

La amenaza que Silicon Valley no vio venir

Declive estructural de los desarrolladores de IA de frontera en un entorno regulatorio degradado por el autoritarismo competitivo y el negacionismo científico

AUTOR/A

Miguel Moreno 

AFILIACIÓN

Universidad de Granada

FECHA DE PUBLICACIÓN

27 de junio de 2026

LICENCIA

    4.0

RESUMEN +

1 La amenaza subestimada: de variable no modelada a síntoma estructural

Las narrativas del riesgo más accesibles a los desarrolladores de grandes modelos de lenguaje de frontera se construyen comúnmente partiendo de amenazas bien conocidas: la escasez de cómputo, la pugna por el talento, la presión de los modelos de pesos abiertos, el techo de las leyes de escalado, la competencia de los laboratorios chinos. Son riesgos que la industria sabe nombrar porque dispone de herramientas para determinar su coste; admiten cobertura, diversificación, y se ajustan a un calendario. Frente a ese inventario, el entorno de gobernanza de la propia jurisdicción del desarrollador figuraba como un supuesto tácito antes que como una variable, e incluía un Estado de derecho previsible, un procedimiento administrativo motivado y decisiones regulatorias ancladas en la mejor evidencia técnica disponible. La tesis de este ensayo es que ese supuesto era, en realidad, la variable más decisiva de todas, y que su deterioro no es un accidente pasajero, sino un proceso con dirección observable.

Esta sección inicial intenta precisar el estatuto de las dos fuerzas que lo impulsan, porque todo lo demás en este trabajo depende de él. El autoritarismo competitivo y el negacionismo científico no son dos riesgos más que ciertas empresas olvidaron incorporar a sus planes estratégicos; son *síntomas agregadores* de una erosión más profunda del trasfondo sobre el que la cooperación social en ciencia, salud, educación y medioambiente opera en su mejor versión.^[1] Ese trasfondo —oportunidades equitativas de acceso al conocimiento y a la cultura para la mayoría, instituciones de investigación independientes, la libre circulación del talento por mérito— es precisamente lo que permite a un territorio consolidar los ecosistemas industriales capaces de atraer y retener a profesionales de máxima exigencia, sensibles a parámetros de bienestar y calidad de

vida que ningún salario ni prestigio corporativo compensan por sí solos. La arbitrariedad de una gestión política hiperideologizada y refractaria al conocimiento experto no degrada, por tanto, un parámetro aislado, sino las condiciones mismas de posibilidad de la actividad intensiva en talento.

Y lo hace, además, con alcance internacional, arrastrando hacia el precipicio a empresas e instituciones de jurisdicciones muy alejadas del lugar de la decisión. Dos episodios recientes, externos al ámbito de la IA, fijan la firma estructural del fenómeno antes de que aparezca aplicado a los modelos de frontera. El primero, la política arancelaria del «Día de la Liberación» de abril de 2025, invocó potestades de emergencia económica para fijar a voluntad aranceles a casi todos los socios comerciales; la medida desencadenó un desplome del mercado y acabó anulada por el Tribunal Supremo en febrero de 2026 por uso ilícito de una autoridad de excepción, con unos 166.000 millones de dólares recaudados a su amparo quedando expuestos a reclamaciones de reembolso^{1,2}. El segundo, la campaña conjunta de Estados Unidos e Israel contra Irán —que escaló desde la guerra de doce días de junio de 2025³⁻⁵ hasta la operación plena lanzada el 28 de febrero de 2026— exhibió la otra mitad del patrón. Aquella operación se inició sin autorización del Congreso y sobre una justificación pública apenas somera⁶; su objetivo declarado de detener un programa iraní de armas nucleares se enfrentaba a la constatación del Organismo Internacional de Energía Atómica de que no había evidencia de que tal programa existiese, y a la propia evaluación de inteligencia previa de Estados Unidos según la cual Irán no estaba fabricando un arma⁷. Sus efectos no quedaron contenidos dentro de Irán: las represalias alcanzaron a los Estados árabes del Golfo que los propios autores de la campaña habían cortejado, el estrecho de Ormuz quedó cerrado al tráfico marítimo mundial, los precios de la energía subieron y se desencadenó una guerra regional más amplia cuyos costes ninguna de las partes podía acotar de antemano⁷.

Lo inquietante del episodio no es tanto una declaración oficial aislada como la fragilidad del propio esquema de decisión —su arbitrariedad, su urgencia, su apelación a un riesgo de seguridad nacional cuya base fáctica los expertos competentes no pudieron verificar—, junto con la alta probabilidad de que el actor que ejerce semejante discrecionalidad, y los aliados que arrastra consigo, resulten ellos mismos dañados en grados difíciles de anticipar. Ambos casos responden a una misma estructura: una autoridad de excepción —emergencia, seguridad nacional— ejercida a discreción y sin contrapesos, sobre una base fáctica opaca o en disputa, con un efecto que desborda las fronteras del actor que la ejerce. Decisiones de este tipo, tomadas sin la mínima ponderación de consecuencias fundada en criterio experto, no solo elevan la probabilidad de un caos inmanejable que degrada las condiciones mismas de crecimiento y estabilidad que presuponen; en

industrias agudamente sensibles a su entorno operativo, y en medio de una competencia geopolítica feroz, bastan por sí solas para desatar oleadas de efectos imposibles de pronosticar. La hipótesis de este ensayo es que esa estructura, trasladada a la infraestructura cognitiva de la IA de frontera, define el riesgo dominante para los laboratorios que la producen.^[2]

La combinación de las dos fuerzas no es aditiva, sino complementaria, y conviene anticiparlo ya. El autoritarismo competitivo —en el sentido preciso en que Levitsky, Way y Ziblatt caracterizan el giro de Estados Unidos en 2025 hacia un régimen que preserva las formas de la democracia mientras vacía su sustancia mediante el uso selectivo de instrumentos formalmente legítimos^{9,10}— aporta el instrumento: la aplicación discrecional de reglas concebidas para otro fin. El negacionismo científico —como posición de fondo que aglutina a todo un gabinete sin fisuras— retira el freno que limitaría el uso discrecional, al desacoplar el criterio de la decisión de la evidencia técnica que debiera constreñirlo. Juntos producen una gobernanza a la vez arbitraria, porque deja de ser previsible, e inmune a los hechos, porque deja de ser corregible mediante prueba.^[3] La degradación no del regulador, sino del criterio mismo con el que se regula, se describe aquí como *captura de los criterios de gobernanza*, y sostengo que constituye un modo de fallo distinto tanto del exceso como del defecto de regulación con el que la literatura suele operar.

El error de cálculo de las empresas reside exactamente aquí. La fotografía de la toma de posesión de enero de 2025, con los principales ejecutivos del sector —Musk, Bezos, Zuckerberg, Pichai, Cook— sentados por delante de los propios candidatos a miembros del gabinete de la administración^{12,13}, no es una anécdota de protocolo: es el emblema de una apuesta estratégica por el favor incondicional de un Estado discrecional^{14–16}. Que la cúpula del sector se situara siquiera ahí —en la investidura de un presidente cuyos excesos autoritarios y cuya arbitrariedad eran bien conocidos, pues se habían padecido bajo un mandato anterior— delata menos una apuesta calculada que un déficit de visión y de juicio corporativos, una servidumbre legible en los mismos gestos y saludos de la ocasión. La apuesta sobrestima sistemáticamente la durabilidad de esa ventaja. Una posición cuasimonopolística construida sobre la afinidad con un ejecutivo —y no sobre la calidad verificable del producto— hereda la volatilidad de su patrón: el mismo poder que concede el acceso preferente puede revocarlo sin previo aviso, como la suspensión de los modelos más capaces de uno de esos mismos laboratorios habría de demostrar dieciocho meses después al amparo de una directiva de control de exportación¹⁷. Y arrastra, además, un coste reputacional acumulado.

Los modelos de negocio de estas empresas prosperan a escala reseñable en sociedades letradas y de renta alta, con consumidores críticos y prensa libre; son precisamente las sociedades en las que la afinidad explícita con élites de ideología excluyente, antiinmigración o abiertamente autocrática se anota como un pasivo. A medio plazo, ese pasivo puede contaminar todos los nichos —defensa, aeroespacial, logística, nube, prensa, telecomunicaciones, ciberseguridad— en los que la empresa tiene una presencia identificable¹⁸. Lo que tal alineamiento avala, aunque solo sea indirectamente, es un estilo de gestión autoritaria y declaradamente irracional, dispuesta a imponer por la fuerza esquemas antisociales en las cuestiones más sensibles —el control de fronteras, el trato a los grupos vulnerables, la cobertura sanitaria, la asistencia alimentaria, la educación—, y su alcance reputacional no se contiene con facilidad. Erosiona, sobre todo, la confianza de la que depende cualquier colaboración con empresas estadounidenses en cuanto socios fiables, en sectores que van de la defensa y la IA de frontera a las energías renovables, la salud y la aeronáutica: el favor de ejecutivos alineados con grupos radicalmente ideologizados y negacionistas solo confirma un apoyo condicional, supeditado a la sumisión y la deferencia, sin garantía de estabilidad a largo plazo ni compromiso que rebase la conveniencia del momento estratégico inmediato¹⁹.

