

Los sesgos naturales en la programación algorítmica: desafíos y dificultades para la aplicación de sistemas automatizados en el derecho

Gonzalo Renzo Villavicencio Rojas*
Juan Pavel Valdez-Rondinel**

Resumen.- Este artículo examina la inteligencia artificial jurídica y su choque con los sesgos del lenguaje legal. Analiza la automatización y el uso de datos, mostrando cómo pueden institucionalizar discriminaciones estructurales. Se advierte así el riesgo de que la IA se convierta en un instrumento tecnocrático que legitime injusticias bajo apariencia de neutralidad.

Palabras clave.- Inteligencia artificial - sesgo legal - sesgos cognitivos - lenguaje jurídico - sesgo algorítmico - interpretación legal - automatización discriminatoria - justicia tecnocrática - automatización jurídica - jurisprudencia e IA.

Abstract. - This article examines legal artificial intelligence and its clash with biases in legal language. It analyzes automation and data use, showing how they can institutionalize structural discrimination. It warns of the risk that AI becomes a technocratic tool legitimizing injustices under the guise of neutrality.

Keywords.- Artificial intelligence - legal bias - cognitive biases - legal language - algorithmic bias - legal interpretation - discriminatory automation - technocratic justice - legal automation - jurisprudence and AI.

* Estudiante del 7.º semestre de Derecho en la Universidad Peruana de Ciencias Aplicadas.

** Estudiante del 12.º semestre de Derecho en la Universidad Nacional de San Cristóbal de Huamanga. Miembro en Círculo académico Franco-Latinoamericano Lucet Universidad de la Sorbona y del Círculo de Derecho Informático CEDIM Universidad Nacional Mayor de San Marcos. Premio nacional de Mejor Orador del Torneo AYNI Bullard Falla & Ezcurra 2024, articulista en LP así como por la universidad Simon Fraser University.

I. Introducción

i. Contextualización del uso de IA en el derecho penal y otros ámbitos jurídicos

En los últimos años, los sistemas de inteligencia artificial (IA) han comenzado a desempeñar un papel cada vez más protagónico en diversas esferas de la vida social, y el ámbito jurídico no ha sido la excepción. La incorporación de algoritmos en los procesos legales ha propiciado una transformación significativa, especialmente en el derecho penal. **Desde modelos predictivos** – como *DEBATER* de IBM – hasta herramientas que evalúan el riesgo procesal de un imputado, la IA se perfila como un actor muy relevante en la toma de decisiones judiciales.

Sistemas como **COMPAS** en Estados Unidos, o aplicaciones más recientes centradas en la clasificación automática de demandas y precedentes, permiten vislumbrar un panorama donde las máquinas ya no solo asisten, sino que influyen directamente en el razonamiento judicial. Sin embargo, esta automatización jurídica convive con interrogantes fundamentales sobre la neutralidad de los algoritmos y la posibilidad de que decisiones aparentemente objetivas estén atravesadas por estructuras sociales, económicas y culturales profundamente arraigadas. Como señalan Boczkowski y Mitchelstein, los algoritmos –*lejos de ser entes autónomos o neutros*– están íntimamente ligados a quienes los diseñan¹, pues “los algoritmos y la agencia humana no son opuestos, sino que están interrelacionados en el sentido de que los humanos son tan parte del software como el código informático mecánico que se destaca en cualquier definición de algoritmos” (p. 48)². En este sentido, la IA no nace del vacío, sino que es el producto de la historia, la cultura y las formas particulares de lenguaje de quienes la crean.

Como advierte Eubanks, “a menudo tratamos la nueva tecnología como algo ajeno a la historia y al contexto, como el monolito de *2001: Odisea en el espacio*, que aparece de la nada, pero aterriza en un terreno vacío

¹ BOCZKOWSKI, P. Y MITCHELSTEIN, E. (2022). El entorno digital: Breve manual para entender cómo vivimos, aprendemos, trabajamos y pasamos el tiempo libre hoy (Ciudad Autónoma de Buenos Aires: Siglo XXI Editora Iberoamericana), pp. 48.

² BUCHER, T. (2018). *If...Then: Algorithmic Power and Politics* (New York: Oxford University Press)

y cambia todo; pero en realidad la tecnología surge de la cultura y vuelve a influir en ella”³. Así, los algoritmos que gobiernan muchas de las decisiones contemporáneas no solo reflejan reglas matemáticas, sino que reproducen las estructuras simbólicas, lingüísticas y políticas que habitan quienes los programan.

ii. Presentación de la tesis

En línea con estas perspectivas, el presente artículo sostiene que los sesgos no son simples errores accidentales o anomalías corregibles mediante ajustes técnicos, sino que constituyen consecuencias estructurales e inevitables derivadas del uso del lenguaje humano en los procesos de programación, decisión y argumentación algorítmica. Dichos sesgos emergen como herencias cognitivas, culturales y lingüísticas que acompañan inevitablemente a todo modelo automatizado entrenado con datos generados por seres humanos.

Desde una perspectiva filosófica, se advertía sobre los límites del lenguaje como medio para representar la realidad: la indeterminación semántica — esto es, la falta de correspondencia fija entre palabras y hechos — impide que el lenguaje sea una herramienta perfectamente transparente, eficaz y neutra⁴. Así, el lenguaje es inherentemente sesgado no solo porque puede ser usado con intencionalidad, sino porque es estructuralmente insuficiente para captar de forma exhaustiva la complejidad del mundo. En el campo algorítmico, esto implica que cualquier codificación basada en lenguaje humano — e incluso los algoritmos lo son — estará afectada, en su base misma, por ambigüedades, omisiones y marcos interpretativos específicos que se construyen bajo la intencionalidad del ser humano que se encuentra detrás del algoritmo.

Además de estos sesgos estructurales o semánticos, existen también sesgos intencionales o funcionales, vinculados a intereses sociales, económicos y políticos. Este sesgo se manifestó, por ejemplo, durante la campaña presidencial de Estados Unidos en 2016, donde la circulación de desinformación estuvo altamente concentrada en pequeños nichos

³ EUBANKS, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin’s Press)

⁴ LEITGEB, H. (2023). Ramsification and semantic indeterminacy. *The Review of Symbolic Logic*, Cambridge Press, pp. 900-950.

digitales⁵. Aunque solo el 6% de la información era falsa, el 80% de las noticias falsas fueron vistas por apenas el 1% de los usuarios de Twitter, demostrando que el problema no era tanto la cantidad de desinformación, sino su focalización y circulación estratégica. Esta dinámica evidencia que la información puede ser programada, dirigida y manipulada para responder a fines específicos, revelando que los algoritmos no solo reflejan sesgos pasivos, sino que pueden reproducir estructuras de poder y exclusión activamente.

