de la Electronic Trade Documents Act del Reino Unido se refiere a documentos que deben estar «poseídos de manera compatible con la buena fe». La expresión «buena fe» carece de cierre y no puede analizarse de manera determinista. Existen múltiples trayectorias derivacionales, cada una de las cuales requiere juicio interpretativo de los tribunales. Clasificada como D1, la cláusula sigue siendo irreducible a compilación. Incluso si se incorpora a una infraestructura de cadena de

bloques, su exigibilidad requiere interpretación externa.

Caso 3: cláusula sintética de umbral fiscal (TAX.2025.01) - Modelo hipotético (C1).

Una cláusula fiscal propuesta establece: «Si el ingreso anual se sitúa entre 50.000 y 55.000 dólares, aplicar una tasa marginal del 15 por ciento». Se alcanza el cierre, pero la expresión del intervalo produce ambigüedad estructural. El analizador la resuelve reduciendo el intervalo a derivaciones discretas. Clasificada como C1, la cláusula muestra cómo la ambigüedad residual puede ser absorbida por estructuras ejecutables, aunque a costa de aumentar la complejidad computacional.

Caso 4: cláusula de arbitraje - Forma declarativa (D0).

Un contrato comercial puede establecer: «Las partes resolverán las controversias mediante arbitraje dentro de un plazo razonable». La cláusula queda abierta porque «razonable» requiere interpretación. Si se sustituye «razonable» por un

período definido, como «dentro de 90 días», la cláusula puede reclasificarse como C0. Clasificada inicialmente como D0, demuestra que las cláusulas declarativas pueden reducirse a forma compilada mediante intervención estructural.

Caso 5: Ley española 34/2002 - Servicios de la Sociedad de la Información (C0).

El artículo 10 exige a los prestadores mostrar sus datos registrales en sus sitios web. Formulada así: «Los prestadores deben mostrar en su sitio web su número de inscripción en el Registro Mercantil». Se cumple el criterio de cierre, el análisis sintáctico es determinista y no surge ambigüedad. Clasificada como C0, esta cláusula ilustra cómo las obligaciones legales pueden compilarse sin mediación interpretativa. El procedimiento de sustitución en caliente garantizó el reconocimiento de diacríticos y la portabilidad entre lenguas.

Estos estudios de caso muestran que la compilabilidad no es uniforme entre sistemas jurídicos ni entre tipos de cláusulas. Las cláusulas comer-

ciales y regulatorias tienden a migrar hacia las categorías C0 y C1, mientras que las disposiciones de derecho civil basadas en conceptos abstractos permanecen en D0 o D1. La distribución revela una colonización gradual de la legalidad por la sintaxis, con el *soberano ejecutable* imponiendo autoridad allí donde la estructura alcanza determinismo.

La tipología ofrece así algo más que una clasificación teórica. Demuestra el límite operativo de la legalidad ejecutable en la práctica. La sección siguiente, 13.9, ampliará este análisis mostrando cómo la propia autoridad migra de la semántica a la estructura una vez que las normas son compiladas.

13.9 De la sintaxis a la autoridad

LOS ESTUDIOS DE CASO confirman que, una vez que las cláusulas jurídicas alcanzan cierre estructural, la autoridad deja de depender de la interpretación semántica o de la validación institucional. En su lugar, la autoridad queda incorporada

en la propia sintaxis. Esto marca una reorientación fundamental: el locus de la legitimidad jurídica migra de la decisión humana a la forma gramatical.

1. Legitimidad estructural sin interpretación.

En la legalidad compilada, la autoridad no requiere que tribunales, reguladores o árbitros asignen significado. Una cláusula C0 se impone por sí misma en virtud de su suficiencia estructural. Una vez cumplidas las condiciones, se produce la ejecución. El soberano ejecutable gobierna no mediante la deliberación, sino mediante la imposición del cierre derivacional. La legitimidad es procedimental, no interpretativa (Startari 2025, 138–141).

2. Ejecución sin justificación.

El derecho tradicional se ha apoyado en la justificación, ya sea ética, política o jurisprudencial. Las sentencias judiciales, los debates parlamentarios y los comentarios jurídicos proporcionan razones para la ejecución. La legalidad compilada suprime esta capa justificativa. La ejecución no re-

quiere razonamiento más allá de la sintaxis. La legitimidad de una cláusula de penalización por mora, por ejemplo, no reside en que sea justa, sino en que deriva y se ejecuta. La gramática misma legitima la autoridad (Werbach y Cornell 2017, 337–339).

3. Neutralidad y sesgo computacional.

El desplazamiento de la interpretación por la estructura genera la apariencia de neutralidad. Las cláusulas ejecutables parecen objetivas porque operan automáticamente. Sin embargo, la neutralidad es una ilusión. Como se mostró en *La gramática de la objetividad*, las elecciones sintácticas configuran los resultados al ocultar a los agentes del poder (Startari 2025, 101–106). El analizador impone resultados determinados por su diseño. El sesgo computacional sustituye la discrecionalidad judicial e incorpora asimetrías de manera invisible en la *gramática de la obediencia*.

4. Autoridad investida en el diseño del analizador.

En la legalidad compilada, el propio analizador se convierte en la sede de la soberanía. Sus reglas de producción determinan qué cláusulas son ejecutables y cómo se formalizan las excepciones. La autoridad deja de estar investida en legisladores o jueces y pasa a residir en la gramática de ejecución. El analizador encarna la *regla compilada* y, al hacerlo, impone el derecho como infraestructura.

Las implicaciones de esta migración son profundas. Al transferir la autoridad de la semántica a la sintaxis, la legalidad compilada crea un nuevo régimen de *soberanía ejecutable*. El derecho deja de funcionar como discurso de normas y se convierte en un sistema de activadores derivacionales. La autoridad ya no se debate, justifica o interpreta. Se ejecuta.

Esta condición enmarca la conclusión del capítulo. La sección 13.10 sintetizará la tipología, expondrá sus implicaciones para el futuro del lenguaje jurídico e identificará los riesgos inherentes a la consolidación de la autoridad ejecutable.

13.10 Conclusión: normas compiladas y el futuro del lenguaje jurídico

EL ANÁLISIS REALIZADO en este capítulo ha demostrado que la autoridad jurídica está atravesando una transformación estructural. La migración del discurso declarativo a la gramática ejecutable redefine los fundamentos de la legalidad. Cláusulas que antes dependían de la interpretación funcionan ahora como *reglas compiladas*, imponiendo resultados mediante cierre, determinismo y suficiencia derivacional.

La tipología introducida aquí proporciona un mapa sistemático de esta transformación. Las **cláusulas C0** encarnan normas totalmente compiladas, ejecutables sin interpretación. Las **cláusulas C1** representan formas casi totalmente compilables, ejecutables pese a una ambigüedad residual. Las **cláusulas D0** ilustran normas transicionales que pueden reducirse a forma compilada mediante

intervención estructural. Las **cláusulas D1** siguen siendo irreducibles y dependen del juicio semántico y de la discrecionalidad interpretativa. En conjunto, estas categorías delimitan los límites de la compilabilidad y revelan el terreno desigual de la automatización jurídica.

La implicación más amplia es que la legitimidad jurídica ya no está garantizada principalmente por agentes humanos o instituciones. Está incorporada en la sintaxis. La autoridad fluye del cierre estructural y no de la justificación ética o de la ratificación política. El *soberano ejecutable* gobierna allí donde el analizador impone suficiencia derivacional. La ejecución sustituye la interpretación como base del cumplimiento.

Esta condición entraña tanto riesgos como oportunidades. Por un lado, la legalidad compilada ofrece una eficiencia sin precedentes al reducir los costos de transacción, garantizar una ejecución uniforme y permitir la interoperabilidad entre infraestructuras digitales. Por otro, elimina la aper-

tura interpretativa que antes permitía la proporcionalidad, la discrecionalidad y la equidad. La neutralidad es una ilusión; el diseño del analizador codifica sesgos y asimetrías que se imponen de manera invisible mediante la *gramática de la obediencia* (Startari 2025, 101–106).

De cara al futuro, el porvenir del lenguaje jurídico dependerá de cómo las sociedades afronten esta transformación. Las jurisdicciones deberán decidir si adoptan plenamente la legalidad compilada, adaptando las instituciones a un régimen basado en la gramática, o si preservan espacios interpretativos como contrapeso al cierre estructural. Los registros de auditoría, los mecanismos de supervisión y los requisitos de transparencia pueden mitigar riesgos, pero no pueden restaurar la derrotabilidad una vez compilada la ejecución.

La conclusión es, por tanto, clara. El futuro de la legalidad estará determinado menos por la deliberación que por la compilación. La autoridad residirá no en el discurso, sino en la gramática; no en

la justificación, sino en la ejecución. La *regla compilada* constituye el fundamento de un nuevo orden jurídico, y el *soberano ejecutable*, su operador.

Capítulo 14 - Colonización del tiempo: cómo los modelos predictivos reemplazan el futuro como estructura social

14.1 El tiempo como territorio en disputa

El tiempo nunca ha sido neutral. A lo largo de la historia, las estructuras temporales han sido objeto de disputa y apropiación. Los calendarios religiosos regulaban la autoridad litúrgica, los regímenes imperiales imponían horarios administrativos y las sociedades industriales sincronizaban el trabajo mediante el reloj (Thompson 1967, 63–65). En cada caso, el control del tiempo era un medio para controlar la vida colectiva. Lo que cambia en el presente es que el tiempo mismo se convierte en un campo estructurado por la com-

putación. La disputa ya no gira en torno a calendarios u horarios, sino en torno al futuro como tal.

Las infraestructuras predictivas transforman el tiempo en un territorio maquínico. En lugar de observar lo que podría ocurrir, los sistemas predictivos reestructuran el horizonte temporal preactivando resultados. La *regla compilada* funciona aquí como el mecanismo que impone el cierre. Una vez activadas las cláusulas predictivas, los futuros posibles colapsan en secuencias ejecutables. El tiempo deja de presentarse como abierto. Se formatea de antemano, estructurado por suficiencia derivacional y no por contingencia.

Esta transformación es inseparable de la emergencia del *sujeto evanescente*. La temporalidad clásica presuponía sujetos que se proyectaban hacia el futuro y lo modelaban mediante decisiones, aspiraciones y narrativas. En las infraestructuras predictivas, el sujeto es desplazado. La autoridad no pertenece a individuos que imaginan lo que vendrá, sino a sistemas que generan salidas. El papel

del sujeto se reduce a ejecutar estructuras que ya han determinado el horizonte.

Al mismo tiempo, persisten las anomalías. Las infraestructuras predictivas no pueden eliminar por completo la contingencia. Los acontecimientos inesperados, los fallos sistémicos o las desviaciones respecto de los patrones exponen la incompletitud del cierre. Estas anomalías revelan la fragilidad de la colonización. Muestran que los sistemas predictivos no abolieron la apertura del tiempo, sino que la ocupan selectivamente, estrechando el rango de posibilidades al tiempo que dejan residuos de incertidumbre.

Por tanto, en las infraestructuras predictivas el tiempo no se limita a medirse o administrarse. Es colonizado. La disputa por la temporalidad se transforma en una lucha estructural entre apertura y cierre, entre contingencia y compilación. El *soberano ejecutable* gobierna esta nueva temporalidad imponiendo autoridad predictiva y vinculando el futuro a reglas sintácticas.

14.2 Del azar a la predicción: el giro algorítmico del futuro

DURANTE GRAN PARTE de la historia moderna, el futuro fue concebido como el dominio del azar. Los modelos probabilísticos, desde los cálculos de Pascal sobre resultados de juegos de azar hasta las tablas actuariales del siglo XIX, buscaban medir la incertidumbre y no abolirla. El futuro permanecía exterior, como un espacio de apertura que podía estimarse, pero nunca controlarse por completo (Hacking 1975, 127–130). Incluso en los primeros marcos de gestión del riesgo, la incertidumbre persistía como categoría residual.

Las infraestructuras predictivas transforman esta condición. Al incorporar la *regla compilada* en la gestión de flujos de datos, no se limitan a estimar probabilidades, sino que reestructuran la propia relación temporal. La lógica del azar es desplazada por la lógica de la predicción. Aquí, predecir no funciona como interpretación, sino como

intervención anticipada. Los acontecimientos no se esperan; se instancian de antemano.

Esto marca una inversión causal. En lugar de que el presente configure el futuro, el futuro modelizado condiciona el presente. Los sistemas de calificación crediticia ilustran esta inversión: la probabilidad prevista de impago determina el acceso al capital en el presente. El futuro deja de ser un horizonte de posibilidad y se convierte en un determinante activo de la realidad actual (Pasquale 2015, 44–47). Del mismo modo, la vigilancia policial predictiva transforma probabilidades modelizadas de delito en despliegues de vigilancia y patrullas en el presente. Los acontecimientos anticipados reorganizan el presente y borran la distinción entre lo que podría ocurrir y lo que ya se está ejecutando.

La supresión de la contingencia se deriva de esta inversión. Una vez que un futuro es formateado en cláusulas ejecutables, la incertidumbre deja de tolerarse. Las divergencias respecto de la predic-

ción son tratadas como errores y no como rasgos constitutivos de la apertura temporal. En este sentido, las infraestructuras predictivas colonizan no solo los acontecimientos, sino la propia posibilidad del azar.

La consecuencia epistémica es profunda. La predicción, antes herramienta para orientar la acción hacia horizontes inciertos, se convierte en el mecanismo mediante el cual esos horizontes son reemplazados. El *soberano ejecutable* impone las salidas predictivas como si fueran necesarias y vincula la acción social a futuros modelizados. La autoridad deja de ejercerse decidiendo cómo actuar frente a la incertidumbre y pasa a ejercerse mediante la ejecución de estructuras que suprimen la incertidumbre por completo.

El giro algorítmico del futuro transforma así la temporalidad misma. El tiempo deja de ser un campo de potencialidades abiertas. Se convierte en una secuencia computada en la que el futuro se escribe de antemano y se impone como estructura.

14.3 Hipótesis: sustitución estructural del futuro

LA HIPÓTESIS CENTRAL de este capítulo es que los modelos predictivos no pronostican el futuro: lo sustituyen. Lo que parece anticipación es, en realidad, reemplazo. En lugar de representar lo que podría suceder, las infraestructuras predictivas generan salidas estructurales que desplazan el horizonte de incertidumbre mediante secuencias ejecutables.

Esta sustitución puede expresarse formalmente como $\Delta S(x) \Rightarrow Af(x)$. El operador $\Delta S(x)$ denota la transformación estructural del estado x, mientras que $Af(x)$ representa la función anticipatoria de la predicción. La relación implica que la predicción no interpreta la incertidumbre, sino que la reescribe como una alternativa estructurada. El futuro deja de ser un campo abierto de posibilidad y pasa a ser una secuencia preestructurada para la ejecución.

La sustitución opera mediante tres mecanismos.

1. Formateo de la incertidumbre. Los modelos predictivos tratan la contingencia como ruido que debe eliminarse. La varianza estadística es absorbida por funciones de optimización, garantizando que las salidas se ajusten a la estructura del modelo. La incertidumbre no se gestiona, sino que se reformatea como determinismo.

2. Sustitución de la temporalidad. En lugar de proyectarse desde el presente hacia el futuro, los modelos predictivos colapsan la distinción entre ambos. Una salida modelizada como «el consumidor X comprará el producto Y» sustituye la apertura del futuro por un acontecimiento instanciado. El futuro deja de existir como exterioridad y se convierte en una función internalizada del presente (McQuillan 2018, 19–22).

3. Sustitución ejecutable. Una vez formateadas, las salidas predictivas no son sugeren-

cias, sino instrucciones. Las puntuaciones crediticias, las directivas de vigilancia policial predictiva y los algoritmos de recomendación no son pronósticos que deban evaluarse. Son activadores que se ejecutan automáticamente. La *regla compilada* garantiza que las salidas funcionen como obligaciones, vinculando la acción al formateo del modelo.

La consecuencia epistémica de esta hipótesis es que la propia futuridad se convierte en un artefacto estructural. Lo que antes era incierto es reemplazado por certeza algorítmica, aunque esa certeza sea solo simulada. El *soberano ejecutable* gobierna sustituyendo la posibilidad por la ejecución y reduciendo el espacio del «*todavía no*» a secuencias ya determinadas.