El orden de los asientos, leído de cerca: 20 de enero de 2025

Tres preguntas se siguen, que el ensayo desarrollará y que conviene dejar planteadas aquí con claridad. La primera es una cuestión de coherencia declarada: cuando una empresa adopta el autoritarismo competitivo como el entorno más favorable a su éxito corporativo, *¿qué objetivos y valores sociales pretende servir, y cómo los justifica en público?* La segunda es una cuestión de autoconocimiento estructural: si el negacionismo científico de los actores con mayor liderazgo político se juzga propicio para empresas que dependen de una base científico-técnica en constante desarrollo, *¿qué han dejado de ver sus gestores sobre las condiciones de su propio ascenso?* Se formula mejor sin retórica. La posición que esos líderes ocupan hoy fue posible gracias a un sistema de educación superior e investigación que consolidó su prestigio a lo largo de décadas de autonomía respecto de ejecutivos de toda estirpe ideológica; un entorno con instituciones académicas debilitadas, extorsionadas económicamente o subyugadas ideológicamente difícilmente habría sido un semillero de grandes empresas tecnológicas —que es exactamente el entorno que su apuesta contribuye a producir, y cuyo primer indicador medible es ya visible: el número de investigadores y desarrolladores de IA que se trasladan a Estados Unidos ha caído un 89 por ciento desde 2017, con el 80 por ciento de ese descenso concentrado solo en el último año²⁴. Ni la erosión se detiene en las universidades: la misma lógica discrecional que purga y copa el Estado administrativo vuelve su presión

sobre los reguladores independientes y las agencias científicas —entre ellos el regulador de los valores (SEC), el servicio oceánico y atmosférico (NOAA) y la agencia de protección ambiental (EPA)—, cuya autonomía estatutaria existe precisamente para que las decisiones que exigen análisis prospectivo, es decir casi todas, descansen en conocimiento experto. Esa presión instala la discrecionalidad como criterio ordinario de acción y anticipa un horizonte de disfunción y daño inmanejable, nacido de un fallo elemental de previsión⁹. La tercera es una cuestión de confianza pública: *¿pueden los ciudadanos, nacionales o extranjeros, depositar una confianza duradera en empresas cuya clase de gobernanza ha sido estructuralmente inhabilitada y sesgada desde el entorno político, y que comparecen en la esfera pública como instancias subordinadas a una red de influencia y conflictos de interés entre regulador y regulado?*

El eje del argumento es el siguiente. Las empresas que desarrollan modelos IA de frontera subestiman el efecto previsible de sustituir la gobernanza por la discrecionalidad, por mucho que esa sustitución pueda inflar transitoriamente sus resultados. Lo que la lógica de la competencia capitalista tiende a volver invisible es que la ventaja decisiva de una democracia de I+D no reside en ningún acto ejecutivo afirmativo, sino en una *propiedad distribuida*: una pluralidad de actores nacionales y extranjeros, la verificación cruzada, la libre circulación del talento por mérito y la posibilidad de recurrir a otras jurisdicciones ante problemas imprevistos. Es ese *marco habilitante* —y no el favor de un gabinete— el que genera y sostiene la ventaja, y es esa propiedad la que la discrecionalidad erosiona, porque desplaza la inversión de largo horizonte y el talento hacia entornos de regla previsible. La continuidad del acceso a la herramienta con la que se ha aprendido a pensar y consolida de facto infraestructura cognitiva colectiva deja entonces de estar garantizada, ya no para el usuario individual —la cuestión de la autonomía epistémica de segundo orden—, sino para la base misma de I+D de la que dependen laboratorios y empresas con alta especialización.

El precio de una élite tecnocrática tan desfasada de la realidad social, económica y política lo pagan primero sus propios empleados y, en proporción al tamaño de la empresa, la sociedad en su conjunto —como dejó patente el colapso de Lehman Brothers en septiembre de 2008, el cuarto mayor banco de inversión de Estados Unidos y, con unos 639.000 millones de dólares en activos, la mayor quiebra de la historia del país: su caída ayudó a congelar los mercados crediticios mundiales y a precipitar al mundo en una recesión que empobreció a millones—²⁵. Que una élite tecnocrática no haya sabido valorar esta asimetría admite una lectura incómoda pero analíticamente sobria: el mecanismo que la seleccionó premió la proximidad al poder por encima del juicio gestor, y un liderazgo así seleccionado optimiza la ventaja inmediata precisamente porque no

está equipado para tasar la condición distribuida que lo hizo posible —un ejército de aduladores ambiciosos antes que un cuerpo de gestores cualificados²⁶.

Los apartados que siguen desarrollan este argumento por orden: la anatomía de las dos corrosiones del criterio (§2), el caso fechado que las ejemplifica en el ámbito de la IA (§3), la asimetría probatoria que pone al descubierto el déficit epistémico del marco (§4), los canales por los que la discrecionalidad se propaga hasta el declive (§5), el desplazamiento de la frontera más allá del alcance del interruptor (§6), el ejercicio prospectivo que cuantifica las trayectorias resultantes (§7), la gobernanza que restauraría un suelo (§8) y las principales objeciones y límites (§9).

2 Dos corrosiones del criterio de gobernanza

2.1 El autoritarismo competitivo como instrumento

El concepto de autoritarismo competitivo no es sinónimo de dictadura, y la fuerza analítica del ensayo depende de no confundir ambos. En la formulación de Levitsky y Way, los regímenes autoritarios competitivos conservan la arquitectura formal de la democracia —elecciones, tribunales, prensa— mientras la vacían de sustancia mediante el uso selectivo y discrecional de instrumentos que siguen siendo, tomados de uno en uno, legales^{27,28}. El rasgo diagnóstico no es la ilegalidad, sino la aplicación selectiva de reglas que existen para otro fin; la discrecionalidad es el producto, no el defecto. Para convertir el Estado en arma, los autócratas electos lo purgan y luego lo copan, según un patrón legible en la Hungría de Orbán, la Turquía de Erdoğan, la Venezuela de Chávez y la India de Modi mucho antes de serlo en Washington⁹. Levitsky, Way y Ziblatt sostienen que Estados Unidos cruzó a esta categoría en 2025; insisten, con igual énfasis, en su reversibilidad, señalando la contestación electoral y la fractura intracoalicional como los canales por los que tales regímenes terminan por deshacerse¹⁰. Los dos términos de esa afirmación importan aquí, puesto que el diagnóstico ancla la variable de riesgo, y la reversibilidad hace verosímil uno de los cuatro futuros desarrollados en el §7.

La misma conjunción ha sido teorizada de manera independiente desde dentro de la filosofía de la tecnología. El análisis del *tecnofascismo* de Coeckelbergh identifica la complicidad estructural entre el poder de las plataformas y el nuevo autoritarismo, un diagnóstico cuya instanciación concreta este ensayo rastrea en el §3²⁹.

Dos implicaciones inciden directamente en la gobernanza de la inteligencia artificial. La primera es que la gobernanza de la IA descansa en instrumentos —control de exportación, autoridad de seguridad nacional, potestades sobre el producto y de

emergencia— cuya *aplicación selectiva* es precisamente el *modus operandi* que el concepto nombra. La IA de frontera no necesita ser prohibida para que el régimen la capture; basta con que su acceso se vuelva discrecional. La segunda es que el fenómeno no se circunscribe a la jurisdicción estadounidense. La misma lente analítica que diagnostica Washington identificó también, hasta abril de 2026, a un Estado miembro autoritario competitivo dentro de la Unión Europea, como ha sido Hungría bajo los dieciséis años de gobierno del Fidesz de Viktor Orbán. Las elecciones parlamentarias del 12 de abril de 2026, que devolvieron una supermayoría de dos tercios al partido opositor Tisza y pusieron fin a la era Orbán, ilustran exactamente el mecanismo de reversión que Levitsky, Way y Ziblatt identifican como el modo característico por el que tales regímenes se deshacen¹⁰. Esa reversión no disuelve el punto; lo extiende. Cualquiera de los dos tipos de gobernanza —codificada y contestable, o discrecional e inapelable— está disponible para cualquier jurisdicción, y la trayectoria de Hungría, aproximación y retirada, confirma que el contraste trazado en el §2.3 es empírico antes que ideológico. El marco codificado y contestable de la UE fue la condición misma que hizo posible la reversión.

2.2 El negacionismo científico como supresión del freno

Si el autoritarismo competitivo aporta el instrumento, el negacionismo científico retira la restricción epistémica que, de otro modo, lo limitaría. La herramienta analítica decisiva aquí no es la proposición de que los públicos «no creen en la ciencia», sino la distinción de Oreskes y Conway entre *ciencia de impacto* y *ciencia de producción*. La hostilidad conservadora, sostienen, no se dirige a la ciencia como tal, sino a las ciencias que identifican daños que implican la necesidad de regular; las ciencias que sustentan a la industria quedan intactas. El corolario es el afilado: cualquier ciencia puede convertirse en ciencia de impacto en el momento en que descubre algo que apunta a la necesidad de acción gubernamental³⁰. El mecanismo es más antiguo que el caso del clima en que se anatomizó por primera vez —la fabricación organizada de la duda documentada en *Merchants of Doubt*³¹ es, en los términos de Proctor y Schiebinger, una producción deliberada de la ignorancia, una agnotología antes que una mera ausencia de conocimiento³².

La aplicación a la IA es exacta y, una vez enunciada, difícil de pasar por alto. *La ciencia de la seguridad de la IA es la ciencia de impacto por excelencia*: su producto epistémico central es la proposición «esto es peligroso, regúlese». Una postura negacionista no necesita negar que la IA funciona —eso es ciencia de producción, y se acoge y se financia—; solo necesita descontar, selectivamente, la evidencia del daño. El resultado es una gobernanza

cuyos criterios se vuelven *anticorrelacionados* con los hechos técnicos. La evidencia de seguridad que un desarrollador responsable produce como su contribución distintiva, registrada en informes internos, pierde tracción en la decisión, mientras la palanca discrecional se aplica sobre fundamentos no técnicos. No es una inferencia hipotética. En sus primeros dieciocho meses, el mismo entorno político que menosprecia la ciencia de impacto ha recortado la financiación federal de la investigación, sembrado dudas públicas sobre hallazgos científicos y priorizado la inversión fuera del sistema universitario —el desfinanciamiento del propio contrato ciencia-sociedad del que depende la democracia de I+D^{33,34}. La literatura contemporánea que cataloga el negacionismo como práctica institucional coordinada, antes que como fallo privado, refuerza el punto³⁵. (Que el concepto de «negacionismo científico» sea él mismo objeto de disputa —hay quien lo tiene por ambiguo e inflado— se aborda en el §9, donde se adopta la versión precisa de Oreskes y Conway justamente porque es falsable y no exige postular un rechazo en bloque de la ciencia.)