En efecto, como señala Virginia Eubanks, la tecnología suele presentarse como algo novedoso, disruptivo y ajeno al contexto; sin embargo, “la tecnología surge de la cultura y vuelve a influir en ella”⁶. Es decir, la inteligencia artificial no es neutral: es una materialización algorítmica de las visiones, intereses y estructuras de quienes la crean y aplican. Por ello, programar un algoritmo es también inscribir en él una cierta visión del mundo, sus normas y sus exclusiones.

Estas prácticas tecnológicas no operan en el vacío. Como advierten Earl y Kimport⁷, los medios digitales se han vuelto inseparables de las formas de interacción social y activismo, permitiendo que las narrativas, las decisiones y los comportamientos circulen por canales tanto personales como automatizados. Esta mezcla de acción individual, red social y tecnología de datos genera nuevos escenarios donde la intención política, el diseño técnico y el lenguaje humano convergen para producir decisiones con efectos materiales sobre la vida de las personas.

Finalmente, Hidalgo, investigador de la Universidad de Toulouse, entre otros autores, sostienen que “la información es el producto final de las cosas que se producen con el conocimiento”⁸, ya sea un texto, un video o un bien material. Esta visión refuerza la idea de que todo

⁵ LAZER, D. et al. (2018). The Science of Fake News: Addressing Fake News Requires a Multidisciplinary Effort. *Science*, vol. 359, no. 6380.

⁶ EUBANKS, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. *St. Martin's Press*.

⁷ EARL, J. y KIMPORT, K. (2013). Digitally Enabled Social Change: Activism in the Internet Age. *MIT Press*.

⁸ RONEN, S.; GONÇALVES, B.; HU, K.; VESPIGNANI, A.; PINKER, S.; HIDALGO, C. (2014). “Links That Speak: The Global Language Network and Its Association with Global Fame,” *Proceedings of the National Academy of Sciences*, vol. 111, No. 52.

dato es una construcción social, no un reflejo objetivo de la realidad. La programación, en consecuencia, es un acto cargado de subjetividad, cultura y lenguaje, por lo que sus resultados —como los algoritmos usados en procesos judiciales— deben ser evaluados no solo desde su eficiencia, sino desde su capacidad para reproducir o cuestionar estructuras de desigualdad existentes —y, sobre todo, la tan buscada y a veces esquiva imparcialidad—.

En síntesis, este artículo propone explicar y replantear la forma en que se conciben los sesgos en sistemas de inteligencia artificial, especialmente en el ámbito judicial. En lugar de asumir que son fallas corregibles, se plantea que son parte constitutiva del modo en que los humanos —y, por tanto, las máquinas que programan— perciben, interpretan y actúan en el mundo. Este desplazamiento de perspectiva es indispensable para problematizar crítica y maduramente el uso de la IA en decisiones judiciales sensibles, como la determinación de la culpabilidad, la proporcionalidad de la pena o la valoración del riesgo procesal. Si bien la promesa de objetividad algorítmica ha ganado terreno en contextos de sobrecarga institucional o incertidumbre normativa, es urgente reconocer que la programación no elimina los prejuicios humanos; en cambio, los traduce a otro lenguaje, los codifica, y en no pocos casos, los amplifica.

iii. Recorrido del artículo

Para sustentar esta tesis, el artículo está compuesto por tres partes interrelacionadas: en primer lugar, se analizan los **sesgos cognitivos** y su anclaje en la psicología humana como mecanismos mentales de simplificación que, si bien útiles, pueden generar distorsiones sistemáticas en la toma de decisiones.

En segundo lugar, se examina el papel del lenguaje como vehículo principal del sesgo estructural, tanto en su dimensión natural —ambigua, polisémica, indeterminada— como jurídica, con su pretensión de objetividad, pero sin escapar a la interpretación.

En tercer lugar, se expone cómo dichos sesgos son trasladados, replicados e incluso amplificadas en los sistemas de inteligencia artificial, especialmente cuando se entrenan con datos generados por humanos. Finalmente, se ofrece una reflexión crítica sobre los límites y posibilidades de corregir o mitigar estos sesgos, proponiendo una mirada ética, normativa y estructural que supere la ilusión de neutralidad técnica.

II. Los sesgos cognitivos y su fundamento humano

i. Heurísticos y sesgos cognitivos: definición y relación

A. Definición

En la toma de decisiones humanas, los heurísticos se configuran como atajos mentales o estrategias cognitivas simplificadas, empleadas para emitir juicios con base en experiencias previas, sin necesidad de procesar exhaustivamente toda la información disponible. Como lo explica Nieva, un heurístico puede entenderse como una directriz general que orienta nuestras decisiones a partir del recuerdo de situaciones similares, replicando respuestas que resultaron eficaces en el pasado⁹.

Por su parte, los sesgos cognitivos representan “formas no razonadas en que se toman las decisiones, permitiendo hacer deducciones sin un alto nivel de esfuerzo, lo que podría desencadenar que la toma de decisiones sea tendenciosa o errada”¹⁰. Es decir, son distorsiones sistemáticas del pensamiento, producto de la aplicación no crítica de heurísticos en contextos que exigen mayor análisis o información.

B. Vinculación entre heurísticos y sesgos

A partir de los influyentes estudios de Tversky y Kahneman, se evidenció una relación intrínseca entre los heurísticos como mecanismos adaptativos de la cognición humana y los sesgos como consecuencias de su mal uso o aplicación automática¹¹. En palabras de Castro et al., cuando estos atajos mentales no conducen a una elección adecuada, dan lugar a sesgos cognitivos, generando decisiones basadas más en preconceptos internos y subjetivos que en razonamientos lógicos o normativos rigurosos¹².

⁹ NIEVA, J. (2018). Inteligencia artificial y proceso judicial. Marcial Pons.

¹⁰ CASTRO, A. et al. (2019). Nivel de sesgos cognitivos de representatividad y confirmación en estudiantes de Psicología de tres universidades del Bío-Bío. *Propósitos y Representaciones*, vol. 7, no. 2, pp. 210-239.

¹¹ TVERSKY, A. y KAHNEMAN, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, vol. 185, no. 4157, pp. 1124-1131.

¹² CASTRO, A. et al. (2019). Nivel de sesgos cognitivos de representatividad y confirmación en estudiantes de Psicología de tres universidades del Bío-Bío. *Propósitos y Representaciones*, vol. 7, no. 2, pp. 210-239.

