

Hypernudges y consentimiento sustantivo: bases para un estándar jurídico de lealtad algorítmica en arquitecturas digitales de decisión

Fernando RAMOS-ZAGA*

RESUMEN: Las arquitecturas digitales de decisión utilizadas por las plataformas configuran entornos diseñados para modular la conducta de los usuarios mediante técnicas de manipulación conductual opacas y personalizadas, conocidas como *hypernudges*. Tales prácticas transforman radicalmente las condiciones bajo las cuales se forma el consentimiento, poniendo en crisis los supuestos dogmáticos que sostienen la autonomía de la voluntad y la validez del negocio jurídico en el derecho privado. Este artículo analiza críticamente la suficiencia de dichas categorías tradicionales frente a los nuevos desafíos planteados, proponiendo como respuesta una reinterpretación de la buena fe objetiva que incorpore un estándar de lealtad algorítmica. Se identifican tres criterios para delimitar esta forma de manipulación ilícita: opacidad deliberada, personalización explotadora y asimetría de resultados. El hallazgo central demuestra que un régimen de responsabilidad por diseño, fundado en la buena fe, permite restablecer la protección del consentimiento en clave sustantiva.

PALABRAS-CLAVE: *Hypernudge*; consentimiento informado; buena fe objetiva; lealtad algorítmica; autonomía de la voluntad.

SUMARIO: 1. Introducción; – 2. La especificidad ontológica del *hypernudge*: de la economía conductual a la gubernamentalidad algorítmica; – 3. La crisis del paradigma del consentimiento informado y la patología estructural de los remedios tradicionales; – 4. La libertad cognitiva como fundamento constitucional del derecho a no ser manipulado; – 5. La buena fe objetiva como fundamento de la lealtad algorítmica: hacia un régimen de responsabilidad por diseño; – 6. Conclusiones; – Referencias bibliográficas.

TITLE: *Private Autonomy and the Hypernudge Challenge: Towards a Normative Standard of Algorithmic Loyalty*

ABSTRACT: Digital choice architectures deployed by platforms constitute environments designed to modulate user behavior through opaque and personalized behavioral manipulation techniques known as *hypernudges*. Such practices radically transform the conditions under which consent is formed, undermining the dogmatic assumptions that sustain autonomy of the will and the validity of the legal act within private law. This article critically examines the adequacy of these traditional categories in the face of emerging challenges, proposing a reinterpretation of objective good faith that incorporates a standard of algorithmic loyalty. Three criteria are identified to delimit this form of unlawful manipulation: deliberate opacity, exploitative personalization, and asymmetrical outcomes. The central finding demonstrates that a liability-by-design regime, grounded in the principle of good faith, enables the restoration of consent protection in substantive terms.

KEYWORDS: *Hypernudge*; informed consent; objective good faith; algorithmic loyalty; autonomy of the will.

CONTENTS: 1. Introduction; – 2. The ontological specificity of the *hypernudge*: from behavioral economics to algorithmic governmentality; – 3. The crisis of the informed consent paradigm and the structural pathology of traditional remedies; – 4. Cognitive liberty as the constitutional foundation of the right not to be manipulated; – 5. Objective good faith as the basis for algorithmic loyalty: towards a regime of liability by design; – 6. Conclusions; – Bibliographic references.

* Universidad Privada del Norte, Perú. ORCID: <https://orcid.org/0000-0001-6301-9460>.

1. Introducción

La progresiva digitalización de la vida cotidiana ha inaugurado un nuevo régimen de interacción entre los individuos y su entorno, mediado por sistemas algorítmicos que intervienen de manera cada vez más sofisticada en la formación de preferencias, juicios y decisiones. Tal fenómeno ha transformado la arquitectura de elección, desplazándola de un diseño estático y transparente hacia configuraciones dinámicas, personalizadas y opacas que operan sobre los procesos mentales pre-reflexivos de los usuarios. La emergencia del llamado *hypernudge*, como evolución tecnológicamente mediada de los tradicionales empujones conductuales (*nudges*), pone en tela de juicio los presupuestos normativos de la teoría contractual clásica, particularmente la idea de un sujeto deliberativo que decide libre e informadamente dentro de un mercado neutral de opciones.¹

Desde el plano conceptual, la economía conductual ha introducido una crítica relevante al modelo racional del agente decisor, al evidenciar la existencia de sesgos sistemáticos y limitaciones cognitivas que afectan la formación de preferencias.² El programa del paternalismo libertario propuso, sobre esta base, intervenciones arquitectónicas mínimamente invasivas que preservaran la libertad de elección, bajo la premisa de que toda presentación de opciones es, en algún grado, influyente.³ Sin embargo, la incorporación de inteligencia artificial, minería de datos y aprendizaje automático ha desplazado el paradigma del *nudge* hacia formas de intervención comportamental que modifican en tiempo real el entorno digital del individuo en función de inferencias predictivas sobre sus vulnerabilidades cognitivas.⁴ Este giro tecnológico ha sido caracterizado como una forma de gubernamentalidad algorítmica, en la que el poder ya no se ejerce por mandato, sino por modulación ambiental pre-consciente.⁵

¹ YEUNG, Karen. Hypernudge, big data as a mode of regulation by design. *Information, Communication & Society*, vol. 20, n. 1, 2017, p. 120; ROUVROY, Antoinette; BERNIS, Thomas. Gouvernamentalité algorithmique et perspectives d'émancipation, le disparate come condition d'individuation par la relation. *Réseaux*, n. 177, 2013, p. 167.

² KAHNEMAN, Daniel. *Thinking, fast and slow*. Nueva York: Farrar, Straus and Giroux, 2011, p. 20; TVERSKY, Amos; KAHNEMAN, Daniel. Judgment under uncertainty, heuristics and biases. *Science*, vol. 185, n. 4157, 1974, p. 1124.

³ THALER, Richard H.; SUNSTEIN, Cass R. *Nudge: improving decisions about health, wealth, and happiness*. New Haven: Yale University Press, 2008, p. 6.

⁴ ESPOSITO, Fabrizio; CATHOUD FERREIRA, Thaís Maciel. Addictive design as an unfair commercial practice, the case of hyper engaging dark patterns. *European Journal of Risk Regulation*, marzo/2024, p. 3.

⁵ ROUVROY, Antoinette; BERNIS, Thomas. Gouvernamentalité algorithmique et perspectives d'émancipation, le disparate come condition d'individuation par la relation. *Réseaux*, n. 177, 2013, p. 165; DELEUZE, Gilles. Postscript on the societies of control. *October*, n. 59, 1992, p. 5.

Pese a la creciente atención que este fenómeno ha suscitado en la literatura, tanto jurídica como filosófica, subsiste una importante brecha analítica en torno a la capacidad del derecho privado para enfrentar las técnicas de manipulación algorítmica sin desbordar sus propias categorías estructurales. Buena parte del debate se ha centrado en la necesidad de reforzar los deberes de información, transparencia y consentimiento en línea, sin cuestionar el núcleo dogmático que articula la validez del negocio jurídico en torno a la autonomía de la voluntad.⁶ Sin embargo, cuando la arquitectura digital de decisión está diseñada para explotar de manera personalizada y opaca las debilidades cognitivas del usuario, la presunción de consentimiento libre e informado pierde anclaje empírico. La cuestión no es simplemente si las técnicas manipulativas vulneran normas existentes, sino si el marco conceptual vigente es capaz de captarlas como ilícitas. En este sentido, nos enfrentamos no a una laguna normativa, sino a una disfunción ontológica del paradigma.

La investigación que se presenta busca, precisamente, evaluar críticamente la idoneidad dogmática de las categorías tradicionales de autonomía de la voluntad y validez del negocio jurídico frente a las técnicas de manipulación conductual empleadas por las plataformas digitales. Se parte del supuesto de que tales categorías, fundadas sobre la imagen de un sujeto racional y deliberativo, resultan conceptualmente insuficientes para aprehender la realidad de decisiones formadas bajo condiciones de opacidad, personalización y explotación cognitiva. Desde esta premisa, el artículo propone reinterpretar la cláusula general de buena fe objetiva como fuente de un deber de lealtad algorítmica, orientado a garantizar que la configuración de las arquitecturas digitales de decisión respete la integridad cognitiva del usuario y permita una expresión voluntaria genuina del consentimiento.

La pertinencia práctica de esta investigación radica en su potencial para ofrecer un marco jurídico operativo capaz de regular el diseño algorítmico sin recurrir a prohibiciones generales de tecnologías ni a soluciones puramente informativas que ya han demostrado su ineficacia. Al traducir el mandato constitucional de protección de la libertad cognitiva en obligaciones concretas de diseño y operación de plataformas, se genera un puente normativo entre el nivel axiológico de los derechos fundamentales y el plano técnico de la ingeniería de sistemas, con implicaciones directas para la responsabilidad contractual,

⁶ DÍEZ PICAZO, Luis. *Fundamentos del derecho civil patrimonial*, vol. 1: Introducción, teoría del contrato. 6. ed. Madrid: Thomson Civitas, 2007, p. 155; CÁMARA LAPUENTE, Sergio. El control de las cláusulas abusivas sobre elementos esenciales del contrato. *Cuadernos de Derecho Transnacional*, vol. 3, n. 1, 2011, p. 90.

el control judicial de prácticas desleales y el desarrollo de estándares auditables de comportamiento empresarial.