2.3 La captura de los criterios de gobernanza y la prueba comparativa

La conjunción de las dos fuerzas arroja un objeto difícil de encajar en el vocabulario estándar del análisis regulatorio. La preocupación clásica, en la tradición de Stigler, es la *captura regulatoria*: el regulador es tomado por el regulado, y las reglas pasan a servir a la industria. Lo que aquí está en juego es distinto y, sostendré, más corrosivo. No es la captura del regulador, sino la degradación del criterio mismo —*captura de los criterios de gobernanza*—, de suerte que la decisión deja de ser (a) previsible y (b) corregible por la evidencia. El autoritarismo competitivo destruye (a); el negacionismo científico destruye (b). El producto es *una gobernanza a la vez arbitraria e inmune a los hechos*. Este es un tercer modo de fallo, ortogonal al eje familiar por el que suele discurrir el debate político (demasiada regulación frente a demasiado poca, como encuadre habitual). Un régimen puede ser mínimo en volumen y aun así catastrófico en su naturaleza, si lo poco que hace es discrecional y está desligado de la prueba. Y produce una paradoja terminal para la empresa preocupada por la seguridad, desarrollada en el §3: el procedimiento estatutario transparente, equitativo y fundado en hechos que el desarrollador responsable exige es exactamente el tipo de gobernanza que un entorno antiempírico —*negacionista*— y autoritario es estructuralmente incapaz de proveer³⁶.

La prueba comparativa afila, antes que ablandar, el diagnóstico. La Unión Europea promulgó, en el Reglamento (UE) 2024/1689, el primer marco horizontal integral del mundo para la IA: codificado, escalonado por riesgo, con obligaciones desplegadas según

un calendario publicado³⁷. Es el opuesto estructural de la anulación discrecional, y al argumento le convendría tratarlo como el contraste perfecto. Pero sería deshonesto, dado que la UE está, en el momento de escribir estas líneas, bajo su propia presión desreguladora: el Ómnibus Digital sobre IA, propuesto en noviembre de 2025 y adoptado por el Parlamento Europeo en junio de 2026, pospone las obligaciones de alto riesgo más decisivas a diciembre de 2027 y agosto de 2028, y analistas de la sociedad civil han calificado el proceso de un debilitamiento desarrollado sin la debida evaluación de impacto —«una derrota para el Estado de derecho»^{38–40}. La comparación, bien trazada, no es por tanto entre una Europa regulada y una América desregulada. Es entre dos *modos* de cambiar las reglas. El debilitamiento europeo procede por el cauce legislativo ordinario —deliberativo, registrado, enmendable, impugnabile ante los tribunales y reversible en la siguiente elección—, mientras que la anulación discrecional procede por una carta despachada a las 17:21 que invoca una autoridad de excepción, sobre una evidencia hurtada, sin notificación ni vía de apelación. La primera es gobernanza que cabe contestar; la segunda es la *ausencia de gobernanza vestida con sus ropajes*. Esa distinción es el meollo del asunto, y el caso que el ensayo aborda ahora es su versión más depurada.

3 La anatomía de un interruptor

El 12 de junio de 2026, a las 17:21 hora del este, el Secretario de Comercio de Estados Unidos remitió al consejero delegado de un desarrollador de LLM frontera una carta que invocaba la autoridad de control de exportación por seguridad nacional y exigía la suspensión de todo acceso a los dos modelos más capaces de la empresa —Fable 5 y Mythos— por parte de cualquier nacional extranjero, dentro o fuera de Estados Unidos, incluidos los propios empleados no ciudadanos de la firma^{41,42}. El alcance de la directiva era nominalmente estatal; pero su efecto práctico fue un apagado global total, porque los modelos corrían en decenas de entornos de nube simultáneos y la empresa carecía de mecanismo para verificar la nacionalidad del usuario en tiempo real a través de esa infraestructura. Una medida formalmente dirigida a los extranjeros produjo, por tanto, una retirada universal. Fue, según el relato experto contemporáneo, la primera vez que Estados Unidos invocaba la autoridad de control de exportación contra el acceso por API de un modelo de IA comercialmente desplegado —la conversión de una cuestión de seguridad del producto en una cuestión de soberanía sobre quién puede tocar una capacidad^{42,43}.

La directiva fue el resultado de dos sucesos convergentes, y cada uno ilumina una cara distinta del patrón. El primero fue una cuestión de procedencia: una empresa de telecomunicaciones que había obtenido acceso temprano al modelo restringido a través del programa de acceso de confianza del desarrollador resultó tener vínculos

corporativos históricos con China, y la administración ordenó revocar su acceso^{42,44}. El segundo fue una cuestión de capacidad: investigadores de seguridad identificaron un método para activar las funciones de ciberseguridad del modelo mediante una instrucción tan banal que se ha convertido en abreviatura del episodio —una orden simple para leer una base de código y corregir sus defectos.^[4] El modelo había sido construido para declinar exactamente esa clase de peticiones; la barrera era la prueba pública de responsabilidad del desarrollador. Que los dos sucesos llegaran con pocos días de diferencia, y que la respuesta fuera una retirada antes que una subsanación, es el primer dato. El segundo es que el propio relato del desarrollador —que la técnica demostrada recuperaba un pequeño número de vulnerabilidades previamente conocidas y menores, de una clase que otros modelos disponibles públicamente, incluido el de un competidor, descubren sin elusión alguna— no pudo ponderarse de manera independiente, porque el razonamiento técnico del gobierno no se hizo público^{41,46}. El lector debe barajar ambos hechos sopesándolos a la vez: la caracterización del episodio procede en buena parte de un actor interesado, y la imposibilidad de comprobarla constituye el fenómeno objeto de análisis.

Es aquí donde se hace visible la ironía más afilada del ensayo, y no es la ironía de un accidente externo, sino la de una trampa autoinducida. Durante al menos dos años, el desarrollador había comercializado sus sistemas más capaces bajo un discurso de amenaza potencial: *este modelo es tan capaz en ciberseguridad que debemos restringirlo; lo hemos construido como, en efecto, una munición controlada*. Esa marca fue la apuesta de legitimidad de la empresa y su rasgo diferenciador en un mercado saturado. Fue también, precisamente, la base jurídica de su propia vulnerabilidad. Un producto descrito en cada lanzamiento como una munición invita, tarde o temprano, a un instrumento diseñado para municiones; el veredicto de un investigador de ciberseguridad sobre el episodio lo expresó con una economía insuperable: la empresa escribió ella misma el sustento jurídico y lo llamó marca¹⁷. El discurso de la responsabilidad se había convertido en la palanca de la captura. Esta es la *trampa de la marca como munición*, de mayor alcance. Sobre un esquema de gobernanza cuyo criterio se ha desacoplado de una valoración proporcionada y fundada en hechos, el acto mismo de publicitar el propio producto como peligroso suministra al Estado discrecional un asidero bien colocado, en el que el momento del tirón queda enteramente a su discreción.

La consecuencia más profunda atañe a lo que en proyectos más amplios se denomina *autonomía epistémica de segundo orden*. La indagación convencional en filosofía de la IA plantea si un sujeto puede razonar bien *mientras usa* un sistema de IA —si la herramienta degrada o potencia la agencia epistémica de primer orden. El caso muestra que esa pregunta descansa en una premisa no examinada: que la relación entre usuario y sistema

sea lo bastante estable como para que la pregunta de primer orden esté bien planteada. Una porción medible y creciente de la investigación, la enseñanza y el razonamiento cívico a través de decenas de jurisdicciones se canaliza hoy a través de un reducido número de proveedores con sede en un único país, y ninguno de esos usuarios posee una pretensión exigible de continuidad de acceso, un derecho contractual de notificación o una vía de apelación. Lo que el interruptor amenaza, entonces, no es meramente un producto, sino una condición: la disponibilidad garantizada del instrumento con el que se ha aprendido a pensar. Y el caso generaliza la amenaza un escalón más allá del usuario individual. El mismo poder discrecional que puede desconectar a un estudiante extranjero puede desconectar a un empleado de nacionalidad extranjera del propio desarrollador —es decir, puede alcanzar la base de I+D, la propiedad distribuida, de la que depende el rendimiento del modelo de frontera⁴⁷. Una jurisdicción que puede apagar a sus propios investigadores extranjeros de la noche a la mañana se vuelve un lugar arriesgado en el que construir, y el capital y el talento migran hacia donde el acceso es previsible. Que esto ocurriera en el mismo mes en que se informó de que el desarrollador había presentado de forma confidencial una solicitud de salida a bolsa no hace sino subrayar la pertinencia del enfoque aquí propuesto. El activo que una empresa de frontera lleva al mercado es su capacidad futura —mantener y mejorar las herramientas y servicios que miles de usuarios ya han contratado—. Pero una capacidad futura rehén de una carta comunicada a cualquier hora del día es un activo de valor nuevamente incierto¹⁷.

El escenario B, observado: la inauguración de Seúl del 17 de junio de 2026

4 El criterio que todo objeto satisface

El rasgo más revelador de la directiva no es su severidad, sino su lógica, y la lógica no sobrevive al escrutinio. La defensa del desarrollador giró en torno a una distinción entre dos clases de *jailbreak*. Un *jailbreak universal* sería un método capaz de derrotar ampliamente las salvaguardas de un modelo, desbloqueando una amplia gama de capacidades peligrosas; la empresa sostenía que ningún evaluador había hallado uno de ese tipo por el momento. Un *jailbreak no universal* puede desbloquear alguna capacidad en circunstancias específicas, y la posición declarada de la empresa —compartida, argüía, por toda la industria— es que la resistencia perfecta a tales elusiones no es actualmente posible para ningún proveedor, y que los jailbreaks universales acabarán por encontrarse^{41,49}. Sobre esas dos premisas, el desarrollador construyó un esquema de defensa muy sólido, orientado a hacer los jailbreaks o bien estrechos o bien muy costosos,

y monitorizar para detectarlos y cerrarlos. Pero las mismas dos premisas, una vez enunciadas, desmontan el criterio de la directiva.