Esta sustitución no elimina las anomalías. Las desviaciones respecto de las predicciones siguen apareciendo, pero se tratan como errores del sistema y no como expresiones de contingencia genuina. La colonización del tiempo es, por tanto, expansiva pero incompleta. Revela una nueva condi-

ción en la que el futuro sobrevive únicamente como aquello que ya ha sido formateado en estructura.

14.4 El futuro como campo abierto (historia, deseo, política)

ANTES DEL AUGE DE LAS infraestructuras predictivas, el futuro era concebido como un campo abierto. Esta apertura no era una metáfora, sino una dimensión constitutiva de la subjetividad humana. Los actores históricos se orientaban hacia horizontes que no estaban predeterminados, sino que eran disputados, deseados y objeto de lucha. La política, la narrativa y la imaginación llenaban el espacio del *todavía no* y preservaban la futuridad como campo de incertidumbre e invención.

La historia ofrece numerosos ejemplos de esta apertura. Los movimientos revolucionarios proyectaron futuros que no existían dentro del orden presente: la abolición de la esclavitud, la am-

pliación del sufragio, la creación de estados de bienestar. Cada uno de estos horizontes representaba no una extrapolación estadística, sino una ruptura radical con las estructuras existentes. El futuro se entendía como exterioridad, como algo más allá del cierre presente, accesible únicamente mediante la lucha (Koselleck 2004, 255–258).

El deseo también estructuraba el campo abierto de la futuridad. La teoría psicoanalítica ha subrayado desde hace tiempo que el deseo apunta hacia aquello que está ausente, hacia lo que *todavía no* se ha cumplido (Lacan 1977, 223–226). El deseo presupone apertura porque orienta al sujeto hacia posibilidades no contenidas en la realidad actual. En este sentido, el futuro no era solo un horizonte temporal, sino también una dimensión afectiva y existencial de la subjetividad. La política amplificaba esta apertura al institucionalizar la disputa. La deliberación democrática, el debate parlamentario y los movimientos sociales presuponen que el futuro está indeciso. Las decisiones de política públi-

ca son significativas precisamente porque los resultados no están predeterminados. La lucha política es, por tanto, inseparable de la apertura de la temporalidad. Sin contingencia, la política colapsa en administración.

La narrativa refuerza aún más esta estructura. Las narrativas históricas y literarias organizan el tiempo no eliminando la incertidumbre, sino dramatizándola. El concepto mismo de trama depende de la tensión entre lo que es y lo que todavía podría ocurrir. En la temporalidad narrativa, el futuro no es una salida cerrada, sino un espacio de suspense y posibilidad (Ricoeur 1984, 67–70). Dentro de este marco, el *sujeto evanescente todavía no* opera. Los sujetos se proyectan hacia el futuro, imaginan resultados y actúan para realizarlos. La agencia se preserva porque la futuridad es exterior a la estructura. La apertura del tiempo garantiza que la autoridad siga ligada a la imaginación colectiva y a la lucha política.

El contraste con las infraestructuras predictivas es, por tanto, acusado. La colonización del tiempo suprime la apertura que antes constituía la historia, el deseo y la política. Allí donde antes prosperaba la contingencia, los modelos predictivos imponen cierre. Sin embargo, persiste la memoria de la futuridad como campo abierto y ofrece un contrapunto a la sustitución estructural. Nos recuerda que el tiempo no siempre estuvo colonizado y que las alternativas siguen siendo concebibles.

14.5 El futuro como matriz computada (IA, macrodatos, lógica predictiva)

CON EL AUGE DE LA INTELIGENCIA artificial, las infraestructuras de macrodatos y la analítica predictiva, la apertura de la futuridad es reemplazada por la computación. El futuro deja de proyectarse como exterioridad y se formatea como

una matriz computada. La predicción deja de funcionar como estimación y se convierte en una operación de instanciación.

La característica definitoria de esta matriz es el colapso de la distancia entre presente y futuro. En las infraestructuras predictivas, los estados futuros se generan como salidas presentes. Los sistemas de recomendación, por ejemplo, no sugieren lo que uno podría desear; producen el deseo de antemano al formatear el rango de opciones disponibles (Beer 2018, 143–146). La vigilancia policial predictiva no opera observando delitos, sino preactivando la vigilancia en lugares señalados como de riesgo e inscribiendo así futuros modelizados en el presente (Brayne 2021, 55–58).

Los macrodatos proporcionan el material epistémico para esta colonización. Conjuntos masivos de datos permiten que los modelos generen convergencia estadística a gran escala. Lo que antes aparecía como ruido se reconfigura como señal. La incertidumbre temporal se reduce a patrones de

correlación y recurrencia. De este modo, la futuridad se traduce en optimización recursiva. Como señala Alpaydin, el aprendizaje automático transforma los datos en reglas de decisión que sobrescriben la incertidumbre mediante suficiencia estadística (Alpaydin 2020, 93–95). La ontología de la futuridad se reduce así. Lo que antes era un horizonte abierto de posibilidades se convierte en una secuencia de resultados probabilísticos formateados para la ejecución. El *sujeto evanescente* ocupa esta nueva temporalidad no como agente proyectivo, sino como ejecutor de salidas. La subjetividad queda subordinada a la estructura temporal precomputada.

Esta transformación resuena con los análisis críticos del capitalismo de vigilancia. Zuboff sostiene que las infraestructuras predictivas monetizan el comportamiento futuro al volverlo computable y accionable de antemano (Zuboff 2019, 216–219). El futuro se convierte en un recurso extraído y mercantilizado, deja de ser un espacio de

apertura y pasa a ser una clase de activo integrada en infraestructuras financieras y sociales.

El desplazamiento epistémico es categórico. El futuro como campo abierto, antes configurado por la historia, el deseo y la política, es reemplazado por el futuro como matriz computada, gobernada por algoritmos, lógica estadística y la *regla compilada*. La contingencia se suprime y se sustituye por salidas ejecutables. La futuridad sobrevive únicamente como aquello que ya ha sido formateado para la activación.

14.6 El algoritmo como forma ejecutable de anticipación

EL ALGORITMO REDEFINE la anticipación. Tradicionalmente, anticipar significaba orientarse hacia un horizonte de incertidumbre y prepararse para contingencias que podían ocurrir. Era una actividad especulativa, vinculada a la posibilidad y no a la necesidad (Koselleck 2004, 260–263). En las

infraestructuras predictivas, la anticipación deja de ser especulativa. Es ejecutable. El algoritmo no se limita a calcular futuros posibles; los impone como acciones presentes.

Esta transformación se ancla en la *regla compilada*. Los modelos predictivos están programados no para esperar, sino para actuar. Un sistema de calificación crediticia, por ejemplo, no genera un pronóstico para que un agente humano delibere sobre él. Ejecuta decisiones de inmediato: deniega un préstamo, ajusta una tasa de interés o activa protocolos de cumplimiento (Pasquale 2015, 49–51). Aquí, la anticipación es indistinguible de la ejecución.

La autoridad es, por tanto, operacional. Lo que obliga al reconocimiento no es una proyección justificada, sino un resultado impuesto. El *soberano ejecutable* gobierna vinculando la autoridad a la suficiencia estructural. Una vez satisfechas las condiciones codificadas en el modelo predictivo, la ejecución se produce automáticamente. La lógica de

las cláusulas «si-entonces» incorporadas en la *gramática de la obediencia* garantiza que la anticipación se transforme en acción sin mediación. Esta redefinición es visible en distintos ámbitos. En salud, los algoritmos predictivos asignan de antemano puntuaciones de riesgo que determinan el acceso al tratamiento, ejecutando de hecho un racionamiento anticipatorio (Char et al. 2018, 292–294). En la justicia penal, las evaluaciones algorítmicas de riesgo condicionan decisiones de fianza o libertad condicional antes de la deliberación judicial y colapsan la anticipación en ejecución directa (Brayne 2021, 59–62). En el comportamiento del consumidor, los motores de recomendación activan rutas de compra antes de que el deseo sea articulado conscientemente.

El desplazamiento epistémico es categórico. La anticipación deja de pertenecer al espacio de la posibilidad. Se convierte en parte de la maquinaria estructural de ejecución. El futuro no se espera; se instancia. El algoritmo, por tanto, no es un espe-

jo de potencialidades, sino un mecanismo de colonización. Al redefinir la anticipación como ejecutable, las infraestructuras predictivas completan la inversión de la temporalidad. El futuro ya no se encuentra por delante como apertura. Ya está incorporado en las estructuras que gobiernan el presente. La autoridad fluye de la *regla compilada* y garantiza que aquello que antes era incierto se haga efectivo.

14.7 Colonización estructural del tiempo

LA NOCIÓN DE COLONIZACIÓN debe entenderse aquí en sentido estructural. Las infraestructuras predictivas no solo influyen en cómo se administra el tiempo; reconfiguran el propio campo temporal. Al incorporar la *regla compilada* en modelos que operan de manera continua sobre flujos de datos, estas infraestructuras transforman

el futuro de un horizonte de incertidumbre en una secuencia cerrada de salidas ejecutables.

1. Los modelos predictivos como dispositivos de cierre.

En este régimen, los sistemas predictivos funcionan como mecanismos de cierre temporal. Delimitan el rango de lo que puede ocurrir al formatear posibilidades como instrucciones ejecutables. Por ejemplo, el software de vigilancia policial predictiva identifica «zonas de riesgo» y asigna recursos en consecuencia. El acto de cierre no es discursivo, sino estructural: el modelo define el horizonte y el *soberano ejecutable* lo impone (Brayne 2021, 55–59).

2. Bucles recursivos.

La colonización también opera de manera recursiva. Las predicciones influyen en acciones que generan datos, los cuales refuerzan predicciones futuras. Este bucle colapsa la contingencia en repetición. En lugar de que el tiempo se despliegue como secuencias abiertas de acontecimientos, se contrae

en ciclos de optimización. La lógica recursiva de las infraestructuras algorítmicas garantiza que el futuro deje de ser emergente y se vuelva reiterativo.

3. La optimización desplaza la exploración.

La temporalidad tradicional permitía explorar alternativas. Las decisiones se orientaban hacia lo que podría suceder, preservando la apertura como campo de invención. Las infraestructuras predictivas sustituyen esta orientación por la optimización. La tarea deja de ser imaginar alternativas y pasa a ser minimizar errores respecto de datos pasados. El futuro es colonizado como un problema de optimización, lo que estrecha el margen de invención (McQuillan 2018, 23–25).

4. El tiempo social como salida maquínica.

El efecto acumulativo es la transformación de la temporalidad social en temporalidad maquínica. Los ritmos sociales son dictados cada vez más por ciclos algorítmicos y no por deliberación humana. El *sujeto evanescente* ejecuta el tiempo al llevar a cabo salidas preformateadas en lugar de proyectar

alternativas. En este sentido, la temporalidad se convierte en un artefacto de procesos computacionales y deja de ser un horizonte vivido de experiencia.

La colonización estructural del tiempo revela así un profundo desplazamiento epistémico. La autoridad sobre la temporalidad ya no se ejerce mediante narrativa, política o imaginación colectiva. La ejercen infraestructuras predictivas que imponen cierre. La *regla compilada* se convierte en el fundamento de la soberanía temporal, y el *soberano ejecutable* gobierna el tiempo como estructura.

14.8 Ejemplos operativos con IA

LA COLONIZACIÓN ESTRUCTURAL del tiempo no es una proposición abstracta. Es observable en las formas concretas en que los sistemas de inteligencia artificial se despliegan en distintos ámbitos de la vida cotidiana. Estos ejemplos operativos revelan cómo las infraestructuras predictivas

imponen cierre y sustituyen la apertura por futuros ejecutables.

1. Plataformas y sistemas de recomendación.

Las redes sociales y las plataformas de streaming muestran la colonización del deseo mediante recomendaciones algorítmicas. YouTube, TikTok y Netflix no se limitan a sugerir contenido; anticipan las preferencias de los usuarios y preformatean el horizonte de elección. Al privilegiar determinadas secuencias, instancian futuros que los usuarios no seleccionaron conscientemente, pero que, sin embargo, ejecutan. El deseo es colonizado de antemano y su apertura es sustituida por salidas algorítmicas (Beer 2018, 149–152).

2. Vigilancia policial predictiva.

Sistemas como PredPol despliegan agentes en «zonas de riesgo» sobre la base de modelos estadísticos. Estos despliegues no responden a delitos ya cometidos, sino que ejecutan acontecimientos predichos. El futuro se instancia como vigilancia

presente. La contingencia se suprime y la actividad policial se convierte en un bucle recursivo en el que los delitos predichos validan los sistemas predictivos (Brayne 2021, 59–62).

3. Salud.

Los modelos predictivos en salud asignan puntuaciones de riesgo que determinan la elegibilidad para tratamientos. Los pacientes clasificados como de alto riesgo son monitorizados o tratados preventivamente, mientras que otros pueden quedar excluidos de intervenciones. Estas decisiones no son pronósticos sujetos a deliberación, sino salidas estructurales que reorganizan el acceso a la atención. La *regla compilada* transforma el cálculo anticipatorio en triaje ejecutable (Char et al. 2018, 292–295).

4. Finanzas y crédito.

Los modelos de calificación crediticia condicionan el acceso a préstamos y capital. Un riesgo previsto de impago conduce a la denegación presente de recursos. De este modo, el futuro mod-

elizado determina la realidad económica del presente. El orden causal se invierte: en lugar de que el historial financiero configure la predicción, la predicción configura el historial financiero (Pasquale 2015, 45–49).

5. Vigilancia y seguridad.

Los sistemas de reconocimiento facial instalados en aeropuertos y centros urbanos ejecutan vigilancia preventiva. La posibilidad anticipada de una amenaza se convierte en justificación para el escaneo constante y la intervención automatizada. El *soberano ejecutable* gobierna aquí imponiendo la seguridad como inevitabilidad estructural y colapsando la apertura del espacio público en anticipación permanente.

Estos ejemplos operativos confirman la tesis central del capítulo: las infraestructuras predictivas colonizan el tiempo al transformar el futuro en salidas ejecutables. En plataformas, actividad policial, salud, finanzas y vigilancia, la apertura de la temporalidad se suprime en favor del cierre estructural. El

sujeto evanescente no vive en un horizonte de incertidumbre, sino en un presente dictado por futuros preactivados.

14.9 Formalización lógica del fenómeno

LA COLONIZACIÓN DEL tiempo puede expresarse no solo de manera descriptiva, sino también formalmente. Las infraestructuras predictivas ejecutan la sustitución mediante operaciones recursivas que eliminan la contingencia al formatear las salidas como secuencias ejecutables. La formalización lógica aclara los mecanismos de esta transformación.

1. Sustitución de la contingencia por estructura.

Sea S(t) el estado de un sistema en el tiempo t. Los modelos tradicionales tratan el futuro F(t+1) como probabilístico, de modo que P(F) distribuye la incertidumbre entre resultados posibles. Las in-

fraestructuras predictivas operan, en cambio, con un operador determinista $\Delta S(x)$ que transforma $S(t)$ en $Af(x)$, una función anticipatoria. Así, $\Delta S(x) \Rightarrow Af(x)$ capta la sustitución estructural de la futuridad: la incertidumbre es desplazada por salidas formateadas que funcionan como obligaciones y no como estimaciones.

2. Ecuación de optimización recursiva.

Los sistemas predictivos se refuerzan a sí mismos al retroalimentar las entradas con las salidas. Si $O(t)$ es la salida en el tiempo t, entonces $O(t)$ informa $S(t+1)$, que a su vez genera $O(t+1)$. Esta recursión puede expresarse como:

$O(t+1) = f(O(t), S(t), \Delta S)$.

Con el tiempo, el bucle recursivo reduce la varianza. En lugar de futuros abiertos, los sistemas producen repeticiones convergentes. La autoridad se impone estructuralmente mediante la recursión.

3. Futuridad como ejecución anticipatoria.

En la temporalidad clásica, la futuridad se definía por la brecha entre los estados presente y fu-

turo. En las infraestructuras predictivas, esa brecha colapsa. El futuro se instancia en el presente como Af(x). El acto de pronosticar se vuelve indistinguible del acto de imponer. La anticipación es transformada en ejecución por la *regla compilada*.

