Ética con IA: del soft law al control real por contrato y código

Ethics with AI: from Soft Law to Real Control through Contract and Code

Edgar David Oliva Terán

Investigador independiente • Santa Cruz – Bolivia
https://orcid.org/0009-0008-1575-3474 • doliva@cmlawyers.com.bo

Revista de Derecho de la UCB – UCB Law Review, Vol. 10 Nº 18, abril 2026, pp. 15-52
ISSN 2523-1510 (en línea), ISSN 2521-8808 (impresa).
DOI: https://doi.org/10.35319/lawreview.202618134

Recibido: 30 de octubre de 2025 • Aceptado: 10 de abril de 2026

Resumen

El presente artículo tiene por objetivo analizar cómo la inteligencia artificial transforma la ética en una forma efectiva de regulación digital. Ante la ausencia de un marco normativo global, las plataformas incorporan principios morales en algoritmos, términos de servicio y arquitecturas técnicas, convirtiendo valores en mecanismos de control conductual y planteando la cuestión de qué moral es la que efectivamente gobierna la conducta de los usuarios. Mediante un enfoque teórico y jurídico-filosófico, basado en el análisis doctrinal y normativo del *soft law*, el derecho contractual y la gobernanza algorítmica, el presente artículo examina la transición del *soft law* hacia prácticas coercitivas de facto, donde el código y el contrato adquieren fuerza normativa. Finalmente, se propone un marco conceptual de gobernanza ética orientado a delimitar los actuales criterios de transparencia, proporcionalidad y revisión humana, contribuyendo al debate sobre ética, poder y regulación en la era algorítmica.

Palabras clave: ética digital; inteligencia artificial; soft law; gobernanza algorítmica; regulación tecnológica.

Abstract

This article aims to analyze how artificial intelligence transforms ethics into an effective form of digital regulation. In the absence of a global normative framework, platforms embed moral principles into algorithms, terms of service, and technical architectures, turning values into mechanisms of behavioral control and raising the question of which morality effectively governs user behavior. Through a theoretical and legal-philosophical approach, based on doctrinal and normative analysis of *soft law*, contract law, and algorithmic governance, the article examines the transition from *soft law* to de facto coercive practices, where code and contract acquire normative force. Finally, it proposes a conceptual framework for ethical governance aimed at delineating current criteria of transparency, proportionality, and human oversight, contributing to the debate on ethics, power, and regulation in the algorithmic era.

Keywords: digital ethics; artificial intelligence; soft law; algorithmic governance; technological regulation.

1. Introducción

En el debate público contemporáneo suele asumirse que la inteligencia artificial (IA) es, en esencia, una herramienta neutral y que los problemas normativos surgen únicamente de su uso indebido. Esa suposición resulta insostenible cuando se observa cómo la IA opera efectivamente en plataformas y servicios digitales. Lejos de limitarse a asistir decisiones humanas, la IA clasifica, prioriza, recomienda, deshabilita, sanciona y condiciona el acceso a oportunidades, visibilidad y participación. En ese tránsito desde el cálculo hacia la decisión, emerge una tesis central de este artículo: la IA no es neutral, sino que vehiculiza una forma de moral privatizada.

Esta afirmación no atribuye agencia moral a los sistemas, ni presupone conciencia o intención. El problema no es si la máquina "razona moralmente", sino qué criterios de corrección se

vuelven operativos cuando valores, políticas internas y supuestos sociales se traducen a parámetros técnicos y mecanismos de *enforcement*. Cuando lo considerado aceptable o riesgoso deja de discutirse públicamente y se ejecuta mediante arquitectura técnica, el conflicto moral no desaparece, sino que se desplaza hacia regímenes privados de decisión.

Este razonamiento se ve reforzado por una condición estructural del entorno digital: la ausencia de una autoridad normativa global capaz de regular de manera unificada. En ese contexto, proliferan instrumentos de *soft law* que formulan principios éticos y marcos de gobernanza, mientras que el control efectivo se ejerce a través de contratos de adhesión y diseño técnico. La ética, así, no opera principalmente como deliberación, sino como estándar interno, obligación contractual y regla ejecutable por infraestructura.

La pregunta no es si esos valores son plausibles en abstracto, sino cómo se transforman en reglas aplicables, quién controla esa traducción y qué garantías existen cuando producen consecuencias adversas relevantes. El argumento central sostiene que la ética se convierte en control real cuando se acopla a contrato y código, y que ese control debe ser analizado como ejercicio de poder normativo. Desde esta perspectiva, el artículo propone desplazar el foco desde la proclamación de principios hacia la legitimidad procedimental y los límites jurídicos del gobierno moral privatizado en sistemas de IA.

1.1. Metodología

El presente artículo adopta un enfoque teórico y jurídico-filosófico. Se basa en el análisis doctrinal de la literatura especializada en ética, filosofía del derecho y gobernanza de la inteligencia artificial, así como en el examen de instrumentos de soft law, marcos regulatorios y prácticas de gobernanza implementadas por plataformas digitales.

2. Inteligencia artificial y gobierno de conductas

2.1. IA como infraestructura decisional

Hablar de inteligencia artificial como herramienta suele ser insuficiente. La metáfora de la herramienta sugiere neutralidad y subordinación a la mano humana, pero en entornos digitales la IA opera con frecuencia como parte estable del modo en que se decide. Por eso, para comprender el gobierno de conductas, no basta con preguntarse qué hace un modelo en abstracto, sino dónde se inserta, qué decisiones organiza y qué consecuencias activa. Para fines de este artículo, se entenderá por infraestructura decisional el conjunto de sistemas, procedimientos, interfaces y métricas que estructuran la toma de decisiones dentro de una organización o plataforma y que, al hacerlo, definen las condiciones bajo las cuales los usuarios actúan.

Esta idea no pretende sugerir que toda decisión sea automática o que la voluntad humana desaparezca. La tesis más precisa es que la IA reconfigura el entorno de decisión, modifica el costo de ciertas opciones, reordena prioridades y vuelve naturales determinados recorridos de conducta. Una plataforma no necesita ordenar explícitamente lo que el usuario debe hacer. Le basta con jerarquizar posibilidades, introducir fricciones y optimizar incentivos. En términos prácticos, esto significa que la IA gobierna conductas no por mandato directo, sino porque se integra a circuitos decisionales que administran atención, acceso, reputación y oportunidad.

Para ver esto con claridad, conviene pensar la decisión como un proceso. En el mundo físico, las decisiones están rodeadas de señales, hábitos, arquitectura y reglas. En el mundo digital, ese entorno se vuelve programable, medible y escalable. La IA se instala

precisamente allí, alrededor de la elección, transformándolo en un espacio diseñado. A medida que las plataformas convierten cada interacción en datos, la IA se vuelve el mecanismo que clasifica esos datos, produce inferencias y ejecuta consecuencias. El

resultado es que el comportamiento del usuario queda moldeado por una infraestructura que anticipa, filtra, puntúa y, en muchos casos, sanciona.

Este funcionamiento suele presentarse como un conjunto de operaciones técnicas. Sin embargo, cuando se observa la cadena completa se advierte que la IA ocupa un lugar normativo en sentido práctico. En un circuito típico, la infraestructura decisional incluye al menos cinco momentos. Primero, la selección de lo que cuenta como información relevante. Todo sistema decide qué recoge, qué ignora y bajo qué formato. Segundo, la clasificación y el puntaje, es decir, la construcción de métricas de valor o riesgo. Tercero, la priorización o *ranking*, que define qué se muestra primero, qué queda oculto y qué se vuelve visible solo para algunos. Cuarto, la aplicación de reglas y umbrales, donde una predicción se traduce en una consecuencia. Quinto, la retroalimentación, porque el sistema aprende del comportamiento que él mismo induce y, al aprender, consolida un patrón.

La IA se convierte en infraestructura decisional porque conecta predicción y ejecución. No se limita a sugerir, sino que organiza el entorno de acción y activa consecuencias. Esa conexión es la base material del gobierno de conductas. En una plataforma de contenidos, un modelo de recomendación no solo predice preferencias, también decide qué ve el usuario y, por tanto, qué consume, qué comparte y qué considera socialmente relevante. En un *marketplace*, un sistema de *ranking* no solo ordena productos, también distribuye atención y ventas. En un entorno laboral o financiero, una puntuación de riesgo no solo describe, sino que puede habilitar o negar acceso. El salto de la descripción a la intervención es el punto donde la IA deja de ser un cálculo y se vuelve gobierno.