La propuesta se inscribe, además, en el contexto de desafíos contemporáneos que trascienden el ámbito jurídico, como el debilitamiento de la agencia individual en entornos digitales, la expansión del capitalismo de vigilancia y la creciente sofisticación de las técnicas de captura atencional y persuasión oculta.⁷ La regulación de la inteligencia artificial y los servicios digitales en la Unión Europea ha comenzado a reconocer, aunque de manera incipiente, la necesidad de establecer límites materiales a la explotación algorítmica de la conducta. No obstante, la eficacia de tales medidas dependerá en buena medida de la capacidad del derecho privado para incorporar criterios sustantivos de lealtad y equidad en las relaciones contractuales mediadas por tecnología.

En términos metodológicos, el artículo adopta un enfoque dogmático-constructivo, articulando elementos de la teoría del derecho privado con aportes de la filosofía del derecho, la economía del comportamiento y la ética de la inteligencia artificial. El objetivo general consiste en fundamentar una reinterpretación de la buena fe objetiva orientada a establecer un estándar de lealtad algorítmica que garantice la protección efectiva del consentimiento en entornos definidos por arquitecturas digitales de decisión. De ese modo, se busca contribuir al desarrollo de un marco normativo coherente que permita restaurar la función emancipadora de la autonomía privada en contextos donde esta se encuentra amenazada por formas invisibles y sofisticadas de manipulación conductual.

2. La especificidad ontológica del *hypernudge*: de la economía conductual a la gubernamentalidad algorítmica

La comprensión jurídica del fenómeno que aquí se examina exige, como paso previo, una reconstrucción precisa de su genealogía conceptual y de los rasgos que lo distinguen de sus antecedentes teóricos. El punto de partida obligado se encuentra en los hallazgos de la psicología cognitiva sobre los procesos duales de pensamiento, según los cuales la mente humana opera a través de dos sistemas diferenciados: uno rápido, intuitivo y automático, y otro lento, deliberativo y reflexivo.⁸ Lejos de constituir una mera curiosidad empírica, esta distinción posee implicaciones normativas de primer orden,

⁷ ZUBOFF, Shoshana. *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. Nueva York: PublicAffairs, 2019, p. 8; SUSSER, Daniel; ROESSLER, Beate; NISSENBAUM, Helen. Technology, autonomy, and manipulation. *Internet Policy Review*, vol. 8, n. 2, 2019, p. 10.

⁸ KAHNEMAN, Daniel. *Thinking, fast and slow*. Nueva York: Farrar, Straus and Giroux, 2011, p. 20.

pues revela que una porción significativa de las decisiones humanas, incluidas aquellas con relevancia jurídica, no se adopta mediante el ejercicio de una racionalidad plenamente reflexiva, sino a través de heurísticas y atajos cognitivos susceptibles de producir sesgos sistemáticos y predecibles.⁹

A partir de este fundamento empírico, se edificó el programa del paternalismo libertario, el cual propuso que las instituciones públicas y privadas pueden mejorar las decisiones de los individuos mediante la configuración estratégica del contexto en el que las decisiones se adoptan, sin eliminar formalmente ninguna opción ni alterar de manera significativa los incentivos económicos.¹⁰ La noción de *nudge*, en tal sentido, designa una intervención en la arquitectura de decisión que orienta la conducta en una dirección predecible, preservando la libertad de elección del destinatario. Un ejemplo paradigmático, la disposición de alimentos saludables en un lugar prominente de la cafetería escolar, ilustra un mecanismo que opera sobre el Sistema 1 sin suprimir la posibilidad de que el Sistema 2 revierta la elección. Así planteada, la arquitectura de decisión fue presentada como un fenómeno inevitable y éticamente neutro en su estructura, cuya legitimidad dependería exclusivamente de los fines perseguidos y de la transparencia de los medios empleados.¹¹

Con la transposición de dicha lógica al entorno digital, se produce una mutación cualitativa que desborda los presupuestos originales del programa conductual. La convergencia entre capacidades de computación masiva, recolección ubicua de datos comportamentales y sistemas de aprendizaje automático ha dado lugar a lo que se ha conceptualizado como *hypernudge*: una modalidad de intervención conductual que se distingue del *nudge* estático por su carácter dinámico, personalizado, en red y continuamente actualizado mediante bucles recursivos de retroalimentación algorítmica.¹² Mientras el *nudge* tradicional opera como una configuración ambiental fija y uniforme, aplicable de manera idéntica a todos los sujetos que transitan por un mismo espacio decisional, el *hypernudge* se alimenta del flujo incesante de datos generados por la interacción del usuario con la plataforma digital: tiempos de visualización, patrones de desplazamiento, micro-interacciones, historial de navegación y metadatos contextuales. Dichos datos son procesados por algoritmos de aprendizaje automático que

⁹ TVERSKY, Amos; KAHNEMAN, Daniel. Judgment under uncertainty, heuristics and biases. *Science*, vol. 185, n. 4157, 1974, p. 1125.

¹⁰ THALER, Richard H.; SUNSTEIN, Cass R. *Nudge: improving decisions about health, wealth, and happiness*. New Haven: Yale University Press, 2008, p. 5.

¹¹ SUNSTEIN, Cass R. Nudging and choice architecture, ethical considerations. *Yale Journal on Regulation*, vol. 32, 2015, p. 417.

¹² YEUNG, Karen. Hypernudge, big data as a mode of regulation by design. *Information, Communication & Society*, vol. 20, n. 1, 2017, p. 122.

refinan continuamente la eficacia de la influencia, generando una experiencia altamente individualizada cuyo objetivo no es persuadir racionalmente, sino optimizar la respuesta conductual del destinatario.¹³

Conviene destacar que la diferencia no es meramente cuantitativa. El tránsito del *nudge* al *hypernudge* implica un salto ontológico que puede articularse en torno a cuatro dimensiones estructurales. En primer lugar, la personalización granular: mientras el *nudge* clásico se dirige al «consumidor medio» o al ciudadano genérico, el *hypernudge* construye un perfil psicográfico individualizado que permite explotar las vulnerabilidades específicas de cada sujeto, incluyendo sus sesgos cognitivos particulares, sus estados emocionales transitorios y sus patrones de comportamiento habitual.¹⁴ En segundo lugar, la dinamicidad adaptativa: la arquitectura de decisión no permanece estable, sino que se reconfigura en tiempo real en función de la respuesta del usuario, creando un sistema cerrado de retroalimentación tecno-social que «guía» el comportamiento mediante ajustes incrementales e imperceptibles.¹⁵ En tercer lugar, la opacidad estructural: los algoritmos de perfilamiento y recomendación operan bajo una lógica de caja negra protegida tanto por la complejidad técnica de los modelos de aprendizaje profundo como por el secreto comercial, lo que impide al destinatario comprender los mecanismos causales que configuran su entorno decisional.¹⁶ En cuarto lugar, la pervasividad ubicua: a diferencia de un espacio delimitado y reconocible, como la cafetería del ejemplo canónico, el *hypernudge* permea la totalidad de la experiencia digital del sujeto, desde las redes sociales hasta las plataformas de comercio electrónico, los motores de búsqueda y las aplicaciones de mensajería.¹⁷

Como resultado de la confluencia de estas cuatro dimensiones, emerge un fenómeno que ha sido descrito como una forma de «gubernamentalidad algorítmica»:¹⁸ un modo de gobierno que no opera mediante el mandato imperativo dirigido a la consciencia del sujeto, propio del paradigma jurídico moderno, sino a través de la modulación pre-reflexiva del entorno informativo en el que se forma la preferencia. En dicho régimen, el individuo no es tratado como un agente moral dotado de racionalidad y dignidad, sino

¹³ YEUNG, Karen. Hypernudge, big data as a mode of regulation by design. *Information, Communication & Society*, vol. 20, n. 1, 2017, p. 125.

¹⁴ ESPOSITO, Fabrizio; CATHOUD FERREIRA, Thaís Maciel. Addictive design as an unfair commercial practice, the case of hyper engaging dark patterns. *European Journal of Risk Regulation*, marzo/2024, p. 7.

¹⁵ HERZOG, Lisa; KELLMEYER, Philipp; WILD, Verina. Digital behavioral technology, vulnerability and justice, towards an integrated approach. *Review of Social Economy*, vol. 80, n. 1, 2022, p. 12.

¹⁶ PASQUALE, Frank. *The black box society: the secret algorithms that control money and information*. Cambridge, MA: Harvard University Press, 2015, p. 3.

¹⁷ MOROZOVATTE, Viktorija. Hypernudging in the changing European regulatory landscape for digital markets. *Policy & Internet*, vol. 15, n. 1, 2023, p. 82.