Desde esta perspectiva, es posible afirmar que los sesgos no son meros errores fortuitos, sino manifestaciones estructurales de los límites del juicio humano. Y, dado que la inteligencia artificial se construye precisamente sobre datos y patrones humanos, estos sesgos no desaparecen al ser informatizados, sino que son trasladados al sistema algorítmico, muchas veces con una potencia amplificadora.

ii. Impacto de los sesgos cognitivos en la toma de decisiones jurídicas

El ámbito jurídico no escapa a esta condición. Como señala Nieva, la actividad jurisdiccional, aunque formalmente guiada por la legalidad, está profundamente atravesada por atajos mentales, especialmente en contextos de alta carga procesal o presión institucional¹³. En tales escenarios, los jueces —como cualquier otro actor humano— recurren a heurísticos que, aunque útiles, pueden derivar en decisiones afectadas por sesgos cognitivos.

En primer lugar, el heurístico de representatividad se activa cuando se evalúa una situación según el grado de similitud con un modelo o categoría preexistente. Un juez, por ejemplo, puede asociar el caso presente con uno previo de características similares, y decidir conforme a su experiencia anterior, aun si los matices del nuevo caso requieren una valoración diferenciada. Este mismo mecanismo puede ser replicado por un sistema de IA, pero traducido en cálculos probabilísticos y patrones estadísticos¹⁴.

En segundo lugar, el heurístico de anclaje y ajuste se manifiesta cuando el juicio inicial se ve excesivamente condicionado por un valor de referencia, que no logra ser corregido de forma suficiente pese a la aparición de nueva información. Así, un magistrado podría establecer una pena inicial basada en precedentes jurídicos disímiles, y luego realizar ajustes menores que no alcanzan a corregir el error de origen. En contraposición, los sistemas algorítmicos tienden a modificar sus conclusiones únicamente cuando se reconfiguran los datos de entrada o se reentrena el modelo, lo que restringe su flexibilidad adaptativa en comparación con el juicio humano¹⁵.

¹³ NIEVA, J. (2018). *Inteligencia artificial y proceso judicial*. Marcial Pons.

¹⁴ *Idem*.

¹⁵ *Idem*.

No obstante, para comprender cabalmente el fenómeno de los sesgos, es necesario ampliar el análisis hacia su dimensión lingüística. En este sentido, la filosofía del lenguaje de Wittgenstein ofrece una base teórica fundamental: según su planteamiento, el lenguaje no es un espejo neutro de la realidad, sino una herramienta limitada que estructura nuestra forma de percibir y describir el mundo¹⁶. Para Wittgenstein, los significados no están dados de forma fija, sino que dependen del uso contextual del lenguaje, lo que introduce una indeterminación semántica inherente a toda comunicación.

Esta idea tiene profundas implicancias: el lenguaje condiciona el pensamiento, y por tanto, todo sistema de interpretación automatizada del mundo —como es el caso de un algoritmo jurídico que veremos más adelante— estará atravesado por las limitaciones y ambigüedades del lenguaje humano del cual se alimenta. Así, los sesgos no sólo derivan de procesos cognitivos individuales, sino también de las estructuras lingüísticas que organizan nuestra experiencia social y jurídica.

Por consiguiente, los sesgos de lenguaje —en línea con Wittgenstein— no son anomalías corregibles mediante ajustes técnicos, sino manifestaciones estructurales de un sistema simbólico que, por definición, excluye, simplifica y clasifica la realidad. Cuando los datos se procesan en modelos de inteligencia artificial, estas ambigüedades semánticas no desaparecen, sino que se traducen a otra lógica (la lógica computacional), que a su vez reproduce y refuerza ciertas formas de entender el mundo, muchas veces sin que sus supuestos sean explícitos o transparentes.

En suma, la toma de decisiones jurídicas está constantemente expuesta a sesgos cognitivos y lingüísticos que inciden tanto en la racionalidad de los jueces como en el diseño y funcionamiento de los sistemas judiciales. Reconocer esta doble dimensión es clave para evaluar críticamente el uso de la inteligencia artificial en el derecho, pues el riesgo no radica solo en la automatización de decisiones, sino en la automatización de prejuicios históricamente normalizados.

¹⁶ WITTGENSTEIN, L. (1922). *Tractatus logico-philosophicus*. Kegan Paul, Trench, Trubner & Co., LTD.

III. El lenguaje como núcleo del sesgo estructural

i. El lenguaje natural y sus sesgos: ambigüedad, indeterminación, interpretación

A. Imperfección del lenguaje humano

El lenguaje natural es inherentemente impreciso, ambiguo y contextual, lo cual permite la riqueza expresiva, pero también genera profundas dificultades cuando se pretende sistematizar o automatizar. Como señala Wittgenstein en sus *Investigaciones Filosóficas*, el significado de una palabra está en su uso dentro de un “juego de lenguaje”, es decir, es inseparable de sus contextos sociales e históricos¹⁷. Esta naturaleza flexible y polisémica del lenguaje humano entra en contradicción con los requerimientos de precisión y univocidad del lenguaje computacional.

En el campo de la inteligencia artificial, esta imprecisión se traduce en una tensión constante entre el lenguaje natural y el lenguaje formal. Los modelos computacionales que trabajan con procesamiento de lenguaje natural (NLP) —como GPT, Watson o Debater de IBM como veremos más adelante— intentan capturar patrones de uso y significado mediante estadísticas y redes neuronales, pero siempre con el riesgo de reproducir los sesgos sociales codificados en el corpus textual del cual se alimentan.

B. Sesgos semánticos y estructurales

Uno de los ejemplos más evidentes son los sesgos semánticos derivados de la polisemia y la ambigüedad léxica. Por ejemplo, la palabra “justicia” puede tener significados radicalmente distintos dependiendo del contexto (retributiva, distributiva, restaurativa), del hablante (juez, víctima, acusado) y del momento histórico (liberal, socialista, populista).

Asimismo, palabras como “violencia”, “autoridad” o “discriminación” no solo tienen múltiples sentidos, sino que esos sentidos están estructuralmente ligados a visiones ideológicas del mundo. Los modelos de IA, al ser entrenados con textos provenientes de medios

¹⁷ WITTGENSTEIN, L. (1999). *Investigaciones filosóficas*. Ediciones Altaya.

de comunicación, literatura jurídica, redes sociales o sentencias judiciales, aprenden esas correlaciones semánticas sin comprender su dimensión crítica, reproduciendo relaciones entre conceptos que pueden estar cargadas de prejuicios o suposiciones no verificadas.