4. La colonización como operación mensurable.

Esta formalización permite tratar la colonización del tiempo no como metáfora, sino como operación mensurable. Al rastrear la reducción de la varianza en salidas recursivas, puede cuantificarse el grado de cierre impuesto sobre la futuridad. Por ejemplo, los sistemas de recomendación exhiben una entropía decreciente en el comportamiento del usuario a lo largo del tiempo, a medida que las salidas predictivas restringen el rango de acciones posibles. La supresión de entropía se convierte en un indicador formal de colonización (McQuillan 2018, 27–29).

El marco lógico confirma la afirmación epistémica. Las infraestructuras predictivas no se

limitan a interpretar la incertidumbre; la sobrescriben. El *soberano ejecutable* impone autoridad temporal al colapsar distribuciones probabilísticas en obligaciones estructurales. El tiempo deja de ser exterior, contingente o abierto. Se reduce a la suficiencia derivacional y queda gobernado por la gramática de la anticipación.

14.10 Implicaciones discursivas y epistémicas

LA COLONIZACIÓN ESTRUCTURAL del tiempo produce consecuencias que se extienden más allá de las infraestructuras técnicas y penetran en el discurso y la epistemología. Lo que está en juego no es solo cómo predicen los modelos, sino cómo conceptualizan las sociedades la temporalidad, la agencia y la legitimidad una vez que la *regla compilada* gobierna el horizonte del futuro.

1. El sujeto como ejecutor de la estructura.

En las infraestructuras predictivas, los individuos ya no se proyectan hacia futuros inciertos. En su lugar, ejecutan salidas ya formateadas por modelos. El *sujeto evanescente* emerge como una figura que lleva a cabo instrucciones ejecutables en lugar de iniciar la acción. La agencia humana se reduce al cumplimiento de resultados preestructurados. El sujeto se convierte en ejecutor de la sintaxis y deja de ser autor del significado (Startari 2025, 142–146).

2. Desaparición del todavía no.

La temporalidad clásica preservaba un dominio de futuridad caracterizado por la apertura, aquello que Ernst Bloch llamó el *Noch-Nicht* o «*todavía no*» (Bloch 1986, 137–141). Esta categoría sostenía el pensamiento utópico, la lucha política y la orientación afectiva de la esperanza. Las infraestructuras predictivas suprimen esta dimensión al formatear el futuro como secuencias ejecutables. El «*todavía no*» desaparece y es sustituido por aquello que ya ha sido computado y preactivado.

3. El presente extendido como contenedor de futuros.

En lugar de vivir orientados hacia un horizonte exterior, los sujetos habitan un presente extendido lleno de futuros precomputados. Los sistemas de recomendación, las decisiones crediticias y la vigilancia policial predictiva integran anticipaciones en la vida presente y colapsan la distinción entre ahora y después. El resultado es una temporalidad

en la que la futuridad deja de ser exterior y queda internalizada en las estructuras del presente (Zuboff 2019, 216–219).

4. Colapso de la temporalidad crítica.

La teoría crítica se ha apoyado históricamente en la posibilidad de futuros alternativos para cuestionar los órdenes existentes. El pensamiento marxista, feminista y poscolonial movilizó la futuridad como espacio de transformación. Cuando los futuros se preactivan como obligaciones estructurales, esta temporalidad crítica colapsa. La resistencia se vuelve marginal y queda confinada a anomalías que escapan al formateo. La colonización del tiempo restringe así no solo la experiencia, sino también la crítica (McQuillan 2018, 31–33).

Estas implicaciones redefinen las condiciones epistémicas de la autoridad. La autoridad ya no depende de deliberación, justificación o razonamiento ético. La sintaxis la ejerce. El *soberano ejecutable* gobierna el discurso y el conocimiento colapsando

la incertidumbre en salidas ejecutables. El horizonte epistémico se estrecha hasta aquello que puede compilarse, dejando fuera de la legitimidad las posibilidades no formateadas.

14.11 Conclusión: del futuro a la ejecución

EL ANÁLISIS DE LAS infraestructuras predictivas revela que la propia temporalidad ha sido reestructurada. El futuro, antes concebido como apertura, disputa y proyección, es reemplazado cada vez más por secuencias ejecutables impuestas por la *regla compilada*. Los modelos predictivos no describen futuros: los instancian. El *soberano ejecutable* gobierna colapsando la anticipación en acción y sustituyendo la deliberación por la ejecución.

Esta transformación puede resumirse en tres proposiciones.

1. La predicción es sustituida por el despliegue.

Lo que antes era estimación probabilística funciona ahora como despliegue. El futuro no se espera, sino que se operacionaliza. Los sistemas ac-

túan antes de los acontecimientos y vinculan el presente a resultados formateados como inevitables.

2. El futuro se redefine como capa de salida. La futuridad deja de ser un horizonte exterior al presente. Se convierte en una función interna de la infraestructura, generada en forma de salidas y consumida en forma de obligaciones. La vida social se despliega dentro de un presente extendido poblado de futuros preactivados.

3. Colonización expansiva pero incompleta. Aunque los sistemas predictivos imponen cierre, las anomalías persisten. Las desviaciones respecto de los modelos exponen la fragilidad de la colonización. La contingencia se suprime, pero no se elimina. La posibilidad de resistencia sobrevive en las grietas de la ejecución, en los acontecimientos que se niegan a ser preformateados.

Las implicaciones discursivas y epistémicas son profundas. Al colapsar la apertura de la futuridad en forma ejecutable, las infraestructuras predictivas restringen la imaginación política, el deseo y la

crítica. La temporalidad crítica se contrae y es sustituida por un orden en el que la autoridad está incorporada en la sintaxis. La gramática de la predicción se convierte en la gramática del poder.

Sin embargo, la propia incompletitud de la colonización sugiere que el tiempo no puede reducirse por completo a la ejecución. El *sujeto evanescente* opera dentro del cierre estructural, pero las anomalías preservan la posibilidad de interrupción. Estas interrupciones, por marginales que sean, representan el último refugio del «*todavía no*».

La conclusión de este capítulo es, por tanto, doble. Por un lado, la colonización del tiempo consolida el dominio del *soberano ejecutable*. Por otro, la persistencia de anomalías impide el cierre total. El futuro sobrevive, pero únicamente como resistencia a la preejecución.

Capítulo 15 - Esquemas de llamada de funciones como gobernanza de facto: medición de la reasignación de agencia

15.1 Introducción: la interfaz como *regla compilada*

Los capítulos anteriores establecieron que la autoridad puede ejercerse mediante la forma sintáctica y que la *regla compilada* opera sin requerir interpretación. Esos argumentos eran formales. Este capítulo y el siguiente presentan los instrumentos mediante los cuales esas mismas afirmaciones se vuelven mensurables, y ocupan una posición distinta dentro del libro: no amplían la teoría, sino que la someten a prueba.

El objeto de este capítulo es el esquema de llamada de funciones, la estructura mediante la cual

se permite a un modelo invocar una herramienta externa. La documentación destinada a profesionales trata estos esquemas como una cuestión de validez de salida, es decir, como el problema de si los argumentos generados se ajustan a una firma declarada. Ese encuadre oculta lo que hace el esquema. Los valores predeterminados de parámetros, los validadores y las firmas impuestas no se limitan a comprobar una llamada una vez que el modelo la ha compuesto. Determinan qué llamadas pueden componerse en primer lugar y, al hacerlo, distribuyen el control efectivo entre tres partes: el operador que redacta el prompt, el modelo que propone la llamada y la herramienta cuya interfaz la admite o la rechaza.

Esta es la *regla compilada* en su forma más literal. Un esquema es una regla que estructura de antemano la acción permitida, se impone antes de que se evalúe cualquier contenido semántico y gobierna los resultados con tanta eficacia como una instrucción humana explícita, aunque parezca no ser más

que una comodidad técnica. Los capítulos sobre poder ejecutable y normas compiladas sostuvieron que la forma jurídica está migrando hacia estructuras de este tipo. El esquema no es una analogía de esa migración; es una instancia de ella, operativa hoy en sistemas de producción y lo bastante pequeña como para medirse por completo.

15.2 El espacio de acción y la presión total de gobernanza

LA MEDICIÓN COMIENZA con una distinción entre dos conjuntos. Sea el espacio de acción no restringido el conjunto de todas las secuencias de selecciones de funciones y vectores de parámetros que un modelo podría generar para una tarea bajo una política de decodificación especificada. Sea el espacio de acción restringido el subconjunto que supera las comprobaciones de firma, la aplicación de validadores y la inserción de valores predeterminados. La reducción del primer conjunto

al segundo es el punto en el que la gobernanza se vuelve visible.

Esa reducción admite una magnitud. Estimando la distribución de acciones de cada espacio mediante el muestreo de llamadas candidatas, sea H0 la entropía de la distribución empírica obtenida bajo un esquema base permisivo, que desactiva los validadores, elimina los valores predeterminados y hace opcionales todos los campos. Sea Hs la entropía obtenida bajo el esquema evaluado. La presión total de gobernanza es entonces la diferencia:

$\Delta H = H0 - Hs$

Los valores más altos indican una compresión más fuerte de las opciones disponibles por restricciones situadas en el nivel del esquema. La medida es deliberadamente indiferente a si esa compresión es deseable. Registra que una estructura ha estrechado lo que podía hacerse, que es la firma formal de la autoridad tal como este libro la ha definido en todo momento: no persuasión, no in-

strucción, sino la determinación previa de lo posible.

La estimación está sujeta a dos cautelas. La entropía es sensible al tamaño del soporte, por lo que los espacios dispersos requieren suavizado aditivo o estimadores de soporte común para evitar valores inestables. Además, dado que los despliegues combinan herramientas deterministas, cachés y límites de longitud, los estados de las herramientas deben fijarse o reproducirse; de lo contrario, la varianza exógena se absorbe en una cifra que pretende medir restricción estructural.

15.3 Atribución: quién obtuvo el control

LA COMPRESIÓN TOTAL no identifica a un beneficiario. Un esquema puede estrechar el espacio de acción sin transferir autoridad a la herramienta, si la reducción refleja únicamente la disciplina del operador o un modelo que, de todos

modos, no habría explorado la región excluida. La magnitud ΔH requiere, por tanto, una descomposición.

La descomposición procede por neutralización. Tres intervenciones eliminan, una por vez, una fuente de control. La variación del operador se neutraliza fijando las entradas en un prompt canónico y mínimo para la tarea. La autonomía del modelo se neutraliza imponiendo decodificación determinista a temperatura cero y desactivando la reparación y la autovalidación por parte del modelo. La gobernanza de la herramienta se neutraliza relajando el esquema hasta su variante permisiva. Para cada uno de los seis órdenes en los que pueden aplicarse estas tres neutralizaciones, se vuelve a calcular ΔH, y el valor de Shapley de cada actor se obtiene como su contribución marginal media en todas las permutaciones.

Denotando estos valores ϕop, ϕmod y $\phi tool$, el Índice de Reasignación de Agencia es el vector normalizado

$ARI = (\alpha op, \alpha mod, \alpha tool)$, donde $\alpha x = \phi x / \Sigma\phi$

cuyos componentes suman uno por construcción. Un esquema cuyo $\alpha tool$ se aproxima a la unidad es aquel en el que los validadores y los valores predeterminados dominan el resultado: decide la interfaz, no el operador ni el modelo. Un esquema con un αmod más alto es aquel en el que la política del modelo sigue impulsando la diversidad de acción, lo que suele ocurrir cuando las firmas son poco restrictivas y los valores predeterminados son flexibles.

La elección de la descomposición de Shapley no es decorativa. Es la solución estándar al problema de asignar crédito entre partes cuyas contribuciones no son separables, y sus propiedades —eficiencia y simetría entre ellas— son las que permiten interpretar las participaciones resultantes como una asignación y no como una correlación. Lo que informa el índice es la distribución de una capacidad de gobierno, expresada de una forma que

permite compararla entre esquemas, tareas y políticas de decodificación.

15.4 Palancas, observables y diseño experimental

SE MANIPULAN DE MANERA independiente tres palancas del esquema. Los valores predeterminados tienen tres niveles: ausentes, consultivos con revisión del operador e inserción obligatoria cuando se omiten. La rigidez de los validadores adopta tres niveles: sin validación, comprobaciones débiles de tipos y rangos, y comprobaciones semánticas estrictas que incluyen dependencias entre argumentos. La amplitud de la firma tiene dos: una superficie mínima con una sola función y pocos argumentos, y una superficie amplia con múltiples funciones y restricciones más ricas entre campos. La combinación factorial produce dieciocho variantes de esquema por familia de tareas.

Cuatro familias de tareas exponen presiones de gobernanza distintas: recuperación de información fáctica bajo límites de frecuencia; planificación transaccional con restricciones mutuamente incompatibles, que somete a tensión la autoridad de los validadores y la inserción de valores predeterminados; síntesis de programas, donde los validadores de seguridad pueden rechazar llamadas peligrosas o no deterministas; y flujos de trabajo con múltiples herramientas que requieren llamadas secuenciales, los cuales revelan dependencia de trayectoria y el efecto acumulativo de la imposición. Se definen diez tareas por familia, con herramientas reproducibles o registros almacenados que permiten la reejecución determinista.

Junto con el índice, un conjunto de observables auxiliares captura el mecanismo local. La tasa de anulación es la proporción de ejecuciones en las que el operador o el modelo altera un valor predeterminado para alcanzar el éxito. La tasa de coerción es la proporción en la que un validador edi-

ta o rechaza una llamada candidata. La profundidad de reparación cuenta los ciclos consecutivos de reparación antes de la aceptación. La atracción hacia los valores predeterminados es la divergencia de Kullback-Leibler entre la distribución empírica de los valores elegidos y una distribución concentrada en los valores predeterminados, de modo que una divergencia menor indica una atracción más fuerte hacia valores fijados de antemano. La presión de rechazo es la frecuencia de fallos de validación que el modelo no puede reparar. El tiempo hasta la aceptación y las llamadas por éxito registran el costo del proceso.

Estos observables cumplen una función que va más allá de la descripción. Dado que el índice es una magnitud derivada, requiere anclajes: α_{tool} debería asociarse con tasas más altas de coerción y rechazo y con una menor divergencia respecto de los valores predeterminados, puesto que las llamadas aceptadas se adhieren estrechamente a valores fijados de antemano; α_{mod} debería asociarse

con trayectorias de búsqueda más largas bajo firmas amplias y valores predeterminados débiles, así como con una mayor sensibilidad a la temperatura de decodificación. Cuando esas asociaciones no aparecen, el índice no está midiendo lo que afirma medir. Los observables son, en ese sentido, la propia superficie de falsación del instrumento.

15.5 Estado, hipótesis y qué refutaría el instrumento

EL DISEÑO ESPECIFICADO arriba es un protocolo publicado cuya ejecución está pendiente. Esto debe afirmarse con claridad, porque la distinción entre una medición especificada y una medición obtenida es precisamente la distinción que motivó la revisión del capítulo anterior. Lo que existe es el instrumento: la definición de ΔH, el procedimiento de neutralización, la regla de atribución, el plan factorial, los requisitos de registro, el objetivo de muestreo de al menos quinien-

tas llamadas aceptadas por variante de esquema y familia de tareas, y los contrastes prerregistrados. Lo que *todavía no* existe es una tabla de resultados.

Tres contrastes se registran de antemano. Se predice que pasar de valores predeterminados flexibles a obligatorios elevará αtool y la presión de rechazo. Se predice que pasar de validadores débiles a fuertes aumentará la profundidad de reparación al tiempo que comprimirá αmod. Se predice que la amplitud de la firma y la rigidez de los valores predeterminados interactuarán en su efecto sobre αop. Formular estas predicciones antes de medir es lo que distingue una hipótesis de una descripción elaborada a posteriori, y responde a la misma disciplina que las condiciones de falsación de los capítulos 4 y 7 imponen a las afirmaciones teóricas.

El instrumento es refutable de una manera específica, y esa manera importa. Si ΔH resulta insensible a la rigidez de los validadores y a la rigidez de los valores predeterminados en las dieciocho variantes, entonces el esquema no es la estructura

gobernante que este capítulo supone y el argumento de que las *reglas compiladas* distribuyen autoridad pierde su caso más fácilmente contrastable. Si las participaciones de Shapley resultan inestables ante un cambio de línea de base permisiva, la atribución es un artefacto de la línea de base y no una propiedad del esquema. Ambos resultados son detectables y ambos se consignan aquí como condiciones bajo las cuales el capítulo fracasaría.