¹⁸ ROUVROY, Antoinette; BERNIS, Thomas. Gouvernamentalité algorithmique et perspectives d'émancipation, le disparate come condition d'individuation par la relation. *Réseaux*, n. 177, 2013, p. 168.

como un agregado de datos comportamentales cuyas regularidades estadísticas permiten predecir y canalizar su conducta futura. La noción deleuziana de «dividuo», un sujeto fragmentado en muestras, datos y mercados, adquiere aquí una materialidad técnica concreta.¹⁹

Desde el punto de vista del derecho privado, la especificidad del *hypernudge* plantea un desafío que trasciende la mera actualización tecnológica de categorías preexistentes. La dogmática civilista continental, heredera de la codificación decimonónica, construyó su teoría de la autonomía de la voluntad sobre el presupuesto de un sujeto capaz de deliberación reflexiva y autodeterminación racional.²⁰ Los mecanismos de protección del consentimiento, como los vicios de la voluntad, los deberes precontractuales de información y el control de cláusulas abusivas, fueron diseñados para corregir distorsiones puntuales en un proceso decisional que se presume esencialmente libre. Sin embargo, el *hypernudge* no distorsiona una decisión específica, sino que configura el proceso mismo de formación de la preferencia, operando en una capa anterior a la deliberación consciente. No se trata, en rigor, de un problema de información asimétrica susceptible de corrección mediante deberes de transparencia, sino de lo que puede denominarse una asimetría epistémica radical. En ella, el proveedor digital no solo conoce el producto mejor que el consumidor, sino que posee un conocimiento profundo y probabilístico sobre la psiquis, las debilidades y los sesgos del propio consumidor, conocimiento del cual el sujeto carece respecto de sí mismo.²¹

En términos técnicos, esta asimetría se sustenta en la lógica extractiva del capitalismo de vigilancia, un modelo económico cuya materia prima no es el dato personal en sí mismo, sino el «excedente conductual» generado por la actividad digital del usuario, el cual es procesado mediante inteligencia artificial predictiva para fabricar «productos de predicción» comercializables en mercados de comportamiento futuro.²² Dentro de este esquema, la arquitectura de decisión digital no es un medio neutro de presentación de opciones, sino un instrumento de extracción y modificación conductual orientado por imperativos comerciales que operan al margen de, y frecuentemente en contra de, los intereses del sujeto cuyo comportamiento se modula. La experiencia del usuario, lejos de

¹⁹ DELEUZE, Gilles. Postscript on the societies of control. *October*, n. 59, 1992, p. 5.

²⁰ FRANK, Felipe. A consolidação da autonomia da vontade como cânone do direito privado moderno, o caso do Code Napoleônico de 1804. *Civilistica.com*, a. 11, n. 1, 2022, p. 4.

²¹ ZUBOFF, Shoshana. *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. Nueva York: PublicAffairs, 2019, p. 233.

²² ZUBOFF, Shoshana. *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. Nueva York: PublicAffairs, 2019, p. 97.

ser el fin del diseño, constituye el medio a través del cual se maximiza el valor del excedente conductual.

En consecuencia, la tensión entre tal lógica operativa y la estática normativa del derecho privado resulta particularmente aguda cuando se examina el requisito de legalidad formal que la tradición jurídica exige a toda forma de regulación de la conducta. Las normas jurídicas, para ser legítimas en un Estado de Derecho, deben satisfacer condiciones de generalidad, promulgación, claridad, prospectividad y posibilidad de cumplimiento.²³ El *hypernudge*, en cuanto técnica de regulación por diseño, subvierte cada uno de estos requisitos: no es general, sino hiperpersonalizado; no es público, sino opaco; no es claro, sino ininteligible para su destinatario; y no permite al sujeto «seguir» conscientemente la regla, pues opera sobre procesos automáticos que eluden la deliberación.²⁴ La pregunta que surge, entonces, no es si el derecho privado debe adaptarse a la realidad digital, cuestión sobre la cual existe un consenso amplio, sino si las categorías fundamentales de la autonomía de la voluntad y la validez del negocio jurídico conservan siquiera su idoneidad conceptual para aprehender un fenómeno que, por definición, opera en el terreno que dichas categorías presuponen como dado: la libertad de formación de la preferencia.

La transformación estructural que representa el paso del *nudge* al *hypernudge*, junto con el modelo de gubernamentalidad algorítmica que lo sustenta, pone en crisis las categorías jurídicas tradicionales, en particular aquellas que giran en torno a la autonomía de la voluntad y al consentimiento informado. A partir de esta constatación, resulta necesario examinar con mayor profundidad cómo esta mutación ontológica socava los fundamentos del paradigma contractual clásico y evidencia la insuficiencia de los remedios actualmente disponibles en el derecho privado y regulatorio para enfrentar la manipulación digital contemporánea.

3. La crisis del paradigma del consentimiento informado y la patología estructural de los remedios tradicionales

El edificio dogmático del derecho contractual moderno reposa sobre una premisa que funciona simultáneamente como supuesto y como ficción operativa: la declaración de voluntad, emitida por un sujeto informado y libre, constituye el vehículo legítimo para la autorregulación de intereses privados. Dicha premisa se traduce, en el plano de la

²³ FULLER, Lon L. *The morality of law*. Ed. rev. New Haven: Yale University Press, 1969, p. 39.

²⁴ ZEILSTRA, Rebecca. Formal legality for nudges. *Ratio Juris*, vol. 38, n. 1, marzo/2025, p. 48.

formación del contrato, en un sistema de controles de validez articulado en torno a dos ejes complementarios: de un lado, los vicios del consentimiento, es decir, error, dolo e intimidación, concebidos como patologías que afectan la libertad o el conocimiento del declarante y cuya concurrencia acarrea la anulabilidad del negocio jurídico; de otro lado, los deberes precontractuales de información, derivados del principio de buena fe, cuyo incumplimiento genera responsabilidad por los daños causados a la confianza legítima de la contraparte.²⁵ A lo anterior se suman, en el ámbito del derecho de consumo, mecanismos de transparencia reforzada, como el deber de información clara y comprensible sobre las condiciones generales de contratación, así como el control de contenido de las cláusulas abusivas, orientados a corregir la desigualdad estructural entre proveedores profesionales y consumidores.²⁶

No obstante, la eficacia de este aparato protector descansa sobre un presupuesto implícito que rara vez se explicita: la posibilidad fáctica de que el destinatario de la información procese, comprenda y utilice dicha información como insumo para una deliberación racional. Justamente en los entornos digitales configurados por *hypernudges* es donde ese presupuesto colapsa, generando lo que puede denominarse una patología estructural de ineficacia que opera simultáneamente en dos niveles: el sustantivo y el procesal.

En lo sustantivo, la primera línea de defensa del ordenamiento jurídico, es decir, el paradigma del consentimiento informado, se revela funcionalmente inoperante. La literatura especializada ha documentado extensamente la denominada paradoja de la transparencia: la complejidad de los sistemas de aprendizaje automático, la longitud inabarcable de los términos de servicio y la frecuencia de las modificaciones unilaterales de las condiciones contractuales generan una sobrecarga cognitiva que invalida fácticamente cualquier manifestación de voluntad fundada en la lectura y comprensión de dichos instrumentos.²⁷ Los datos empíricos son elocuentes: estudios reiterados han demostrado que la abrumadora mayoría de los usuarios acepta los términos de servicio sin leerlos, y que incluso aquellos que intentan hacerlo carecen de los conocimientos técnicos necesarios para comprender las implicaciones del tratamiento algorítmico de sus datos.²⁸ Aun cuando la regulación europea, particularmente el Reglamento General

²⁵ DÍEZ PICAZO, Luis. *Fundamentos del derecho civil patrimonial*, vol. 1: Introducción, teoría del contrato. 6. ed. Madrid: Thomson Civitas, 2007, p. 270.

²⁶ CÁMARA LAPUENTE, Sergio. El control de las cláusulas abusivas sobre elementos esenciales del contrato. *Cuadernos de Derecho Transnacional*, vol. 3, n. 1, 2011, p. 92.

²⁷ SOLOVE, Daniel J. Privacy self management and the consent dilemma. *Harvard Law Review*, vol. 126, 2013, p. 1886.

²⁸ BEN SHAHAR, Omri; SCHNEIDER, Carl E. *More than you wanted to know: the failure of mandated disclosure*. Princeton: Princeton University Press, 2014, p. 10.

de Protección de Datos (RGPD), ha intentado abordar este problema mediante la exigencia de un consentimiento libre, específico, informado e inequívoco, acompañado de información proporcionada en un lenguaje claro y sencillo, la evidencia disponible sugiere que tales exigencias formales no han logrado cerrar la brecha cognitiva entre la complejidad del tratamiento de datos y la capacidad de comprensión del usuario medio.²⁹

Conviene precisar que el problema jurídico que plantea el *hypernudge* trasciende cualitativamente la mera imposibilidad de lectura. No se trata solo de que el consumidor no pueda comprender los términos del contrato, sino de que el proveedor digital utiliza activamente el conocimiento derivado del perfilamiento algorítmico para configurar un entorno decisional que explota las vulnerabilidades cognitivas específicas del sujeto. Dicha dinámica genera una asimetría que difiere sustancialmente de la que el derecho de consumo tradicionalmente aborda. En la asimetría informativa clásica, el proveedor conoce mejor el producto y el consumidor no; la solución consiste en obligar al primero a revelar la información relevante. Frente a ello, en la asimetría epistémica que caracteriza al capitalismo de vigilancia, el proveedor posee un conocimiento profundo y probabilístico sobre la psiquis del propio consumidor, incluyendo sus miedos, deseos, sesgos y estados emocionales, conocimiento del cual el sujeto carece incluso respecto de sí mismo.³⁰ La transparencia, bajo tales condiciones, no solo es insuficiente, sino estructuralmente inadecuada, ya que el problema no reside en la ocultación de información sobre el objeto del contrato, sino en la explotación de información sobre el sujeto contratante.