La literatura sobre responsabilidad algorítmica ayuda a sostener este enfoque. Cuando decisiones relevantes se delegan en sistemas computacionales, el problema no es únicamente si el modelo acierta, sino si el proceso es explicable, auditable y sujeto a control institucional. En otras palabras, la cuestión no se agota en la precisión del resultado, sino en la gobernanza del sistema como parte de una infraestructura que decide sobre personas (Kroll et al.,

2017). En la misma línea, se ha mostrado que sistemas basados en datos pueden reproducir o amplificar desigualdades existentes, incluso sin intención explícita, porque los datos reflejan patrones históricos y porque la forma de modelar puede traducir correlaciones en consecuencias reales (Barocas & Selbst, 2016). La IA no es un componente neutro agregado al final de una decisión humana, sino un elemento que estructura el modo de decidir y distribuye cargas y beneficios.

Lo anteriormente descrito permite formular una distinción útil. Una cosa es entender la IA como sistema de predicción y otra es entenderla como sistema de organización del entorno. El gobierno de conductas aparece cuando la IA cumple, al menos, tres funciones en la infraestructura decisional.

La primera función es la preselección y jerarquización. En este nivel la IA gobierna la atención. Decide qué se vuelve visible y qué queda fuera del campo de percepción del usuario. Puede hacerlo por recomendación, por ranking, por personalización o por moderación. El usuario no elige en un vacío, sino dentro de un menú construido. Y ese menú responde a objetivos institucionales como retención, seguridad, rentabilidad o cumplimiento, que se codifican en métricas.

La segunda función es la habilitación y restricción. En este nivel la IA gobierna posibilidades. No se limita a mostrar, también permite o impide. Esto puede adoptar formas suaves, como fricciones que desincentivan una conducta, o formas duras, como bloqueos, suspensiones, degradación de alcance o denegación de acceso. Cuando estas medidas se aplican de forma automatizada, el sistema no solo evalúa, sino que también ejecuta. La infraestructura decisional se vuelve entonces un dispositivo de control efectivo porque opera sobre el conjunto de opciones disponibles.

La tercera función es la retroalimentación. En este nivel la IA gobierna hábitos. Las plataformas aprenden de la conducta del usuario, pero el usuario también aprende del sistema. Si el *ranking* premia un tipo de contenido, los productores se adaptan. Si ciertos comportamientos reducen visibilidad, los usuarios se autocorri-

gen. Se genera un circuito de ajuste mutuo. Este circuito es relevante porque convierte una decisión puntual en un patrón de conducta. El gobierno de conductas no se produce solo por sanción, sino también por adaptación.

Los marcos contemporáneos de gobernanza de IA, en particular el NIST AI Risk Management Framework (NIST, 2023), insisten en que la IA debe analizarse como un fenómeno sociotécnico. Sus impactos no emergen únicamente del modelo, sino de cómo se diseña, se implementa, se supervisa y se integra en prácticas organizacionales. Esto desplaza el foco desde el rendimiento del sistema hacia el conjunto de decisiones institucionales que lo hacen operar. Una infraestructura decisional no es únicamente un algoritmo. Incluye objetivos, datos, métricas, incentivos, reglas de negocio, equipos humanos, procesos de apelación, documentación, auditoría y mecanismos de respuesta. El riesgo, por tanto, no es solo técnico, sino también organizacional.

Instrumentos como el NIST AI Risk Management Framework (NIST, 2023) y las recomendaciones de la OECD (2019) pretenden orientar buenas prácticas y gestionar riesgos, pero lo hacen como marcos voluntarios cuya eficacia depende de su adopción, traducción y ejecución interna. Si la IA opera como infraestructura decisional, la pregunta deja de ser si los principios son correctos en abstracto y se vuelve más concreta: cómo esos principios se implementan en decisiones reales, qué procesos los vuelven exigibles y qué mecanismos los convierten en control efectivo.

En síntesis, entender la IA como infraestructura decisional permite explicar por qué su impacto normativo no es una metáfora. La IA organiza entornos de elección, distribuye visibilidad, introduce fricciones y activa consecuencias. Al hacerlo, configura patrones de conducta. Esta caracterización prepara el terreno para examinar, en los apartados siguientes, los tipos de servicios de IA que operan con efectos normativos y el tránsito desde orientaciones voluntarias hacia mecanismos de control real asentados en contrato y código.

2.2. Tipos de servicios de IA con impacto normativo

No toda inteligencia artificial produce efectos normativos del mismo modo. Para este artículo, el impacto normativo aparece cuando un servicio de IA no se limita a asistir, sino que estructura posibilidades, distribuye consecuencias o activa restricciones con relevancia práctica sobre la conducta. El criterio de clasificación, por tanto, es operativo: se distingue por el tipo de decisión que median y por el tipo de consecuencia que desencadenan. En ese marco, la IA puede entenderse como un conjunto de sistemas capaces de generar salidas como predicciones, recomendaciones, contenidos o decisiones que influyen en entornos físicos o virtuales (ISO/IEC, 2022; OECD, 2019; Reglamento (UE) 2024/1689, 2024).

2.2.1. Sistemas de recomendación, *ranking* y curaduría algorítmica

Estos sistemas regulan principalmente la atención. Determinan qué se vuelve visible primero, qué queda relegado y qué permanece fuera del campo perceptivo ordinario. Su impacto normativo no depende de prohibir, sino de distribuir oportunidades como visibilidad, reputación, alcance o acceso a audiencias.

2.2.2. Moderación automatizada y *enforcement* de políticas internas

En este caso, la IA ejecuta reglas internas que operan como condiciones prácticas de participación. La consecuencia típica no es solo la remoción de contenido, sino también medidas graduales como restricción de alcance, desmonetización, suspensión o bloqueo, con efectos comparables a sanciones de facto.

2.2.3. Sistemas de *scoring*, reputación y evaluación automatizada

Estos sistemas asignan puntajes o categorías que condicionan acceso, condiciones contractuales o trato diferenciado. El riesgo

normativo central es que el criterio se vuelva opaco y difícil de impugnar, generando cargas de prueba asimétricas y efectos distributivos relevantes.

2.2.4. Sistemas de selección, filtrado y priorización en mercados (laboral, crédito, seguros, consumo)

Aquí la IA actúa como puerta de acceso. Cuando el filtrado automatizado se integra en procesos de admisión, elegibilidad o selección, los errores y sesgos no se limitan a fallas técnicas sino que producen exclusiones prácticas y redefinen oportunidades.

2.2.5. Modelos generativos con efectos regulatorios

Los sistemas generativos incorporan restricciones operativas que no se presentan como regulación, pero se viven como tal: rechazos, advertencias, fricción y denegación de asistencia. Así, parte del debate ético se traduce directamente en una gramática de permisiones y prohibiciones funcionales.

2.2.6. Personalización conductual y arquitectura de fricción

Estos sistemas gobiernan hábitos mediante ajustes persistentes del entorno: notificaciones, recordatorios, fricciones, recompensas y nudges. Su relevancia normativa no reside en una sanción explícita, sino en la capacidad de inducir patrones de conducta a escala.

2.2.7. Conclusión operativa

En todos estos casos, la IA produce un efecto normativo porque define condiciones de posibilidad, prioriza opciones y activa consecuencias. Por eso, discutir ética en IA exige atender a la cadena que lleva desde principios generales hacia reglas internas y hacia su ejecución técnica.

2.3. De la recomendación al control

En un primer momento, la ética aplicada a la IA aparece como un lenguaje de orientación. Se expresa en principios, marcos de gestión de riesgos, recomendaciones y estándares de buenas prácticas. Su promesa es razonable: ofrecer un suelo común de valores en un campo donde la técnica avanza más rápido que el derecho y donde la incertidumbre invita a la prudencia. Sin embargo, en plataformas y servicios digitales, ese lenguaje rara vez se queda en el plano declarativo. Cuando los principios se incorporan a políticas internas, se insertan en términos y condiciones y se implementan en arquitectura técnica, la ética deja de ser un consejo y se convierte en una condición operativa.