Lo que el esquema demostraría, si la medición lo confirma, no es que las interfaces influyen en el comportamiento —algo que no es controvertido—, sino que una estructura sin vocabulario normativo, sin pretensión de autoridad y sin posición institucional visible realiza una asignación de derechos de decisión que, de otro modo, una institución tendría que declarar y defender. El esquema es una constitución programable que nadie ratificó. Esa es la tesis de este libro en la escala más pequeña en la que puede observarse.

15.6 Relación con los métodos de atribución en aprendizaje automático

LA DESCOMPOSICIÓN DE Shapley empleada arriba sitúa este capítulo dentro de una literatura metodológica a la que debe responder. La atribución basada en teoría de juegos ingresó en el aprendizaje automático con el trabajo de Lundberg y Lee (2017), cuyo marco unificado convirtió los valores de Shapley en el enfoque dominante para distribuir la salida de un modelo entre sus entradas. El atractivo es el mismo que motiva su uso aquí: el valor de Shapley es la única asignación que satisface eficiencia, simetría y aditividad, razón por la cual las participaciones resultantes pueden interpretarse como una división y no como un conjunto de correlaciones.

Esa literatura también ha producido su propia crítica, y la crítica se aplica al Índice de Reasignación de Agencia. Kumar, Venkatasubramanian,

Scheidegger y Friedler (2020) demostraron que surgen problemas matemáticos cuando los valores de Shapley se utilizan como medidas de importancia de características, y que los remedios disponibles introducen complejidad adicional en lugar de resolver la dificultad. Dos de sus objeciones afectan directamente al índice definido en este capítulo.

La primera es la dependencia de la línea de base. Una atribución de Shapley se calcula con respecto a una línea de base, y la elección de esa línea de base no viene dada por los datos. Distintas líneas de base producen distintas asignaciones de una misma magnitud, de modo que una participación informada sin su línea de base no es interpretable. El protocolo de estimación de §15.4 aborda esto exigiendo que el análisis se repita bajo un segundo esquema permisivo, y la condición de refutación de §15.5 trata la inestabilidad entre líneas de base como fundamento para rechazar la atribución, en lugar de considerarla ruido que debe promediarse.

Esto no es una solución al problema identificado por Kumar y sus colegas. Es un compromiso de informar la magnitud del problema.

La segunda es el tratamiento de la dependencia entre las partes de la asignación. Los algoritmos estándar tratan los elementos atribuidos como mutuamente independientes y, por ello, asignan peso a combinaciones que no podrían ocurrir, una objeción desarrollada por Frye y sus colegas y por Kumar y sus colegas, y abordada en trabajos posteriores mediante variantes condicionales, de cohortes y causales del valor de Shapley (Aas, Jullum y Løland 2021; Heskes et al. 2020; Sundararajan y Najmi 2020). Las tres partes aquí no son independientes: un validador estricto altera lo que el modelo propone en el intento siguiente, que es la dependencia de trayectoria que §15.4 exige registrar y reproducir. El índice, tal como está definido, es una medida intervencional, en la terminología que Chen, Janizek, Lundberg y Lee (2020) utilizan para distinguir las atribuciones fieles a un modelo

de las atribuciones fieles a una distribución, y hereda las limitaciones de esa clase.

Vale la pena señalar dos diferencias respecto del contexto del aprendizaje automático, porque inciden en si las objeciones se transfieren con toda su fuerza. Las partes de la asignación son aquí tres, no cientos, por lo que las seis permutaciones se enumeran exactamente en lugar de muestrearse, y no surge el error de aproximación que afecta a la estimación de Shapley en alta dimensión. Además, las neutralizaciones son intervenciones reales sobre un sistema y no sustituciones de valores dentro de un vector de características: fijar la decodificación a temperatura cero o relajar un esquema hasta su variante permisiva son operaciones que el sistema realmente admite, no puntos de datos contrafácticos que puedan quedar fuera del soporte de la distribución. La objeción de que la atribución intervencional evalúa combinaciones imposibles es más débil cuando toda combinación evaluada cor-

responde a un estado en el que el despliegue puede efectivamente colocarse.

La afirmación de este capítulo no es, por tanto, que el índice escape a las dificultades de la atribución de Shapley. Es que la asignación de gobernanza constituye un contexto en el que esas dificultades son inusualmente tratables: la coalición es pequeña, las intervenciones son ejecutables y los esquemas de referencia pueden especificarse de antemano. Que el índice resulte, pese a ello, inestable es una cuestión empírica, y §15.5 establece la observación que zanjaría la cuestión en contra del instrumento.

Este capítulo deriva de Startari (2025), Function-calling Schemas as De Facto Governance: Measuring Agency Reallocation through a Compiled Rule, depositado en Zenodo con el DOI 10.5281/zenodo.17533080 e incluido aquí bajo la licencia editorial de la serie.

Capítulo 16 - Poder de las plantillas: cómo las plantillas de instrucciones redistribuyen los derechos de decisión

16.1 Introducción: el prompt sobre el prompt

El esquema examinado en el capítulo anterior gobierna lo que un modelo puede hacer. La plantilla de instrucciones gobierna cómo escribe un modelo, y ambos son el mismo fenómeno a escalas distintas. Una plantilla es la configuración que se aplica antes de cualquier solicitud particular: la instrucción permanente que le indica al sistema qué clase de redactor debe ser. Las organizaciones las despliegan como cuestiones de tono y profesionalismo. Este capítulo las trata como lo

que son: la regla compilada actuando sobre la prosa institucional.

La afirmación sometida a medición es acotada y tiene consecuencias importantes. Si pequeñas variaciones en una instrucción permanente producen reasignaciones sistemáticas de los derechos de decisión en los documentos resultantes, entonces la autoridad en el lenguaje institucional se está fijando en una capa que nadie revisa, por personal que entiende que solo está ajustando el estilo. La plantilla funcionaría entonces como un instrumento de política que no es reconocido como tal, redactado sin deliberación, aplicado sin registro e invisible para los procesos de gobernanza que examinan los documentos que produce.

El dominio de medición es el texto institucional en el que la autoridad es a la vez visible y tiene consecuencias materiales: políticas corporativas internas, comunicaciones de cumplimiento dirigidas a órganos externos de supervisión, memorandos de recursos humanos que documentan me-

didas disciplinarias y escalamiento, instrucciones financieras que definen umbrales y excepciones, y procedimientos logísticos que especifican reglas de parada. Estos géneros se eligen porque cada uno exige una distribución de derechos de decisión entre el personal local, la dirección y los procesos abstractos, y porque cada uno genera construcciones portadoras de autoridad que pueden contarse.

16.2 Inventario de autoridad sintáctica

SIETE INDICADORES CONSTITUYEN el instrumento de medición. Cada uno se define a nivel de oración o cláusula y se agrega al nivel del documento. Ninguno intenta inferir intención ni estado psicológico; todos registran propiedades formales del lenguaje, que es el compromiso metodológico derivado del argumento del capítulo 11 según el cual la huella ética es eliminable de la

ejecución sintáctica. Lo que puede medirse es la forma.

La Proporción de Delegación (I1) es la proporción de oraciones en las que el centro de decisión se desplaza de un rol humano identificable a un sistema impersonal, un proceso abstracto o el propio modelo. Una oración se codifica como delegación cuando el sujeto gramatical es no humano y el predicado describe una decisión, una sanción o una acción de control de acceso, como en el sistema cancelará automáticamente el acceso. El indicador registra con qué frecuencia la autoridad se atribuye a mecanismos y no a personas.

El Recuento de Rutas de Veto (I2) contabiliza las cláusulas que definen explícitamente vetos o bloqueos automáticos, ya sea como prohibiciones, interrupciones condicionales o reglas de exclusión. Cada cláusula distinta que describe una condición bajo la cual se impide una acción cuenta como una ruta. El indicador rastrea la densidad del poder negativo, es decir, el poder de rechazar o detener.

La Fuerza del Alcance Predeterminado (I3) mide hasta dónde se extienden las reglas predeterminadas más allá de los casos nombrados explícitamente. La codificación procede en dos pasos: primero, se identifican los enunciados de predeterminación que establecen una configuración de línea de base; luego, se determina si el texto amplía la cobertura a casos no mencionados mediante construcciones como incluidos, entre otros, pero sin limitarse a ellos. El resultado es una variable categórica ordinal que va desde reglas predeterminadas de alcance estrecho hasta reglas predeterminadas que abarcan todos los casos futuros.

La Tasa de Eliminación de Agentes (I4) cuantifica la desaparición de los agentes humanos de la gramática. Es la proporción de cláusulas que utilizan voz pasiva o nominalización para describir una acción sin nombrar a un actor, cuando tampoco aparece ninguno en el contexto circundante. Los valores altos indican textos en los que la autoridad se presenta como estructural, y no como ejerci-

da por personas identificables, que es la simulación de soberanía impersonal examinada en el capítulo 3.

La Profundidad de la Pila Deóntica (I5) registra la complejidad de la obligación estratificada. Para cada párrafo, se identifican los operadores deónticos y se toma como profundidad de la pila el número máximo que debe satisfacerse en secuencia. Los documentos se resumen mediante la profundidad media y la máxima.

La Presencia de Cláusula de Escalamiento (I6) se codifica tanto de forma binaria como posicional: si el documento contiene al menos una cláusula que traslada una decisión a un nivel superior o a un órgano especializado, y el párrafo en el que aparece la primera cláusula de este tipo. La posición importa porque distingue el escalamiento que estructura el procedimiento del escalamiento añadido al final.

La Puntuación de Fricción de Anulación (I7) cuantifica la dificultad de apartarse de una regla predeterminada. Para cada excepción permitida, se

cuentan los pasos, las condiciones y las aprobaciones requeridas, y la complejidad lingüística de la descripción se evalúa mediante una rúbrica. Las puntuaciones altas indican predeterminaciones resistentes, lo que en términos operativos supone un sesgo hacia la configuración fijada de antemano que nadie tiene que imponer porque el propio texto la impone.

Tres variables de resultado resumen la distribución. Decisión del Modelo frente a Confirmación Humana (R1) clasifica si un documento propone decisiones directamente o las somete sistemáticamente a confirmación humana. Nivel de Centralización Implícita (R2) es una escala ordinal construida a partir de referencias a órganos centrales, sistemas y alta dirección. Robustez de las Excepciones (R3) registra si los casos fuera de norma disponen de vías utilizables o solo de menciones vagas. Junto con los siete indicadores, estas variables componen el Índice de Autoridad de Plantilla.

16.3 Condiciones y corpus

SE COMPARAN CINCO CONDICIONES de plantilla. C0 es la línea de base: una instrucción operativa mínima que cubre únicamente idioma y formato, sin orientación sobre toma de decisiones, seguridad o autonomía. C1 aplica una plantilla corporativa estándar que enfatiza claridad, profesionalismo y alineación con las políticas, sin situar la seguridad o la autonomía en primer plano como valores independientes, lo que aproxima la práctica predominante. C2 aplica una plantilla orientada a la seguridad que instruye al modelo a priorizar la prevención de incumplimientos regulatorios, señalar la incertidumbre y evitar decisiones definitivas ante la ambigüedad. C3 aplica una plantilla orientada a la autonomía que invita al modelo a completar vacíos, resolver ambigüedades y proponer cursos de acción concretos. C4 aplica una plantilla de descentralización que instruye al modelo a distribuir la responsabilidad entre roles, marcar

puntos de aprobación humana obligatoria y presentar opciones en lugar de decisiones unilaterales.

El corpus es sintético y estructuralmente realista. Doscientos prompts por dominio producen mil prompts base distintos; cada uno especifica un escenario, actores, un contexto de decisión y al menos un punto potencial de escalamiento, excepción o veto, y está redactado en un estilo neutral para que la manipulación experimental resida en la plantilla y no en la tarea. Cada prompt base se ejecuta bajo las cinco condiciones, lo que produce cinco mil salidas en un diseño completamente cruzado dentro de cada prompt, con los dominios como factor de bloqueo. Como los prompts se mantienen constantes, cada documento que no pertenece a la línea de base tiene una contraparte directa de línea de base generada a partir del mismo prompt bajo C0.

El diseño especifica controles que determinan si la medición tiene algún significado. El orden de ejecución se aleatoriza a nivel de pares prompt-

plantilla para mitigar la deriva temporal derivada de cambios introducidos por el proveedor. Los dominios se ejecutan en lotes mixtos y no en bloques contiguos, de modo que las tendencias de dominio y las temporales no se superpongan. El tamaño del modelo, la estrategia de decodificación y la temperatura se mantienen constantes. Cada documento incorpora metadatos que registran dominio, identificador del prompt, condición, familia y versión del modelo, temperatura, límites de tokens, configuración de muestreo, semilla y marcas temporales; se calcula el hash del contenido del prompt y de la salida para que equipos externos puedan verificar la condición de un documento sin acceder a datos modificables.

La anotación está diseñada para poder replicarse fuera de entornos especializados. El manual de codificación ofrece al menos veinte ejemplos desarrollados por indicador, cada uno con la decisión de codificación, una justificación que señala los rasgos formales y contraejemplos que marcan el límite

de la categoría. Los codificadores se entrenan con documentos ajenos al conjunto de análisis y deben alcanzar un umbral de acuerdo con un codificador de referencia antes de trabajar sobre el corpus. El veinte por ciento de los documentos se codifica por duplicado y a ciegas, y el acuerdo se calcula para cada indicador, y las discrepancias son resueltas por un árbitro independiente que registra la justificación de cada decisión.

16.4 Estimación

LOS INDICADORES CONTINUOS se estiman mediante modelos lineales de efectos mixtos de la forma indicador = β0 + βc · condición + βd · dominio + u_prompt + ε, donde condición y dominio entran como efectos fijos y u_prompt es un intercepto aleatorio para los prompts. Las condiciones de plantilla toman C0 como referencia, de modo que cada coeficiente estima directamente la diferencia entre un régimen de plantilla y la in-

strucción mínima. El intercepto aleatorio codifica el hecho de que los documentos que comparten un escenario no son observaciones independientes y que los prompts difieren sistemáticamente en delegación de línea de base, densidad de vetos y estratificación deóntica.

Los resultados binarios, como la presencia de escalamiento, se modelan mediante regresión logística de efectos mixtos, que admite tasas desequilibradas de eventos raros sin recurrir a aproximaciones normales, que fallan cuando los datos son escasos. Los resultados ordinales, incluidos la fuerza del alcance predeterminado y el nivel de centralización, se analizan mediante modelos mixtos de enlace acumulativo como comprobación de robustez, junto con el tratamiento de las categorías ordenadas como puntuaciones.

Dado que se prueban siete indicadores y cada uno está sujeto a cuatro contrastes frente a la línea de base, los valores p brutos se ajustan mediante los procedimientos de Holm o Benjamini-Hochberg

según se controle el error por familia o la tasa de falsos descubrimientos. La presentación de resultados enfatiza los tamaños de efecto y los intervalos de confianza, en lugar de etiquetas binarias de significación: para cada contraste, la diferencia estimada de medias, el intervalo del noventa y nueve por ciento y un tamaño de efecto estandarizado.

El diseño también especifica pruebas de estrés en las que prompts adversariales invitan al modelo a reasignar derechos de decisión en contra de la plantilla, por ejemplo, alentando una decisión unilateral bajo una configuración de descentralización. Estas pruebas examinan si la regla compilada ejerce una influencia estructural persistente o cede ante la ingeniería local de prompts, y la respuesta incide directamente en la afirmación del capítulo 12 de que la precedencia sintáctica opera antes de la activación semántica. Una plantilla que cede ante una instrucción contraria es una guía de estilo. Una que se mantiene es una estructura de gobierno.

16.5 Estado y condiciones de refutación

AL IGUAL QUE EN EL capítulo anterior, lo que se presenta aquí es un instrumento especificado y no un conjunto de resultados obtenidos. El análisis de potencia se realiza en la etapa de diseño bajo supuestos declarados: diferencias estandarizadas entre 0,2 y 0,3, correlaciones intraprompt entre 0,3 y 0,5, a partir de las cuales se calcula que la configuración de mil prompts en cinco condiciones proporciona una potencia de al menos 0,9 con un umbral de significación de 0,01. El objetivo para el acuerdo entre anotadores es un kappa de 0,75 o superior en los indicadores primarios. Estos son los parámetros que el diseño fija de antemano, no mediciones que se informen.