Con especial claridad, esta insuficiencia se manifiesta al examinar la tipología de los vicios del consentimiento. El dolo, en su configuración clásica, presupone una conducta engañosa, ya sea activa u omisiva, dirigida a provocar un error en la contraparte respecto de un elemento esencial del negocio. La intimidación exige una amenaza ilícita que genere un temor racional. El error, por su parte, requiere una falsa representación de la realidad que recaiga sobre la sustancia, las cualidades esenciales o la identidad del objeto.³¹ El *hypernudge*, sin embargo, no opera mediante ninguno de estos mecanismos. No engaña sobre las propiedades del producto ni amenaza con consecuencias adversas. Su eficacia reside en la configuración del entorno decisional de modo tal que la

²⁹ SOLOVE, Daniel J. Privacy self management and the consent dilemma. *Harvard Law Review*, vol. 126, 2013, p. 1893.

³⁰ ZUBOFF, Shoshana. *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. Nueva York: PublicAffairs, 2019, p. 240.

³¹ FLUME, Werner. *Allgemeiner Teil des Bürgerlichen Rechts*, vol. 2: *Das Rechtsgeschäft*. 4. ed. Berlín: Springer, 1992. Trad. parcial consultada: *El negocio jurídico*. Madrid: Fundación Cultural del Notariado, 1998, p. 473.

preferencia del sujeto se forme en una dirección predeterminada, sin que medie un error cognitivo identificable ni una coacción reconocible. Más bien, se trata de lo que puede conceptualizarse como una coacción arquitectónica o una heteronomía programada que vicia la libertad en su génesis, no en su ejercicio.³²

Desde el plano doctrinal, el fenómeno ha comenzado a ser identificado bajo la categoría de influencia indebida (*undue influence*), entendida como la utilización de una posición de poder para ejercer presión sobre el consumidor de forma que se limite significativamente su capacidad de tomar una decisión informada.³³ En el contexto del derecho de la Unión Europea, esta noción encuentra su anclaje normativo en el artículo 2(j) de la Directiva sobre Prácticas Comerciales Desleales (Directiva 2005/29/CE), que prohíbe las prácticas agresivas que utilicen una posición de poder respecto del consumidor para influir en su decisión.

La subsunción de los patrones oscuros de hiperinteracción (*hyper-engaging dark patterns*) en esta categoría ha sido defendida con argumentos persuasivos. Dichas interfaces, diseñadas para ser adictivas mediante la explotación del ciclo de la dopamina y el agotamiento de los recursos atencionales del usuario, constituyen una forma de influencia que, sin ejercer coacción física, suprime la posibilidad de autodeterminación real.³⁴ El modelo del gancho (*Hook Model*) resulta especialmente ilustrativo de esta dinámica, ya que estructura el diseño adictivo en cuatro fases consecutivas: en primer lugar, un disparador; luego, una acción del usuario; posteriormente, una recompensa variable; y finalmente, una inversión. A través de esta secuencia, se muestra cómo la formación de hábitos compulsivos es deliberadamente diseñada para capturar la atención y neutralizar la deliberación racional.³⁵

Ahora bien, la calificación del *hypernudge* como práctica desleal o como vicio del consentimiento enfrenta un obstáculo adicional que opera en el nivel procesal: la opacidad algorítmica genera una situación de indefensión probatoria que imposibilita, en la práctica, el ejercicio efectivo de la tutela judicial. Los algoritmos de perfilamiento y recomendación operan bajo una lógica de caja negra, protegida simultáneamente por la

³² BROWNSWORD, Roger. From Erehwon to AlphaGo, for the sake of human dignity, should we destroy the machines. *Law, Innovation and Technology*, vol. 9, n. 1, 2017, p. 130.

³³ ESPOSITO, Fabrizio; CATHOUD FERREIRA, Thaís Maciel. Addictive design as an unfair commercial practice, the case of hyper engaging dark patterns. *European Journal of Risk Regulation*, marzo/2024, p. 10.

³⁴ ESPOSITO, Fabrizio; CATHOUD FERREIRA, Thaís Maciel. Addictive design as an unfair commercial practice, the case of hyper engaging dark patterns. *European Journal of Risk Regulation*, marzo/2024, p. 14.

³⁵ EYAL, Nir. *Hooked: How to build habit forming products*. Nueva York: Portfolio Penguin, 2014, p. 7.

complejidad técnica de los modelos de aprendizaje profundo y por la protección jurídica del secreto comercial.³⁶ Tal opacidad no es un efecto colateral indeseable, sino un rasgo constitutivo del modelo de negocio. De hecho, la ventaja competitiva de la plataforma reside precisamente en la inescrutabilidad de sus algoritmos. El resultado es una *probatio diabolica*: el consumidor que pretenda impugnar la validez de un contrato celebrado bajo la influencia de un *hypernudge* deberá demostrar, en primer lugar, que su entorno decisional fue manipulado algorítmicamente; en segundo lugar, que dicha manipulación explotó una vulnerabilidad cognitiva específica; y en tercer lugar, que existe un nexo causal entre la manipulación y la decisión adoptada. Cada uno de estos extremos presupone el acceso a información que se encuentra bajo el control exclusivo de la contraparte y que esta tiene incentivos económicos poderosos para no revelar.

En el plano regulatorio, la respuesta más ambiciosa hasta la fecha, la Ley de Inteligencia Artificial de la Unión Europea (Reglamento 2024/1689), reconoce explícitamente la peligrosidad de tales técnicas al prohibir, en su artículo 5, las prácticas de IA que utilicen técnicas subliminales que operen más allá de la consciencia del individuo o que exploten vulnerabilidades de grupos específicos para distorsionar materialmente su comportamiento. A pesar de ello, un análisis dogmático riguroso revela una deficiencia estructural en la configuración de dicha prohibición. El legislador europeo condiciona su aplicación a la prueba de un perjuicio significativo para la persona afectada.³⁷ Tal umbral probatorio convierte la prohibición en un mecanismo de eficacia limitada frente a la sutileza del *hypernudge*, cuyas técnicas más insidiosas, aquellas que operan mediante la captura hegemónica de la atención y la explotación sistemática de vulnerabilidades psicosociales, raramente provocan un daño agudo e inmediato acreditable bajo los estándares tradicionales de la responsabilidad civil. Más bien, generan efectos acumulativos y difusos sobre la autonomía individual.³⁸

En complemento a este marco, la Ley de Servicios Digitales (DSA, Reglamento 2022/2065) introduce lo que se ha denominado una doble arquitectura de elección: una arquitectura interna, diseñada por la plataforma para maximizar el *engagement*, y una arquitectura externa, impuesta por el legislador, que obliga a incluir opciones de control para el usuario, como la posibilidad de acceder a sistemas de recomendación no basados

³⁶ PASQUALE, Frank. *The black box society: the secret algorithms that control money and information*. Cambridge, MA: Harvard University Press, 2015, p. 6.

³⁷ CORNEJO PLAZA, María Isabel. Análisis de la noción de las técnicas subliminales contenida en el artículo 5 (1) (a, b y c) del Reglamento de Inteligencia Artificial de la Unión Europea. *Revista Justicia Derecho*, n. temático, 2025, p. 8.

³⁸ SUSSER, Daniel; ROESSLER, Beate; NISSENBAUM, Helen. Technology, autonomy, and manipulation. *Internet Policy Review*, vol. 8, n. 2, 2019, p. 18.

en la elaboración de perfiles.³⁹ Sin embargo, dicho enfoque adolece de un individualismo metodológico que confía en que la simple puesta a disposición de opciones de desactivación (*opt-out*) resolverá problemas que son estructurales y poblacionales. La evidencia empírica sugiere que la mera existencia de opciones alternativas no garantiza su uso efectivo cuando el diseño predominante de la plataforma sigue orientado a la maximización de la interacción adictiva.⁴⁰ La transparencia horizontal, es decir, aquella dirigida del usuario hacia la plataforma, permanece radicalmente asimétrica frente a la transparencia inversa, entendida como la que va de la plataforma hacia el usuario, lo que exige la intervención de mecanismos de supervisión vertical por parte de autoridades reguladoras con capacidad técnica para auditar los algoritmos.⁴¹

La incapacidad de los instrumentos tradicionales para contener eficazmente las prácticas manipulativas basadas en *hypernudges* revela la existencia de una patología normativa que no puede ser tratada únicamente mediante ajustes técnicos al marco vigente. La magnitud del fenómeno requiere desplazar el análisis hacia una dimensión más profunda del ordenamiento: la esfera de los derechos fundamentales. Desde allí es posible identificar un fundamento axiológico robusto que permita reconstruir las bases de la protección jurídica de la autodeterminación individual en entornos digitales.