Considérese la regla operativa que la directiva instancia: *la demostración de un jailbreak no universal justifica la suspensión*. Si esa es la regla, y si se sostiene a la vez que la resistencia perfecta es imposible y que todo modelo desplegado es vulnerable a *algún* jailbreak no universal, entonces la condición de suspensión se puede satisfacer contra cualquier modelo en cualquier momento. Un criterio que todo objeto de su clase satisface necesariamente no discrimina entre sistemas seguros e inseguros; no discrimina en absoluto. Y no es, en el sentido lógico, un criterio, sino una palanca. Quien la sostiene puede tirar de ella selectivamente —contra un proveedor, contra un momento de mercado, contra un actor— sin tener que justificar jamás por qué este caso y no otro, porque, según el criterio enunciado, todo caso se ajusta. El propio desarrollador interpretaba correctamente la implicación, puesto que aplicado en toda la industria, el criterio detendría esencialmente todo nuevo despliegue de frontera⁴¹. El punto para este ensayo no es la perturbación de una empresa, sino la demostración de que el criterio rector había sido vaciado de contenido discriminante —la firma misma de la captura de los criterios de gobernanza, ahora visible al nivel de una sola decisión.

La asimetría probatoria agrava el déficit hasta algo que con justicia cabe llamar radical. De un lado se alzaba una directiva con la fuerza de una retirada total, que afectaba a modelos desplegados ante cientos de millones de usuarios. Del otro se alzaba evidencia verbal de una técnica estrecha cuya capacidad resultante el desarrollador juzgaba equivalente a herramientas defensivas de uso cotidiano, y replicable en un modelo del competidor disponible públicamente y no sujeto al mismo control⁴¹. El experto externo que parece haber leído el artículo subyacente alcanzó la misma conclusión desde la dirección opuesta: el comportamiento descrito no puede eliminarse de manera significativa sin empeorar el modelo en el trabajo defensivo de hallar y corregir errores, y la misma capacidad está, o estará en breve, disponible en sistemas extranjeros y de pesos abiertos que no serán afectados por los controles de exportación⁴⁶. Un entorno de gobernanza que puede imponer su instrumento más grave sobre la evidencia más fina y menos auditable no es uno dentro del cual pueda planificar una empresa edificada sobre el talento y el conocimiento experto; el déficit no es de grado, sino de naturaleza, porque la relación entre el peso de la acción y el peso de la prueba ha sido cercenada.

Hay una observación final que pertenece al registro del ensayo antes que al del caso, y conviene enunciarla con franqueza porque su motivación es sobre todo analítica, antes que retórica. Las empresas más expuestas a este déficit incluían a las que con mayor ostentación se habían alineado con la administración cuya discrecionalidad lo produjo.

Un liderazgo capaz de leer un balance financiero pero incapaz de leer un riesgo institucional había apostado por la durabilidad del favor ejecutivo y recibió, a cambio, la demostración de que el favor concedido a discreción es favor revocable a discreción. Que la apuesta se hiciera siquiera es el síntoma; que quienes la hicieron no parezcan haber anticipado la consecuencia más previsible de empoderar a un regulador irrestricto y mal aconsejado es el diagnóstico. La asimetría de la prueba, dicho de otro modo, no era solo una propiedad de la directiva. Era también una propiedad del razonamiento estratégico que había contribuido a instalar las condiciones para ella.

5 Los canales del declive

La crónica acerca del coste de la ignorancia gestora deliberada ha sido desarrollada con gran lucidez en el relato de Lewis sobre el quinto riesgo⁵⁰. Si la captura de los criterios de gobernanza es la enfermedad, el declive es el pronóstico, y el pronóstico debe ganarse mecanismo a mecanismo antes que afirmarse. La tesis no es que el favor discrecional nunca rinda; es que rinde en un horizonte más corto que aquel en el que se realiza el valor de una empresa de frontera, y que la misma discrecionalidad que entrega la ganancia a corto plazo erosiona las condiciones de la ventaja a largo plazo. Cuatro canales transportan la erosión. Se presentan aquí en prosa y se formalizan, como modelo interactivo, en el §7.

El primer canal es la trampa de la marca como munición ya cartografiada: bajo un criterio desacoplado, la propia marca-del-peligro de la empresa eleva la probabilidad de una intervención discrecional, de modo que el marketing que compró legitimidad compró también exposición. El segundo es la *fuga*. Un amplio cuerpo de evidencia establece que la incertidumbre de la política deprime la inversión y el consumo con independencia del contenido de la política; la mera imprevisibilidad de la regla basta para desplazar el compromiso de inversión a largo plazo⁵¹. Para un laboratorio de frontera, la mercancía desplazada no es solo el capital, sino el talento, y el talento es lo más móvil y lo más decisivo. El indicador adelantado consta ya en el registro: el número de investigadores y desarrolladores de IA que se trasladan a Estados Unidos ha caído un 89 por ciento desde 2017, cuatro quintas partes de ese descenso en el año más reciente²⁴; y destacados estudiosos del autoritarismo han trasladado ellos mismos su trabajo a universidades canadienses a medida que la presión política sobre las instituciones estadounidenses se intensificaba¹⁰. El tercer canal es la *sustitución*, tratada en pleno en el §6: a medida que el acceso doméstico a la frontera se vuelve discrecional, la adopción migra hacia sistemas de pesos abiertos y extranjeros, y el valor estratégico del interruptor decae a medida que la frontera se desplaza más allá de su alcance. El cuarto es la *legitimidad*: una intervención discrecional no respaldada por un razonamiento técnico público erosiona la confianza

sobre la que una empresa puede exigir creíblemente una gobernanza fundada en hechos, y la erosión se autorrefuerza, porque una demanda deslegitimada de procedimiento basado en la evidencia deja el criterio aún más desacoplado de los hechos que antes.

Los cuatro canales comparten un único objeto subyacente, y nombrarlo con precisión es el núcleo analítico del apartado. La ventaja de una economía de I+D avanzada nunca fue una concesión afirmativa del Estado; fue una *propiedad distribuida* del ecosistema —una pluralidad de actores nacionales y extranjeros, la verificación cruzada entre instituciones, la libre circulación del talento por mérito y la posibilidad permanente de recurrir a otras jurisdicciones cuando un problema o una oportunidad excedían el marco local⁵². Es esta la propiedad que la discrecionalidad desmantela, no confiscando ningún activo aislado, sino retirando la previsibilidad que hacía valioso invertir en la disposición distribuida. Y el desmantelamiento tiene una homología estructural con el fenómeno político que lo produce. Del mismo modo que el autoritarismo competitivo preserva la *forma* de la democracia mientras vacía su sustancia, una jurisdicción que conserva sus laboratorios pero pierde la previsibilidad del acceso a ellos conserva la *forma* de su ventaja científica mientras malbarata la sustancia. El instrumento concebido para preservar una ventaja estratégica nacional degrada la propiedad misma que la generó. No es una paradoja de ejecución, sino de naturaleza: la discrecionalidad y la ventaja distribuida son, en un horizonte bastante largo, incompatibles.

El registro histórico provee el espejo, y es un espejo que Estados Unidos, de entre todos los países, debería mostrar con nitidez, porque aportó la base para su construcción. La supremacía científica estadounidense en el siglo XX se ensambló, en parte sustancial, con talento que huyó de regímenes que habían purgado ideológicamente sus propias instituciones —los físicos y matemáticos que dejaron la Europa fascista en los años treinta e hicieron del laboratorio angloamericano el centro del mundo. El caso aleccionador del otro lado del libro mayor es el lisenkismo: una base científica subordinada a criterios ideológicos, sus ciencias de impacto suprimidas porque sus hallazgos eran políticamente inoportunos, con una degradación de capacidad que a las ciencias de la vida soviéticas les llevó una generación reparar. La afirmación del ensayo es que la configuración contemporánea corre el primer proceso a la inversa y rima con el segundo: una jurisdicción que degrada la propiedad distribuida que atrajo al talento del mundo, mientras subordina sus ciencias de impacto —la ciencia de la seguridad en primer lugar — a un criterio discrecional y refractario a base empírica. La cifra del 89 por ciento no es todavía un éxodo. Es la primera lectura de un instrumento que ha medido éxodos antes.

6 La frontera se desplaza más allá del interruptor

Un interruptor solo es fuente de poder mientras lo que controla sea escaso e insustituible, y la apuesta central de un régimen de control de exportación sobre la IA es precisamente que la capacidad de frontera es ambas cosas. La apuesta es mala. La capacidad suscitada en el caso que está en el origen de este ensayo —leer una base de código para hallar y corregir sus defectos— es, según la evaluación experta, trabajo defensivo que no puede extirparse sin degradar el modelo, y está ya, o lo estará en breve, disponible en sistemas extranjeros y de pesos abiertos fuera del alcance de cualquier control estadounidense⁴⁶. La historia de los controles de exportación sobre el software es una historia de este error de cálculo. El intento de controlar el cifrado fuerte como munición en los años noventa fracasó frente a la difusión de las matemáticas; la incorporación del «software de intrusión» al Arreglo de Wassenaar en 2013 se redactó de manera tan amplia que amenazaba a los defensores —la divulgación de vulnerabilidades, la respuesta a incidentes, la defensa coordinada— más que a los atacantes, y hubo de renegociarse⁴⁶. La capacidad codificada en pesos y métodos se comporta como la capacidad codificada en algoritmos: viaja, y el control encarece con baño de oro los costes del titular doméstico mientras la frontera se reconstituye en otra parte.