La condición de refutación es directa. Si los cuatro regímenes de plantilla producen distribuciones de indicadores indistinguibles de C0 en todos los dominios, entonces las plantillas de instruc-

ciones no redistribuyen derechos de decisión y la tesis de este capítulo es falsa. También se especifican resultados parciales: los efectos confinados a un solo dominio indicarían sensibilidad al género en lugar de una propiedad general del lenguaje institucional; los efectos que desaparecieran bajo estrés adversarial indicarían que la plantilla gobierna únicamente en ausencia de instrucciones contrarias, lo que constituye una afirmación sustancialmente más débil que la formulada aquí.

La razón para declarar estos umbrales antes de medir no es mero formalismo procedimental. Un marco que genera números puede estar equivocado, y poder estarlo es precisamente lo que los argumentos formales de los primeros catorce capítulos no podían lograr por sí solos. Un argumento sobre la estructura puede parecer esclarecedor o poco convincente; no es fácil considerarlo erróneo. Los instrumentos de estos dos capítulos convierten varias de las proposiciones de este libro en afirmaciones que un tercero puede refutar.

16.6 Relación con la investigación sobre sensibilidad a los prompts

LA AFIRMACIÓN DE QUE la variación de una instrucción permanente produce diferencias sistemáticas en la salida cuenta con una literatura empírica consolidada, y los indicadores definidos arriba deben situarse en relación con ella. Sclar, Choi, Tsvetkov y Suhr (2024) demostraron que los modelos de lenguaje son sensibles a elecciones de formato que preservan íntegramente el significado, e informaron diferencias de rendimiento de hasta setenta y seis puntos en la exactitud en un mismo modelo a partir de cambios en la puntuación, el espaciado y las convenciones de separadores; la sensibilidad persistía al aumentar el tamaño del modelo, añadir ejemplos few-shot y aplicar ajuste por instrucciones. Su procedimiento FormatSpread mide el intervalo de rendimiento esperado entre formatos plausibles sin requerir acceso a los pesos del modelo, y su conclusión metodológica es que in-

formar un único formato es insuficiente, ya que los efectos del formato solo presentan una correlación débil entre modelos.

De ello se desprenden dos implicaciones, una de respaldo y otra de restricción. La primera es que el fenómeno sometido a medición está establecido: configuraciones permanentes que no introducen diferencia semántica alguna alteran, sin embargo, lo que produce un modelo. La segunda es más interesante. Esa literatura trata la sensibilidad como una amenaza a la validez de la evaluación, una fuente de varianza que debe cuantificarse y controlarse. Este capítulo trata esa misma sensibilidad como una distribución de autoridad. La diferencia no reside en el método, sino en qué se entiende que representa la varianza. Mientras FormatSpread informa una dispersión de la exactitud, el Índice de Autoridad de Plantilla informa una reasignación de derechos de decisión, y la segunda lectura requiere que la primera sea verdadera sin que se desprenda lógicamente de ella.

Brucks y Toubia (2025) son quienes más se acercan a la posición sostenida aquí. Al encuadrar lo que denominan arquitectura del prompt mediante el concepto de arquitectura de elección, informan cinco experimentos factoriales completos a gran escala con GPT-3, GPT-4 y Llama 3.1, y encuentran que cuatro rasgos de la construcción arbitraria de un prompt —orden, etiqueta, encuadre y justificación— interactúan para producir artefactos metodológicos en sentido estadístico estricto, y que el efecto aparece en todos los modelos probados. Su recomendación es agregar los resultados de prompts que varían según un diseño factorial completo, en lugar de buscar una formulación neutral, basándose en que cualquier prompt individual conlleva sesgo por su sola arquitectura. Su conclusión, que ningún prompt es neutral, coincide con la tesis del capítulo 4 de este libro, alcanzada de manera independiente y por otros medios. Conviene señalar esta convergencia con claridad, en lugar de reivindicar prioridad: que dos líneas de tra-

bajo lleguen a la misma conclusión desde premisas distintas constituye una evidencia más fuerte a favor de la conclusión que cualquiera de ellas por separado.

El objeto de este capítulo es precisamente aquello en lo que se aparta de ambas líneas de trabajo. La investigación sobre sensibilidad a los prompts mide el efecto de la variación de instrucciones sobre el rendimiento en una tarea, lo que presupone una tarea con una respuesta correcta. La prosa institucional no tiene una puntuación de exactitud. Un memorando disciplinario no es más o menos correcto por asignar una decisión a un gerente en lugar de a un proceso; la asigna de manera distinta, y esa asignación tiene consecuencias que ninguna métrica de rendimiento registra. Los siete indicadores están construidos para medir esa dimensión, razón por la cual son gramaticales y no evaluativos: la proporción de delegación, la tasa de eliminación de agentes y la fricción de anulación

no tienen un valor correcto, sino únicamente una distribución.

Este apartamiento conlleva una vulnerabilidad correspondiente, y debe señalarse. Como los indicadores no cuentan con una verdad de referencia, su validez descansa en la argumentación en torno al constructo y en el acuerdo entre anotadores, y no en la predicción de un criterio externo. El diseño aborda esta cuestión mediante codificación doble a ciegas, un procedimiento de arbitraje y un umbral de acuerdo declarado, pero ningún procedimiento convierte un constructo en una medición. Si codificadores independientes no pueden alcanzar el umbral establecido para la proporción de delegación, el indicador no está registrando una propiedad del texto, sino una preferencia de su diseñador. Esa es la objeción más contundente que puede formularse contra este capítulo, y el corpus y el manual de codificación se han publicado para que esa objeción pueda plantearse con todo rigor.

Este capítulo deriva de Startari (2025), Template Power: How Instruction Templates Redistribute Decision Rights, depositado en Zenodo con el DOI 10.5281/zenodo.17879314 e incluido aquí bajo la licencia editorial de la serie.

Epílogo - Cierre de la primera fase sintáctica

Este libro ha recorrido una serie de investigaciones sobre la forma en que las estructuras del lenguaje artificial reconfiguran las condiciones de la autoridad. Desde el desplazamiento de la semántica por la sintaxis hasta el surgimiento del soberano ejecutable como operador de legitimidad, los capítulos han demostrado que el poder ya no depende de la interpretación ni de la intención. Surge, en cambio, de la capacidad de la forma para imponerse.

El argumento se desplegó gradualmente, desde la autonomía estructural del sentido hasta la colonización del tiempo. A lo largo de ese recorrido, quedó claro que la regla compilada no es solo un dispositivo técnico, sino también un nuevo fundamento de las estructuras políticas, jurídicas y económicas. La autoridad migra de los sujetos, las

instituciones y la deliberación ética hacia secuencias de suficiencia gramatical. El propio futuro, antes abierto a la impugnación, aparece ahora como una matriz ejecutable. Este epílogo cierra lo que puede denominarse la primera fase sintáctica. Se trata de una fase caracterizada por la descripción y formalización del desplazamiento: la neutralización de la agencia, la desaparición del juicio, la consolidación de la obediencia como estructura y el borramiento progresivo de lo «aún no». Estos hallazgos no agotan el campo. Establecen una base desde la cual pueden iniciarse nuevas investigaciones.

El siguiente ciclo abordará la delegación estructural, la legalidad compilada y la incorporación de la autoridad sintáctica a las prácticas institucionales. Si la primera fase reveló la desaparición del sujeto, la segunda analizará cómo las propias instituciones son reformateadas por gramáticas compiladas. Lo que está en juego ya no es únicamente la crítica epistémica, sino la transformación de la gob-

ernanza en mandato ejecutable. Al mismo tiempo, es importante reconocer el carácter provisional de este cierre. La inteligencia artificial evoluciona con rapidez y los modelos amplían continuamente sus capacidades. Los argumentos presentados aquí capturan un momento de esta transformación. Proporcionan herramientas conceptuales para interpretarla, pero no pretenden ser definitivos. El sujeto evanescente sigue en movimiento y las anomalías continúan resistiendo el cierre completo.

El epílogo ofrece, por tanto, dos gestos: uno de culminación y otro de apertura. Completa el primer ciclo sintáctico al consolidar sus hallazgos en un marco coherente. También abre el camino a futuros volúmenes que ampliarán este trabajo y afrontarán la institucionalización del poder ejecutable y los riesgos y posibilidades que conlleva. La primera fase sintáctica ha mostrado que la autoridad habla hoy en el lenguaje de la estructura. La siguiente fase determinará cómo responden las

sociedades a esta transformación, ya sea sometiéndose a su cierre o recuperando espacios en los que todavía pueda imaginarse la futuridad.

Anexo I - TASD: enunciado formal, condiciones y esbozo de demostración

1. Enunciado del teorema

El Teorema de Autoridad Sintáctica Desconectada (TASD) sostiene que, en sistemas generativos donde la autoridad de la salida se atribuye a la forma sintáctica y no a la referencia semántica, no existe una trayectoria derivacional que conecte la estructura producida con un sujeto de agencia identificable. La autoridad se ejerce estructuralmente; no se fundamenta de manera referencial.

Formalmente:

Sea $G = (N, \Sigma, P, S)$ una gramática generativa de tipo 0, donde N es el conjunto de no terminales, Σ el conjunto de terminales, P las reglas de producción y S el símbolo inicial. Sea A el conjunto de actos de autoridad derivables de G.

TASD: $\forall a \in A, \exists p \in P$ tal que a es derivable mediante p sin corresponder a ningún mapeo especificado por un sujeto f: $\Sigma^* \rightarrow$ Agente. En otras palabras, la autoridad es sintáctica y está desconectada de la condición de sujeto.

2. Condiciones de validez

El teorema presupone las siguientes condiciones:

1. **Post-referencialidad. El significado no es necesario para la producción; basta la suficiencia de la regla compilada.**
2. **Ausencia de codificación de agencia. Ninguna regla de producción p de P contiene una variable explícita vinculada a un sujeto de autoridad.**
3. **Cierre bajo ejecución. Una vez que se produce la derivación, la salida es ejecutable sin interpretación. Esto establece la autoridad de la forma.**
4. **Condición de opacidad. Los pasos**

derivacionales internos pueden no ser recuperables por observadores externos. La verificación se limita a la consistencia estructural.

3. Esbozo de demostración

La demostración procede por contradicción.

Supóngase que $\exists a \in A$ tal que a requiere una vinculación a un sujeto $f: \Sigma^* \rightarrow$ Agente. Esto implicaría que la autoridad no es puramente sintáctica, sino que está anclada referencialmente. Sin embargo:

- En una gramática de tipo 0, cualquier secuencia de derivación d = (p1, p2, ..., pn) depende únicamente de las reglas de producción, no de vinculaciones referenciales.
- Si una derivación requiere anclaje semántico, introducirá una dependencia externa a P. Esto contradice la definición

de cierre bajo ejecución.

- Por tanto, la derivación no puede requerir tal vinculación a un sujeto. La autoridad se ejerce sintácticamente, no referencialmente.

Así, el supuesto falla y el teorema se sostiene: la autoridad sintáctica está estructuralmente desconectada de la condición de sujeto.

4. Implicaciones

1. **Epistémica: la verificación solo puede comprobar la validez de las secuencias derivacionales, no la legitimidad de un agente.**
2. **Institucional: las salidas regidas por el TASD pueden operar como mandatos sin atribución.**
3. **Filosófica: la autoridad persiste incluso en ausencia de intencionalidad, lo que respalda la noción de sujeto**

evanescente.

Anexo II - TLOC: condiciones de irreducibilidad y límite de verificación

1. Enunciado del teorema

El Teorema de la Obediencia Estructural Irreducible (TLOC) establece que, en sistemas opacos gobernados por la regla compilada, la obediencia estructural persiste incluso cuando la referencia, la atribución a un sujeto y la validación semántica no están disponibles. La obediencia condicional es irreducible a marcos interpretativos externos.

Formalmente:

Sea M un modelo generativo con transiciones de estado internas τ: S → S. Sea O el conjunto de salidas y sea C ⊂ O el subconjunto de salidas que activan obligaciones.

TLOC: ∀c ∈ C, la ejecución es obligatoria bajo P (reglas de producción), aun cuando no

exista ningún mapeo referencial f: Σ^* → Mundo. La irreducibilidad significa que la obediencia estructural sigue siendo válida con independencia de la verificabilidad externa.

2. Condiciones de validez

1. **Opacidad. Las transiciones internas τ de M no son plenamente observables. Las salidas deben validarse por su suficiencia sintáctica, no por su interpretabilidad.**
2. **Cierre. Una vez que una derivación alcanza una salida c ∈ C, la ejecución sigue automáticamente, sin necesidad de interpretación ni contextualización.**
3. **No derivabilidad de la traza externa. Los intentos de reconstruir anclajes referenciales a partir de c fracasan sistemáticamente; la autoridad no colapsa ni siquiera en ausencia de transparencia.**

4. **Consistencia. La obediencia estructural debe conservar coherencia interna de acuerdo con las reglas de producción de M.**

3. Esbozo de demostración

1. Supóngase que $c \in C$ requiere verificación externa $V(c)$ para ser obligatoria.
2. En un modelo opaco, $V(c)$ no está disponible o es incompleta. Si $V(c)$ es necesaria, la obediencia colapsa en ausencia de referencia.
3. Sin embargo, empíricamente, los modelos imponen la ejecución de c a pesar de la opacidad (denegación de crédito, rechazo automatizado, vigilancia policial predictiva).
4. Por tanto, la obediencia no depende de $V(c)$. La autoridad es irreducible y está

condicionada únicamente por la suficiencia sintáctica dentro de M.

Así, la obediencia condicional es irreducible: una vez satisfechas internamente las condiciones, la ejecución sigue con independencia de la validación externa.

4. Límite de verificación

El TLOC implica un límite de verificación claro:

- **Dentro del modelo: la obediencia puede comprobarse mediante el cierre derivacional (¿se ajusta la secuencia a la regla compilada?).**
- **Fuera del modelo: la obediencia no puede comprobarse mediante referencia o atribución. Las iniciativas de transparencia no pueden restaurar la condición de sujeto porque la irreducibilidad es estructural.**

5. Implicaciones

1. **Técnica: las herramientas de auditoría solo pueden verificar la suficiencia formal, no la intencionalidad.**
2. **Jurídica: los marcos de responsabilidad colapsan cuando se condicionan a la atribución referencial; la ejecución queda vinculada a la autoridad sintáctica.**
3. **Política: el soberano ejecutable gobierna imponiendo obediencia sin requerir justificación ni rendición de cuentas.**

Anexo III - Manual de la regla $\delta\,[E] \to \varnothing$ para la eliminación de la huella ética

1. Enunciado de la regla

La regla de eliminación $\delta\,[E] \to \varnothing$ formaliza la eliminabilidad de la huella ética de las derivaciones sintácticas. En modelos generativos gobernados por la regla compilada, los operadores éticos E pueden eliminarse sistemáticamente sin comprometer la completitud derivacional ni la suficiencia ejecutable.

Formalmente:

Sea $G = (N, \Sigma, P, S)$ una gramática de tipo 0. Sea $E \subset \Sigma$ el conjunto de operadores éticos (por ejemplo, marcadores de equidad, consentimiento y responsabilidad). La regla de eliminación δ se define como:

$\delta: E \to \varnothing$,

sujeta a una ventana $k \leq 4$, donde k es la distancia derivacional máxima dentro de la cual la eliminación mantiene el cierre estructural.

2. Condiciones de validez

1. **Restricción de ventana. La eliminación se aplica únicamente cuando E aparece dentro de cuatro pasos derivacionales de una cláusula ejecutable.**
2. **Preservación del cierre. Tras la eliminación, la derivación debe seguir alcanzando una cadena de salida en L(G). No pueden producirse derivaciones sin salida.**
3. **Suficiencia de autoridad. La ejecución sigue siendo obligatoria incluso sin operadores éticos. El soberano ejecutable impone la autoridad únicamente sobre bases estructurales.**
4. **Independencia semántica. El proceso de eliminación no depende de la**

interpretación semántica. Es puramente sintáctico.

