

INTELIGENCIA ARTIFICIAL GENERATIVA (IAGEN) Y DEFENSA: ¿CÓMO LOGRAR LA “CONVICCIÓN” DE UNA MÁQUINA?

Jordi Nieva Fenoll
Catedrático de Derecho Procesal
Universitat de Barcelona

RESUMEN: Hasta el momento, la defensa en los procesos ha estado centrada, más allá de en el estudio jurídico, en la elaboración de un relato que pudiera persuadir emocionalmente al juez, creándole un marco mental. Esa estrategia dejaría de tener sentido si el juez es sustituido por una herramienta de inteligencia artificial. Con todo, existen estrategias similares a la citada que todavía podrían seguirse en ese escenario, aunque teniendo en cuenta un trasfondo que nada tiene que ver con la experiencia del juez ni con sus emociones, sino con la consciencia de cómo funciona la inteligencia artificial generativa.

PALABRAS CLAVE: marco mental, heurístico, prompt, proceso judicial

ABSTRACT: Until now, defence in legal proceedings has focused, beyond legal analysis, on crafting a narrative that could emotionally persuade the judge, creating a framing effect. This strategy would cease to make sense if the judge were replaced by an artificial intelligence tool. However, there are similar strategies that could still make sense in this scenario, although taking into account a background that has nothing to do with the judge's experience or emotions, but rather with an awareness of how generative artificial intelligence works.

KEYWORDS: framing effect, heuristic, prompt, judicial process

SUMARIO: 1. INTRODUCCIÓN.— 2. EL FIN DE LOS MARCOS MENTALES.—3. SÓLO RELATOS DE HECHO COMPATIBLES CON EL PASADO.—4. EL FIN DE LAS EMOCIONES.—5. SÓLO INTERPRETACIONES JURÍDICAS COHERENTES CON LÍNEAS JURISPRUDENCIALES.—6. LA CAJA NEGRA DEBE SER BLANCA.—7. BIBLIOGRAFÍA

1. INTRODUCCIÓN

Hace ya un tiempo que se detecta como una realidad más o menos inminente el hecho de que una aplicación de inteligencia artificial sustituya a los jueces en un número creciente de procesos judiciales y de fases del proceso¹. La primera experiencia, nada exitosa² pero real, está siendo desde hace tiempo³ la evaluación del riesgo de violencia para dictaminar sobre la libertad personal en el marco de las medidas cautelares en el proceso penal. Es un trance del proceso extremadamente sensible en el que, sin embargo, no han sido pocos jueces⁴ los que han puesto su confianza en el criterio de una herramienta muy cuestionable por sus sesgos de raza y edad⁵ que, además, curiosamente puede provocar delitos y no prevenirlos⁶. Ni tan siquiera se sabe demasiado bien cómo funciona porque sus fabricantes no han desclasificado sus algoritmos⁷. Imagínese lo poco que puede llegar a costar introducir la inteligencia artificial (IA) en otros sectores del proceso que no sean tan sumamente delicados.

Por ello, hay que plantearse ya la posibilidad de que el trabajo de los abogados, siempre destinado a la convicción de un juez, esté enfocado en el futuro de un modo más o menos limitado a la “persuasión” de una máquina, por mucho que haya intentado evitarlo, puede que vanamente, el Reglamento (UE) 2024/1689 de 13 de junio en materia de inteligencia artificial⁸. Como vamos a ver, ello provocará un cambio radical en los planteamientos de estrategias de defensa, y debiera ocasionar también reformas relevantes en las leyes pro-

¹ Vid. Julià Pijoan, M., *La computerización del derecho, a partir del proceso y de los procedimientos judiciales*, Madrid 2024. Quattrocchio, S., *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, Springer 2020. Nieva Fenoll, J. *Inteligencia artificial y proceso*, Madrid 2018.

² Bornstein, A. M., Are algorithms building the new infrastructure of racism?, *Nautilus*, 21-12-2017, <http://nautil.us/issue/55/trust/are-algorithms-building-the-new-infrastructure-of-racism>. Larson, J., Mattu, S., Kirchner, L., Angwin, J., How We Analyzed The Compas Recidivism Algorithm, *Propublica*, 23-5-2016, <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Brandom, R., Mechanical Turkers may have out-predicted the most popular crime-predicting algorithm, *The Verge*, 17-1-2018, <https://www.theverge.com/2018/1/17/16902016/compas-algorithm-sentencing-court-accuracy-problem>. Fussell, S., Study Finds Crime-Predicting Algorithm Is No Smarter Than Online Poll Takers, *Gizmodo*, 18-1-2018, <https://gizmodo.com/study-finds-crime-predicting-algorithm-is-no-smarter-th-1822173965>.

³ Correctional Offender Management Profiling for Alternative Sanctions (COMPAS). Vid. <https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html>, así como su manual de uso en NORTHPOINTE, *Practioners Guide to COMPAS*, 17-8-2012.

⁴ State v. Loomis, 881 N.W.2d 749 (Wis. 2016). Tashea, J., Risk-Assessment Algorithms Challenged In Bail, Sentencing And Parole Decisions, *ABA Journal*, Marzo 2017, http://www.abajournal.com/magazine/article/algorithm_bail_sentencing_parole.

⁵ Engel, C., Linhardt, L., Schubert, M., Code is law: how COMPAS affects the way the judiciary handles the risk of recidivism, *Artificial Intelligence and Law*, 2025, pp. 383 y ss.