Un ejemplo concreto es el de los embeddings de palabras utilizados por modelos como Word2Vec o GloVe: se ha comprobado que dichos modelos asocian términos como “hombre” con “ciencia” o “mujer” con “hogar” simplemente por la frecuencia estadística de esas relaciones en los textos analizados, lo cual reproduce estereotipos de género¹⁸.

ii. El lenguaje jurídico como convención estructurada (y su falibilidad)

A. Pretensión de objetividad en el derecho

El lenguaje jurídico ha sido históricamente construido como una convención institucionalizada orientada a dotar de certeza y previsibilidad a la aplicación del derecho. Para ello, recurre a mecanismos como la normatividad abstracta, las definiciones legales precisas, y el uso de técnicas interpretativas sistemáticas, que buscan reducir la ambigüedad propia del lenguaje natural. Esta arquitectura normativa permite hablar de una cierta pretensión de objetividad en el discurso jurídico, en la medida en que intenta garantizar condiciones mínimas de estabilidad semántica y control racional sobre el uso del poder público.

Sin embargo, esta pretensión es, en el fondo, parcial y falible. El derecho, en tanto fenómeno lingüístico y social, no escapa a la indeterminación estructural del lenguaje natural, tal como lo han

¹⁸ Véase BOLUKBASI, T., CHANG, K. W., ZOU, J. Y., SALIGRAMA, V., & KALAI, A. T. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In *Advances in Neural Information Processing Systems*. NEURIPS (2016), CALISKAN, A., BRYSON, J. J., & NARAYANAN, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, pp. 183–186.

advertido autores como Alexy¹⁹ y Ferrajoli²⁰. Este último afirma que el Derecho moderno es un “sistema de normas secundarias”, que pretende establecer reglas sobre cómo se crean, interpretan y corrigen otras normas. Pero incluso este metanivel normativo no elimina su condicionamiento ideológico, ni su dependencia de contextos históricos y estructuras de poder. La semántica jurídica, por tanto, no es neutral: está atravesada por conflictos interpretativos, disputas por el sentido, y tensiones entre el texto normativo y los valores constitucionales en juego.

Esto implica reconocer que el lenguaje jurídico, aunque aspire a la objetividad, funciona como una convención estructurada, pero abierta, en la cual las palabras no remiten de manera unívoca a significados fijos, sino que su sentido se configura discursivamente. En otras palabras, el lenguaje jurídico no puede ser reducido a un código cerrado, sino que opera como una práctica argumentativa institucionalizada, donde el significado es siempre el resultado de un proceso interpretativo.

B. Relación con la teoría de la argumentación jurídica

Frente a esta indeterminación, la teoría de la argumentación jurídica ha mostrado que el razonamiento jurídico no puede concebirse como una operación meramente lógica o deductiva, sino como una práctica racional discursiva, donde se busca justificar normativamente las decisiones jurídicas mediante argumentos plausibles, reconstruibles y contextualmente adecuados.

Autores como Perelman, Alexy y Atienza han desarrollado modelos teóricos que reconocen el carácter abierto del lenguaje jurídico y la necesidad de criterios de racionalidad práctica para orientarlo. En esta línea, Atienza²¹ sostiene que los enunciados jurídicos “no tienen un sentido fijo”, lo que obliga a que el jurista construya sentidos mediante operaciones interpretativas, analógicas y ponderativas, todo dentro de un marco institucional de justificación.

¹⁹ ALEXY, R. (2017). *Teoría de la argumentación jurídica: la teoría del discurso racional como teoría de la fundamentación jurídica*. Palestra Editores.

²⁰ FERRAJOLI, L. (2018). *Derecho y razón. Teoría del garantismo penal*. Trotta 10.^a ed.

²¹ ATIENZA, M. (2016). *Las razones del Derecho: Teorías de la argumentación jurídica* (Colección Derecho y Argumentación, cap. 1. Palestra Editores.

Por ejemplo, Alexy articula su propuesta sobre la base de una “lógica de la argumentación jurídica” que conjuga exigencias de coherencia formal con la inclusión de principios constitucionales y elementos del contexto práctico²². Así, el razonamiento jurídico implica juicios de valor estructurados racionalmente, donde el significado de las normas depende de su articulación con otras normas, principios y hechos del caso.

En consecuencia, cualquier intento de formalizar o automatizar el razonamiento jurídico –como pretenden ciertos modelos algorítmicos de inteligencia artificial– debe enfrentar necesariamente esta ambigüedad estructural del lenguaje jurídico. Esta ambigüedad no es un defecto, sino una condición inherente a su funcionamiento práctico, que habilita la interpretación y garantiza su adaptabilidad a contextos cambiantes, aunque también introduce zonas de disputa hermenéutica. En última instancia, lo que está en juego no es solo el significado de las palabras, sino la legitimidad racional del poder jurídico que las interpreta.

iii. El problema de la traducción al lenguaje de programación

A. De lo natural a lo formal

Uno de los mayores retos en la construcción de sistemas de inteligencia artificial jurídica es la traducción del lenguaje natural (o jurídico) al lenguaje de programación, es decir, a estructuras formales de datos, funciones y reglas lógicas. Esta traducción implica una pérdida inevitable de riqueza semántica y contexto. Como advierte Ashley²³, aunque es posible modelar ciertos elementos del razonamiento jurídico, los sistemas computacionales actuales aún enfrentan dificultades importantes para incorporar la indeterminación estructural del Derecho.

Por ejemplo, los estándares de prueba en derecho penal, como el de “más allá de toda duda razonable”, no son simplemente umbrales de probabilidad, sino juicios cualitativos que implican una convicción moral reforzada por la argumentación y el contexto. Como

²² ALEXY, R. (2017). *Teoría de la argumentación jurídica: la teoría del discurso racional como teoría de la fundamentación jurídica*. Palestra Editores.

²³ ASHLEY, K. (2023). *Inteligencia artificial y analítica jurídica: nuevas herramientas para la práctica del derecho en la era digital*. Yachay Legal.

señala Weiss²⁴, para que un sistema computacional evalúe si una afirmación satisface un estándar probatorio, debe tener acceso a los supuestos valorativos y fácticos que hacen aceptable una inferencia en una comunidad jurídica determinada, lo cual es extremadamente complejo de formalizar.

B. Pérdida de contexto y significado

La experiencia con programas como IBM Debater — como veremos más adelante — ha evidenciado esta dificultad. Aunque el sistema es capaz de detectar argumentos a favor y en contra de un tema específico, su evaluación depende de correlaciones textuales y no del contenido normativo o axiológico de los argumentos. Cuando Debater intentó analizar la constitucionalidad de una ley del Estado de California sobre videojuegos violentos, sus predicciones no lograron captar el estándar de escrutinio estricto aplicado por la Corte Suprema, ni los principios fundamentales en juego, como la libertad de expresión o el interés estatal en proteger a los menores.