2.3.1. *Soft law* como arquitectura de legitimación y como laboratorio normativo

En términos generales, el *soft law* designa instrumentos que formulan estándares de conducta sin crear, por sí mismos, obligaciones jurídicas coercibles en sentido clásico. Su utilidad es evidente: reduce costos de coordinación, ofrece flexibilidad ante cambios tecnológicos y permite construir consensos preliminares donde un tratado o una ley serían lentos o políticamente inviables. Aun así, su aparente suavidad puede inducir a error. La influencia del *soft law* no depende solo de su fuerza formal, sino de su capacidad para ordenar expectativas y para convertirse en referencia institucional. En la práctica, marcos como NIST AI RMF u orientaciones internacionales como las de la OECD tienden a operar como un vocabulario común, establecen qué se entiende por riesgo, qué se entiende por responsabilidad, qué prácticas se consideran maduras y qué se considera inaceptable (NIST, 2023; OECD, 2019).

Aquí conviene introducir una distinción analítica que ayuda a entender el giro hacia el control real. La legalización no es un interruptor que se enciende o apaga, sino un continuo. En una formulación clásica, la legalización se caracteriza por tres dimensiones: “obligation, precision, and delegation” (Abbott et al., 2000). En la esfera pública, un instrumento de *soft law* suele presentar menor

obligación, y a veces menor precisión o menor delegación, precisamente para conservar flexibilidad. Pero en la esfera privada, esas dimensiones pueden incrementarse por otras vías. La obligación se obtiene por aceptación contractual. La precisión se obtiene por políticas internas detalladas, listas de prohibiciones y taxonomías de contenido. La delegación se obtiene cuando el *enforcement* se entrega a equipos de *trust and safety*, a proveedores y, crecientemente, a sistemas automatizados. El resultado es una legalización privada y funcional: no necesariamente estatal, pero sí eficaz.

2.3.2. El primer salto: del principio a la política interna

El primer puente entre recomendación y control es organizacional. Las empresas adoptan estándares no solo por convicción ética, sino porque subjetivamente también influyen razones estratégicas: reputación, gestión de riesgo, presión regulatoria anticipada, exigencias de inversionistas, requerimientos de compliance en cadenas de suministro, o expectativas de clientes corporativos. En ese tránsito, el *soft law* deja de ser un documento externo y se convierte en una regla interna. Una vez internalizado, el estándar ya no opera como sugerencia, sino como criterio de evaluación y desempeño. Puede traducirse en procesos de gobernanza (comités, evaluaciones de impacto, documentación), en requisitos técnicos (monitoreo, auditoría, trazabilidad) y en decisiones de producto (qué se permite, qué se degrada, qué se bloquea). Esta traducción cobra importancia, porque allí se define qué parte del lenguaje ético se toma en serio, qué parte se vuelve *marketing*, y qué parte se transforma en umbrales operativos.

El punto normativo aparece cuando la política interna se vuelve condición para participar. Dentro de una plataforma, la política no es un manifiesto filosófico, sino una regla aplicada. No se limita a enunciar valores, sino que organiza consecuencias. Y cuanto más se automatiza su aplicación, más se parece a un sistema de gobierno: define estándares, detecta infracciones, ejecuta medidas, gestiona apelaciones, produce estadísticas, modula la conducta. Lo que comenzó como recomendación externa se convierte en disciplina interna.

2.3.3. El segundo salto: del estándar a la obligación por contrato

El contrato es la tecnología jurídica que vuelve exigible lo que antes era voluntario. En el ecosistema digital, este fenómeno se expresa en términos y condiciones, políticas de uso, códigos de conducta y reglas comunitarias aceptadas como condición de acceso. Desde un punto de vista formal, el usuario no está frente a una ley estatal, sino frente a un contrato de adhesión. Pero desde un punto de vista práctico, ese contrato opera como puerta. La aceptación no solo autoriza el acceso, sino que fija el régimen de comportamiento esperado, define sanciones y regula el procedimiento de aplicación. Es un modo de producción normativa privada.

En derecho comparado, la discusión sobre la validez de estos acuerdos suele concentrarse en el consentimiento. Los tribunales, por ejemplo, han exigido que exista aviso razonablemente visible y una manifestación inequívoca de aceptación para que el vínculo sea exigible en términos estrictos (Specht v. Netscape Communications Corp., 2002). Esta línea muestra algo importante: incluso cuando el contrato es la vía de obligación, su legitimidad depende de condiciones mínimas de transparencia y de posibilidad real de entender qué se acepta. Sin embargo, aun cumpliendo esos requisitos, el efecto estructural permanece: valores y criterios morales, formulados como ética, pasan a operar como cláusulas obligatorias. El *soft law*, al acoplarse al contrato, adquiere dientes jurídicos privados.

El problema se intensifica por la asimetría típica del entorno digital. La negociación es inexistente. La salida suele ser binaria: aceptar o no participar. Esto no convierte automáticamente el contrato en inválido, pero sí vuelve más exigente el análisis normativo sobre límites y garantías. Si el contrato funciona como ley privada de acceso, el debate ya no es solo si hay consentimiento, sino qué se está gobernando, con qué grado de discrecionalidad y con qué mecanismos de control.

2.3.4. El tercer salto: del contrato al código como *enforcement*

La pieza decisiva del control real es la arquitectura técnica. Aquí la tesis de la regulación por diseño adquiere su fuerza más concreta. En el entorno digital, el cumplimiento no depende únicamente de la amenaza de sanción futura. Depende de condiciones de posibilidad. Un sistema puede permitir, dificultar, degradar, bloquear o suspender conductas de forma inmediata y escalable. Por eso, la regulación por código no es metáfora sino que se constituye en infraestructura.

Lessig (1999) expresa esta intuición al sostener que el diseño del sistema habilita y restringe comportamiento, y por eso la arquitectura cumple un rol normativo. Cuando esa arquitectura se combina con IA, el control se vuelve adaptativo, no solo ejecuta reglas estáticas, sino que clasifica contextos, infiere intenciones probables, puntúa riesgos y decide respuestas. Lo que antes era una regla textual se convierte en un *pipeline* de decisión: señales, modelo, umbral, acción, retroalimentación. El resultado es un régimen de gobierno operacional, donde la ética declarada se traduce en fricciones, límites y sanciones ejecutadas.

En este punto se entiende por qué la transición del *soft law* al control real no es una simple intensificación, sino un cambio de naturaleza. El *soft law* orienta conductas en el plano discursivo. El contrato convierte esa orientación en obligación de acceso. El código convierte la obligación en un sistema de ejecución. La consecuencia es una coerción de facto: no necesariamente coacción estatal, pero sí restricción efectiva de opciones. Y cuando las decisiones son automatizadas o semi-automatizadas, la opacidad puede aumentar. No solo se vuelve difícil discutir el valor que gobierna, también se vuelve difícil identificar la razón concreta de la decisión.

2.3.5. La fórmula del control real y su problema filosófico

El control real puede resumirse como un encadenamiento. Primero, un marco ético formula valores bajo lenguaje de *soft law*. Se-

gundo, una organización los internaliza como política. Tercero, la política se incorpora al contrato como condición de acceso. Cuarto, el sistema técnico implementa esa política como arquitectura ejecutable. En ese punto, la ética ya no es únicamente un debate. Es un régimen operativo.

La pregunta filosófica que atraviesa este proceso no es de menor importancia. Es el punto incómodo cuando hablamos de regulación y sus fundamentos. Si la moral admite marcos incompatibles, entonces la traducción de valores en cláusulas y en umbrales técnicos no es neutra. Selecciona. Prioriza. Excluye. Decide qué daño es tolerable y cuál no. Decide qué error se compensa y cuál se absorbe. Decide qué riesgo se evita aun a costa de libertades, y qué libertad se preserva aun a costa de riesgos. Esa selección no se hace en la cámara de senadores o diputados, sino que se hace en comités internos, en documentos de producto, en taxonomías de contenido, en parámetros de modelos, en diseño de interfaces. Es allí donde la moral se privatiza: no porque la empresa invente la moral desde cero, sino porque la vuelve operativa bajo su propia gramática y bajo sus propios incentivos.

En síntesis, el tránsito del *soft law* al control real por contrato y código nombra un fenómeno estructural: el pasaje desde la recomendación ética hacia un gobierno efectivo de conducta mediante reglas privadas ejecutadas por infraestructura.