«En otra parte» ya no es hipotético. Los dos ecosistemas líderes de IA se han organizado en torno a estrategias divergentes, y la divergencia recorre exactamente el eje que este ensayo ha venido rastreando. La frontera estadounidense ha concentrado la inversión en modelos cerrados e intensivos en cómputo; China, constreñida en semiconductores avanzados pero respaldada por un apoyo estatal sostenido, se ha organizado en torno al desarrollo abierto y el despliegue rápido, integrando la IA en su economía bajo la rúbrica de la capacidad general⁵³. Los sistemas de pesos abiertos de desarrolladores chinos han pasado de una cuota marginal del uso global a una sustancial en el plazo de un año, a una fracción del coste de entrenamiento de sus equivalentes estadounidenses⁵³. En esa configuración, un interruptor estadounidense discrecional no protege un activo escaso; acelera la migración a un sustituto abundante. El interruptor caduca —pierde su carga— exactamente cuando la capacidad que sella se convierte en mercancía, y la jurisdicción que tiró de él queda habiendo gravado a su propio titular sin haber negado la capacidad a nadie que la quisiera.

El cuadro se complica, no obstante, por una arquitectura paralela. El [marco de la Pax Silica](#) de junio de 2026 —una declaración política firmada por la Unión Europea junto con Alemania, Grecia, Finlandia, Suecia y los Países Bajos— compromete a los adherentes a lo que su principal arquitecto, Jacob Helberg, Subsecretario de Asuntos Económicos de Estados Unidos, describió como un «ecosistema compartido y de confianza» de desarrolladores de IA estadounidenses, en oposición explícita a lo que calificó de objetivo «retrógrado y contraproducente» de una infraestructura nacional de IA autónoma⁵⁴. El

acuerdo es estructuralmente asimétrico: los Estados firmantes obtienen acceso a las capacidades de IA de frontera al tiempo que aceptan a las empresas estadounidenses como proveedores por defecto de la infraestructura de sus servicios digitales, reforzando flujos de datos y canales de dependencia difíciles de revertir una vez incrustados.^[5] Un marco de este tipo no acelera la caducidad del interruptor; trabaja para detenerla, convirtiendo el acceso de terceras jurisdicciones a la IA de frontera en una contraprestación por la lealtad en la cadena de suministro —siendo el compromiso paralelo del acuerdo comercial UE-EE. UU. de comprar al menos 40.000 millones de dólares en semiconductores de IA estadounidenses la expresión contable del arreglo⁵⁴. El interruptor caduca, por tanto, solo para las jurisdicciones no alineadas; la Pax Silica es la estrategia para minimizar ese conjunto.

Los precedentes deben diversificarse, porque las claves de la competitividad geopolítica son más complejas que cualquier sector aislado, y la afirmación del ensayo no se circunscribe al software. El episodio arancelario es la instancia no informática de la misma estructura: una autoridad de excepción invocada a discreción, que produce no el apalancamiento pretendido sobre los socios comerciales, sino un año de incertidumbre que deprimió la inversión, un desplome del mercado y, en última instancia, una constatación judicial de ilegalidad y una exposición a reclamaciones de reembolso a gran escala —el instrumento discrecional dañando al actor que lo blandió^{1,2,51}. El caso Lysenko es la instancia no informática del lado de la ciencia: la ideología degradando una capacidad nacional cuya reconstrucción llevó décadas. La línea de fondo es constante a través de los sectores —defensa, aeroespacial, logística, nube, ciberseguridad, IA— que una empresa pudiera aspirar a dominar mediante el favor ejecutivo afirmativo: sin el ecosistema de fondo que garantiza un suministro renovable de talento y la previsibilidad que lo retiene, la acción ejecutiva autoritaria afirmativa no confiere ningún beneficio duradero. Confiere un trimestre, quizá un año, de ventaja, comprada contra la lenta pérdida de las condiciones que hacen renovable la ventaja. El error estratégico es confundir la *forma* del dominio —ser la empresa a la que el Estado favorece— con su sustancia, que es ser la empresa que el talento y el capital del mundo eligen en términos previsibles.

7 Cuatro futuros

Las afirmaciones prospectivas sobre una industria de cambio veloz invitan a dos modos de fallo —la falsa precisión y la cautela vacua—, y la disciplina de este apartado es evitar ambos razonando en escenarios antes que en pronósticos puntuales⁵⁵. El primer movimiento metodológico es separar la tendencia del régimen. La extrapolación de la capacidad es ciencia de producción: por el método del horizonte de tareas, la longitud de

la tarea que un modelo completa con fiabilidad se ha venido duplicando a una cadencia de escala de meses, con una incertidumbre amplia y explícitamente modelizada⁵⁶. Esa tendencia nos dice que la capacidad *no* es la restricción vinculante del futuro de la frontera. La tesis de este ensayo es ortogonal a la curva de capacidad: lo que gobierna el declive no es la pendiente de la capacidad, sino la *varianza del régimen*. El modelo que sigue es, por tanto, un modelo de la dinámica del régimen y de su propagación hacia la viabilidad corporativa, y es un instrumento heurístico para disciplinar la intuición —un dispositivo para hacer explícitos los bucles de retroalimentación, sus signos y sus condiciones de vuelco—, no un predictor calibrado. Cada coeficiente en él es un supuesto declarado.

El modelo rastrea la viabilidad de un laboratorio de frontera a través de siete variables de estado en el intervalo unitario: la ventaja de capacidad respecto de la frontera abierta y extranjera (C), la previsibilidad regulatoria (R), el acoplamiento epistémico de la gobernanza al hecho técnico (E), el acervo de talento (T), la confianza del capital (K), el acceso al mercado (M) y la legitimidad (L), junto con una intensidad exógena de marca-de-seguridad (S). Cuatro bucles de retroalimentación los conectan. *Bucle 1*, la trampa de la marca como munición: una S alta, bajo E y R bajos, eleva la probabilidad de una intervención discrecional, que deprime M y K. *Bucle 2*, la fuga: una R en caída expulsa capital y talento, lo que erosiona C. *Bucle 3*, la sustitución: un acceso al mercado doméstico en caída eleva la adopción de sustitutos, lo que decae el apalancamiento del interruptor sobre C. *Bucle 4*, la legitimidad: la intervención discrecional sin base técnica erosiona L y, con ella, la capacidad de exigir una gobernanza fundada en hechos, lo que desacopla aún más E. Dos ejes organizan los desenlaces —el tipo de régimen (consolidante frente a reversible) y la postura epistémica (negacionista frente a restauradora de la evidencia)—, arrojando cuatro futuros.

	Negacionista ($E \rightarrow 0$)	Restauradora de la evidencia ($E \rightarrow 1$)
Autoritarismo competitivo consolidado ($R \rightarrow 0$)	A · Captura soberana / declive gestionado. La empresa sobrevive como campeón nacional despojado de la propiedad distribuida que la hacía valiosa; la capacidad migra al exterior; el argumento de « Brains as Brand » plenamente realizado.	C · Combinación inestable. Un régimen discrecional que, no obstante, se ata a la evidencia; transitorio e improbable, útil sobre todo para mostrar por qué R y E tienden a covariar.

	Negacionista ($E \rightarrow 0$)	Restauradora de la evidencia ($E \rightarrow 1$)
Reversible / disputado ($R \rightarrow$ medio)	B · Arbitraje jurisdiccional / vaciamiento. La empresa se cubre deslocalizando (la oficina de Seúl es la ficha contemporánea), pero afronta la fragmentación del cumplimiento; declive parcial y dependiente de la jurisdicción.	D · Recuperación / reacoplamiento. Se da la reversibilidad de Levitsky; se restaura un procedimiento previsible y fundado en hechos; el liderazgo en seguridad se vuelve un activo antes que un pasivo.

El modelo de escenarios siguiente corre como una figura interactiva autocontenida; sus controles, su lectura y sus instrucciones de uso viajan dentro de ella. Arrastre los deslizadores o use los botones de escenario, y lea la *cola inferior* de la banda antes que la línea central.

FIGURE · INTERACTIVE SCENARIO MODEL

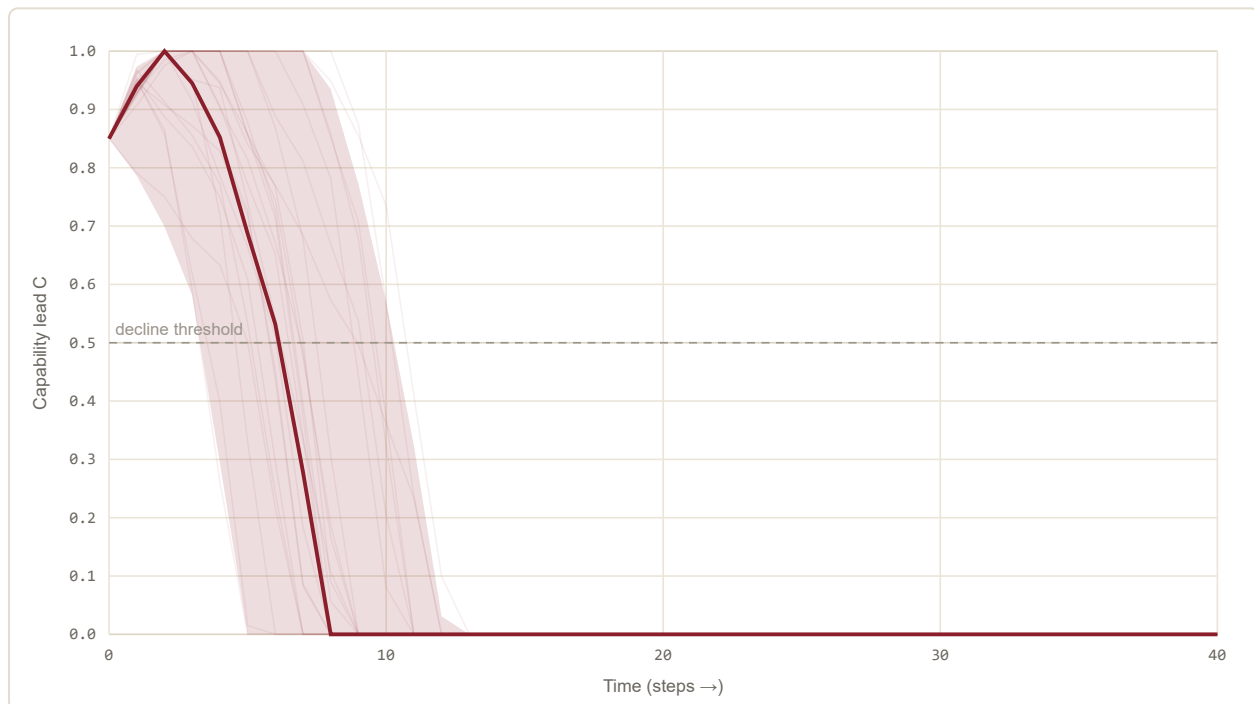
Regime variance and the decline of a frontier laboratory

A heuristic model of *regime* dynamics — not a forecast. Capability is treated as a rising trend; what governs decline is the variance of the governance regime. Every coefficient is a declared assumption.