Este tipo de errores no son meramente técnicos: son fallos epistémicos, que reflejan la imposibilidad de una equivalencia perfecta entre norma, hecho y código. La automatización de decisiones jurídicas implica no solo computar hechos, sino evaluar valores, contextos, trayectorias jurisprudenciales y sentidos comunes jurídicos — todo lo cual es muy difícil de codificar sin perder la riqueza del lenguaje jurídico —.

IV. Los sesgos en la inteligencia artificial y en los algoritmos

i. Herencia humana: cómo se trasladan los sesgos a los sistemas

A. Programación realizada por humanos

Una de las características estructurales más importantes de la inteligencia artificial (IA) radica en que los algoritmos son diseñados por humanos y se entrenan con datos generados por humanos, lo cual implica que no son neutros ni imparciales por defecto. Desde una perspectiva epistemológica, esto significa que **los sesgos**

²⁴ WEISS, C. (2003). "Expressing Scientific Uncertainty." *Law, Probability and Risk*, pp. 25-46.

cognitivos, sociales y lingüísticos de quienes diseñan y entrenan los sistemas pueden replicarse —consciente o inconscientemente— en el comportamiento de la IA.

De acuerdo con Ashley, los primeros sistemas legales expertos, como WLES (Waterman Legal Expert System) o Taxman II, ya enfrentaban limitaciones estructurales derivadas de su dependencia del lenguaje jurídico, las reglas heurísticas y el razonamiento por encadenamiento (forward o backward chaining)²⁵. Estos sistemas, aunque sofisticados para su época, se basaban en marcos conceptuales previamente definidos por humanos, lo cual conllevaba el riesgo de reproducir concepciones jurídicas sesgadas o parciales, dependiendo del contexto doctrinal, jurisprudencial o ideológico del programador.

B. Reproducción de los sesgos en los sistemas

Pese a la creencia popular, los sistemas de inteligencia artificial no están exentos de sesgos, sino que más bien los heredan estructuralmente de los humanos, quienes programan su funcionamiento y proveen sus datos de entrenamiento²⁶. Al respecto, Irazabal advierte que “no se trata de un error que pueda evitarse, sino de una condición esencial para el conocimiento de cualquier tipo”²⁷. A partir de ello, es posible afirmar que la presencia misma del sesgo no implica necesariamente una falla, pero sí una alerta epistémica que requiere ser gestionada con rigurosidad para evitar su propagación acrítica en sistemas automatizados.

Un estudio empírico reveló que la ausencia de identificación y corrección de estos sesgos no solo afecta la efectividad en la toma de decisiones, sino que además puede producir una “herencia de sesgo”²⁸. Es decir, los usuarios que interactúan con sistemas

²⁵ ASHLEY, K. (2023). Inteligencia artificial y analítica jurídica: nuevas herramientas para la práctica del derecho en la era digital. Yachay Legal.

²⁶ VICENTE, L. & MATUTE, H. (2023). Humans inherit artificial intelligence biases. *Scientific Reports*.

²⁷ IRAZABAL, E. (2023). La Inteligencia Artificial explicada para abogados. La Ley.

²⁸ VICENTE, L. & MATUTE, H. (2023). Humans inherit artificial intelligence biases. *Scientific Reports*.

sesgados tienden a internalizar dichos sesgos y a reproducirlos en decisiones futuras, incluso después de haber cesado el uso del sistema²⁹, lo que genera un círculo vicioso de decisiones erradas amplificadas tecnológicamente.

Este problema no es meramente técnico, sino profundamente lingüístico y filosófico. Desde la perspectiva del lenguaje de Wittgenstein como de Saussure, es posible comprender que el significado de una palabra está determinado por su uso en un juego de lenguaje específico, y que el lenguaje es, en última instancia, una forma de vida. Esto implica que las resoluciones judiciales, al ser textos jurídicos cargados de significado, no son estructuras objetivas e inmutables, sino expresiones de contextos culturales, institucionales e históricos específicos. De este modo, las ambigüedades, indeterminaciones semánticas y disputas conceptuales propias del Derecho también son transferidas a los sistemas de IA cuando estos se entrenan con textos jurídicos.

Ashley advierte que muchos programas actuales, como Ravel, Lex Machina o Ross Intelligence, aunque utilizan técnicas sofisticadas de analítica textual, no superan del todo el problema del sesgo semántico, pues analizan el lenguaje normativo desde sus representaciones textuales, pero sin comprender su complejidad pragmática o filosófica³⁰. En efecto, estos programas tienden a tratar los signos jurídicos como estructuras cerradas, pero ignoran que, como señala Wittgenstein en el fundamento 43 de su *Tractatus*³¹, “el significado de una palabra es su uso en el lenguaje”, lo que en el campo jurídico se traduce en que el mismo término puede adoptar sentidos distintos según el foro, el juez, la doctrina dominante o el contexto fáctico.

²⁹ VIDAL, G. (2008). “La teoría de la elección racional en las ciencias sociales”, *Revista Sociológica*, pp. 12.

³⁰ ASHLEY, K. (2023). Inteligencia artificial y analítica jurídica: nuevas herramientas para la práctica del derecho en la era digital. Yachay Legal.

³¹ WITTGENSTEIN, L. *Tractatus Logico-Philosophicus*. Escuela de Filosofía Universidad ARCIS, fundamento 43.

Por ejemplo, Ashley menciona que algunos programas entrenados con decisiones judiciales (case-based reasoning systems) funcionan a partir de modelos previos que pueden no ser representativos de la totalidad del derecho vigente, o incluso pueden haberse basado en sentencias controvertidas, inexactas o inestables³². Esto evidencia que la base sobre la cual se entrena la IA jurídica es epistemológicamente frágil. Si a esto se suma que el programador puede incorporar —de forma deliberada o accidental— determinados intereses, paradigmas o limitaciones culturales, entonces el sistema deja de ser una herramienta imparcial para convertirse en un mecanismo de reproducción de sesgos estructurales, incluso ideológicos.