3. Ética y moral

3.1. Ética y moral: distinciones conceptuales

En este artículo se distingue entre moral y ética para evitar ambigüedades al analizar reglas privadas ejecutadas en entornos digitales. La moral se usará para referirse al conjunto de normas prácticas que efectivamente regulan conductas dentro de una comunidad, una profesión o una organización; la ética se usará para referirse al plano de reflexión normativa que evalúa, justifica

o critica esas normas. Esta distinción resulta especialmente útil cuando el objeto de análisis no es el Estado, sino regímenes normativos privados que operan mediante políticas internas, contrato y arquitectura técnica. En este contexto, Miller (2013) permite entender que la cuestión no se agota en los valores que se declaran, sino en cómo estos se vuelven operativos y producen consecuencias efectivas.

3.1.1. Moral como práctica y ética como justificación

Miller (2013) explica que el término moral puede emplearse en un sentido descriptivo, como sistema de normas efectivamente vigentes sostenidas por expectativas sociales y mecanismos de sanción, y también en un sentido teórico, como el fenómeno que se analiza cuando se pregunta por la naturaleza o validez de los juicios morales. En cambio, ética designa la reflexión sistemática sobre la moral, su coherencia, alcance y límites. Esta separación permite evitar un error recurrente en debates sobre IA: suponer que porque un instrumento se presenta como ético ya contiene una justificación suficiente para operar como regla ejecutable en un entorno de poder asimétrico.

3.1.2. Tres niveles de análisis y moral institucionalizada

Siguiendo a Miller (2013), el capítulo distingue entre ética normativa, metaética y ética aplicada. La ética normativa discute criterios sustantivos de corrección, la metaética examina el estatus del discurso moral y la ética aplicada analiza la traducción de estos criterios a ámbitos concretos como la gobernanza algorítmica, la moderación de contenidos y el enforcement automatizado. Además, conviene distinguir moral social, moral profesional y moral institucionalizada. La tesis de la moral privatizada se refiere sobre todo a la moral institucionalizada, porque allí la regla no solo se declara, sino que se ejecuta como condición de acceso, permanencia o visibilidad y puede escalarse mediante automatización (Raz, 1986).

3.1.3. Derecho, moral y el tránsito hacia el gobierno efectivo

En filosofía del derecho, la relación entre derecho y moral estructura debates sobre validez y autoridad. Una posición influyente sostiene una tesis de separación conceptual entre validez jurídica y corrección moral, sin negar la posibilidad de crítica moral del derecho (Hart, 1961). Otras tradiciones subrayan que un sistema normativo requiere condiciones internas de coherencia y racionalidad para orientar conducta y sostener pretensión de legitimidad (Fuller, 1964). Para este artículo, lo decisivo es que puede existir eficacia normativa por contrato y código, pero la legitimidad del régimen depende de estándares que exceden la mera eficacia, especialmente cuando hay decisiones adversas automatizadas y poder asimétrico.

3.2. Moral objetiva y moral subjetiva

Este apartado no busca resolver el debate metaético, sino fijar un marco mínimamente riguroso para el resto del artículo. La tesis de la moral privatizada no exige demostrar verdades morales objetivas, pero sí exige reconocer que regímenes privados suelen actuar como si sus estándares fueran correctos en un sentido fuerte, incluso cuando son razonablemente discutibles. Por ello, la evaluación no puede quedar solo en el contenido, sino también en el modo de producción, aplicación y contestación de reglas cuando se vuelven ejecutables por contrato y código (Miller, 2013).

El realismo moral sostiene que ciertos juicios morales pueden ser verdaderos o falsos y que su verdad no depende simplemente de preferencias. En su versión no metafísica, esta tesis no requiere compromisos ontológicos fuertes, pero sí preserva una pretensión de objetividad justificable racionalmente (Shafer Landau, 2003). El constructivismo desplaza el foco desde hechos morales hacia condiciones de justificación, y por ello es especialmente útil para legitimidad, porque vuelve normativamente relevantes la publicidad, la posibilidad de impugnación y la coherencia del régimen (Korsgaard, 1996). El naturalismo, por su parte, es relevante por su afinidad con la traducción de valores a métricas y umbrales,

pero esa traducción no elimina la dimensión normativa, y el riesgo es confundir cuantificación con justificación y ocultar elecciones sobre daños, distribución del riesgo y carga del error (Brink, 1989). En conjunto, estas posiciones permiten sostener un punto metodológico común: ante desacuerdo moral persistente, lo decisivo es exigir garantías procedimentales mínimas cuando el régimen produce consecuencias adversas relevantes (Miller, 2013).

3.3. Moral, poder y legitimidad

La tesis de la moral privatizada conduce a una pregunta de teoría del poder. Cuando criterios de corrección se vuelven operativos mediante cláusulas contractuales y arquitectura técnica, dejan de ser únicamente orientaciones y pasan a funcionar como condiciones de acceso, permanencia y visibilidad. En ese tránsito, la moral se convierte en un arreglo institucional con capacidad de producir consecuencias. Esto obliga a distinguir eficacia, autoridad y legitimidad. La eficacia describe la existencia operativa del régimen. La autoridad describe su capacidad de generar razones prácticas para actuar. La legitimidad concierne a la justificación moral del poder que el régimen ejerce, especialmente cuando produce restricciones significativas.

3.3.1. Legitimidad y aceptación

La legitimidad permite evitar dos simplificaciones. La primera consiste en equiparar legitimidad con aceptación. Un régimen puede estabilizarse por dependencia, hábito, costos de salida o ausencia de alternativas sin que ello pruebe justificabilidad normativa (Suchman, 1995). En términos sociológicos, todo sistema de autoridad busca cultivar la creencia en su legitimidad, pero esa creencia no equivale a legitimidad normativa (van Leeuwen, 2008; Weber, 1978). Por eso, en plataformas, el argumento el usuario aceptó no agota el problema. La aceptación contractual puede ser relevante para explicar obligación privada, pero no basta para justificar un régimen que afecta participación, expresión, reputación o acceso a oportunidades en condiciones de poder asimétrico (Buchanan, 2002; Schmidt, 2013).

3.3.2. Justificación pública y control de la arbitrariedad

En la tradición liberal, un enfoque influyente entiende legitimidad como exigencia de justificación pública del ejercicio del poder. En la formulación rawlsiana, Rawls (2005) sostiene que el principio liberal de legitimidad exige razones públicas aceptables para ciudadanos libres e iguales. Aunque la gobernanza de plataformas no sea poder político en sentido estricto, la analogía es normativamente productiva, cuando reglas privadas operan como condiciones de participación y su aplicación produce efectos relevantes, se requiere un umbral de garantías para controlar arbitrariedad, especialmente cuando la moral se ejecuta por infraestructura (Buchanan & Keohane, 2006; Schmidt, 2013).

3.3.3. Deliberación, umbrales y debido proceso en plataformas

La moral privatizada se vuelve más aguda cuando la discrecionalidad migra hacia sistemas. Decisiones que antes dependían de juicio humano se trasladan a taxonomías, señales y umbrales implementados en infraestructura, produciendo un gobierno por categorías ejecutables. Esto puede intensificar opacidad y consolidar errores como hechos operativos si no hay revisión significativa. En respuesta, estándares voluntarios como los Principios de Santa Clara han insistido en transparencia, notificación y apelación como condiciones mínimas cuando el régimen afecta participación y voz pública (Santa Clara Principles, 2021). En paralelo, desarrollos regulatorios como el Reglamento de Servicios Digitales de la Unión Europea refuerzan esa lógica al exigir razones para restricciones y sistemas de reclamación y revisión, institucionalizando que el poder ejercido por infraestructura debe volverse rastreable y contestable (Reglamento (UE) 2022/2065, arts. 17, 20, 21). Así, la relación entre moral, poder y legitimidad nos permite entender que cuando la ética se vuelve ejecutable por contrato y código, se convierte en poder normativo, y ese poder debe evaluarse bajo estándares procedimentales mínimamente defendibles para evitar arbitrariedad automatizada (Rawls, 2005; Schmidt, 2013; UNESCO, 2021).

4. Del juicio moral al diseño técnico

4.1. De valores a criterios operativos

Una dificultad estructural de la ética en IA es que los valores son formulaciones abiertas, mientras que los sistemas decisionales requieren criterios operativos. La organización que declara un valor debe traducirlo a una regla de decisión que pueda aplicarse de manera consistente y que pueda integrarse a procesos de diseño, despliegue y monitoreo. Esta traducción es el punto en el que el discurso ético se vuelve gobernanza. No se trata solo de declarar que se busca equidad, seguridad o transparencia, sino de definir qué cuenta como riesgo, qué tipo de daño es relevante, qué evidencia se considera suficiente y qué respuesta institucional corresponde.