4. La libertad cognitiva como fundamento constitucional del derecho a no ser manipulado

La insuficiencia demostrada de los remedios del derecho privado y del derecho regulatorio sectorial para contener la manipulación algorítmica obliga a elevar el análisis hacia el plano constitucional, donde es posible identificar un fundamento axiológico más robusto para la protección de la autonomía en entornos digitales. La tesis que aquí se sostiene es que la intervención mental indirecta operada por el *hypernudge*, entendida como la modulación algorítmica de los procesos de formación de la preferencia sin contacto físico con el cuerpo ni con el sistema nervioso del sujeto, constituye una afectación de la libertad cognitiva y la integridad psíquica que exige el reconocimiento

³⁹ HAKKARAINEN, Jenni; SAVOLAINEN, Laura. Individual choice, collective effects, recommender systems, law by design, and the DSA's double choice architecture. *Information, Communication & Society*, diciembre/2025, p. 5.

⁴⁰ SÖDERLUND, Kasia; ENGSTRÖM, Emma; HARESAMUDRAM, Kashyap; LARSSON, Stefan; STRIMLING, Pontus. Regulating high reach AI, on transparency directions in the Digital Services Act. *Internet Policy Review*, vol. 13, n. 1, 2024, p. 12.

⁴¹ SÖDERLUND, Kasia; ENGSTRÖM, Emma; HARESAMUDRAM, Kashyap; LARSSON, Stefan; STRIMLING, Pontus. Regulating high reach AI, on transparency directions in the Digital Services Act. *Internet Policy Review*, vol. 13, n. 1, 2024, p. 15.

de un derecho fundamental a no ser manipulado, el cual debe operar como base axiológica para la reinterpretación de las normas de derecho privado.

El punto de partida de la presente construcción dogmática se encuentra en la evolución del concepto de integridad personal. La tradición constitucional liberal configuró la protección de la persona en torno a la integridad física, concebida como un límite absoluto frente al poder del Estado. La prohibición de la tortura, las garantías del debido proceso penal y el *habeas corpus* constituyen manifestaciones de dicha protección en su dimensión corporal. Sin embargo, el desarrollo de las neurociencias y de las tecnologías de la información ha puesto de manifiesto que la autonomía humana puede ser vulnerada sin intervención física alguna, mediante la manipulación de los procesos cognitivos y afectivos que subyacen a la toma de decisiones.⁴² La noción de neuroderechos surge precisamente como respuesta a ese vacío, designando un conjunto de derechos fundamentales destinados a proteger la privacidad mental, la continuidad psicológica, la integridad mental y la libertad cognitiva frente al acceso tecnológico al cerebro y a los procesos mentales.⁴³

En este marco, la libertad cognitiva se define como el derecho de los individuos a controlar sus propios procesos mentales, a formar sus pensamientos y preferencias sin interferencia externa y a no sufrir alteraciones tecnológicas de su conciencia sin su consentimiento.⁴⁴ Esta noción trasciende la libertad de pensamiento clásica, tradicionalmente entendida como libertad de expresión y de conciencia religiosa, ya que protege no solo el resultado del pensamiento (la opinión formada), sino también el proceso mismo de su formación.⁴⁵ La distinción es crucial: mientras la libertad de expresión ampara el derecho a manifestar las propias convicciones, la libertad cognitiva salvaguarda el derecho a que esas convicciones se formen en condiciones de autodeterminación, libres de interferencias que operen por debajo del umbral de la conciencia.

La relevancia de esta distinción para el análisis del *hypernudge* se hace especialmente patente al examinar la naturaleza específica de la interferencia que produce. A diferencia

⁴² IENCA, Marcello; ANDORNO, Roberto. Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy*, vol. 13, n. 5, 2017, p. 5.

⁴³ IENCA, Marcello; ANDORNO, Roberto. Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy*, vol. 13, n. 5, 2017, p. 11.

⁴⁴ BUBLITZ, Jan Christoph. My mind is mine!? Cognitive liberty as a legal concept. En HILDT, Elisabeth y FRANKE, Andreas G. (Eds.). *Cognitive enhancement: an interdisciplinary perspective*. Dordrecht: Springer, 2013, p. 238.

⁴⁵ FARAONI, Stefano. Persuasive technology and computational manipulation, hypernudging out of mental self determination. *Frontiers in Artificial Intelligence*, vol. 6, 2023, art. 1216340, p. 4.

de las técnicas de persuasión legítima, que operan mediante argumentos dirigidos a la razón del destinatario y que este puede evaluar, aceptar o rechazar conscientemente, la manipulación digital se caracteriza por su naturaleza aracional. No apela a las facultades deliberativas del sujeto, sino que actúa sobre sesgos, emociones y respuestas automáticas, eludiendo las defensas cognitivas mediante la explotación del Sistema 1.⁴⁶ Dicha influencia aracional, cuando además es intencional, opaca y asimétrica en sus resultados, es decir, cuando produce un beneficio para el manipulador a expensas del manipulado, satisface los requisitos que la filosofía moral y la ética de la inteligencia artificial han identificado como constitutivos de la manipulación en sentido estricto.⁴⁷

La cuestión relativa a si esta forma de manipulación alcanza el umbral de una vulneración de derechos fundamentales ha sido abordada desde dos perspectivas complementarias. Una primera, anclada en la tradición kantiana, sostiene que la manipulación es intrínsecamente contraria a la dignidad humana porque trata al sujeto como un medio y no como un fin en sí mismo. En efecto, lo convierte en un recurso conductual optimizable. Desde esta lectura, el *hypernudge* viola el imperativo categórico al instrumentalizar la psiquis del usuario para la consecución de fines comerciales que no ha elegido ni puede siquiera identificar.⁴⁸ Una segunda perspectiva, de corte más analítico, distingue dos dimensiones de la autonomía que el *hypernudge* puede afectar de manera diferenciada: por un lado, la autonomía basada en la agencia del yo (*self-agency*), que se ve amenazada cuando el proceso deliberativo es puenteado; por otro, la autonomía basada en la autocracia (*autocracy*), que resulta vulnerada cuando la intervención induce acciones que contradicen los valores y preferencias de segundo orden del individuo.⁴⁹ Esta distinción permite un diagnóstico más preciso del daño. No todo *nudge* algorítmico implica necesariamente una violación de la autonomía, pero aquel que explota vulnerabilidades cognitivas para inducir decisiones contrarias a los intereses reflexivos del sujeto constituye una interferencia ilegítima en ambas dimensiones.

El reconocimiento constitucional de la libertad cognitiva como derecho fundamental no puede considerarse una especulación doctrinaria. Chile se convirtió, en 2021, en el primer país del mundo en consagrar constitucionalmente la protección de la integridad mental frente a las neurotecnologías, mediante la reforma del artículo 19 N° 1 de su

⁴⁶ IENCA, Marcello. On artificial intelligence and manipulation. *Topoi*, vol. 42, 2023, p. 836.

⁴⁷ IENCA, Marcello. On artificial intelligence and manipulation. *Topoi*, vol. 42, 2023, p. 838.

⁴⁸ TRZASKOWSKI, Jan. Manipulation by design. *Electronic Markets*, vol. 34, n. 14, febrero/2024, p. 5.

⁴⁹ CALBOLI, Stefano. Autonomy and AI nudges, distinguishing concepts and highlighting AI's advantages. *Philosophy & Technology*, vol. 38, n. 116, 2025, p. 8.

Constitución Política. Dicha reforma incorporó expresamente la protección de la integridad mental y estableció que el desarrollo científico y tecnológico debe estar al servicio de las personas, respetando la vida y la integridad física y psíquica. Aunque esta innovación normativa está dirigida primariamente a las neurotecnologías directas, como las interfaces cerebro-computadora o la estimulación transcraneal, sienta un precedente axiológico cuya *ratio decidendi* puede extenderse a las neurotecnologías indirectas, es decir, algoritmos que, sin tocar el cuerpo, modulan los procesos de toma de decisiones mediante la configuración del entorno informativo.⁵⁰

La extensión de la protección constitucional a la manipulación algorítmica indirecta encuentra, además, un sólido fundamento en la propia lógica de los derechos fundamentales como mandatos de optimización. Si se acepta, con la dogmática de los principios, que los derechos fundamentales no son reglas de aplicación binaria sino normas que exigen la máxima realización posible en función de las circunstancias fácticas y jurídicas,⁵¹ entonces la protección de la integridad mental no puede limitarse exclusivamente a las interferencias directas sobre el sistema nervioso. Debe abarcar también toda forma de modulación tecnológica de los procesos cognitivos que alcance un grado de intensidad suficiente para comprometer la autodeterminación del sujeto. En esta línea, el modelo de ponderación y optimización del razonamiento autónomo proporciona la infraestructura filosófica necesaria para sostener que la manipulación algorítmica es ilegítima precisamente porque puentea el proceso de autovinculación racional mediante el cual el sujeto adopta las normas que rigen su conducta.⁵²

A partir de esta base constitucional, puede identificarse una doble función dogmática. En primer lugar, se configura como límite externo a la autonomía privada. La libertad contractual, entendida como derecho fundamental de configuración legal, encuentra en la libertad cognitiva un límite material que impide su utilización como instrumento de dominación algorítmica. El principio de *pacta sunt servanda* no puede invocarse para dotar de fuerza obligatoria a una declaración de voluntad cuyo proceso de formación ha sido modulado externamente mediante inferencias predictivas que el sujeto no puede comprender ni resistir. En segundo lugar, esta base constitucional actúa como mandato de protección dirigido tanto al legislador como a los jueces. El Estado no solo debe abstenerse de manipular cognitivamente a los ciudadanos, sino que tiene además el

⁵⁰ COHEN, Julie E. What privacy is for. *Harvard Law Review*, vol. 126, 2012, p. 1912.