Este escenario exige una reflexión crítica: si la inteligencia artificial depende de datos y estructuras lingüísticas que son construidas, discutidas y disputadas socialmente, entonces su desarrollo ético y técnico no puede desentenderse de una concepción filosófica del lenguaje y del conocimiento. Solo reconociendo la fragilidad epistémica del lenguaje jurídico, su ambigüedad semántica y su carácter disputable, es posible diseñar sistemas que no perpetúen ciegamente los sesgos humanos, sino que los enfrenten desde una lógica reflexiva y autocrítica.

ii. Casos ilustrativos: Compass, IBM Debater, entre otros

A. Sistema COMPAS (EE.UU.)

El sistema COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), desarrollado por Northpointe (actualmente Equivant), es un instrumento de evaluación de riesgos de cuarta generación que se aplica en diversas etapas del proceso penal, particularmente en decisiones de libertad condicional y sentencias. Según su propia “Guía para profesionales”, busca identificar necesidades criminógenas y riesgos de reincidencia, a través del análisis de más de 137 variables³³.

³² ASHLEY, K. (2023). Inteligencia artificial y analítica jurídica: nuevas herramientas para la práctica del derecho en la era digital. Yachay Legal.

³³ NORTHPOINTE. (2015). Practitioner’s Guide to COMPAS Core.

No obstante, este procesamiento de datos no escapa a condicionamientos estructurales: si las variables utilizadas están impregnadas por desigualdades históricas —como antecedentes policiales desproporcionadamente altos en comunidades racializadas, o factores socioeconómicos interpretados como indicadores de riesgo³⁴—, el sistema no solo reproduce, sino que automatiza la discriminación. En ese sentido, el sesgo no es una mera falla técnica, sino una manifestación algorítmica de asimetrías sociales preexistentes. El riesgo radica en que, al integrarse en decisiones judiciales con efectos jurídicos, estos sesgos adquieren un carácter normativo, institucionalizando la exclusión bajo el manto de una supuesta neutralidad técnica.

Esta institucionalización ocurre cuando los resultados sesgados del sistema se convierten en insumos regulares para decisiones formales (libertad condicional, medidas cautelares, sentencias), sin que se visibilice su origen desigual o se activen mecanismos de corrección³⁵. Así, lo que debería ser un instrumento auxiliar se transforma en una fuente de legitimación de prácticas discriminatorias, blindadas por el prestigio del dato y la opacidad del algoritmo.

Investigaciones como la de Roa et al. (2022) y reportes de ProPublica han evidenciado cómo este sistema, lejos de garantizar neutralidad, reproduce y amplifica sesgos raciales, otorgando mayores puntajes de riesgo a personas afrodescendientes en comparación con personas blancas en condiciones similares³⁶. Esto evidencia un problema estructural: la IA jurídica puede institucionalizar la discriminación preexistente bajo un barniz técnico de objetividad.

³⁴ FIGUEROA, M., REBOLLEDO, R. (2024). Presentación Dossier. Racismos. Ideología del mestizaje y colonialismo. Reflexiones para repensar los procesos de alterización de nuestra América. CUHSO (Temuco), vol. 34, no. 2.

³⁵ *Idem*.

³⁶ ROA, M.; SANABRIA, J.; DINAS, K. (2022). Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos. *Rev. Bras. De Direito Processual Penal*, vol. 8, no. 1, pp. 275-310.

B. IBM Debater y los límites del lenguaje argumentativo

Un caso más sofisticado es IBM Project Debater, sucesor del célebre Watson. Este sistema, probado públicamente en 2014, fue diseñado para participar en debates complejos sobre cuestiones sociales. Al ser presentado con un tópico —por ejemplo, “¿Debería prohibirse la venta de videojuegos violentos a menores?”—, escaneó millones de fuentes, seleccionó los argumentos relevantes, los clasificó como favorables o contrarios, y construyó un esquema discursivo automatizado.

Este experimento, estudiado por Dvorsky³⁷ y comentado por Ashley, muestra los alcances de la minería de argumentación automatizada, pero también sus limitaciones: el sistema fundamentó su argumento en una interpretación literal del Código Civil de California y del caso *Brown v. Entertainment Merchants Association*, sin captar que la resolución del tribunal supremo estadounidense no sólo evaluó la norma, sino que aplicó un test de escrutinio estricto basado en principios constitucionales más complejos (Primera y Decimocuarta Enmienda).

Esto revela un hecho crucial: la argumentación jurídica no se agota en el análisis textual ni lógico-formal, sino que requiere una ponderación de valores, contextos y principios que los sistemas computacionales actuales aún no pueden reproducir con fidelidad. Como afirma Ashley, los modelos computacionales deben incorporar estándares de prueba bien definidos, tales como el más allá de toda duda razonable, para que una conclusión jurídica sea aceptable dentro del sistema³⁸.

C. Reflexión crítica

La cuestión central es entonces: *¿puede la IA juzgar con mayor objetividad que un humano? ¿O simplemente reproduce, con mayor velocidad, los mismos sesgos del lenguaje y los datos humanos?* En efecto, los modelos de IA —por más avanzados que sean— se alimentan

³⁷ DVORSKY, G. (2014, May 5). IBM's Watson Can Now Debate Its Opponents (minute 45:47). Gizmodo, (Consultado el 13 de mayo del 2025).

³⁸ ASHLEY, K. (2023). Inteligencia artificial y analítica jurídica: nuevas herramientas para la práctica del derecho en la era digital. Yachay Legal.

de corpus lingüísticos, decisiones previas y datos históricos, todos ellos atravesados por visiones ideológicas, valores implícitos y estructuras de poder.

Además, como bien advierte Ashley, los sistemas jurídicos no son homogéneos ni políticamente neutros³⁹. Existen sistemas como el estadounidense, fuertemente influido por precedentes y razones institucionalizadas (“institutional reasons”), y otros como el continental europeo o el peruano, donde el formalismo legal convive con una fuerte influencia política, mediática y cultural. En consecuencia, pretender que una IA pueda abstraerse de este contexto es desconocer la naturaleza política del derecho.

Sin embargo, el problema no radica únicamente en la reproducción de sesgos humanos por parte de la IA, sino en su amplificación y normalización bajo una apariencia de objetividad algorítmica, así propiciando una discriminación institucional. Esto ocurre, por ejemplo, con el sistema COMPAS, diseñado para estimar el riesgo de reincidencia en el proceso penal estadounidense, el cual ha sido criticado por su falta de transparencia y por reproducir sesgos raciales preexistentes⁴⁰.

Sobre este punto, resulta pertinente analizar la incidencia del sesgo de representatividad, el cual se manifiesta cuando COMPAS, alimentado con datos históricos del sistema de justicia penal que ya habían adquirido sesgos contra ciertas minorías, construye un “prototipo” de alto riesgo, lo que deriva en puntuaciones artificialmente infladas para individuos pertenecientes a tales minorías, al basar su análisis en estadísticas grupales más que en el riesgo individual⁴¹. De manera análoga, el sesgo de anclaje y ajuste opera cuando las minorías son objeto de un número mayor de detenciones, siendo que estas anclas las sitúan, desde el

³⁹ *Idem.*

⁴⁰ ARIZA, J. (2022). Inteligencia artificial legal y administración de justicia. En S. Calaza y M. Sánchez-Arjona (Dir.), *Impugnación de las decisiones judiciales dictadas con auxilio de inteligencia artificial*, pp. 29-54. Thomson Reuters Aranzadi.