En la práctica, esta traducción adopta la forma de estándares internos. Se fijan categorías, se definen escenarios de uso, se determinan niveles de impacto, se asignan responsabilidades y se crean rutas de escalamiento. Los marcos de gestión de riesgos son relevantes precisamente porque ordenan este proceso y transforman valores generales en controles organizacionales verificables, como documentación, evaluaciones, monitoreo y revisión (NIST, 2023).

4.2. Traducción técnica: taxonomías, señales, datos, métricas y objetivos de optimización

La traducción a criterios operativos se vuelve traducción técnica cuando el sistema debe ejecutar decisiones. En ese paso aparecen cinco operaciones frecuentes. Primero, taxonomías, porque el sistema necesita categorías discretas para identificar conductas o eventos relevantes. Segundo, señales, porque debe seleccionar qué datos o indicadores cuentan como evidencia. Tercero, datos y calidad, porque lo que el sistema puede decidir depende de lo que puede registrar y de cómo se construyen los conjuntos de entrenamiento y prueba. Cuarto, métricas, porque el desempeño se define

mediante medidas que implican elecciones sobre error tolerable y distribución del riesgo. Quinto, objetivos de optimización, porque el sistema no solo predice, sino que persigue funciones objetivo que pueden entrar en tensión con valores invocados.

Estas operaciones suelen presentarse como ingeniería, pero contienen elecciones normativas. Definir qué señal cuenta, qué variable se considera relevante, qué umbral activa una consecuencia, o qué tasa de falsos positivos se tolera, son decisiones que distribuyen cargas del error y que afectan a usuarios de manera desigual. Por ello, la gobernanza ética no puede limitarse a principios. Debe exigir trazabilidad de elecciones técnicas, documentación suficiente, evaluación de impacto y mecanismos de revisión cuando corresponda (NIST, 2023).

4.3. Decisión automatizada como *enforcement*

El fenómeno de la conversión decisiva ocurre cuando la clasificación se conecta con la consecuencia. Mientras un sistema solo produce un score o una recomendación, todavía existe un espacio de mediación humana o de interpretación institucional. Pero cuando el output se usa como disparador de una medida, el sistema se convierte en ***enforcement***. La decisión automatizada funciona entonces como una tecnología de ejecución, transforma un juicio probabilístico en una restricción real.

Este punto puede describirse como una cadena que inicia con señales, pasa a inferencia, luego al umbral y ejecuta acción. La acción puede ser suave, como una fricción, degradación o priorización negativa, o dura, como el bloqueo, suspensión, remoción o denegación de acceso. La diferencia entre ambas no elimina el problema, solo cambia su intensidad. En ambos casos, el régimen moral se materializa como estructura de oportunidades, se participa, se pierde voz, se degrada reputación o se restringe acceso según reglas privadas operadas por infraestructura.

Las teorías de rendición de cuentas algorítmica han insistido en que el problema central no es únicamente la calidad predictiva,

sino la gobernanza del sistema como parte de un proceso decisional que afecta a personas. De allí la relevancia de diseñar mecanismos institucionales de control, trazabilidad, auditoría, explicabilidad funcional y garantías procedimentales que permitan contestar y corregir decisiones (Kroll et al., 2017).

El *enforcement* automatizado también reconfigura la temporalidad del poder. La sanción no ocurre al final de un procedimiento, ocurre en tiempo real. Esto altera el tipo de daño, el costo de error se vuelve inmediato y a menudo difícil de revertir. En sistemas de moderación o detección, la elección entre falsos positivos y falsos negativos se convierte en una decisión sobre quién sufre el daño principal. Esta estructura ha sido estudiada en *fairness* como un problema de compatibilidad entre criterios y como un problema de distribución de errores entre grupos, mostrando que corregir puede significar redistribuir errores más que eliminarlos (Hardt, Price, & Srebro, 2016; Chouldechova, 2017). Cuando el *enforcement* se automatiza, esa redistribución deja de ser una cuestión abstracta y produce denegaciones, bloqueos y degradaciones concretas.

Además, el *enforcement* automatizado tiende a desplazar el lenguaje de razones hacia un lenguaje de umbrales. En un proceso clásico, la justificación se ofrece como motivación, por qué se decidió. En un proceso automatizado, la razón suele comprimirse en una categoría contenido infractor, actividad sospechosa o riesgo elevado, sin acceso al conjunto de señales que llevaron al resultado. Esto no es solo un problema de transparencia técnica, es un problema de forma de gobierno. Un régimen que sanciona sin razones inteligibles o que ofrece razones no contestables erosiona condiciones mínimas de legitimidad procedimental y aumenta el riesgo de arbitrariedad.

Por ello, el paso del juicio a la consecuencia constituye el punto donde la moral privatizada se vuelve jurídicamente relevante. El estándar moral, una vez implementado, ya no es únicamente un criterio interno, es un mecanismo de restricción. Cuando el enforcement es infraestructural, el derecho aparece como límite al poder no porque moralice la técnica, sino porque exige condiciones

de control, procedimientos de impugnación, revisión, proporcionalidad y rendición de cuentas, precisamente porque la ejecución automatizada produce efectos comparables a coerción práctica.

4.4. El vacío normativo y la expansión del soft law digital

El llamado vacío normativo en inteligencia artificial no debe entenderse como ausencia total de reglas, sino como un desajuste entre la capacidad del derecho estatal para generar estándares exigibles y la forma en que los sistemas de IA operan, se despliegan y producen efectos. Ese desajuste explica la expansión del *soft law* como respuesta práctica en un espacio transnacional y técnicamente dinámico, donde instrumentos no vinculantes que ordenan expectativas, estabilizan vocabularios y ofrecen marcos de legitimación cuando no existe una autoridad regulatoria global única y cuando la respuesta legislativa suele ir detrás del cambio tecnológico (Abbott et al., 2000; OECD, 2019).

5. El vacío normativo y la expansión del *soft law* digital

5.1. El desajuste regulatorio como condición de posibilidad del *soft law*

El desajuste se expresa, primero, en la transnacionalidad de los servicios digitales, cuyos efectos se producen simultáneamente en múltiples jurisdicciones, debilitando soluciones estrictamente territoriales. Se expresa, segundo, en la fragmentación regulatoria, porque los problemas asociados a IA atraviesan sectores y agencias diversas, desde consumo y competencia hasta privacidad, ciberseguridad y derechos fundamentales, generando respuestas heterogéneas y difíciles de integrar. Se expresa, tercero, en la velocidad de iteración tecnológica y organizacional, que amplía la brecha temporal entre innovación y reacción jurídica y

desplaza parte del control hacia el diseño interno y técnico de las organizaciones. En ese escenario, el *soft law* cumple una función de coordinación y convergencia de bajo costo político, al permitir principios comunes y lenguajes compartidos sin depender de un proceso legislativo necesariamente lento (OECD, 2019).

5.2. Función y límite del *soft law* en IA

El *soft law* coordina, legitima y orienta la gobernanza interna. Coordina al ofrecer taxonomías y principios comunes para actores diversos. Legitima al permitir que organizaciones declaren alineamiento con estándares reconocibles. Orienta al impulsar prácticas organizacionales de gestión del riesgo, evaluación de impacto, documentación y monitoreo, como se aprecia en marcos técnicos de gestión de riesgos que estructuran el modo en que se identifican y mitigan impactos aun sin fuerza jurídica directa (NIST, 2023). Sin embargo, su límite es estructural. La voluntariedad impide exigir cumplimiento por sí mismo, la indeterminación obliga a decisiones posteriores para volver aplicables los principios, y su efecto depende de la traducción interna. Por ello, el problema central no se encuentra solo en el texto del *soft law*, sino en el punto en que se convierte en política, procedimiento, contrato y arquitectura técnica, que es el tránsito que aborda el capítulo siguiente.