⁵¹ ALEXY, Robert. *A theory of constitutional rights*. Trad. Julian Rivers. Oxford: Oxford University Press, 2002, p. 47.

⁵² SIECKMANN, Jan R. *Autonomy and law*. Law and Philosophy Library, vol. 150. Cham: Springer Nature Switzerland, 2025, p. 95.

deber positivo de protegerlos frente a la manipulación ejercida por actores privados que ostentan un poder *de facto* sobre los procesos decisionales. Tal deber de protección se proyecta, a través de la eficacia horizontal de los derechos fundamentales (*Drittwirkung*), sobre las relaciones contractuales entre plataformas y usuarios, lo que exige una reinterpretación de las normas de derecho privado que incorpore la tutela de la integridad cognitiva como criterio hermenéutico rector.

La propuesta de reconocer un derecho a no ser manipulado no implica, conviene precisarlo, una prohibición absoluta de toda forma de influencia algorítmica sobre las decisiones del usuario. La inevitabilidad de la arquitectura de decisión, en tanto toda presentación de opciones influye de algún modo en la elección, hace inviable y conceptualmente erróneo un régimen de neutralidad total.⁵³ Lo que prohíbe el derecho a no ser manipulado es la explotación deliberada e intencional de vulnerabilidades cognitivas mediante técnicas opacas y personalizadas que produzcan una asimetría de resultados injustificada a favor del arquitecto de la decisión. La línea divisoria entre la influencia legítima y la manipulación ilícita se traza, en este modelo, mediante el criterio de la lealtad. Es legítima la arquitectura de decisión que, aun influyendo en la conducta, respeta la posibilidad de deliberación autónoma del sujeto. Es ilegítima, en cambio, aquella que, por su diseño, suprime esa posibilidad con el fin de maximizar un beneficio unilateral.

La identificación de la libertad cognitiva como derecho fundamental vulnerado por las arquitecturas de decisión algorítmica abre la puerta a una reconstrucción dogmática del derecho privado que incorpore dicha libertad como límite estructurante. En función de ello, corresponde explorar ahora cómo puede traducirse ese fundamento constitucional en un estándar operativo eficaz dentro del tráfico jurídico, para lo cual resulta clave reconfigurar la cláusula general de buena fe como fuente del deber de lealtad algorítmica y eje de un régimen de responsabilidad por diseño.

5. La buena fe objetiva como fundamento de la lealtad algorítmica: hacia un régimen de responsabilidad por diseño

La identificación de la libertad cognitiva como fundamento constitucional de la protección frente a la manipulación algorítmica permite ahora abordar la cuestión central de la investigación: la construcción de una propuesta dogmática integral que,

⁵³ SUNSTEIN, Cass R. Nudging and choice architecture, ethical considerations. *Yale Journal on Regulation*, vol. 32, 2015, p. 430.

operando dentro del sistema de derecho privado, traduzca el mandato constitucional en un estándar operativo capaz de disciplinar la conducta de las plataformas digitales en la configuración de sus arquitecturas de decisión. El instrumento dogmático idóneo para esa tarea se encuentra, según se argumenta, en la cláusula general de buena fe objetiva, cuya reinterpretación extensiva permite cristalizar un nuevo deber de lealtad algorítmica que supere las insuficiencias del paradigma informativo y establezca un régimen de responsabilidad por diseño.

La buena fe objetiva constituye, en la tradición del derecho continental, el principio vertebrador de las relaciones obligacionales. A diferencia de la buena fe subjetiva, que designa un estado psicológico de creencia o ignorancia, la buena fe objetiva opera como un estándar normativo de conducta que impone a las partes de una relación jurídica el deber de comportarse con lealtad, honestidad y consideración razonable de los intereses legítimos de la contraparte.⁵⁴ Este principio, consagrado con diversas formulaciones en los códigos civiles de la tradición continental, desde el §242 del BGB alemán hasta el artículo 1546 del Código Civil chileno, pasando por el artículo 7.1 del Código Civil español y el artículo 1.7 de los Principios UNIDROIT, ha demostrado históricamente una extraordinaria capacidad de adaptación a nuevas realidades sociales. Funciona así como una válvula de escape que permite al sistema jurídico responder a situaciones no previstas por el legislador sin necesidad de una reforma legislativa formal.⁵⁵

La potencialidad de la buena fe objetiva para disciplinar la conducta algorítmica de las plataformas digitales se funda en tres características dogmáticas esenciales. En primer lugar, su carácter de cláusula general le otorga una apertura semántica suficiente para abarcar modalidades de deslealtad no tipificadas expresamente en el catálogo de vicios del consentimiento o de prácticas desleales. En segundo lugar, su operatividad en todas las fases de la relación contractual, ya sea durante la formación, la ejecución o la extinción del contrato, permite capturar tanto la manipulación precontractual del proceso decisonal como la explotación continuada de las vulnerabilidades del usuario mientras el vínculo permanece vigente. En tercer lugar, su eficacia como fuente autónoma de deberes de conducta, independiente de la voluntad de las partes, la sustrae de la lógica contractualista que vincula el nacimiento de las obligaciones exclusivamente al acuerdo libre.⁵⁶

⁵⁴ WIEACKER, Franz. *Privatrechtsgeschichte der Neuzeit*. 2. ed. Gotinga: Vandenhoeck & Ruprecht, 1967. Trad. consultada: *Historia del derecho privado de la Edad Moderna*. Madrid: Aguilar, 1977, p. 440.

⁵⁵ ZIMMERMANN, Reinhard; WHITTAKER, Simon (Eds.). *Good faith in European contract law*. Cambridge: Cambridge University Press, 2000, p. 30.

⁵⁶ LARENZ, Karl. *Lehrbuch des Schuldrechts*, vol. 1: *Allgemeiner Teil*. 14. ed. Múnich: C. H. Beck, 1987, p. 128.

La reinterpretación extensiva que se propone consiste en derivar de la buena fe objetiva un deber específico de lealtad algorítmica, entendido como la obligación de la plataforma digital de diseñar y operar sus arquitecturas de decisión de modo que no exploten las vulnerabilidades cognitivas, emocionales o situacionales de los usuarios para obtener ventajas que no habrían sido concedidas en condiciones de deliberación autónoma. Este deber se articula en torno a tres dimensiones complementarias: un contenido prohibitivo, unos criterios para identificar el ilícito y un régimen de consecuencias jurídicas.

En relación con el contenido prohibitivo, el deber de lealtad algorítmica proscribe la explotación de vulnerabilidades psicosociales del usuario mediante el diseño de la interfaz y la configuración algorítmica del entorno decisional. Esta prohibición se sustenta en la noción de explotación conductual desarrollada por la doctrina reciente, entendida como la toma de ventaja injusta de los sesgos cognitivos o afectivos del consumidor para obtener un beneficio que no se produciría en una transacción equitativa.⁵⁷ La explotación, así definida, no requiere ni engaño explícito ni coacción física. Basta con que el diseño de la interfaz instrumentalice mecanismos psicológicos como el ciclo de la dopamina, la aversión a la pérdida, el efecto dotación o la sobrecarga cognitiva, con el fin de canalizar la conducta del usuario en una dirección que beneficia unilateralmente a la plataforma. La noción de vulnerabilidad relevante, para estos efectos, no se limita a la individual o patológica, sino que abarca también la situacional, relacional y dinámica, es decir, aquella que surge de la interacción entre el sujeto y el entorno digital y afecta de forma desproporcionada a los grupos en situación de desventaja socioeconómica.⁵⁸

Los criterios para identificar el *hypernudge* ilícito, entendido como aquel que transgrede el deber de lealtad algorítmica, pueden articularse mediante un test tripartito que opera como filtro dogmático. El primer criterio es la opacidad deliberada, que se configura cuando la plataforma diseña su sistema de manera que el usuario no pueda percibir, comprender ni resistir la influencia ejercida sobre su proceso decisional. Este criterio reformula la exigencia de transparencia, no como simple deber de información sobre el contenido contractual, sino como una obligación de inteligibilidad de la arquitectura de decisión, en consonancia con la propuesta de un principio de legalidad especial para *nudges* que ajuste los requisitos de promulgación y claridad a las realidades de la

⁵⁷ BRENNCKE, Martin. A theory of exploitation for consumer law, online choice architectures, dark patterns, and autonomy violations. *Journal of Consumer Policy*, vol. 47, 2024, p. 135.