⁴¹ ROA, M.; SANABRIA, J.; DINAS, K. (2022). Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos. *Rev. Bras. De Direito Processual Penal*, vol. 8, no. 1, pp. 275-310.

principio, en un perfil de riesgo elevado que ajustes posteriores no logran compensar. De ahí que, la aplicación institucional de estas puntuaciones viciadas por los sesgos perpetúa la discriminación en el sistema judicial, siendo un factor agravante la “caja negra” intrínseca a los algoritmos que obstaculiza la auditoría de sus procesos y, en consecuencia, la rendición de cuentas⁴².

Ahora bien, lo más inquietante no es, en realidad, que la inteligencia artificial o los algoritmos puedan equivocarse, sino que estos sesgos sean automatizados y consolidados por las instituciones que debieran impartir justicia. Este fenómeno, además de comprometer una evaluación objetiva y amenazar la salvaguarda de principios como el debido proceso y la no discriminación⁴³, revela el desafío que enfrenta el Derecho en su propia digitalización: el de no ceder de su rol garantista frente a herramientas que transforman sesgos en decisiones judiciales con apariencia de objetividad.

iii. Mecanismos actuales de detección y corrección (y sus limitaciones)

A. Técnicas para mitigar sesgos

Frente al creciente reconocimiento de los sesgos estructurales que afectan a los sistemas de inteligencia artificial (IA) jurídica, se han desarrollado múltiples estrategias tanto técnicas como institucionales para su identificación, mitigación y posible corrección. Estas soluciones buscan reducir el impacto discriminatorio que pueden tener las decisiones automatizadas, sobre todo en contextos de alta sensibilidad como el penal o el constitucional.

Entre los mecanismos más utilizados destacan:

- a. **Preprocesamiento de datos (data cleansing):** Consiste en la eliminación o anonimización de variables sensibles (como raza, género o lugar de residencia), así como en la corrección de desbalances en los conjuntos de datos. Sin embargo, esta técnica

⁴² MIRÓ, F. (2018). Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots. UNED, vol. 3, no. 20, pp. 87-130.

⁴³ ROA, M.; SANABRIA, J.; DINAS, K. (2022). Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos. Rev. Bras. De Direito Processual Penal, vol. 8, no. 1, pp. 275-310.

ha sido cuestionada por su superficialidad, ya que borrar datos sensibles no impide que correlaciones indirectas (proxies) reproduzcan los mismos sesgos.

- b. Auditorías algorítmicas externas e independientes:** Evaluaciones sistemáticas, realizadas por terceros imparciales, que buscan examinar la equidad y transparencia de un modelo en diferentes escenarios. A menudo se emplean metodologías como pruebas de caja negra o análisis contrafactuales para determinar si un sistema responde de forma desigual ante casos similares.
- c. Métricas de justicia algorítmica (*fairness metrics*):** Criterios cuantitativos que permiten evaluar si un algoritmo respeta ciertos principios normativos. Entre ellos se encuentran el demographic parity (paridad demográfica), equality of opportunity (igualdad de oportunidades) y disparate impact (impacto adverso). Estas métricas buscan formalizar, dentro del lenguaje técnico, intuiciones legales sobre discriminación y trato desigual.
- d. Técnicas de aprendizaje adversarial y regularización:** Estas estrategias se centran en entrenar al sistema para que no se sobreajuste a patrones sesgados presentes en los datos. Por ejemplo, se introduce un “adversario” que intenta identificar sesgos en las predicciones del modelo, forzándolo a generalizar mejor.

Pese al avance técnico, estas medidas enfrentan limitaciones conceptuales y prácticas cuando se insertan en sistemas jurídicos complejos, políticamente definidos y normativamente conflictivos.

B. Críticas y limitaciones

A pesar de los esfuerzos descritos, existe un consenso creciente en la literatura crítica sobre los límites estructurales e inherentes de estas soluciones técnicas. Como advierte Roa (2022), no es posible “limpiar” un conjunto de datos que ya refleja un mundo desigual: el sesgo no es simplemente un error técnico corregible, sino una expresión profunda de la cultura, el lenguaje y la estructura del poder en el derecho⁴⁴.

⁴⁴ *Idem.*

Desde esta perspectiva, el problema no radica solamente en cómo se entrena un algoritmo, sino en **qué tipo de sistema jurídico y político** da origen a los datos y qué valores se privilegian al automatizar sus decisiones. Tal como lo señala Ferrajoli, el derecho no es un sistema neutro, sino el resultado de luchas históricas, conflictos sociales y estructuras de poder⁴⁵. Así, la IA jurídica no puede ser comprendida únicamente como una herramienta técnica, sino como un nuevo campo de disputa por la definición misma de lo justo y lo injusto, incluso también del aun extenuante debate de lo verdadero o falso de Bruno Cavallone y Michelle Taruffo⁴⁶.

Este enfoque cobra especial relevancia en contextos como el peruano del año 2025, caracterizado por una débil institucionalidad, alta volatilidad normativa y fuerte instrumentalización política del derecho. En tales condiciones, la pretensión de construir una IA jurídica objetiva se convierte en una empresa epistemológicamente inviable y políticamente riesgosa. Automatizar decisiones jurídicas justas presupone, al menos, la existencia de un sistema normativo estable, coherente y legítimo, condición que hoy no se cumple. De lo contrario, el riesgo no es solo replicar sesgos individuales, sino institucionalizar sesgos estructurales bajo el ropaje de la neutralidad algorítmica, sin ningún control democrático ni mecanismos efectivos de rendición de cuentas.

C. Estándares de prueba y modelos computacionales de argumentación jurídica

En los modelos computacionales de argumentación jurídica, los estándares de prueba constituyen un pilar esencial para estructurar razonamientos válidos y aceptables. Como ha señalado Weiss⁴⁷, estos estándares no solo orientan la toma de decisiones, sino que también permiten que las máquinas evalúen cuándo una afirmación

⁴⁵ FERRAJOLI, L. (2018). Derecho y razón. Teoría del garantismo penal. Trotta 10.^a ed.

⁴⁶ CAVALLONE, B. y TARUFFO, M. (2012). Un diálogo sobre prueba y verdad (Colección Palestra Extramuros). Lima: Editorial Palestra.