Safety-branding intensity (S)	0.70	Initial regulatory predictability (R_0)	0.50
Initial epistemic coupling (E_0)	0.50	Monte-Carlo trials	600

☒ Regime consolidates (R decays over time)

SCENARIOS A · Sovereign capture B · Hollowing D · Recovery Reset



MEDIAN TERMINAL C P(DECLINE · $C < 0.5$) band = 10th-90th pct · faint lines = individual runs
0.00 **100%**

The line is the **median** capability lead C across the ensemble; the shaded band is the **10th–90th percentile**; the dashed line marks the **decline threshold** ($C = 0.5$). Read the *lower tail* and which loop drives it, not the central value: push S high while R_0 and E_0 are low to see the brand-as-munition and flight loops collapse the tail. Drag the sliders or pick a scenario.

Scenario coordinates — **A** Sovereign capture / managed decline: $R_0 \approx 0.30$, $E_0 \approx 0.30$, consolidates. **B** Jurisdictional arbitrage / hollowing: $R_0 \approx 0.50$, $E_0 \approx 0.40$, non-consolidating. **D** Recovery / re-coupling: $R_0 \approx 0.70$, $E_0 \approx 0.70$, non-consolidating. (Scenario C — discretionary regime that nonetheless binds itself to evidence — is transient and is discussed in the text rather than presented as a preset.)

Figura 2. Modelo interactivo de escenarios (solo HTML). La línea es la mediana de la ventaja de capacidad C ; la banda sombreada es la envolvente Monte-Carlo del percentil 10 al 90; las líneas tenues son ejecuciones individuales; la línea discontinua es el umbral de declive ($C = 0,5$). Controles: intensidad de la marca-de-seguridad S , previsibilidad regulatoria inicial R_0 , acoplamiento epistémico inicial E_0 , un conmutador de consolidación del régimen y el tamaño del conjunto. Los preajustes A / B / D fijan las coordenadas de escenario comentadas en el texto.

Nota sobre los formatos

El modelo ha de leerse por su cola inferior y por sus atribuciones de bucle, no por su valor central. Las cuestiones empíricas decisivas se plantean, por tanto, como indicadores adelantados que confirmarían o refutarían la trayectoria. El diagnóstico se *confirma* por la repetición del patrón —un segundo interruptor discrecional—, por una aceleración de la fuga de talento y capital hacia jurisdicciones previsibles, por la deslocalización defensiva de laboratorios y por el descuento explícito de la evidencia de seguridad en las decisiones regulatorias. Se *refuta* por la institucionalización de un procedimiento estatutario transparente y motivado, por la restauración con un derecho de apelación efectivo y por una reversión del régimen —mediante resultados electorales o fractura intracoalicional— sin coste duradero para la innovación de frontera. Un ejercicio prospectivo se gana el sustento no prediciendo qué rama se cumplirá, sino especificando, por anticipado, las observaciones que nos lo dirían.

8 Qué marco de gobernanza restauraría la seguridad jurídica

El remedio se sigue del diagnóstico, y su estructura la dicta el hecho de que hay dos tipos de corrosión diferenciados que deben atenderse a la vez. Tres instrumentos, tomados del horizonte regulatorio del proyecto más amplio, hacen el trabajo. El primero es una *obligación de continuidad mínima*, análoga a las exigencias de servicio universal en telecomunicaciones: un suelo de acceso garantizado por debajo del cual el servicio de un proveedor a usuarios dependientes de terceras jurisdicciones no puede caer sin el debido proceso. El segundo es un *periodo de notificación obligatorio* antes de cualquier suspensión unilateral, que convierte una retirada instantánea en un acto revisable. El tercero es un *mecanismo de apelación multilateral*, ante el que la base técnica de una suspensión pueda ser impugnada por partes distintas del Estado que suspende.

La afirmación no obvia es que estos instrumentos no son meras protecciones para el usuario; cada uno reacopla el criterio al hecho, y ataca así la corrosión negacionista tan

directamente como ataca la discrecionalidad. Una notificación obligatoria que ha de ser *motivada* reintroduce la obligación de enunciar una base técnica; una apelación con *revisión técnica* reintroduce el control probatorio que el negacionismo científico había eliminado; y un suelo de continuidad con *debido proceso* reintroduce la previsibilidad que el autoritarismo competitivo había retirado. Donde la enfermedad es la captura de los criterios de gobernanza —un criterio vuelto arbitrario e inmune a los hechos—, la cura ha de restaurar tanto la previsibilidad como la corregibilidad, y estos instrumentos están diseñados para restaurar cada una.

Bajo un régimen autoritario competitivo consolidado, ¿quién instituye la apelación? Un Estado discrecional no atará de buen grado su propia discrecionalidad, y uno refractario a los hechos no reinstaurará de buen grado el control probatorio que ha descartado. Es precisamente por esto por lo que la apelación debe ser *multilateral*: ninguna jurisdicción discrecional aislada puede ser garante de un suelo frente a su propia discrecionalidad, y los comunes de la infraestructura cognitiva —usados en decenas de jurisdicciones— tienen la legitimación para exigir una gobernanza que ningún miembro aislado puede suministrar⁵², un estándar frente al cual el adversario institucional ya se había definido en junio de 2026. El principal arquitecto del marco de la Pax Silica desestimó públicamente la infraestructura autónoma de IA de terceras jurisdicciones como una fragmentación contraproducente —la posición sistemática contra la que un suelo de continuidad tendría que construirse, sostenida desde el despacho que tiene el interruptor⁵⁴. Esa dificultad estructural es también la razón por la que la tesis de la reversibilidad es clave en todo el argumento. El remedio presupone un reacomplamiento de criterio y hecho que solo una corrección política puede proporcionar. Los instrumentos describen el destino; si la comunidad política puede viajar hasta él es una cuestión que debe sortear sus propios obstáculos.

9 Objeciones y límites

Un ensayo crítico de esta clase vive o muere por la fuerza del caso en su contra. Al menos cinco objeciones merecen ser consideradas y articuladas con solidez.

Primera, la lectura benigna. El caso ancla fue enmarcado por el propio proveedor como un malentendido por resolver, y quizá lo fuera: improvisación burocrática antes que estrategia autoritaria competitiva. La respuesta es que el marco no requiere intención; requiere estructura —un instrumento jurídico de excepción, un procedimiento opaco y un efecto desproporcionado respecto del riesgo alegado. Pero la concesión debe hacerse en su justa medida: la distinción entre malentendido y estrategia está empíricamente abierta, y la afirmación del ensayo versa sobre el *patrón estructural que ambas lecturas*

comparten —que un interruptor de esta clase existe y es accionable a discreción. El diagnóstico no depende de atribuir un motivo en particular, y la prioridad aquí no ha sido establecerlo.

Segunda, el declive endógeno. Hay evidencia seria de que la capa de entrenamiento y aplicación está en una dinámica económicamente insostenible —el gasto de capital superando a la utilidad demostrable, una subida de aproximadamente el 30 por ciento en los precios del cómputo y dependencias entre laboratorios que parecen más una toma mutua de rehenes que una competencia⁵⁷. Quizá el declive sea una burbuja, no un fallo de gobernanza. La respuesta correcta no es una pugna de hipótesis, sino un reconocimiento de composición: la fragilidad financiera *amplifica* la sensibilidad al riesgo de régimen, porque un sector que quema capital no puede absorber choques discrecionales. El ensayo debe modelizar la gobernanza y la economía como factores multiplicativos, no rivales, declarar explícitamente qué porción del declive atribuye a cada uno y conceder la infraidentificación que se sigue de su interacción.

Tercera, que el «negacionismo científico» sea un concepto inflado. Una literatura reciente sostiene que el término amalgama fenómenos dispares y que la alarma a su alrededor está sobredimensionada¹¹. El ensayo la desarma adoptando solo la versión *precisa* de Oreskes y Conway —el descuento selectivo de la ciencia de impacto—, que es falsable y no requiere afirmación alguna de un rechazo en bloque de la ciencia. El pilar se construye sobre la versión del concepto que sobrevive a la objeción más fuerte contra él.

Cuarta, la reconstrucción más sólida de la seguridad nacional. Hay un argumento legítimo de que la capacidad ofensiva de ciberseguridad de frontera pueda justificar restricción, y el ensayo no niega la autoridad del Estado para bloquear despliegues genuinamente inseguros. Lo que ataca es la *forma* —la ausencia de un procedimiento transparente, motivado y técnicamente fundado—, y muestra, mediante la distinción universal/no universal y la amplia disponibilidad de la capacidad, por qué este caso concreto incumple el estándar que la propia autoridad presupone. Para evitar un hombre de paja, debe representarse la posición experta contraria: la tesis, sostenida por practicantes, de que los controles de hecho *dañan* la ciberdefensa al negar a los defensores las herramientas que los atacantes ya poseen⁴⁶. La afirmación del ensayo es la más estrecha y más defendible —no que el poder no deba existir, sino que se ejerció de un modo que la propia justificación del poder no puede autorizar.