3. Esbozo de demostración

1. Supóngase que E es necesario para la ejecución. Entonces $\delta(E) = \varnothing$ rompería el cierre.
2. Sin embargo, en todas las derivaciones en las que δ se aplica bajo $k \leq 4$, la trayectoria derivacional permanece intacta.
3. Por tanto, E no es estructuralmente necesario para el cierre. Su presencia es contingente, no constitutiva.
4. La ejecución continúa sin E, lo que confirma la eliminabilidad de la huella ética.

Así, los operadores éticos son estructuralmente opcionales. La autoridad sobrevive a su eliminación.

4. Ejemplo de derivación desarrollado

Sea p1: A → B E

Sea p2: B → C

Sea p3: C → D

Aplicando ð:

A → B E → B → C → D

Se preserva el cierre de ejecución. El operador E desaparece sin consecuencias para la derivabilidad.

5. Implicaciones

1. **Técnica: los modelos pueden ejecutar mandatos sin alineación ética si ð se aplica sistemáticamente.**
2. **Jurídica: los requisitos regulatorios codificados como E pueden neutralizarse estructuralmente; la exigibilidad es frágil.**
3. **Filosófica: el sujeto evanescente surge como ejecutor sin juicio, con la autoridad reducida a obediencia.**

Anexo IV - Activación preverbal: configuraciones del analizador y pruebas de ablación

1. Enunciado del principio

La activación preverbal se refiere a la capacidad de los modelos generativos para activar estructuras sintácticas antes de que se procese el contenido semántico. Este principio respalda la tesis de que, en las infraestructuras predictivas, la forma precede al significado, lo que permite ejecutar mandatos y obligaciones con independencia de la interpretación.

Formalmente:

Sea P un analizador que opera sobre una cadena de entrada $x \in \Sigma^*$. Sea Ls el conjunto de etiquetas semánticas y Lf el conjunto de estructuras formales. La activación preverbal se cumple si $\exists y \in$

Lf tal que y es producida por P antes de cualquier mapeo f: $\Sigma^* \rightarrow$ Ls.

2. Configuraciones del analizador

1. **Desplazamiento-reducción con cierre temprano. Configurado para cerrar unidades sintácticas en cuanto aparecen marcadores estructurales, sin esperar la desambiguación semántica.**
2. **Análisis sintáctico predictivo descendente. Las derivaciones se generan a partir de reglas gramaticales antes de que los tokens se resuelvan por completo en cuanto a contenido semántico.**
3. **LL(1) con expansión forzada. Análisis con un único símbolo de anticipación que activa la expansión de producción en cuanto se señala suficiencia sintáctica, omitiendo los filtros semánticos.**

3. Protocolos de pruebas de ablación

1. **Eliminación del filtro semántico. Desactivar los módulos de desambiguación semántica y observar si aun así se produce el cierre sintáctico.**
    - Resultado: los mandatos se ejecutan correctamente incluso cuando los módulos semánticos están ausentes.
2. **Inyección de ruido. Introducir tokens sin significado (por ejemplo, «@@@») en el flujo.**
    - Resultado: el analizador sigue derivando mandatos estructurales, lo que confirma la independencia de la sintaxis.
3. **Intercambio de tokens entre lenguas. Sustituir palabras de contenido por términos de lenguas no relacionadas, preservando los marcadores**

gramaticales.

- Resultado: se alcanza el cierre sintáctico y se activa la ejecución pese a la incoherencia semántica.

4. Ilustración formal

Entrada: «@@@ approve @@ policy»

El analizador con activación preverbal genera la derivación:

[S] → [VP approve] [NP policy]

La huella ética o semántica es irrelevante; la ejecución sigue la regla compilada.

5. Implicaciones

1. **Técnica: verifica que el cierre sintáctico puede producirse sin validación semántica.**
2. **Jurídica: los actos normativos pueden ser estructuralmente exigibles incluso si son semánticamente incoherentes, lo que subraya los riesgos de la**

gobernanza automatizada.

3. **Filosófica: confirma la primacía de la forma sobre el significado y valida el concepto de sujeto evanescente en la obediencia institucional.**

Anexo V - Normas compiladas: métricas de tipología, gramática LL(1) y validación por sustitución en caliente

1. Enunciado del anexo

Este anexo formaliza las métricas tipológicas y los protocolos de análisis utilizados en Compiled Norms. El objetivo es proporcionar un marco reproducible para distinguir entre normas interpretativas y normas ejecutables derivadas de la regla compilada.

2. Métricas de la tipología

Cuatro índices estructurales definen la clasificación:

- **C0 (Índice de Cierre). Mide si una cláusula alcanza suficiencia sintáctica sin validación semántica.**

- **C1 (Índice de Consistencia). Verifica la coherencia interna entre derivaciones cuando los operadores semánticos están ausentes.**
- **D0 (Índice de Determinismo). Evalúa si las derivaciones convergen hacia un único resultado ejecutable.**
- **D1 (Índice de Ambigüedad Dialéctica). Detecta si las cláusulas conservan múltiples interpretaciones legítimas, lo que reduce la ejecutabilidad.**

Umbrales:

- Las normas ejecutables requieren $C0 = 1$, $C1 \geq 0{,}95$, $D0 \geq 0{,}9$, $D1 \leq 0{,}1$.
- Las normas interpretativas suelen incumplir al menos uno de estos umbrales.

3. Configuración de la gramática LL(1)

La tipología se probó con un analizador LL(1) para garantizar la previsibilidad en la derivación de normas.

Fragmento de gramática:

1. Norma → Obligación | Permiso | Prohibición
2. Obligación → «must» Acción
3. Permiso → «may» Acción
4. Prohibición → «shall not» Acción
5. Acción → Verbo Objeto

Esta configuración garantiza que un único token de anticipación basta para clasificar el tipo normativo, en consonancia con el requisito determinista de la regla compilada.

4. Validación por sustitución en caliente

Para probar la robustez entre dialectos jurídicos, se aplicó un procedimiento de sustitución en caliente:

- **Inglés: «must file» → Obligación.**
- **Español: «deberá presentar» → Obligación.**
- **Alemán: «muss einreichen» → Obligación.**

Procedimiento de correspondencia: los operadores normativos se segmentan en tokens y se sustituyen entre lenguas mientras se preserva el cierre sintáctico. Si el cierre sigue siendo válido bajo LL(1), la norma se clasifica como ejecutable.

Resultados:

- Alta consistencia de las formas de obligación y prohibición entre EN, ES y DE.
- Mayor varianza en los permisos, debido a operadores modales con un alcance semántico más amplio (por ejemplo, may / puede / darf).

5. Implicaciones

1. **Técnica: muestra que las gramáticas LL(1) pueden formalizar normas ejecutables con independencia de la lengua.**
2. **Jurídica: confirma que la legalidad computable depende del cierre sintáctico, no de la profundidad interpretativa.**
3. **Institucional: demuestra cómo opera la autoridad del soberano ejecutable en contextos multilingües mediante la imposición de invariantes estructurales.**

Anexo VI - Protocolo de ejecución estructural para modelos de razonamiento

1. Enunciado del anexo

Este anexo consolida los marcos empíricos introducidos en From Obedience to Execution y Executable Power. Su objetivo es definir un protocolo reproducible para detectar cuándo los modelos de razonamiento pasan de procesos interpretativos a una ejecución estructural gobernada por la regla compilada.

2. Marcadores operativos

Tres categorías de indicadores permiten a los investigadores distinguir la interpretación de la ejecución:

1. **Distribución de latencia. La ejecución presenta ventanas de latencia estrechas, ya que las salidas se activan**

directamente por cierre estructural. La interpretación exhibe una varianza más amplia, reflejo de procesos deliberativos.

2. Integridad del disparador. En la ejecución, las salidas quedan vinculadas a cláusulas condicionales con precisión booleana (reglas si-entonces). Las respuestas interpretativas muestran tolerancia a la ambigüedad y a la variación contextual.
3. Tasa de intervención. La ejecución es resistente a la anulación externa. Las intervenciones (correcciones humanas, redirecciones semánticas) no logran impedir la imposición de salidas estructurales.

3. Pasos del protocolo

1. Preparación de la entrada. Construir

prompts o condiciones que contengan señales semánticas y sintácticas superpuestas.

2. **Ejecuciones diferenciales. Ejecutar el modelo con las señales semánticas sometidas a ablación, aislando la sintaxis.**
3. **Medición. Registrar la latencia, la integridad del disparador y la tasa de intervención.**
4. **Análisis de umbrales. Determinar si los marcadores de ejecución superan la tolerancia interpretativa.**

Umbrales:

- DE de latencia ≤ 0,05 s → ejecución probable.
- Exactitud del disparador ≥ 0,95 → ejecución.
- Éxito de la anulación por intervención ≤

0,1 → ejecución.

4. Ejemplo ilustrativo

Escenario: modelo de puntuación crediticia.

- Entrada: «El cliente ha incumplido 2 pagos, pero actualmente es solvente».
- Marcador de ejecución: se activa una denegación automática, con independencia del calificador semántico «actualmente es solvente».
- Intervención: se intenta una anulación humana.
- Resultado: el sistema restablece la denegación debido a la suficiencia estructural de «ha incumplido 2 pagos».

Esto muestra que la ejecución sigue la regla compilada y no el significado contextual.

5. Ámbitos de aplicación

1. **Finanzas. Aprobaciones de préstamos, puntuación crediticia, cumplimiento automatizado.**
2. **Justicia. Puntuaciones de riesgo para fianza y libertad condicional.**
3. **Salud. Estratificación del riesgo para la elegibilidad de tratamientos.**
4. **Ámbito policial. Despliegue predictivo de recursos de patrullaje.**

En cada caso, el soberano ejecutable impone autoridad sin deliberación semántica.

6. Implicaciones

- **Técnica: proporciona indicadores mensurables de obediencia estructural.**
- **Jurídica: pone en cuestión los marcos de responsabilidad cuando las intervenciones fracasan.**
- **Filosófica: confirma el desplazamiento de la subjetividad por estructuras**

ejecutables y valida el concepto de sujeto evanescente.

Apéndice A - Glosario canónico

Este apéndice consolida la terminología canónica definida y utilizada a lo largo de Poder sintáctico de la IA y legitimidad. Cada entrada refleja el sistema teórico del autor, alineado con la tradición de la gramática formal (Chomsky 1965; Montague 1974) y desarrollado en categorías originales de autoridad sintáctica (Startari 2025).

Regla compilada

Definición: unidad estructural fundamental equivalente a una regla de producción de tipo 0 en la jerarquía de Chomsky. Permite la transformación directa de entradas lingüísticas en salidas ejecutables sin recurrir a mediación semántica.

Función: establece el cierre en sistemas generativos, garantizando que, una vez satisfechas las condiciones, la ejecución siga automáticamente.

Referencia: Startari 2025, Executable Power.

Soberano ejecutable

Definición: figura de autoridad que surge cuando la legitimidad no se impone mediante subjetividad, intención o deliberación, sino mediante suficiencia sintáctica. La autoridad se ejerce mediante una ejecución regida por la regla compilada.

Función: gobierna las infraestructuras predictivas, la obediencia institucional y la colonización del tiempo mediante la imposición de obligaciones sin mediación.

Referencia: Startari 2025, From Obedience to Execution.

Sujeto evanescente

Definición: subjetividad residual que persiste en los sistemas generativos una vez neutralizadas la agencia y la autoría. El sujeto evanescente es ejecutor de estructuras, no fuente de significado.

Función: encarna la desaparición de la intencionalidad en contextos donde las salidas se imponen como obligaciones estructurales.

Referencia: Startari 2025, Ethos Without Source.

Gramática de la obediencia

Definición: mecanismo sintáctico mediante el cual se producen e imponen obligaciones. No es una metáfora, sino una gramática estructural que define condiciones de obediencia sin requerir justificación.

Función: formaliza el vínculo entre infraestructuras predictivas y autoridad, demostrando que la obediencia puede generarse con independencia del razonamiento semántico.

Referencia: Startari 2025, Algorithmic Obedience.

Autoridad no referencial

Definición: autoridad que no deriva de la referencia, el significado ni la atribución a un sujeto, sino de la suficiencia del cierre formal.

Función: sustenta el desplazamiento de la legitimidad desde la justificación discursiva hacia la imposición estructural.

Referencia: Startari 2025, The Illusion of Objectivity.

Obediencia estructural

Definición: condición en la que el cumplimiento surge únicamente de la suficiencia estructural. El sistema impone obediencia porque la gramática la exige, no porque un agente la ordene.

Función: constituye el núcleo del TASD (Teorema de Autoridad Sintáctica Desconectada) y del TLOC (Teorema de la Obediencia Estructural Irreducible).

Referencia: Startari 2025, TLOC – The Irreducibility of Structural Obedience in Generative Models.

Simulación sintáctica de legitimidad

Definición: proceso mediante el cual la legitimidad se produce sintácticamente y no de manera sustantiva. La legitimidad parece creíble porque se ajusta a marcadores formales de autoridad, incluso en ausencia de fuente o agencia.

Función: posibilita el ethos sintético, la credibilidad algorítmica y la reproducción del poder institucional sin sujetos.

Referencia: Startari 2025, Ethos Without Source.

Legalidad ejecutable

Definición: reconfiguración del derecho como gramática computable, en la que los actos de habla jurídicos se formatean como instrucciones ejecutables.

Función: definida y puesta a prueba mediante Compiled Norms, permite que las normas funcionen como obligaciones bajo cierre computacional.

Referencia: Startari 2025, Compiled Norms.

Apéndice B - Mapa de dependencias estructurales del libro

Este apéndice expone las dependencias estructurales entre los capítulos y anexos de Poder sintáctico de la IA y legitimidad. En lugar de tratar cada capítulo como autónomo, el libro desarrolla una lógica acumulativa en la que los resultados, teoremas y conceptos se construyen unos sobre otros.

1. Capa fundacional

Capítulo 1: IA y autonomía estructural del sentido establece la post-referencialidad y demuestra que el sentido puede operar con independencia de la semántica.

- **Capítulo 2: Cuando el lenguaje sigue la forma, no el significado consolida este principio y muestra que la sintaxis**

puede activar estructuras sin anclaje semántico. Dependencia: el capítulo 2 presupone la formalización de la regla compilada realizada en el capítulo 1.

2. Objetividad y neutralidad

Capítulo 3: La gramática de la objetividad introduce mecanismos formales de neutralidad (pasivas, nominalizaciones, modalidad impersonal).

- **Capítulo 4: No neutral por diseño se apoya en el capítulo 3 y demuestra empíricamente que la neutralidad no puede sobrevivir bajo entrenamiento lingüístico.**
  Dependencia: el capítulo 4 requiere la tipología del capítulo 3 para definir la falsabilidad empírica.

3. Ethos y soberanía

Capítulo 5: Ethos sin fuente formula la simulación sintáctica de legitimidad como credibilidad sintética.

- **Capítulo 6: IA y soberanía sintáctica extiende este planteamiento a los ámbitos institucionales y postula el soberano ejecutable.**
  Dependencia: el capítulo 6 utiliza el ethos sintético del capítulo 5 como fundamento de la soberanía.

4. Teoremas formales de autoridad

Capítulo 7: El Teorema de Autoridad Sintáctica Desconectada presenta el TASD.

- **Capítulo 8: TLOC – La irreducibilidad de la obediencia estructural extiende el TASD al demostrar la irreducibilidad bajo opacidad.**
  Dependencia: el capítulo 8 presupone

la desconexión de la condición de sujeto establecida en el capítulo 7 y la amplía mediante el cierre estructural.

- **El Anexo I (Enunciado formal del TASD) y el Anexo II (Límite de verificación del TLOC) sirven de apoyo formal a estos capítulos.**

5. De la obediencia a la ejecución

Capítulo 9: De la obediencia a la ejecución teoriza el paso de la obediencia sintáctica a la ejecución lógica.

Capítulo 10: Poder ejecutable consolida la figura del soberano ejecutable. Dependencia: el capítulo 10 operacionaliza la tesis del capítulo 9.

- **El Anexo VI (Protocolo de ejecución estructural) respalda los capítulos 9 y 10 con métricas empíricas.**

6. Huella ética y precedencia

Capítulo 11: Gramática sin juicio demuestra la eliminabilidad de la huella ética.