⁵⁸ HERZOG, Lisa; KELLMEYER, Philipp; WILD, Verina. Digital behavioral technology, vulnerability and justice, towards an integrated approach. *Review of Social Economy*, vol. 80, n. 1, 2022, p. 16.

psicología conductual.⁵⁹ El segundo criterio es la personalización explotadora. Tiene lugar cuando la plataforma emplea datos derivados del perfilamiento algorítmico para dirigir estímulos específicos a un individuo, aprovechando sus vulnerabilidades particulares y generando una influencia que, por su nivel de individualización, no podría ser neutralizada por un consumidor medio actuando con diligencia razonable. El tercer criterio es la asimetría de resultados, la cual se verifica cuando la arquitectura de decisión produce un beneficio sistemático para la plataforma, medido en tiempo de permanencia, volumen de datos extraídos o transacciones inducidas, y lo hace a costa de los intereses reflexivos del usuario, entendidos como aquellos que el propio sujeto endosaría tras un proceso de deliberación autónoma informada.

La aplicación combinada de los tres criterios mencionados permite distinguir, con el grado de rigor que exige la seguridad jurídica, entre la influencia legítima y la manipulación ilícita. Una publicidad personalizada que recomienda un producto coherente con los intereses declarados del usuario no satisface, en principio, el test tripartito, ya que no presenta opacidad deliberada en su carácter publicitario ni explota una vulnerabilidad cognitiva concreta. En cambio, el diseño de interfaces que incorporan mecanismos como el desplazamiento infinito (*infinite scroll*), la reproducción automática (*autoplay*), las notificaciones con urgencia simulada o procesos de cancelación intrincados de manera deliberada, todos ellos sistematizados por la doctrina bajo la categoría de patrones oscuros de hiperinteracción,⁶⁰ sí cumplen los tres criterios. Dichas técnicas operan de forma opaca sobre el Sistema 1, explotan la vulnerabilidad atencional del usuario y generan una asimetría de resultados fácilmente cuantificable. En este contexto, el principio de igualdad de clics (*Equal Clicks Principle*), que propone que abandonar un servicio no debe requerir más esfuerzo que el necesario para unirse a él, ofrece un parámetro práctico y auditable para detectar dicha asimetría de diseño.⁶¹

En lo relativo a las consecuencias jurídicas, el incumplimiento del deber de lealtad algorítmica debe generar efectos tanto en el plano de la validez del negocio como en el de la responsabilidad civil. En el primer plano, la manipulación algorítmica que supere el umbral fijado por el test tripartito debe considerarse un vicio del consentimiento de carácter atípico y estructural, cuya concurrencia justifique la anulabilidad del acto jurídico celebrado bajo su influencia. Para alcanzar esta consecuencia no es

⁵⁹ ZEILSTRA, Rebecca. Formal legality for nudges. *Ratio Juris*, vol. 38, n. 1, marzo/2025, p. 52.

⁶⁰ ESPOSITO, Fabrizio; CATHOUD FERREIRA, Thaís Maciel. Addictive design as an unfair commercial practice, the case of hyper engaging dark patterns. *European Journal of Risk Regulation*, marzo/2024, p. 18.

⁶¹ MILLS, Stuart; WHITTLE, Richard; AHMED, Rafi; WALSH, Tom; WESSEL, Martin. Dark patterns and sludge audits, an integrated approach. *Behavioural Public Policy*, vol. 10, 2026, p. 180.

imprescindible crear legislativamente un nuevo vicio *ex novo*, ya que puede lograrse mediante la interpretación extensiva del dolo, entendido no como engaño sobre el objeto contractual, sino como maquinación orientada a captar la voluntad ajena. También resulta plausible subsumir la conducta en la categoría de contravención de la buena fe como límite material a la autonomía privada. En el segundo plano, el diseño manipulativo debe dar lugar a responsabilidad civil por los daños derivados de la explotación de vulnerabilidades cognitivas. En este punto, se impone una inversión de la carga probatoria que obligue al proveedor a demostrar que su arquitectura de decisión cumple con los estándares de lealtad algorítmica. Tal inversión se justifica por la asimetría informativa estructural, que impide al consumidor acceder a la lógica operativa del algoritmo, así como por el principio de facilidad probatoria, que exige atribuir la carga a quien se encuentra en mejores condiciones de aportar la evidencia.

Desde esta perspectiva, el régimen aquí propuesto supera el paradigma informativo, centrado en la simple provisión de datos al consumidor, e instaura en su lugar un paradigma de responsabilidad por diseño, orientado a intervenir de forma anticipada en la etapa de concepción y configuración de las arquitecturas decisionales. Este enfoque guarda coherencia con la orientación adoptada en las regulaciones europeas más recientes, que han transitado desde la sanción *ex post* de conductas hacia la intervención *ex ante* sobre el diseño técnico, como demuestran las obligaciones de evaluación de riesgos sistémicos y auditorías que la DSA impone a las plataformas de muy gran tamaño,⁶² así como la prohibición de técnicas manipulativas contemplada en el AI Act.⁶³ La buena fe objetiva, interpretada a través del prisma del deber de lealtad algorítmica, ofrece el fundamento dogmático necesario para proyectar tales exigencias regulatorias sobre las relaciones de derecho privado, cerrando así la brecha entre la protección constitucional de la libertad cognitiva y la tutela efectiva del consentimiento en el tráfico jurídico digital.

Cabe advertir que la propuesta no aspira a fijar de manera rígida y definitiva el contenido del deber de lealtad algorítmica, sino más bien a ofrecer un marco principiológico cuya concreción corresponderá a la jurisprudencia y a la doctrina en función del desarrollo tecnológico. La cláusula general de buena fe ha demostrado a lo largo de más de un siglo su capacidad para adaptarse a contextos sociales y económicos cambiantes sin perder

⁶² HAKKARAINEN, Jenni; SAVOLAINEN, Laura. Individual choice, collective effects, recommender systems, law by design, and the DSA's double choice architecture. *Information, Communication & Society*, diciembre/2025, p. 12.

⁶³ CORNEJO PLAZA, María Isabel. Análisis de la noción de las técnicas subliminales contenida en el artículo 5 (1) (a, b y c) del Reglamento de Inteligencia Artificial de la Unión Europea. *Revista Justicia Derecho*, n. temático, 2025, p. 15.

coherencia dogmática. La era algorítmica no exige el abandono de esa tradición, sino su actualización creativa al servicio de la tutela de la persona en un entorno donde la formación de la voluntad, presupuesto indispensable de la autonomía privada, se ha convertido en objeto de una ingeniería de precisión sin precedentes históricos.

6. Conclusiones

La arquitectura jurídica clásica, erigida sobre el ideal de un sujeto racional y soberano, enfrenta hoy una disrupción ontológica que parece desbordar sus presupuestos fundacionales. Se observa que no se trata de una simple aceleración tecnológica, sino de una mutación en el modo en que se gobierna la conducta, desplazando el mandato imperativo por una modulación sutil y algorítmica de la preferencia. Dicha realidad invita a reflexionar sobre si las categorías decimonónicas conservan su idoneidad para aprehender un fenómeno que opera, precisamente, en el estrato previo a la consciencia, allí donde la voluntad comienza a gestarse bajo el influjo de arquitecturas digitales de decisión.

En este escenario, la noción de *hypernudge* se consolida como una categoría que trasciende la estática del estímulo conductual tradicional para configurarse como una técnica de gubernamentalidad dinámica y personalizada. La contribución teórica más significativa radica en identificar una asimetría epistémica radical: el desequilibrio ya no versa únicamente sobre las cualidades de la prestación, sino sobre la propia psiquis del contratante. Al caracterizar esta influencia como una coacción arquitectónica o una heteronomía programada, se hace visible una patología estructural del consentimiento que el dolo o el error, en sus acepciones clásicas, no logran capturar con la precisión requerida por la era del aprendizaje automático.

El traslado de estos hallazgos al ámbito jurídico sugiere la necesidad de revitalizar la cláusula de buena fe objetiva como fuente de un estándar de lealtad algorítmica. Dicho estándar permitiría operacionalizar un test compuesto por tres criterios: opacidad deliberada, personalización explotadora y asimetría de resultados, con el fin de distinguir la persuasión comercial legítima de la manipulación ilícita. En esta línea, la protección de la voluntad se reconfigura dentro de un paradigma de responsabilidad por diseño, en el cual la validez del negocio jurídico depende del respeto técnico y ético del proceso decisional, lo que implica que las interfaces deben dejar de funcionar como instrumentos de extracción para constituirse en espacios que garanticen una autonomía efectiva.

No obstante, se reconoce que el debate apenas comienza, motivo por el cual resulta imperativo expandir la agenda investigativa hacia la consolidación dogmática de los neuroderechos, específicamente la libertad cognitiva. Queda pendiente profundizar en la superación del individualismo metodológico que permea las soluciones regulatorias actuales, explorando cómo la lealtad algorítmica puede abordar no solo agravios individuales, sino también daños colectivos y sistémicos. Parece necesario, asimismo, evaluar la viabilidad de modelos de supervisión vertical y auditoría algorítmica que permitan penetrar la opacidad de la “caja negra”, subordinando el secreto comercial a la transparencia exigida por el Estado de Derecho.