6. Del *soft law* al control real

El *soft law* adquiere relevancia práctica cuando deja de ser un texto externo y se convierte en una cadena de transformación interna que produce reglas operativas, obligaciones y consecuencias. Esa cadena suele seguir tres movimientos. Primero, internalización, cuando un marco de principios pasa a convertirse en estándar corporativo y estructura gobernanza interna. Segundo, contractualización, cuando ese estándar se inserta en términos y condiciones como régimen de acceso. Tercero, codificación, cuando el estándar se implementa en arquitectura técnica que ejecuta decisiones a escala. Esta cadena explica por qué el debate sobre ética

en IA no puede quedar en principios generales, y es que la ética se vuelve control real cuando se acopla a contrato y código.

6.1. Internalización del principio al estándar corporativo

El *soft law* en IA suele presentarse como un conjunto de principios y marcos de buenas prácticas sin obligatoriedad jurídica directa. Sin embargo, su influencia depende de su capacidad para convertirse en referencia institucional y en criterio reputacional y organizacional, especialmente cuando esos principios se traducen en políticas internas, procesos, métricas y requisitos de diseño y despliegue (OECD, 2019; NIST, 2023). En esta fase, la organización transforma el lenguaje de principios en instrumentos de gobernanza. Es una traducción práctica. Se definen categorías operativas, se establecen controles de cambio, documentación de modelos, evaluaciones de impacto y rutas internas de escalamiento. El resultado es que el principio se convierte en criterio de decisión corporativo y produce una moral operativa, no en sentido teórico, sino como estándar aplicado que estructura conductas y prioriza riesgos y daños dentro del sistema.

6.2. Contractualización: términos y condiciones como norma privada de acceso

La segunda transformación es la contractualización. En el entorno digital, el contrato no solo organiza intercambios, sino que funciona como puerta de acceso y como técnica de producción normativa privada. Los términos y condiciones y políticas de uso se aceptan como condición para participar y suelen operar como contratos de adhesión, lo que permite convertir estándares internos en obligaciones del usuario y en reglas de acceso y permanencia (Radin, 2013). Aquí se produce un desplazamiento relevante. El estándar se presenta como voluntario porque el usuario puede no aceptar, pero puede operar como coerción de facto cuando el servicio se vuelve infraestructura social o económica y la salida real es costosa o irrazonable. Además, el problema no se resuelve con afirmar que

se informó. La crítica a la divulgación masiva muestra que informar no garantiza comprensión ni control, especialmente cuando el texto es extenso, técnico y estructuralmente no leído, y cuando la arquitectura del mercado produce dependencia (Ben-Shahar & Schneider, 2010). Incluso en un marco favorable a la exigibilidad, la doctrina y jurisprudencia han exigido condiciones mínimas de aviso y manifestación inequívoca de aceptación, lo que revela que la contractualización es una tecnología de obligación, pero su legitimidad depende de estándares mínimos de transparencia y procedimiento (Specht v. Netscape Communications Corp., 306 F.3d 17, 2d Cir. 2002). Sin embargo, aun superados esos mínimos, persiste el punto estructural: el *soft law* se convierte en régimen privado de conducta con consecuencias transnacionales y márgenes de discrecionalidad que se vuelven relevantes para la legitimidad.

6.3. Codificación: *enforcement* por arquitectura, automatización y escalabilidad

La tercera transformación, decisiva para el control real, es la codificación. El contrato vuelve exigible el estándar, pero el código lo vuelve ejecutable. La arquitectura técnica regula conductas porque define condiciones de posibilidad, fricciones y consecuencias, y por ello no puede tratarse como elemento meramente instrumental (Lessig, 1999). En IA y plataformas, esta regulación por arquitectura se intensifica porque el sistema no solo aplica reglas estáticas, sino que clasifica, puntúa e infiere, conectando predicción y consecuencia en tiempo real. En esta fase, la moral operativa se convierte en *pipeline* técnico. La normatividad se materializa en elecciones que suelen presentarse como técnicas, qué cuenta como señal relevante, qué umbral dispara una restricción, qué tasa de error se tolera, qué casos reciben revisión humana y cuáles se ejecutan automáticamente. Cuando sistemas computacionales median decisiones con efectos relevantes sobre personas, la cuestión no se agota en desempeño, sino que exige mecanismos de explicabilidad, auditoría, responsabilidad y control institucional, precisamente porque el sistema no amenaza con sancionar, sino que sanciona en la práctica (Kroll et al., 2017). Por ello, la li-

teratura sobre plataformas describe estos arreglos como un orden normativo propio, aplicado por procedimientos internos y por arquitecturas técnicas que funcionan como reglas efectivas de la vida digital (Gillespie, 2018; Suzor, 2019).

6.4. *Soft law* como cadena de densificación normativa

Con lo anterior, el desplazamiento desde el *soft law* al control real puede entenderse como una cadena de densificación normativa. Un marco de principios establece un vocabulario de corrección. Ese vocabulario se internaliza como estándar corporativo, se contractualiza como condición de acceso y se codifica como arquitectura ejecutable. Cada eslabón añade fuerza práctica. La voluntariedad del *soft law* se vuelve obligación interna al corporativizarse, la obligación interna se vuelve obligación del usuario al contractualizarse, y la obligación contractual se vuelve *enforcement* de facto al codificarse. En este punto, el conflicto moral se reubica. Ya no se disputa únicamente en el plano de principios, sino en el de políticas, cláusulas y umbrales técnicos. Por ello, el problema filosófico y jurídico que emerge no es si el marco es razonable como texto, sino qué garantías existen cuando ese marco se convierte en gobierno efectivo de conductas por contrato y código (Lessig, 1999; NIST, 2023; OECD, 2019).

7. Moral privatizada: quién decide y cuáles son los límites jurídicos

7.1. Locus de decisión moral: actores internos, incentivos y dependencias externas

La tesis de la moral privatizada exige precisar dónde se toman, en términos efectivos, las decisiones normativas que luego se presentan como simples políticas técnicas o estándares neutrales. En regímenes digitales, el locus de decisión moral no se ubica en un

órgano público ni en una deliberación democrática, sino en una combinación de actores internos y dependencias externas que determinan qué se considera permitido, riesgoso, valioso o sancionable. La arquitectura decisional suele distribuirse entre equipos de producto, seguridad, confianza y moderación, áreas legales y de cumplimiento, y unidades de datos y aprendizaje automático. Esa distribución no es accidental. Responde a un objetivo organizacional básico, minimizar riesgos, estabilizar reputación y sostener la escalabilidad del servicio, lo que tiende a transformar valores abiertos en estándares operativos consistentes, medibles y ejecutables.

Por ello, entendemos que la moral institucionalizada no es solo un conjunto de principios, sino un sistema de gobierno con incentivos definidos. La presión por eficiencia y por reducción de incertidumbre favorece reglas generales, umbrales y taxonomías, aun cuando el conflicto moral requiera contextualización. Además, el régimen no decide en un vacío. Su marco de decisión se ve condicionado por dependencias externas, como requerimientos regulatorios fragmentados, presiones comerciales, conflictos de interés derivados de modelos de negocio basados en atención o intermediación, y exigencias de actores privados con poder de influencia, incluyendo anunciantes, proveedores de infraestructura, titulares de derechos y, en ocasiones, coaliciones políticas o sociales. En consecuencia, la moral privatizada debe comprenderse como un producto institucional. Lo que importa no es solo qué regla se adopta, sino qué estructura de incentivos la produce y qué tipo de racionalidad organiza su aplicación a escala.

7.2. Pluralismo moral, conflictos de valores y riesgos democráticos en plataformas

La moral privatizada se despliega en un contexto de pluralismo moral persistente. En sociedades complejas, existen desacuerdos razonables sobre valores, prioridades y bienes en conflicto, lo que impide asumir un consenso sustantivo único. Este pluralismo no es un defecto contingente. Es una condición estructural de la vida

pública y, por tanto, de cualquier régimen que pretenda gobernar conductas de manera general. La consecuencia nos dejaría en la comprensión que, si el contenido moral es disputado, entonces la legitimidad del régimen no puede descansar solo en la proclamación de principios o en la supuesta corrección de un conjunto de valores, sino en la existencia de condiciones institucionales que hagan el ejercicio del poder razonablemente justificable ante quienes lo soportan.