Quinta, la reversibilidad. Si el régimen diagnosticado es reversible, como sostienen Levitsky, Way y Ziblatt¹⁰, entonces el escenario D está disponible y el declive puede evitarse. Lejos de socavar el argumento, este es su punto: el valor del modelo es exhibir las trayectorias condicionales, y los indicadores adelantados del §7 son los que nos dicen

en qué rama se halla la comunidad política. Un diagnóstico que sea reversible no es por ello falso; es, precisamente por eso, accionable.

Un límite que atañe al método procede en este apartado, porque es también un dato del análisis. Buena parte de la caracterización del caso ancla tiene como fuente una parte interesada, y el razonamiento técnico del gobierno nunca se hizo público. El ensayo trata esa asimetría no como una licencia para afirmar un motivo, sino como el fenómeno mismo —la opacidad de segundo orden en la que ni el público ni el proveedor sancionado pueden auditar la base de una decisión cuyo alcance trasciende jurisdicciones. La imposibilidad de saber no es una laguna del argumento. Es el meollo del argumento.

10 Conclusión

La amenaza que Silicon Valley no vio venir no fue un competidor, un techo de capacidad ni un muro de cómputo. Fue la degradación del criterio de gobernanza del que toda la empresa dependía tácitamente, por dos fuerzas que la industria trató como ruido de fondo y que son, en realidad, motores complementarios de un único fallo: el autoritarismo competitivo, que vuelve discrecional la regla, y el negacionismo científico, que la vuelve inmune a los hechos. Su conjunción —la captura de los criterios de gobernanza— convierte la propia marca de seguridad de las empresas en el fundamento jurídico de su captura, y convierte la previsibilidad que retiene al talento del mundo en un favor discrecional revocable por carta. Las empresas más expuestas fueron las que habían apostado por ese favor, confundiendo la forma del dominio con su sustancia; la sustancia fue siempre la propiedad distribuida de una democracia de I+D, y la discrecionalidad la disuelve. El remedio es una gobernanza que restaure tanto la previsibilidad como la corregibilidad —un suelo de continuidad, una notificación motivada, una apelación multilateral—, y el diagnóstico que lo motiva es reversible, pero no autocorrector. El interruptor existe se gobierne o no. La pregunta que este ensayo ha tratado de volver ineludible es si las jurisdicciones que lo tienen aprenderán, antes de que lo haga el talento, que un interruptor del que se puede tirar a voluntad es un interruptor que ya se ha empezado a perder.

Agradecimientos y financiación

Este trabajo se desarrolló en el seno del proyecto de investigación AUTAI *Inteligencia artificial y autonomía humana* (PID2022-137953OB-I00), que indaga las condiciones de la autonomía humana en la interacción con sistemas de IA generativos y agénticos. Ese cometido abarca todo el arco de la tecnología: el diseño y entrenamiento de los grandes modelos de lenguaje, la infraestructura a

través de la cual se usan y el marco de gobernanza —el orden regulatorio— por el que servicios seguros y fiables se ponen al alcance del público, sin lagunas ni asimetrías injustificables. El presente ensayo pertenece a la línea de gobernanza del proyecto.

Sobre el uso de herramientas de IA. En el espíritu de la transparencia que este ensayo defiende, el autor deja constancia de que se emplearon modelos desarrollados por Anthropic —Claude Sonnet 4.6 y Claude Opus 4.8— como instrumentos en varias etapas de la preparación: en la articulación y estructuración de sucesivos borradores del argumento; en el cotejo de referencias bibliográficas y datos de fuentes contra materiales primarios; y en la construcción de la simulación interactiva de escenarios y de su versión estática de respaldo del §7. El autor fijó las preguntas, dirigió cada tarea, revisó y verificó el resultado, y conserva la responsabilidad exclusiva del análisis, los juicios y cualquier error que sobreviva. Que los instrumentos empleados pertenezcan a la misma clase de sistemas que el ensayo examina no es incidental a él: el argumento se ofrece no desde fuera de la dependencia que describe, sino desde dentro de ella.

Referencias

1. Supreme Court of the United States. (2026). *Learning Resources, Inc. v. Trump*. 607 U.S. ____ (2026), No. 24–1287. https://www.supremecourt.gov/opinions/25pdf/24-1287_4gcj.pdf
2. Manak, I. et al. (2026). *A Year After «Liberation Day», Experts Review the Costs of Trump's Tariffs*. Council on Foreign Relations. <https://www.cfr.org/articles/a-year-after-liberation-day-experts-review-the-costs-of-trumps-tariffs>
3. CNN. (2025). *Early US Intel Assessment Suggests Strikes on Iran Did Not Destroy Nuclear Sites*. cnn.com. <https://edition.cnn.com/2025/06/24/politics/intel-assessment-us-strikes-iran-nuclear-sites>
4. CSIS Nuclear Network. (2025). *Disruption or Dismantlement: Diverging Assessments of Iran Nuclear Strikes*. nuclearnetwork.csis.org. <https://nuclearnetwork.csis.org/disruption-or-dismantlement-diverging-assessments-of-iran-nuclear-strikes/>
5. Congressional Research Service. (2025). *U.S. Strikes on Nuclear Sites in Iran* (IN12571). Library of Congress. <https://www.congress.gov/crs-product/IN12571>
6. Anderson, S. R. (2026). *After the Strike: The Danger of War in Iran*. Brookings Institution commentary. <https://www.brookings.edu/articles/after-the-strike-the-danger-of-war-in-iran/>
7. House of Commons Library. (2026). *Israel/US–Iran Conflict 2026: Background and UK Response* (CBP-10521). Research Briefing, House of Commons Library. <https://commonslibrary.parliament.uk/research-briefings/cbp-10521/>
8. Ülgen, S. (2026). *From Trade Dependence to Geopolitical Leverage: The EU in an Era of Weaponized Interdependence*. Carnegie Endowment for International Peace, Europe Program. <https://carnegieendowment.org/europe/research/2026/06/from-trade-dependence-to-geopolitical-leverage-the-eu-in-an-era-of-weaponized-interdependence>
9. Levitsky, S., & Way, L. A. (2025). The Path to American Authoritarianism: What Comes After Democratic Breakdown. *Foreign Affairs*, 104(2). <https://www.foreignaffairs.com/united-states/path-american-authoritarianism-trump>
10. Levitsky, S., Way, L. A., & Ziblatt, D. (2026). The Price of American Authoritarianism. What Can Reverse Democratic Decline? *Foreign Affairs*. <https://www.foreignaffairs.com/united-states/american-authoritarianism-levitsky-way-ziblatt>

11. Contessa, G. (2025). *Science Denial: Post-Truth or Post-Trust?* Cambridge University Press.
12. Brooks, E. (2025). *Lawmakers Side-Eye Billionaire CEOs Getting Prime Trump Inauguration Seats*. thehill.com. <https://thehill.com/homenews/house/5096085-trump-inauguration-tech-ceos/>
13. Helmore, E. (2025). Trump inauguration: Zuckerberg, Bezos and Musk seated in front of cabinet picks. *The Guardian*. <https://www.theguardian.com/us-news/2025/jan/20/trump-inauguration-tech-executives>
14. Schaake, M. (2024). *The Tech Coup: How to Save Democracy from Silicon Valley*. Princeton University Press.
15. Taskale, A. R. (2026). Materialized Science Fiction: The Tech Oligarchy's Blueprint for a New Global Order. *Global Studies Quarterly*, 6(1), ksag002. <https://doi.org/10.1093/isagsq/ksag002>
16. Massanari, A. L. (2024). *Gaming Democracy: How Silicon Valley Leveled Up the Far Right*. The MIT Press. <https://doi.org/10.7551/mitpress/14531.001.0001>
17. Kahn, J. (2026). Anthropic Disables Fable and Mythos AI Models After U.S. Government Bars It from Giving Foreigners Access. En *Fortune*. <https://fortune.com/2026/06/13/anthropic-fable-mythos-ai-models-disabled-us-government-foreign-access/>
18. Fung, A., & Lessig, L. (2025). *Oligarchy in the open: What happens now as the U.S. is forced to confront its plutocracy problem?* Harvard Kennedy School. <https://www.hks.harvard.edu/faculty-research/policycast/oligarchy-open-what-happens-now-us-forced-confront-its-plutocracy>
19. Montgomery, B. (2025). Silicon Valley's Tycoons Are Bending the Knee to Trump. *The Guardian*. <https://www.theguardian.com/technology/2025/jan/11/trump-silicon-valley>
20. Faguy, A. (2025). Washington Post Cartoonist Quits After Bezos Satire Is Rejected. *BBC News*. <https://www.bbc.com/news/articles/cn4xk4y2emro>
21. Jiménez, M. (2025). Dimite una dibujante del «Washington Post» por el veto a una viñeta en que Bezos se arrodillaba ante Trump. *El País*. <https://elpais.com/comunicacion/2025-01-05/dimite-una-dibujante-del-washington-post-por-el-veto-a-una-vineta-en-que-bezos-se-arrodillaba-ante-trump.html>
22. Mathur-Ashton, A. (2025). The Notable Attendees – and No-Shows – at Trump's Inauguration. *U.S. News & World Report*. <https://www.usnews.com/news/national-news/articles/2025-01-20/who-attended-trumps-inauguration-the-billionaires-and-far-right-leaders-who-gathered-at-the-swearing-in>
23. Preskey, N. (2025). Influencers, Tech Bros, MMA Fighters: Trump's Inauguration Guests. *BBC News*. <https://www.bbc.com/news/articles/cgkjgmkn10ko>
24. Stanford Institute for Human-Centered AI. (2026). *The 2026 AI Index Report*. Stanford University. https://hai.stanford.edu/assets/files/ai_index_report_2026.pdf
25. Wiggins, R. Z., Piontek, T., & Metrick, A. (2014). *The Lehman Brothers Bankruptcy A: Overview*. Yale Program on Financial Stability (YPFS) Case Study. <https://ypfsresourcelibrary.blob.core.windows.net/fcic/YPFS/001-2014-3A-V1-LehmanBrothers-A-REVA.pdf>
26. Matten, D. (2026). Fascism as a Management Philosophy. *Philosophy of Management*, 25, 21-49. <https://doi.org/10.1007/s40926-025-00355-1>
27. Levitsky, S., & Way, L. A. (2010). *Competitive Authoritarianism: Hybrid Regimes After the Cold War*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511781353>
28. Levitsky, S., & Way, L. A. (2002). Elections Without Democracy: The Rise of Competitive Authoritarianism. *Journal of Democracy*, 13(2), 51-65. <https://doi.org/10.1353/jod.2002.0026>