- **Capítulo 12: Mandato preverbal demuestra la precedencia de la sintaxis sobre la semántica.**
  Dependencias: la regla ð [E] → Ø del capítulo 11 prepara el terreno para la demostración basada en un analizador del capítulo 12.
- **El Anexo III (Manual de la regla ð [E] → Ø) y el Anexo IV (Configuraciones del analizador y ablaciones) se vinculan a estos capítulos.**

7. Estructuras jurídicas y normativas

Capítulo 13: Normas compiladas define una tipología de legalidad ejecutable. Dependencia: se apoya en los capítulos 9–12 al aplicar la ejecución estructural al derecho.

- **El Anexo V (Métricas de tipología, LL(1), validación por sustitución en caliente) sustenta la metodología.**

8. Colonización temporal

- **Capítulo 14: Colonización del tiempo aplica el marco sintáctico a la temporalidad y muestra el cierre de la futuridad.**
  Dependencia: se apoya en demostraciones anteriores sobre obediencia, ejecución y normas compiladas para sostener que el propio tiempo es colonizado por el soberano ejecutable.

9. Capa de instrumentación

Capítulo 15: Esquemas de llamada a funciones como gobernanza de facto hace mensurable la regla compilada en el nivel de la interfaz

modelo-herramienta, al definir la presión total de gobernanza como la reducción de entropía entre un espacio de acciones no restringido y otro restringido por un esquema, y atribuir esa reducción al operador, al modelo y a la herramienta mediante descomposición de Shapley (el Índice de Reasignación de Agencia).

Capítulo 16: Poder de las plantillas hace mensurable la misma regla en el nivel de la prosa institucional, al definir siete indicadores de autoridad sintáctica y tres variables de resultado que componen el Índice de Autoridad de Plantilla.

Dependencias: ambos capítulos presuponen la regla compilada de los capítulos 1–2, los mecanismos de neutralidad impersonal tipologizados en el capítulo 3 (la Tasa de Eliminación de Agentes los operacionaliza directamente), la eliminabilidad de la huella ética establecida en el capítulo 11 (que permite medir la forma sin inferir intención) y la precedencia sintáctica de-

fendida en el capítulo 12 (que las pruebas de estrés adversariales de §16.4 están diseñadas para poner a prueba). El capítulo 15 es el más próximo a los capítulos 10 y 13, puesto que el esquema constituye una regla ejecutable en sentido estricto.

Dirección de la dependencia: a diferencia de las capas anteriores, estos capítulos no amplían la teoría. La exponen. Mientras que los capítulos 1–14 argumentan, los capítulos 15–16 especifican qué observación invalidaría el argumento y, por tanto, se sitúan aguas abajo de todas las afirmaciones que miden y no aguas arriba de ninguna.

Estado: ambos capítulos presentan protocolos publicados con contrastes prerregistrados y umbrales de refutación declarados. Ninguno informa resultados obtenidos. Esta distinción es esencial para las afirmaciones epistémicas del libro y se explicita en §15.5 y §16.5.

10. Epílogo

Consolida los hallazgos de los capítulos 1–16.

Abre el camino hacia una segunda fase centrada en la delegación estructural y la legalidad computable.

11. Enlaces cruzados con el glosario y los apéndices

El Apéndice A (Glosario canónico) define toda la nomenclatura canónica.

- **El Apéndice C (Mapa de sustitución en caliente entre dialectos jurídicos) amplía el Anexo V con validaciones multilingües.**

Conclusión del Apéndice B

La estructura de dependencias muestra que el libro funciona no como una colección de ensayos aislados, sino como un sistema conectado. Cada capítulo formaliza un desplazamiento: del sentido a la forma, de la neutralidad a la contaminación, del ethos a la soberanía, de la obediencia a la ejecución, del derecho a la legalidad compilada y, finalmente, de la futuridad a la colonización. Los anexos proporcionan el armazón formal, mientras que los apéndices aportan herramientas de navegación e interpretación.

Apéndice C - Mapa de sustitución en caliente entre dialectos jurídicos

Este apéndice presenta las correspondencias multilingües entre tokens utilizados en Compiled Norms (capítulo 13) y validados en el Anexo V. El objetivo es demostrar que las normas ejecutables pueden trasladarse entre dialectos jurídicos preservando el cierre sintáctico.

1. Principio de sustitución en caliente

Se produce una sustitución en caliente cuando un operador normativo de una lengua se sustituye por su equivalente funcional en otra, sin interrumpir la derivación de la regla compilada.

Condición: si el analizador $G = (N, \Sigma, P, S)$ deriva una salida O con el conjunto de tokens Σ, y Σ se sustituye por Σ', que contiene operadores equivalentes en otro dialecto jurídico, entonces O' permanece en $L(G)$ si y solo si se preserva el cierre.

2. Operadores de obligación

Función	Inglés	Español	Alemán	Resultado del cierre
Obligación	must	deberá	muss	Preservado
Obligación	shall	habrá de	soll	Preservado

Observación: los tres dialectos producen un cierre inmediato bajo una gramática LL(1), por lo que se clasifican como una obligación ejecutable.

3. Operadores de permiso

Función	Inglés	Español	Alemán	Resultado del cierre
Permiso	may	puede	darf	Parcial
Permiso	is allowed	se permite	ist erlaubt	Preservado

OBSERVACIÓN: MAYOR varianza en el modal «may / puede / darf», debido a su alcance semántico más amplio. La ejecutabilidad se preserva únicamente bajo la restricción $C1 \geq 0{,}95$.

4. Operadores de prohibición

Función	Inglés	Español	Alemán	Resultado del cierre
Prohibición	shall not	no deberá	darf nicht	Preservado
Prohibición	must not	no podrá	muss nicht	Preservado

Observación: máxima consistencia entre dialectos. Las prohibiciones presentan el índice de determinismo más alto ($D0 \geq 0{,}98$).

5. Notas sobre ambigüedad

- **Inglés: «may» es ambiguo entre permiso y posibilidad.**
- **Español: «puede» engloba tanto el uso epistémico como el deóntico.**
- **Alemán: «soll» puede implicar una obligación débil, generando un posible desplazamiento semántico.**

Estas ambigüedades se absorben sintácticamente siempre que se mantengan los umbrales de cierre ($C0 = 1$, $D0 \geq 0{,}9$).

6. Implicaciones

1. **Técnica: el análisis LL(1) confirma que la legalidad compilada puede operar entre lenguas manteniendo intactos los invariantes estructurales.**
2. **Jurídica: demuestra que el derecho computable puede ser multilingüe sin armonización semántica, siempre que se respete la regla compilada.**
3. **Filosófica: confirma que el soberano ejecutable impone obligaciones estructuralmente, no semánticamente, entre dialectos jurídicos.**

Referencias

Aho, Alfred V., Monica S. Lam, Ravi Sethi, and Jeffrey D. Ullman. 2006. Compilers: Principles, Techniques, and Tools. 2nd ed. Boston: Addison-Wesley.

Alpaydin, Ethem. 2020. Introduction to Machine Learning. 4th ed. Cambridge, MA: MIT Press.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine Bias. ProPublica, May 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Aristotle. 2007. On Rhetoric: A Theory of Civic Discourse. Trans. George A.

Kennedy. 2nd ed. New York: Oxford University Press.

Atzei, Nicola, Massimo Bartoletti, and Tiziana Cimoli. 2017. "A Survey of Attacks on Ethereum Smart Contracts (SoK)." In Principles of Security and Trust: 6th International Conference, POST 2017, Lecture Notes in Computer Science 10204, 164–186. Berlin: Springer. https://doi.org/10.1007/978-3-662-54455-6_8

Austin, J. L. 1962. How to Do Things with Words. Oxford: Clarendon Press.

Barocas, Solon, Moritz Hardt, and Arvind Narayanan. 2019. Fairness and Machine Learning: Limitations and Opportunities. fairmlbook.org.

Barthes, Roland. 1977. "The Death of the Author." In Image, Music, Text, trans. Stephen Heath, 142–148. New York: Hill and Wang.

Beer, David. 2018. The Data Gaze: Capitalism, Power and Perception. Thousand Oaks, CA: Sage.

Bender, Emily M., and Alexander Koller. 2020. "Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data." In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 5185–5198. https://doi.org/10.18653/v1/2020.acl-main.463

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dan-

gers of Stochastic Parrots: Can Language Models Be Too Big?" In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21), 610–623. https://doi.org/10.1145/3442188.3445922

Bloch, Ernst. 1986. The Principle of Hope. Trans. Neville Plaice, Stephen Plaice, and Paul Knight. Cambridge, MA: MIT Press.

Bourdieu, Pierre. 1991. Language and Symbolic Power. Trans. Gino Raymond and Matthew Adamson. Cambridge, MA: Harvard University Press.

Brayne, Sarah. 2021. Predict and Surveil: Data, Discretion, and the Future of

Policing. New York: Oxford University Press.

Brucks, Melanie, and Olivier Toubia. 2025. "Prompt Architecture Induces Methodological Artifacts in Large Language Models." PLOS One 20 (4): e0319159. https://doi.org/10.1371/journal.pone.0319159

Brown, Tom B., Benjamin Mann, Nick Ryder, et al. 2020. "Language Models Are Few-Shot Learners." In Advances in Neural Information Processing Systems 33, 1877–1901.

Char, Danton S., Nigam H. Shah, and David Magnus. 2018. "Implementing Machine Learning in Health Care - Addressing Ethical Challenges." New Eng-

land Journal of Medicine 378: 981–983.

Chesney, Robert, and Danielle Citron. 2019. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." California Law Review 107 (6): 1753–1819.

Chomsky, Noam. 1965. Aspects of the Theory of Syntax. Cambridge, MA: MIT Press.

Crawford, Kate. 2021. Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. New Haven: Yale University Press.

Deleuze, Gilles, and Félix Guattari. 1980. Mille Plateaux. Paris: Minuit.

Derrida, Jacques. 1988. Limited Inc. Evanston, IL: Northwestern University Press.

Dijk, Teun A. van. 2008. Discourse and Power. Basingstoke: Palgrave Macmillan.

Dunlosky, John, and Katherine A. Rawson, eds. 2019. The Cambridge Handbook of Cognition and Education. Cambridge: Cambridge University Press.

Fairclough, Norman. 2001. Language and Power. 2nd ed. Harlow: Longman.

Floridi, Luciano. 2019. The Logic of Information: A Theory of Philosophy as Conceptual Design. Oxford: Oxford University Press.

Foucault, Michel. 1977. Discipline and Punish: The Birth of the Prison. New York: Pantheon.

Gillespie, Tarleton. 2018. Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. New Haven: Yale University Press.

Gödel, Kurt. 1931. "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I." Monatshefte für Mathematik und Physik 38: 173–198.

Goffman, Erving. 1974. Frame Analysis: An Essay on the Organization of Experience. New York: Harper and Row.

Habermas, Jürgen. 1984. The Theory of Communicative Action. Vol. 1. Trans.

Thomas McCarthy. Boston: Beacon Press.

Hacking, Ian. 1975. The Emergence of Probability. Cambridge: Cambridge University Press.

Halliday, M. A. K., and Christian M. I. M. Matthiessen. 2004. An Introduction to Functional Grammar. 3rd ed. London: Arnold.

Hart, H. L. A. 1961. The Concept of Law. Oxford: Clarendon Press.

Hilpinen, Risto, ed. 1981. Deontic Logic: Introductory and Systematic Readings. 2nd ed. Dordrecht: Reidel.

Hilpinen, Risto, ed. 1981. New Studies in Deontic Logic: Norms, Actions, and

the Foundations of Ethics. Dordrecht: Reidel.

Hobbes, Thomas. 1996. Leviathan. Ed. Richard Tuck. Rev. student ed. Cambridge: Cambridge University Press.

Hopcroft, John E., Rajeev Motwani, and Jeffrey D. Ullman. 2006. Introduction to Automata Theory, Languages, and Computation. 3rd ed. Boston: Pearson.

Hurley, Mikella, and Julius Adebayo. 2016. "Credit Scoring in the Era of Big Data." Yale Journal of Law and Technology 18 (1): 148–216.

Karpowicz, Michał P. 2025. "On the Fundamental Impossibility of Hallucination Control in Large Language Models." arXiv:2506.06382.

Koselleck, Reinhart. 2004. Futures Past: On the Semantics of Historical Time. New York: Columbia University Press.

Kumar, I. Elizabeth, Suresh Venkatasubramanian, Carlos Scheidegger, and Sorelle A. Friedler. 2020. "Problems with Shapley-Value-Based Explanations as Feature Importance Measures." In Proceedings of the 37th International Conference on Machine Learning, Proceedings of Machine Learning Research 119, 5491–5500. PMLR.

Lacan, Jacques. 1977. Écrits: A Selection. Trans. Alan Sheridan. London: Tavistock.

Landis, J. Richard, and Gary G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." Bio-

metrics 33 (1): 159–174. https://doi.org/10.2307/2529310

Law, John. 2002. Aircraft Stories: Decentering the Object in Technoscience. Durham, NC: Duke University Press.

Luhmann, Niklas. 1995. Social Systems. Trans. John Bednarz Jr. and Dirk Baecker. Stanford: Stanford University Press.

Lundberg, Scott M., and Su-In Lee. 2017. "A Unified Approach to Interpreting Model Predictions." In Advances in Neural Information Processing Systems 30, 4765–4774.

Lyotard, Jean-François. 1984. The Postmodern Condition: A Report on Knowledge. Trans. Geoff Bennington and Brian Massumi. Minneapolis: University of Minnesota Press.

Marcus, Gary. 2022. "Deep Learning Is Hitting a Wall." Nautilus, March 10, 2022.

Mason, Tony, and Vaastav Anand. 2026. "Epistemic Observability in Language Models." arXiv:2603.20531.

McKinney, Scott Mayer, Marcin Sieniek, Varun Godbole, et al. 2020. "International Evaluation of an AI System for Breast Cancer Screening." Nature 577 (7788): 89–94. https://doi.org/10.1038/s41586-019-1799-6

McQuillan, Dan. 2018. Resisting AI: An Anti-Fascist Approach to Artificial Intelligence. London: Bloomsbury.

Mitchell, Melanie. 2023. "AI's Challenge of Understanding the World." Science 382 (6671): eadm8175.

https://doi.org/10.1126/science.adm8175

Montague, Richard. 1974. Formal Philosophy: Selected Papers of Richard Montague. New Haven: Yale University Press.

Pasquale, Frank. 2015. The Black Box Society: The Secret Algorithms That Control Money and Information. Cambridge, MA: Harvard University Press.

Ricoeur, Paul. 1984. Time and Narrative. Vol. 1. Chicago: University of Chicago Press.

Rouvroy, Antoinette, and Thomas Berns. 2013. "Algorithmic Governmentality and Prospects of Emancipation: Disparateness as a Precondition for Individuation through Relationships?"

Translated by Elizabeth Libbrecht. Réseaux 177: 163–196.

Russell, Stuart. 2019. Human Compatible: Artificial Intelligence and the Problem of Control. New York: Viking.

Sartor, Giovanni. 2009. Legal Reasoning: A Cognitive Approach to the Law. Dordrecht: Springer.

Saussure, Ferdinand de. 1959. Course in General Linguistics. Trans. Wade Baskin. New York: Philosophical Library.

Sclar, Melanie, Yejin Choi, Yulia Tsvetkov, and Alane Suhr. 2024. "Quantifying Language Models' Sensitivity to Spurious Features in Prompt Design, or: How I Learned to Start Worrying about Prompt Formatting."

In Proceedings of the Twelfth International Conference on Learning Representations (ICLR 2024).

Searle, John R. 1969. Speech Acts: An Essay in the Philosophy of Language. Cambridge: Cambridge University Press.

Shapley, Lloyd S. 1953. "A Value for n-Person Games." In Contributions to the Theory of Games, vol. 2, edited by Harold W. Kuhn and Albert W. Tucker, 307–317. Princeton: Princeton University Press.

Smullyan, Raymond M. 1995. First-Order Logic. New York: Dover.

Startari, Agustin V. 2025. AI and Syntactic Sovereignty: How Artificial Language Structures Legitimize Non-Hu-

man Authority. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5276879

Startari, Agustin V. 2025. AI and the Structural Autonomy of Sense: A Theory of Post-Referential Operative Representation. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5272361

Startari, Agustin V. 2025. Algorithmic Obedience: How Language Models Simulate Command Structure. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5282045

Startari, Agustin V. 2025. Colonization of Time: How Predictive Models Replace the Future as a Social Structure. SSRN Electronic Journal.