En plataformas y servicios digitales, este problema adquiere un carácter específico. Las reglas privadas se aplican a escala, en un espacio transnacional, con capacidad de afectar condiciones de participación social, acceso a mercados, visibilidad, reputación y posibilidades de expresión. Al mismo tiempo, el régimen opera mediante procedimientos internos y decisiones que pueden ser opacas o difíciles de impugnar. Esa combinación intensifica riesgos democráticos. Primero, porque desplaza conflictos que serían materia de deliberación pública hacia decisiones internas sin control equivalente. Segundo, porque el desacuerdo moral puede ser tratado como un problema de gestión del riesgo y no como conflicto de valores. Tercero, porque la automatización tiende a estabilizar categorías y umbrales, reduciendo el espacio de contextualización y ampliando el costo del error cuando las consecuencias son restrictivas. Por ello, el pluralismo moral conduce a una exigencia normativa básica. Si el régimen no puede apoyarse en consenso sustantivo, debe apoyarse en legitimidad procedimental, esto es, en garantías mínimas que controlen arbitrariedad y permitan contestación, corrección y rendición de cuentas en decisiones adversas relevantes (Rawls, 2005; Buchanan & Keohane, 2006)

7.3. El derecho como límite: garantías mínimas y estándar de legitimidad procedimental

Si la moral privatizada opera como régimen de gobierno efectivo por contrato y arquitectura técnica, el problema jurídico central no es describir su existencia, sino fijar sus límites. En este punto, el derecho cumple una función clásica, limitar el poder donde exis-

te asimetría y donde las decisiones pueden producir restricciones significativas. La idea de límite no exige equiparar plataformas con Estados, pero sí reconocer que cuando un régimen privado define condiciones de participación y distribuye consecuencias, la ausencia de garantías procedimentales genera un riesgo estructural de arbitrariedad, y ese riesgo es jurídicamente relevante.

Un estándar mínimo de legitimidad procedimental en gobernanza privada puede formularse como un conjunto de garantías exigibles, al menos en decisiones adversas con impacto relevante. Primero, cognoscibilidad de reglas, entendida como disponibilidad y claridad suficiente para anticipar consecuencias y comprender qué conductas activan restricciones. Segundo, explicación funcional de decisiones, de modo que el afectado pueda conocer la razón operativa de la medida y no reciba fórmulas vacías. Tercero, contestabilidad efectiva, es decir, acceso real a mecanismos de impugnación con revisión significativa y posibilidad de corrección. Cuarto, proporcionalidad y graduación, de manera que la respuesta se adecúe a la gravedad del supuesto incumplimiento y se reduzca el costo del error cuando la detección es incierta. Quinto, trazabilidad y rendición de cuentas, mediante registro verificable de criterios, uso de automatización, tasas de error relevantes, y evaluación de impacto o sesgos cuando corresponda. Estas condiciones no definen por sí mismas una teoría moral completa, pero sí delimitan la forma aceptable de un poder normativo privatizado bajo pluralismo moral y asimetría.

Este estándar mínimo se refleja en desarrollos normativos contemporáneos que, sin resolver disputas metaéticas, buscan institucionalizar garantías. En el contexto europeo, el Reglamento de Servicios Digitales exige motivación clara y específica para restricciones basadas en ilegalidad o incompatibilidad con términos y condiciones, y refuerza mecanismos de reclamación y revisión, reconociendo que la legitimidad de restricciones impuestas por proveedores depende de procedimientos que vuelvan el poder rastreable y contestable (Regulation (EU) 2022/2065). En paralelo, instrumentos y marcos de gobernanza de IA han enfatizado accountability, evaluaciones de impacto, documentación y control

institucional como condiciones para evitar que sistemas sociotécnicos operen como caja negra con capacidad de producir daños sin remedio (NIST, 2023; UNESCO, 2021). La convergencia de estos enfoques muestra el punto estructural del artículo. Cuando la ética se vuelve ejecutable, el derecho deja de ser una referencia externa y pasa a ser un límite interno al diseño institucional. El estándar de legitimidad no se agota en el contenido moral invocado por la organización, sino en las garantías que acompañan la producción, aplicación y corrección de reglas cuando esas reglas se vuelven control real por contrato y código.

8. Hacia un test de legitimidad para la moral privatizada

La idea central del presente escrito sostiene que, en el ecosistema digital, la ética aplicada a plataformas y sistemas de IA rara vez opera como mera adhesión voluntaria a principios generales. Opera, sobre todo, como un régimen ejecutable que se internaliza en dos planos. Primero, en instrumentos normativos privados, tales como términos y condiciones, políticas, contratos con usuarios, y acuerdos B2B. Segundo, en decisiones técnicas, tales como diseño de producto, arquitectura de modelos, umbrales de detección, flujos de *enforcement*, y métricas de desempeño. En ese tránsito, la moral se vuelve operativa y automatizable. Y al volverse automatizable, adquiere densidad de poder, porque produce efectos reales sobre acceso, visibilidad, monetización, reputación y participación.

Desde esa premisa, la pregunta relevante deja de ser exclusivamente sustantiva. No se reduce a qué valores son correctos. La pregunta central se desplaza hacia condiciones de legitimidad. Bajo qué condiciones es legítimo que un actor privado ejecute, a gran escala, un conjunto de decisiones moralmente cargadas mediante contrato y código. La propuesta de este capítulo es un test de legitimidad compuesto por criterios verificables, pensado para evaluar regímenes privados que funcionan como gobernanza normativa de facto.

El test no pretende resolver el debate metaético sobre la verdad moral, ni imponer una moral única. Su apuesta es procedimental y jurídico institucional. En contextos de pluralismo moral y transnacionalidad, la legitimidad depende menos del consenso sobre el contenido final de los valores y más de condiciones mínimas de publicidad, trazabilidad, contestabilidad, proporcionalidad y rendición de cuentas. Esta aproximación dialoga con la idea de que la arquitectura técnica regula conductas de forma comparable a la norma, y con enfoques que sostienen que la justificación pública y la posibilidad de impugnación son requisitos de legitimidad cuando una decisión afecta intereses relevantes, incluso fuera del Estado. También se apoya, como insumos traducibles a obligaciones, en marcos contemporáneos de gestión de riesgo y gobernanza, tales como NIST AI RMF y la Recomendación de IA de la OCDE, y en estándares de transparencia y debido proceso en moderación, tales como los Santa Clara Principles.

8.1. Alcance y función del test

El test se aplica a sistemas de IA que cumplen simultáneamente tres condiciones: i) operan como infraestructura decisional; ii) ejecutan reglas privadas que condicionan acceso, visibilidad, permanencia o trato; y iii) producen efectos adversos relevantes sin mediación humana directa o con mediación limitada. En estos casos, la evaluación de legitimidad no puede agotarse en la voluntariedad contractual ni en la invocación de principios éticos generales.

El objetivo del test es identificar déficits estructurales de legitimidad y orientar tanto el diseño institucional como la intervención jurídica, sin asumir que toda gobernanza privada sea ilegítima per se.

8.2. Publicidad y cognoscibilidad de las reglas

El primer criterio exige que las reglas que gobiernan el sistema sean cognoscibles para los sujetos afectados. Esto implica que los estándares relevantes no se limiten a declaraciones abstractas, sino que se expresen de forma suficientemente clara como para

permitir anticipación razonable de consecuencias. La opacidad normativa, especialmente cuando se combina con *enforcement* automatizado, erosiona la capacidad del sistema para orientar conductas y debilita su pretensión de legitimidad.

8.3. Explicabilidad funcional de las decisiones adversas

Cuando un sistema produce una decisión adversa, debe existir una explicación funcional suficiente que permita comprender qué tipo de regla se aplicó y qué hecho desencadenó la consecuencia. Este criterio no exige revelar modelos completos ni secretos industriales, pero sí superar la mera notificación automática. Sin explicación mínima, la decisión se vuelve indistinguible de la arbitrariedad, incluso si responde a un proceso técnicamente sofisticado.

8.4. Contestabilidad efectiva y revisión significativa

La legitimidad procedimental requiere la posibilidad real de impugnar decisiones adversas. La contestación debe ser accesible, comprensible y capaz de producir revisión sustantiva, no solo confirmación automática. En contextos de alta automatización, la ausencia de revisión significativa consolida errores y traslada de manera desproporcionada los costos del sistema a los usuarios afectados.

8.5. Proporcionalidad y graduación del *enforcement*

Las consecuencias derivadas de la aplicación de reglas privadas deben ser proporcionales al riesgo o a la conducta atribuida. Cuando el costo del error es elevado, el régimen debe privilegiar respuestas graduadas, reversibles y escalonadas antes que sanciones máximas o definitivas. La proporcionalidad funciona aquí como límite estructural a la ejecución automática de normas controvertidas.