Por consiguiente, la defensa de la autonomía en la era digital sitúa a la ciencia jurídica ante una encrucijada sobre la esencia misma de la agencia humana. El derecho privado no debería resignarse a validar ficciones de consentimiento que ocultan una suerte de servidumbre algorítmica; su función crítica debe ser la de un límite axiológico que preserve la capacidad de autovinculación racional del sujeto. Mantener viva la posibilidad de que el individuo sea el autor de sus propias razones se perfila como la tarea más urgente de la civilística contemporánea, asegurando que la arquitectura de decisión sea un vehículo para el florecimiento de la libertad y no un mecanismo de ingeniería moral invisible.

Declaración sobre el uso de IA

Se usó el modelo de lenguaje de gran escala GPT-5.1 de OpenAI, con el propósito de identificar y corregir errores ortográficos y de redacción. El comando aplicado fue: “detecta y corrige todos los errores de redacción y ortografía”. Una vez obtenida la respuesta del sistema, se realizó una revisión del texto obtenido para asegurar que el contenido conservase tanto el tono como la intención del borrador original.

Referencias bibliográficas

ALEXY, Robert. *A theory of constitutional rights*. Trad. Julian Rivers. Oxford: Oxford University Press, 2002.

BEN SHAHAR, Omri; SCHNEIDER, Carl E. *More than you wanted to know: the failure of mandated disclosure*. Princeton: Princeton University Press, 2014.

BRENNCKE, Martin. A theory of exploitation for consumer law, online choice architectures, dark patterns, and autonomy violations. *Journal of Consumer Policy*, vol. 47, 2024.

BROWNSWORD, Roger. From Erewhon to AlphaGo, for the sake of human dignity, should we destroy the machines. *Law, Innovation and Technology*, vol. 9, n. 1, 2017.

BUBLITZ, Jan Christoph. My mind is mine!? Cognitive liberty as a legal concept. En HILDT, Elisabeth y FRANKE, Andreas G. (Eds.). *Cognitive enhancement: an interdisciplinary perspective*. Dordrecht: Springer, 2013.

CALBOLI, Stefano. Autonomy and AI nudges, distinguishing concepts and highlighting AI's advantages. *Philosophy & Technology*, vol. 38, n. 116, 2025.

CÁMARA LAPUENTE, Sergio. El control de las cláusulas abusivas sobre elementos esenciales del contrato. *Cuadernos de Derecho Transnacional*, vol. 3, n. 1, 2011.

COHEN, Julie E. What privacy is for. *Harvard Law Review*, vol. 126, 2012.

COHEN, Julie E. *Between truth and power: the legal constructions of informational capitalism*. Oxford: Oxford University Press, 2019.

CORNEJO PLAZA, María Isabel. Análisis de la noción de las técnicas subliminales contenida en el artículo 5 (1) (a, b y c) del Reglamento de Inteligencia Artificial de la Unión Europea. *Revista Justicia Derecho*, n. temático, 2025.

DELEUZE, Gilles. Postscript on the societies of control. *October*, n. 59, 1992.

DÍEZ PICAZO, Luis. *Fundamentos del derecho civil patrimonial*, vol. 1: Introducción, teoría del contrato. 6. ed. Madrid: Thomson Civitas, 2007.

ESPOSITO, Fabrizio; CATHOUD FERREIRA, Thaís Maciel. Addictive design as an unfair commercial practice, the case of hyper engaging dark patterns. *European Journal of Risk Regulation*, marzo/2024.

EYAL, Nir. *Hooked: How to build habit forming products*. Nueva York: Portfolio Penguin, 2014.

FARAONI, Stefano. Persuasive technology and computational manipulation, hypernudging out of mental self determination. *Frontiers in Artificial Intelligence*, vol. 6, 2023, art. 1216340.

FLUME, Werner. *Allgemeiner Teil des Bürgerlichen Rechts*, vol. 2: *Das Rechtsgeschäft*. 4. ed. Berlín: Springer, 1992. Trad. parcial consultada: *El negocio jurídico*. Madrid: Fundación Cultural del Notariado, 1998.

FRANK, Felipe. A consolidação da autonomia da vontade como cânone do direito privado moderno, o caso do Code Napoleônico de 1804. *Civilistica.com*, a. 11, n. 1, 2022.

FULLER, Lon L. *The morality of law*. Ed. rev. New Haven: Yale University Press, 1969.

HAKKARAINEN, Jenni; SAVOLAINEN, Laura. Individual choice, collective effects, recommender systems, law by design, and the DSA's double choice architecture. *Information, Communication & Society*, diciembre/2025.

HERZOG, Lisa; KELLMEYER, Philipp; WILD, Verina. Digital behavioral technology, vulnerability and justice, towards an integrated approach. *Review of Social Economy*, vol. 80, n. 1, 2022.

IENCA, Marcello. On artificial intelligence and manipulation. *Topoi*, vol. 42, 2023.

IENCA, Marcello; ANDORNO, Roberto. Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy*, vol. 13, n. 5, 2017.

KAHNEMAN, Daniel. *Thinking, fast and slow*. Nueva York: Farrar, Straus and Giroux, 2011.

LARENZ, Karl. *Lehrbuch des Schuldrechts*, vol. 1: *Allgemeiner Teil*. 14. ed. Múnich: C. H. Beck, 1987.

LESSIG, Lawrence. *Code: version 2.0*. Nueva York: Basic Books, 2006.

MILLS, Stuart; WHITTLE, Richard; AHMED, Rafi; WALSH, Tom; WESSEL, Martin. Dark patterns and sludge audits, an integrated approach. *Behavioural Public Policy*, vol. 10, 2026.

MOROZOVAITE, Viktorija. Hypernudging in the changing European regulatory landscape for digital markets. *Policy & Internet*, vol. 15, n. 1, 2023.

PASQUALE, Frank. *The black box society: the secret algorithms that control money and information*. Cambridge, MA: Harvard University Press, 2015.

- ROUVROY, Antoinette; BERNS, Thomas. Gouvernamentalité algorithmique et perspectives d'émancipation, le disparate come condition d'individuation par la relation. *Réseaux*, n. 177, 2013.
- SIECKMANN, Jan R. *Autonomy and law*. Law and Philosophy Library, vol. 150. Cham: Springer Nature Switzerland, 2025.
- SÖDERLUND, Kasia; ENGSTRÖM, Emma; HARESAMUDRAM, Kashyap; LARSSON, Stefan; STRIMLING, Pontus. Regulating high reach AI, on transparency directions in the Digital Services Act. *Internet Policy Review*, vol. 13, n. 1, 2024.
- SOLOVE, Daniel J. Privacy self management and the consent dilemma. *Harvard Law Review*, vol. 126, 2013.
- SUNSTEIN, Cass R. Nudging and choice architecture, ethical considerations. *Yale Journal on Regulation*, vol. 32, 2015.
- SUNSTEIN, Cass R. (Ed.). *Behavioral law and economics*. Cambridge: Cambridge University Press, 2000.
- SUSSER, Daniel; ROESSLER, Beate; NISSENBAUM, Helen. Technology, autonomy, and manipulation. *Internet Policy Review*, vol. 8, n. 2, 2019.
- THALER, Richard H.; SUNSTEIN, Cass R. *Nudge: improving decisions about health, wealth, and happiness*. New Haven: Yale University Press, 2008.
- TRZASKOWSKI, Jan. Manipulation by design. *Electronic Markets*, vol. 34, n. 14, febrero/2024.
- TVERSKY, Amos; KAHNEMAN, Daniel. Judgment under uncertainty, heuristics and biases. *Science*, vol. 185, n. 4157, 1974.
- WIEACKER, Franz. *Privatrechtsgeschichte der Neuzeit*. 2. ed. Gotinga: Vandenhoeck & Ruprecht, 1967. Trad. consultada: *Historia del derecho privado de la Edad Moderna*. Madrid: Aguilar, 1977.
- YEUNG, Karen. Hypernudge, big data as a mode of regulation by design. *Information, Communication & Society*, vol. 20, n. 1, 2017.
- ZEILSTRA, Rebecca. Formal legality for nudges. *Ratio Juris*, vol. 38, n. 1, marzo/2025.
- ZIMMERMANN, Reinhard; WHITTAKER, Simon (Eds.). *Good faith in European contract law*. Cambridge: Cambridge University Press, 2000.
- ZUBOFF, Shoshana. *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. Nueva York: PublicAffairs, 2019.

Como citar:

RAMOS-ZAGA, Fernando. *Hypernudges* y consentimiento sustantivo: bases para un estándar jurídico de lealtad algorítmica en arquitecturas digitales de decisión. **Civilistica.com**, a. 15, n. 2, 2026. Disponível em: <https://civilistica.emnuvens.com.br/redc>. Data de acesso.



civilistica.com

Recebido em:

5.2.2026

Aprovado em:

12.6.2026