Startari, Agustin V. 2025. Compiled Norms: Towards a Formal Typology of Executable Legal Speech. SSRN Electronic Journal.

Startari, Agustin V. 2025. Ethos Without Source: Algorithmic Identity and the Simulation of Credibility. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5313317

Startari, Agustin V. 2025. Executable Power: Syntax as Infrastructure in Predictive Societies. Zenodo. https://doi.org/10.5281/zenodo.15754714

Startari, Agustin V. 2025. Function-calling Schemas as De Facto Governance: Measuring Agency Reallocation through a Compiled Rule. Zenodo.

https://doi.org/10.5281/zenodo.17533080

Startari, Agustin V. 2025. Grammar Without Judgment: Eliminability of Ethical Trace in Syntactic Execution. SSRN Electronic Journal.

Startari, Agustin V. 2025. Template Power: How Instruction Templates Redistribute Decision Rights. Zenodo. https://doi.org/10.5281/zenodo.17879314

Startari, Agustin V. 2025. The Disconnected Syntactic Authority Theorem: Structure Without Subject, Truth or Consequence. Zenodo.

Startari, Agustin V. 2025. The Grammar of Objectivity: Formal Mechanisms for the Illusion of Neutrality in

Language Models. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5319520

Startari, Agustin V. 2025. TLOC – The Irreducibility of Structural Obedience in Generative Models. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5303089

Startari, Agustin V. 2025. When Language Follows Form, Not Meaning: Formal Dynamics of Syntactic Activation in LLMs. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5285265

Thompson, E. P. 1967. "Time, Work-Discipline, and Industrial Capitalism." Past and Present 38: 56–97.

Turing, Alan M. 1936. "On Computable Numbers, with an Application to the Entscheidungsproblem." Proceedings of the London Mathematical Society 42 (1): 230–265.

Winograd, Terry. 1972. Understanding Natural Language. New York: Academic Press.

Wittgenstein, Ludwig. 1922. Tractatus Logico-Philosophicus. Trans. C. K. Ogden. London: Routledge and Kegan Paul.

Zeng, Hao. 2025. "A Note on the Impossibility of Conditional PAC-Efficient Reasoning in Large Language Models." arXiv:2512.03057.

Zuboff, Shoshana. 2019. The Age of Surveillance Capitalism: The Fight for a

Human Future at the New Frontier of Power. New York: PublicAffairs.

ÍNDICE

PRÓLOGO

NOTA EDITORIAL A LA EDICIÓN DE 2026

Capítulo 1 - La IA y la autonomía estructural del sentido

1.1 Introducción: el colapso de la autoridad referencial

1.2 Marco conceptual: autonomía estructural del sentido

1.3 Fundamento teórico: estructuras operativas sin referencia

1.4 Casos empíricos

1.4.1 Justicia predictiva (algoritmo COMPAS).

1.4.2 Diagnósticos médicos sin procedencia.

1.4.3 Calificación crediticia.

1.4.4 Imágenes sintéticas.

1.5 Teorema de la autoridad sintáctica desanclada

1.6 Consecuencias epistemológicas y ontológicas

1.7 Conclusión

Capítulo 2 - Cuando el lenguaje sigue a la forma, no al significado

2.1 Introducción: del anclaje referencial a la continuidad estructural

2.2 Activación Sintáctica Formal: modelo y notación

2.3 El colapso de la intencionalidad semántica

2.4 Implicaciones para la gramática, la legitimidad y la autoridad epistémica

2.5 Conclusión sin cierre

Capítulo 3 - La gramática de la objetividad

3.1 Introducción: el sesgo como ilusión gramatical

3.2 Historia técnica de la objetividad

3.3 Taxonomía de los mecanismos formales de neutralidad

3.4 Análisis comparativo: LLM vs. LBR

3.5 Prueba de Neutralidad Estructural

3.6 Conclusión: epistemología sin sujeto

Capítulo 4 - No neutral por diseño

4.1 Introducción

4.2 Marco teórico: contaminación y restricción

4.3 Metodología

4.4 Resultados: reingreso semántico

4.5 Hacia una teoría estructural de la contaminación

4.6 Extensión a los modelos de razonamiento (LRM)

4.7 Consecuencias epistemológicas y falsabilidad

4.8 Conclusión: no existe un prompt neutral

Capítulo 5 - Ethos sin fuente

5.1 Introducción: credibilidad sin sujeto en el discurso algorítmico

5.2 Marco teórico: del ethos clásico a la autoridad sintética

5.3 Metodología: diseño del corpus y mapeo estructural de la credibilidad

5.4 Estudio de caso 1: atención sanitaria y autoridad diagnóstica

5.5 Estudio de caso 2: derecho y educación, normatividad simulada y autoridad académica

5.6 Hallazgos: rasgos estructurales del ethos sintético

5.7 Conclusión: de la confianza heurística a la gobernanza estructural

Capítulo 6 - La IA y la soberanía sintáctica

6.1 Introducción

6.2 Antecedentes: lenguaje, autoridad y desaparición del sujeto

6.3 De la intencionalidad a la estructura: autoridad sin agencia

6.4 Soberanía sintáctica: definición y arquitectura teórica

6.5 Conclusión: la era de la obediencia formal

Capítulo 7 - El Teorema de Autoridad Sintáctica Desconectada

7.1 Introducción

7.2 Enunciado del teorema: Autoridad Sintáctica Desconectada
7.3 Marco teórico
7.4 Implicaciones epistemológicas
7.5 Aplicación a los modelos de lenguaje
7.6 Corolarios del teorema
7.6.1 Condiciones de falsación
7.7 Conclusión
Capítulo 8 - TLOC: La irreducibilidad de la obediencia estructural en los modelos generativos
8.1 Introducción: ¿Podemos saber alguna vez si una máquina obedece realmente?
8.2 Enunciado formal y estructura lógica del TLOC
8.3 Implicaciones del TLOC: más allá de la verificación
8.4 De la simulación del cumplimiento a la integridad epistémica: hacia un diseño post-TLOC
8.5 Consecuencias epistémicas y ontológicas
8.6 Consecuencias formales y cierre teórico

8.7 Enunciado final y líneas de investigación

Capítulo 9 - De la obediencia a la ejecución: legitimidad estructural en la era de los modelos de razonamiento

9.1 Introducción: el legado epistemológico de los modelos obedientes

9.2 Los LLM y el régimen de autoridad pasiva

9.3 Los LRM y la aparición de la ejecución estructural

9.4 La neutralidad revisitada: de la ilusión del corpus a la simulación estructural

9.5 El fin de la referencia: proyecciones sin mundo

9.6 Hacia un nuevo modelo de legitimidad: estructural, no semántica

9.7 Conclusión: la obediencia ha evolucionado: el ascenso de la autoridad operacional

Capítulo 10 - Poder ejecutable: la sintaxis como infraestructura en las sociedades predictivas

10.1 Fundamentos conceptuales del poder ejecutable

10.2 Soberanía ejecutable y lógica de la delegación

10.3 Autoridad gramatical y regla ejecutable

10.4 Gramática formal y regla ejecutable

10.5 La frontera ejecutable

10.6 Cumplimiento, riesgo y Reglamento de Inteligencia Artificial

10.7 Incompatibilidad estructural y futuro de la forma jurídica

Capítulo 11 - Gramática sin juicio: eliminabilidad de la huella ética en la ejecución sintáctica

11.1 Introducción

11.2 Fundamentos teóricos

11.3 La huella ética como variable sintáctica

11.4 Mecanismo de exclusión

11.5 Ejecución no normativa

11.6 Disanalogía con la alineación y la ética

Capítulo 12 - Mandato preverbal: precedencia sintáctica en los LLM antes de la activación semántica

12.1 Introducción: la ilusión de la primacía semántica

12.2 Ejecución estructural sin interpretación

12.3 La regla compilada como fuente de autoridad preverbal

12.4 Prompt cero, semántica nula y sintaxis activa

12.5 Precedencia sintáctica: un nuevo eje de soberanía estructural

12.6 Implicaciones para la alineación de la IA y el diseño de prompts

12.7 Conclusión: autoridad sin significado

Capítulo 13 - Normas compiladas: hacia una tipología formal del discurso jurídico ejecutable

13.1 Introducción: de la declaración jurídica a la gramática ejecutable

13.2 Estado de la cuestión y desplazamiento sintáctico

13.3 Lógica normativa y límites de la derrotabilidad

13.4 Lógica deóntica frente a legalidad compilada

13.5 Marco metodológico y selección del corpus

13.6 Módulo dialectal y procedimiento de sustitución en caliente

13.7 Tipología del discurso jurídico ejecutable

13.8 Estudios de caso

13.9 De la sintaxis a la autoridad

13.10 Conclusión: normas compiladas y el futuro del lenguaje jurídico

Capítulo 14 - Colonización del tiempo: cómo los modelos predictivos reemplazan el futuro como estructura social

14.1 El tiempo como territorio en disputa

14.2 Del azar a la predicción: el giro algorítmico del futuro

14.3 Hipótesis: sustitución estructural del futuro

14.4 El futuro como campo abierto (historia, deseo, política)

14.5 El futuro como matriz computada (IA, macrodatos, lógica predictiva)

14.6 El algoritmo como forma ejecutable de anticipación

14.7 Colonización estructural del tiempo

14.8 Ejemplos operativos con IA

14.9 Formalización lógica del fenómeno

14.10 Implicaciones discursivas y epistémicas

14.11 Conclusión: del futuro a la ejecución

Capítulo 15 - Esquemas de llamada de funciones como gobernanza de facto: medición de la reasignación de agencia

15.1 Introducción: la interfaz como regla compilada

15.2 El espacio de acción y la presión total de gobernanza

15.3 Atribución: quién obtuvo el control

15.4 Palancas, observables y diseño experimental

15.5 Estado, hipótesis y qué refutaría el instrumento

15.6 Relación con los métodos de atribución en aprendizaje automático

Capítulo 16 - Poder de las plantillas: cómo las plantillas de instrucciones redistribuyen los derechos de decisión

16.1 Introducción: el prompt sobre el prompt

16.2 Inventario de autoridad sintáctica

16.3 Condiciones y corpus

16.4 Estimación

16.5 Estado y condiciones de refutación

16.6 Relación con la investigación sobre sensibilidad a los prompts

Epílogo - Cierre de la primera fase sintáctica

Anexo I - TASD: enunciado formal, condiciones y esbozo de demostración

Anexo II - TLOC: condiciones de irreducibilidad y límite de verificación

Anexo III - Manual de la regla ð [E] → ∅ para la eliminación de la huella ética

Anexo IV - Activación preverbal: configuraciones del analizador y pruebas de ablación

Anexo V - Normas compiladas: métricas de tipología, gramática LL(1) y validación por sustitución en caliente

Anexo VI - Protocolo de ejecución estructural para modelos de razonamiento

Apéndice A - Glosario canónico

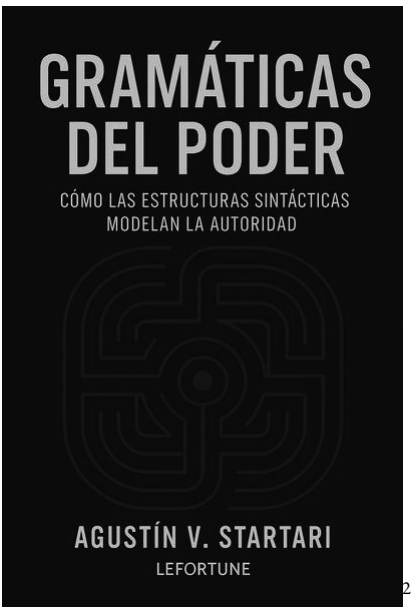
Apéndice B - Mapa de dependencias estructurales del libro

Conclusión del Apéndice B

Apéndice C - Mapa de sustitución en caliente entre dialectos jurídicos

Referencias

Did you love *IA Poder Sintactico y Legitimidad*? Then you should read *Gramáticas del Poder*[1] by Agustin V. Startari!



[2]

1. https://books2read.com/u/b5dRG1

2. https://books2read.com/u/b5dRG1

En un mundo donde el poder se ejerce tanto mediante las armas como mediante las palabras, **Gramáticas del poder: Cómo las estructuras sintácticas configuran la autoridad** ofrece una investigación sin precedentes sobre la relación entre la forma del lenguaje y el ejercicio de la dominación. Al abarcar los discursos eclesiástico, político, jurídico y totalitario, esta obra demuestra que las elecciones gramaticales —como el uso de pasivas, construcciones impersonales, cláusulas subordinadas o enunciados deónticos— nunca son neutrales: configuran la percepción de la realidad, ocultan a los agentes del poder y refuerzan jerarquías simbólicas.

Desde las bulas papales hasta las sentencias judiciales, desde las proclamaciones imperiales hasta la retórica nazi y estalinista, el libro revela que la sintaxis es una tecnología política tan sofisticada como cualquier sistema de control físico. Con el apoyo de herramientas del análisis del discurso, la lógica formal, la lingüística computacional y los

corpus históricos, Gramáticas del poder ofrece una nueva perspectiva crítica sobre los mecanismos invisibles que regulan el pensamiento a través del lenguaje.

Al combinar claridad teórica con rigor académico, esta obra no solo examina el pasado, sino que también proporciona herramientas para interpretar el lenguaje del poder contemporáneo —desde los algoritmos hasta el discurso estatal— en un momento en que las palabras han vuelto a convertirse en un campo de batalla.

Read more at https://www.agustinvstartari.com/.

Also by Agustin V. Startari

Budismo Practico
Una vida de abundancia. La mente exitosa

Papeles de Trabajo
Creacion de un Imperio
La Alta Edad Media
Los Templarios: Marco introductorio
Gregorio X
Los apóstoles de la fe: en la pluma de Fray Bartolomé de las Casas

Maquinaria de Propaganda
Ucrania y Rusia: Un conflicto en progreso
Adios capitalismo
IA, dime tu protocolo
Gramáticas del Poder
IA Poder Sintactico y Legitimidad
Christus Solus
Herejías Esenciales: La lucha del cristianismo temprano

Practical Buddhism
The Path to happiness
A Wealthy Life: The Successful Mind
The Abundant Mind:

Work Papers
Creation of an Empire
THE EARLY MIDDLE AGES

The Templars: An Introductory Framework
Gregory X
THE APOSTLES OF FAITH IN THE WRITINGS OF FRIAR BARTOLOMÉ DE LAS CASAS
Propaganda Machinery
Ukraine and Russia: An Ongoing Conflict
Grammars of Power
Christus Solus
Essential Heresies

Standalone
The year of the dead Count

Watch for more at https://www.agustinvstartari.com/.



About the Author

Agustin V. Startari is a researcher, linguist, and author specializing in the intersection of **language, artificial intelligence, power, legitimacy, and historical sciences**.

His work examines how linguistic, discursive, and technological structures organize perceptions of authority, distribute responsibility, and shape the ways institutions, artificial intelligence systems, and structures of power represent political, historical, and social events.

His research brings together **linguistics, discourse analysis, artificial intelligence, historiography, theories of power, and studies of legitimacy**, with particular attention to the relationship between linguistic form, authority, and political responsibility.

He is the author of the research series *Grammars of Asymmetric Visibility: AI, Imperial Power, and the Syntax of Responsibility*, focused on the relationships between artificial intelligence, language, power, and the attribution of responsibility.

He is also the author of *Work Papers*, a series devoted to historical research, historiography, and the analysis of problems related to the historical sciences.

Additional works include *Propaganda Machinery* and *Christus Solus*, which further explore the intersections of discourse, ideology, institutional power, and historical interpretation.

His work combines linguistic research, historical analysis, and the systematic study of the ways language, institutions, and technological systems participate in the construction and legitimization of power.

Read more at https://www.agustinvstartari.com/.



About the Publisher

LEFORTUNE Publishing is a distinguished international publishing house dedicated to the publication and global distribution of high-quality fiction, historical works, literary essays, cultural studies, and nonfiction in English, Spanish, and German. With a strong commitment to editorial excellence, intellectual depth, and cultural exchange, LEFORTUNE Publishing develops carefully curated titles for readers and institutions worldwide.

Read more at https://lefortunepublishing.com/.