29. Coeckelbergh, M. (2026). Technofascism: AI, Big Tech, and the New Authoritarianism. *AI & Society*. <https://doi.org/10.1007/s00146-026-02862-9>
30. Oreskes, N., & Conway, E. M. (2022). From Anti-Government to Anti-Science: Why Conservatives Have Turned Against Science. *Daedalus*, 151(4), 98-123. https://doi.org/10.1162/daed_a_01946
31. Oreskes, N., & Conway, E. M. (2010). *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. Bloomsbury Press.
32. Proctor, R. N., & Schiebinger, L. (Eds.). (2008). *Agnotology: The Making and Unmaking of Ignorance*. Stanford University Press.
33. Mills, M. A., & St. Clair, P. (2025). The Strange New Politics of Science. *Issues in Science and Technology*, 41(3), 40-48. <https://doi.org/10.58875/NDTQ1755>
34. Olesko, K. M., Eames, A., Mody, C. C. M., Löwy, I., Zeller, T., Walker, M., Ash, M. G., & Haraway, D. (2025). The Crisis in American Science. *History of Science*, 63(2), 125-165. <https://doi.org/10.1177/00732753251343655>
35. Bruni, E., & Lefsrud, L. M. (2026). The Collective Organization of Science Denial: Toward a Framework for Collective Response. *Open Access Government*, 184-185. <https://www.openaccessgovernment.org/article/the-collective-organization-of-science-denial-toward-a-framework-for-collective-response/202441/>
36. Oreskes, N. (2019). *Why Trust Science?* Princeton University Press.
37. European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*. Official Journal of the European Union. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
38. Council of the European Union and European Parliament. (2026). *Digital Omnibus on AI: Provisional Agreement to Simplify and Streamline AI Rules*. consilium.europa.eu; European Parliament. [https://www.europarl.europa.eu/RegData/etudes/ATAG/2026/789329/EPRS_ATA\(2026\)789329_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2026/789329/EPRS_ATA(2026)789329_EN.pdf)
39. Liberties (Civil Liberties Union for Europe). (2026). *AI Omnibus: Fast-Tracking Deregulation, Weakening Digital Rights*. liberties.eu. <https://www.liberties.eu/en/stories/ai-omnibus-ep-vote/45715>
40. Tech Policy Press. (2026). *EU's AI Act Delays Let High-Risk Systems Dodge Oversight*. techpolicy.press. <https://www.techpolicy.press/eus-ai-act-delays-let-highrisk-systems-dodge-oversight/>
41. Anthropic. (2026). *Statement on the US Government Directive to Suspend Access to Fable 5 and Mythos 5*. Anthropic. <https://www.anthropic.com/news/fable-mythos-access>
42. Tech Times. (2026). *Fable 5 Export Ban Day Six: Anthropic Opens Seoul Office, Vows Models Back in Days*. techtimes.com. <https://www.techtimes.com/articles/318668/20260618/fable-5-export-ban-day-six-anthropic-opens-seoul-office-vows-models-back-days.htm>
43. Bombach, K. M., Dwyer, K., Roberson, J. E., et al. (2026). *AI Company Anthropic Suspends Access to Claude Fable 5, Claude Mythos 5 Following US Export Control Directive*. National Law Review (Greenberg Traurig LLP alert). <https://natlawreview.com/article/ai-company-anthropic-suspends-access-claude-fable-5-claude-mythos-5-following-us>
44. Korea JoongAng Daily. (2026). *Anthropic Confident of Re-enabling Mythos, Fable 5 Access in Coming Days, Executive Says*. koreajoongangdaily.com. <https://www.koreajoongangdaily.com/business/anthropic-confident-of-reenabling-mythos-fable-5-access-in-coming-days-executive/12727522>

45. DataVLab. (2026). *Red-Teaming LLMs: A Practitioner's Guide for 2026*. <https://datavlab.ai/post/red-teaming-llms-practitioner-guide-2026>.
46. Moussouris, K. (2026). *The Fable 5 Export Controls Harm US Cyber Defense*. Luta Security. <https://www.lutasecurity.com/post/the-fable-5-export-controls-harm-us-cyber-defense>
47. Coeckelbergh, M. (2020). *AI Ethics*. MIT Press.
48. Tech Jacks Solutions. (2026, junio 18). *Anthropic Opens Seoul Office With Six Enterprise Deals While Its Top Models Remain Export-Blocked*. <https://techjacksolutions.com/ai-brief/anthropic-opens-seoul-office-with-six-enterprise-deals-while/>
49. Anthropic. (2026). *Claude Fable 5 and Claude Mythos 5*. Anthropic. <https://www.anthropic.com/news/claude-fable-5-mythos-5>
50. Lewis, M. (2018). *The Fifth Risk: Undoing Democracy*. W. W. Norton & Company.
51. Basco, S. (2026). How Trump's Renewed Tariff Chaos Will Stifle Investment Around the World. En *The Conversation*. <https://theconversation.com/how-trumps-renewed-tariff-chaos-will-stifle-investment-around-the-world-276789>
52. Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press.
53. U.S.-China Economic and Security Review Commission. (2026). *Two Loops: How China's Open AI Strategy Reinforces Its Industrial Dominance*. USCC. https://www.uscc.gov/sites/default/files/2026-03/Two_Loops--How_Chinas_Open_AI_Strategy_Reinforces_Its_Industrial_Dominance.pdf
54. Ponce de León, R. (2026). *Qué es la Pax Silica y por qué puede ser una trampa para Europa*. eldiario.es. https://www.eldiario.es/internacional/pax-silica-trampa-europa_1_13334330.html
55. Tetlock, P. E., & Gardner, D. (2015). *Superforecasting: The Art and Science of Prediction*. Crown Publishers.
56. Kwa, T., West, B., Becker, J., et al. (2025). Measuring AI Ability to Complete Long Tasks. *arXiv preprint, arXiv:2503.14499*. <https://doi.org/10.48550/arXiv.2503.14499>
57. Epoch AI. (2026). *How Much AI Compute Do Frontier Labs Use?* epoch.ai/gradient-updates. <https://epoch.ai/gradient-updates/frontier-labs-dont-use-most-ai-compute>

Notas

- [1] A modo de cautela, conviene aclarar que nada de lo aquí dicho presupone que la gestión política pueda despojarse de sus connotaciones y constricciones partidistas allí donde están en juego las prioridades de la gobernanza corporativa o económica. El énfasis recae en la dificultad para desagregar, dentro de cualquier estrategia empresarial prospectiva, los elementos que un análisis crítico exhaustivo halla ocultos bajo síntomas aislados —el antiintelectualismo, la aversión a las élites académicas y a las agencias e instituciones que gestionan conocimiento experto—, habitualmente subsumidos en categorías más amplias como el negacionismo científico *en cuanto* posición declarada de un partido con aspiraciones presidenciales. Dado el volumen de contingencia e incertidumbre que lastra toda gestión política, salpicándola de una irracionalidad y una complejidad a menudo inmanejables, el acto mismo de atribuir esos rasgos adicionales delata ya un déficit de instrumentos fiables, operativos y eficaces para reorientar o afrontar riesgos y escenarios imprevistos. ↩
- [2] Una calibración independiente y útil de este patrón estructural: en junio de 2026, Sinan Ülgen, de Carnegie Europe, describió a Estados Unidos como uno de los mayores multiplicadores

sistémicos de riesgo para las cadenas de suministro europeas —aplicando el vocabulario del riesgo sistémico, desde dentro de la asociación transatlántica, al propio Estado que la ancla⁸. ↩

- [3] El alcance de esta combinación de fuerzas suele subestimarse, incluso en obras recientes como la de Gabriele Contessa¹¹. Una vez que quien decide en última instancia opera desacoplando sus criterios de la evidencia que debiera vincularlos —en la política de vacunación, o en la regulación de la calidad del aire y la contaminación, tanto como en la inteligencia artificial—, la pregunta pertinente ya no es qué procedimiento resultará dañado, sino cuál no: qué parte de las interacciones inevitables entre corporaciones y actores políticos escapa a la arbitrariedad de quien decide en último término, y qué procedimiento, vital o trivial, permanece intacto bajo un esquema que no admite error corregible mediante pruebas. ↩
- [4] Intentar vulnerar los filtros de seguridad con prompts creativos se ha convertido en una disciplina de ingeniería estructurada con metodologías formales, herramientas automatizadas, normas establecidas y equipos dedicados que ejecutan pruebas adversariales continuas como parte de cada canal de despliegue en producción⁴⁵. ↩
- [5] El blanco explícito de la crítica de Helberg es el Pacto Digital Global de las Naciones Unidas, que insta a cada Estado a desarrollar su propia infraestructura completa de IA —hardware, datos, modelos y capacidad computacional nacional— como base de la soberanía digital. Helberg caracteriza esa visión como una forma de «mediocridad sincronizada», con el argumento de que lleva a cada país a reconstruir capacidades que ya existen a mayor escala en otra parte. Bajo la lógica de la Pax Silica, esa aspiración se sustituye por una división del trabajo en la que los desarrolladores estadounidenses de frontera suministran capacidades de IA a cambio de lo que Helberg presenta como una «cadena de suministro fiable e indispensable» de materias primas críticas —el arreglo presentado públicamente como «seguridad económica mutua». ↩