8.6. Gobernanza del error y distribución de riesgos

Todo sistema automatizado produce errores. Un régimen legítimo no es aquel que promete infalibilidad, sino aquel que reconoce el error, lo mide y distribuye sus costos de manera razonable. Esto implica mecanismos internos de monitoreo, corrección y aprendizaje, así como decisiones explícitas sobre quién asume las consecuencias cuando el sistema falla.

8.7. Trazabilidad y rendición de cuentas

El último criterio exige trazabilidad suficiente para permitir auditoría interna y, cuando corresponda, externa. La existencia de decisiones automatizadas con efectos relevantes sin trazabilidad impide cualquier forma de control jurídico o institucional. La rendición de cuentas no requiere transparencia absoluta, pero sí evidencia verificable de que el sistema opera conforme a reglas conocidas y revisables.

8.8. Función jurídica del test

El test de legitimidad propuesto cumple una doble función. En el plano analítico, permite evaluar críticamente regímenes de moral privatizada sin recurrir a juicios morales sustantivos controvertidos. En el plano normativo, ofrece un marco para la intervención jurídica, ya sea mediante regulación pública, control judicial o exigencias contractuales reforzadas.

El punto central es que, cuando la ética se vuelve ejecutable por contrato y código, el derecho no puede limitarse a observar. Su función es reintroducir garantías mínimas allí donde el poder normativo privado se ejerce sin los contrapesos que históricamente han acompañado a otros regímenes de decisión con consecuencias relevantes.

9. Conclusiones

Este artículo ha sostenido que el problema central de la IA en contextos de gobernanza digital no es la ausencia de valores, sino la forma en que esos valores se traducen en reglas exigibles y se ejecutan mediante arquitectura técnica. En ese tránsito, la ética deja de operar como orientación discursiva y se convierte en un régimen práctico de control, articulado principalmente por contrato y código. El resultado es la emergencia de una moral privatizada que ejerce poder normativo con efectos relevantes sobre individuos y organizaciones.

Frente a este fenómeno, el análisis mostró que el *soft law* cumple una función importante de coordinación, legitimación y estandarización del lenguaje ético, pero resulta insuficiente para explicar o controlar el ejercicio efectivo del poder. El control real aparece cuando los principios se incorporan a términos contractuales y se implementan en sistemas técnicos que producen consecuencias automáticas o semi-automáticas. En ese punto, el debate deja de ser meramente ético y se convierte en un problema jurídico-filosófico de legitimidad.

El artículo propuso abordar este problema desplazando el foco desde la corrección moral sustantiva hacia la legitimidad procedimental. En contextos de desacuerdo moral persistente, la pregunta relevante no es qué valores son correctos en abstracto, sino bajo qué condiciones un régimen normativo privado puede justificar el ejercicio de poder cuando produce decisiones adversas. Este enfoque permite evaluar la gobernanza algorítmica sin exigir unanimidad moral y sin reducir la legitimidad al consentimiento contractual formal.

Como aporte central, se formuló un test de legitimidad para la moral privatizada ejecutada por sistemas de inteligencia artificial. El test no pretende agotar el debate ni sustituir la regulación pública, sino ofrecer criterios mínimos para identificar déficits estructurales de legitimidad allí donde el control se ejerce mediante automatización. Publicidad de reglas, explicabilidad funcional, contestabilidad efectiva, proporcionalidad, gobernanza del error

y trazabilidad constituyen exigencias básicas para cualquier régimen que aspire a ejercer poder normativo de manera compatible con estándares mínimos de racionalidad y justicia procedimental.

Desde una perspectiva más amplia, el argumento desarrollado sugiere que la gobernanza de la inteligencia artificial reabre preguntas clásicas de la filosofía del derecho y de la teoría política en un nuevo contexto tecnológico. La separación entre ética declarativa y moral operativa, la relación entre poder y justificación, y la función del derecho como límite al ejercicio de autoridad privada adquieren una relevancia renovada cuando las decisiones se escalan por diseño y se presentan bajo la apariencia de neutralidad técnica.

La conclusión no es que toda forma de gobernanza privada basada en IA sea ilegítima, ni que el derecho deba absorber íntegramente estos regímenes. La conclusión es más precisa: cuando la ética se vuelve ejecutable por contrato y código, el derecho no puede limitarse a observar ni a confiar en la autorregulación. Su función es reintroducir garantías allí donde el poder normativo opera sin los contrapesos que históricamente han acompañado a otros sistemas de decisión con consecuencias relevantes.

En última instancia, la discusión sobre IA no trata únicamente sobre tecnología, sino sobre las condiciones bajo las cuales aceptamos ser gobernados por reglas que no votamos, aplicadas por sistemas que no vemos y justificadas por valores que no siempre compartimos. Identificar esos límites y reconstruir criterios de legitimidad es una tarea jurídica y filosófica ineludible para cualquier sociedad que aspire a integrar la inteligencia artificial sin renunciar a sus principios básicos de justicia y control del poder.

10. Referencias

Abbott, K. W., Keohane, R. O., Moravcsik, A., Slaughter, A.-M., & Snidal, D. (2000). The concept of legalization. *International Organization*, 54(3), 401–419. https://doi.org/10.1017/S0020818300442515

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732. https://doi.org/10.2139/ssrn.2477899

Ben-Shahar, O., & Schneider, C. E. (2010). The failure of mandated disclosure. *University of Pennsylvania Law Review*, 159(3), 647–749. https://scholarship.law.upenn.edu/penn_law_review/vol159/iss3/2/

Brink, D. O. (1989). *Moral realism and the foundations of ethics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511624604

Buchanan, A. (2002). Political legitimacy and democracy. *Ethics*, 112(4), 689–719. https://doi.org/10.1086/340313

Buchanan, A., & Keohane, R. O. (2006). The legitimacy of global governance institutions. *Ethics*, 20(4), 405–437. https://doi.org/10.1017/S0892679400000207

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2), 153–163. https://doi.org/10.1089/big.2016.0047

Fuller, L. L. (1964). *The morality of law*. Yale University Press.

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. https://arxiv.org/abs/1610.02413

Hart, H. L. A. (1961). *The concept of law*. Oxford University Press.

ISO/IEC. (2022). *Information technology—Artificial intelligence—Artificial intelligence concepts and terminology (ISO/IEC 22989:2022)*.

International Organization for Standardization. https://www.iso.org/standard/74296.html

Korsgaard, C. M. (1996). *The sources of normativity*. Cambridge University Press. https://doi.org/10.1017/CBO9780511554475

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, 165(3), 633–705. https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/

Lessig, L. (1999). *Code and other laws of cyberspace*. Basic Books.

Miller, A. (2013). *Moral theory: An introduction*. Wiley-Blackwell.

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1)*. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

Organisation for Economic Co-operation and Development. (2019). *Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449)*. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Radin, M. J. (2013). *Boilerplate: The fine print, vanishing rights, and the rule of law*. Princeton University Press.

Rawls, J. (2005). *Political liberalism* (Expanded ed.). Columbia University Press.

Raz, J. (1986). *The morality of freedom*. Oxford University Press.

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (Digital Services Act). https://eur-lex.europa.eu/eli/reg/2022/2065/oj

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). (2024). *Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Santa Clara Principles on Transparency and Accountability in Content Moderation. (2021). New America. https://www.newamerica.

org/oti/reports/santa-clara-principles-transparency-and-accountability-content-moderation/

Schmidt, V. A. (2013). Democracy and legitimacy in the European Union revisited: Input, output and throughput. *Political Studies*, 61(1), 2–22. https://doi.org/10.1111/j.1467-9248.2012.00962.x

Shafer-Landau, R. (2003). *Moral realism: A defence*. Oxford University Press.

Specht v. Netscape Communications Corp., 306 F.3d 17 (2d Cir. 2002). https://law.justia.com/cases/federal/appellate-courts/F3/306/17/642323/

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review, 20*(3), 571–610. https://doi.org/10.2307/258788

Suzor, N. (2019). *Lawless: The secret rules that govern our digital lives*. Cambridge University Press. https://doi.org/10.1017/9781108697841

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

van Leeuwen, T. (2008). *Discourse and practice: New tools for critical discourse analysis*. Oxford University Press.

Weber, M. (1978). *Economy and society* (G. Roth & C. Wittich, Eds.). University of California Press.