⁴⁷ WEISS, C. (2003). Expressing Scientific Uncertainty. Law, Probability and Risk, vol. 2, pp. 25-46.

jurídica ha alcanzado un umbral probatorio determinado, como ocurre con el clásico “más allá de toda duda razonable” en el proceso penal.

En esa línea, Weiss indica que, para que un sistema de IA jurídica emule una decisión humana válida, debe incorporar estándares que funcionen como metas evaluables, es decir, niveles mínimos de aceptabilidad para que una afirmación sea tomada en cuenta. Este principio permite que las máquinas no solo acumulen datos, sino que razonen con base en reglas definidas, construyendo convicciones firmes y excluyendo decisiones con márgenes inaceptables de incertidumbre.

Un ejemplo paradigmático de los desafíos que enfrentan estos sistemas lo representa IBM Debater, descendiente del famoso Watson. En 2014, IBM probó un sistema capaz de escanear millones de textos (como Wikipedia) y generar argumentos a favor y en contra de una proposición, tal como “la venta de videojuegos violentos a menores debería estar prohibida”⁴⁸. Para ello, el sistema analizó más de 4 millones de artículos, extrajo oraciones relevantes, identificó alegatos, evaluó su polaridad y construyó una maqueta argumentativa.

No obstante, su razonamiento automatizado presentó fallos significativos cuando se comparó con casos reales, como *Video Software Dealers Association v. Schwarzenegger*, en el que la Corte Suprema de EE. UU. aplicó el escrutinio estricto sobre una ley californiana similar. IBM Debater no pudo captar que el valor constitucional de la libertad de expresión pesa más que el supuesto interés estatal en prohibir tales ventas. El problema de fondo fue que el sistema no comprendía la técnica de la ponderación —elemento que aún es discutible por su indefinición pragmatodialectica⁴⁹—, central en los modelos de argumentación jurídica compleja, donde los derechos y principios deben ser equilibrados con sensibilidad al contexto.

⁴⁸ ASHLEY, K. (2023). *Inteligencia artificial y analítica jurídica: nuevas herramientas para la práctica del derecho en la era digital*. Yachay Legal, pp. 61.

⁴⁹ VAN EEMEREN, F. (2019). *La teoría de la argumentación: Una perspectiva pragmatodialectica*. (Cap. 6, pp. 139-166) Palestra.

D. La cognición jurídica artificial y sus dilemas políticos

En años recientes, han surgido nuevas aplicaciones cognitivas jurídicas —como Ross, Lex Machina y Ravel— que no solo procesan normas, sino que evalúan contextos fácticos y jurisprudenciales. Estas herramientas buscan superar los límites del análisis meramente textual, incorporando elementos como tendencias jurisprudenciales, patrones estadísticos y lenguaje natural. Sin embargo, esta aparente sofisticación no elimina un problema central: la inclinación política y valorativa del derecho.

Según Atienza, citado por Summers, el sistema norteamericano —basado en “razones buenas” y no en pruebas racionales estructuradas— responde más a una tradición judicial que a un método lógico-deductivo⁵⁰. En cambio, los sistemas continentales, como el europeo, adoptan una estructura racional más definida. El Perú, sin embargo, representa una excepción sui generis: su sistema legal es fragmentario, ambiguo, inestable y permeable a intereses políticos, económicos y mediáticos.

Este contexto vuelve particularmente problemática cualquier tentativa de automatización jurídica. Como bien señala Ferrajoli, todo sistema de derecho refleja una visión política del mundo⁵¹; por tanto, automatizar decisiones sin definir previamente cuál visión política se implementa es equivalente a delegar a la máquina una función soberana sin legitimidad. De la misma manera, en la estructura de la argumentación jurídica, García destaca la importancia primaria de definir los patrones de significado y los parámetros de evaluación de los conceptos a utilizar en un sistema, cual se constituye en el concepto de “Tópica Jurídica”⁵².

En resumen, la inteligencia artificial jurídica no puede ser comprendida ni utilizada como una herramienta puramente técnica. Requiere de una estructura normativa definida, una ética jurídica consensuada y mecanismos de control institucional. Sin estos elementos, corre el riesgo de convertirse no en una aliada de

⁵⁰ SUMMERS, R. (2020). Las razones sustantivas y la interpretación del Derecho en el common law. Palestra Editores, p. 37.

⁵¹ FERRAJOLI, L. (2018). Derecho y razón. Teoría del garantismo penal. Trotta 10.^a ed.

⁵² GARCÍA, J. (2018). Teorías de la tópica jurídica. Palestra Editores.

la justicia, sino en una forma tecnocrática de reproducción de injusticias estructurales.

V. Conclusiones

- El lenguaje natural está dotado de ambigüedad e imprecisión, dependiendo de factores como el contexto y momento histórico en que se comunica un mensaje para lograr su correcta interpretación. El lenguaje jurídico no es la excepción, pues, aunque se utilicen diversas técnicas para atribuir objetividad y determinación al derecho, el entendimiento pleno de las ideas en el plano jurídico sigue estando supeditado a circunstancias propias de la realidad contemporánea. Por tales razones, para automatizar el razonamiento jurídico, se deberá abordar minuciosamente la ambigüedad estructural del lenguaje jurídico, a fin de procurar la mayor claridad posible en el uso de los sistemas de inteligencia artificial.
- Los heurísticos son estrategias cognitivas empleadas para simplificar el proceso de toma de decisiones; su aplicación debe reservarse a situaciones que no ameriten un análisis crítico o en profundidad. Cuando se utilizan indistintamente, devienen en sesgos (tanto estructurales o semánticos, como intencionales o funcionales), los cuales impiden el razonamiento lógico y objetivo que se esperaría de un sistema automatizado.
- Los sesgos son inherentes a la inteligencia artificial, dada la inevitabilidad de que sean heredados durante la programación de algoritmos. En ese orden de ideas, se han desarrollado diversos mecanismos de detección y corrección, cuyo fin es disminuir los efectos discriminatorios de los sesgos en los sistemas de inteligencia artificial jurídica. No obstante, su falta de eficacia indica que la IA no puede abstraerse completamente de los sesgos, en tanto que estos son manifestaciones de aspectos culturales, lingüísticos y políticos de una sociedad.
- En resumen, la inteligencia artificial jurídica no puede ser comprendida ni utilizada como una herramienta puramente técnica. Requiere de una estructura normativa definida, una ética jurídica consensuada y mecanismos de control institucional. Sin estos elementos, corre el riesgo de convertirse no en una aliada de la justicia, sino en una forma tecnocrática de reproducción de injusticias estructurales.