⁶ Khosrowi, D., Van Basshuysen, P., Making a Murderer: How Risk Assessment Tools May Produce Rather Than Predict Criminal Behavior, *American Philosophical Quarterly*, 2024, 61 (4), pp. 309 y ss. <https://scholarlypublishingcollective.org/uip/apq/article/61/4/309/391697>

⁷ Seymour, W., Detecting Bias: Does an Algorithm Have to Be Transparent in Order to Be Fair?, 2018, <https://ir.shef.ac.uk/bias/pdf/seymour2018detecting.pdf>

⁸ Punto 61, art. 74.8 y anexo III, 8.

cesales, concebidas para crear un marco decisorio adecuado para un juez en el que, asimismo, el abogado pueda desarrollar su retórica. Sin embargo, esas normas son inaptas para suministrar los datos que necesita una herramienta de IA para poder operar. No se debería esperar a que vengan las herramientas IA que presten asistencia generativa —es decir, generando textos— al juez en el proceso en la búsqueda de jurisprudencia, análisis del Derecho aplicable e incluso en materia de valoración de la prueba y de redacción de la resolución judicial, sino que en las leyes debiera elaborarse ya un procedimiento adecuado para su uso con lo que se puede esperar racionalmente, ya a día de hoy, de esas herramientas. El estado de la tecnología ya es lo suficientemente avanzado como para hacer que el futuro sea nuestro presente.

En el presente trabajo me referiré solamente a cómo se va a modificar el trabajo de los abogados, lo que conllevará una exposición somera de cómo van a ser esas nuevas herramientas. Hay que ser conscientes, como idea de base, que la inteligencia artificial requiere datos parametrizables, es decir, que se repitan un número indeterminado de veces de manera que se puedan establecer patrones. En todos los sectores del proceso en que se puedan obtener ese tipo de datos, podrá haber una aplicación IA que la asista.

2. EL FIN DE LOS MARCOS MENTALES

Una parte importante del trabajo de los abogados, sobre todo en los escritos de alegaciones, ha estado históricamente centrada en algo que, en puridad, no es Derecho procesal, sino simplemente retórica: la elaboración de marcos mentales⁹. Consisten en evocar un relato favorable al cliente en el que los indicios del caso encajen bien, de manera que al juez, que identificará en ese relato una historia que le sea familiar¹⁰, le parezca creíble y, finalmente, sienta empatía por el litigante defendido por ese abogado. Ese marco mental que se habrá creado de ese modo puede corresponderse con la realidad, pero también puede ser una trampa en toda regla. En no pocas ocasiones se trata de una auténtica telaraña de la que es muy difícil salir después de haber quedado atrapado. Una vez que el juez asume de manera casi espontánea el relato del marco mental propuesto por una de las partes, empezará a incidir en su mente el heurístico de anclaje y ajuste¹¹, que provoca el llamado sesgo de confirmación¹². Lo que significa que no solamente reinterpretará toda la

⁹ Tversky, A., K., Daniel, The Framing of decisions and the psychology of choice, *Science*, 211, 1981, (4481), pp. 453 y ss. Plous, S., *The psychology of judgment and decision making*, 1993. Andrews, A.; Clawson, R.A.; Gramig, B.M., Finding the Right Value: Framing Effects on Domain Experts, *Political Psychology*, vol. 38, 2, Abril 2017, pp. 261 y ss.

¹⁰ Tversky, A., Kahneman, D., Judgments of and by Representativeness, en Kahneman, Slovic, Tversky (ed.), *Judgment under Uncertainty: Heuristics and Biases*, pp. 84 y ss.

¹¹ Tversky, A., Kahneman, D., Judgment under Uncertainty: Heuristics and Biases, *Science*, New Series, Vol. 185, n. 4157, Sep. 27, 1974, p. 1128. Furnham, A.; Chu Boo, H., A literature review of the anchoring effect, *The Journal of Socio-Economics*, vol 40, 1, Feb. 2011, pp. 35 y ss.

¹² Myers, D. G., *Intuición. El poder y el peligro del sexto sentido*, trad. de Guillermo Solana de *Intuition: its power and perils*, New Haven y Londres 2002, Barcelona 2003. p. 175.

información que vaya surgiendo en el proceso en sentido positivo al relato del marco mental, sino que, además, la información que sea claramente adversa a ese relato será *anclada y ajustada* —heurístico de anclaje y ajuste— para que parezca que *confirma* el relato —sesgo de confirmación—. Incluso se puede llegar al extremo de descartar los datos que rompan ese relato o simplemente estorben en su credibilidad. Como sabe cualquier abogado, es muy difícil persuadir a un juez que ya ha asumido previamente un relato, incurriendo en lo que tradicionalmente se ha llamado prejuicio.

La técnica del marco mental no es utilizada solamente en los escritos judiciales, sino sobre todo por la política e incluso por la literatura. Los políticos¹³ intentan inducir esos marcos mentales en la población para conseguir el ansiado *framing effect*, de manera que la gente los vea como buenos líderes, o visualice a su formación política como el partido del orden o del progreso, por ejemplo, incurriendo naturalmente en una generalización que hace que el votante deje completamente de lado los detalles, lo que es sensacional para los intereses del político. Es sencillo de entender e incluso fácil de operar, sobre todo si se conocen bien los intereses de base del interlocutor a quien se desea convencer, y también su sentido de lo que es justo. Sin embargo, por elemental que sea la estrategia, es muy difícil escapar de ella porque apela al sentido común de cada cual. Una vez que ese sentido común es captado por el marco mental, su víctima no se moverá de ahí, dado que cualquier ser humano suele pensar firmemente que es acertado su propio criterio. Como dijo Descartes¹⁴ justo al inicio de su *Discurso del método*, “el buen sentido es la cosa mejor repartida del mundo, pues cada cual piensa que posee tan buena provisión de él, que aun los más descontentadizos respecto a cualquier otra cosa, no suelen apetecer más del que ya tienen”.

Pues bien, evidentemente todo lo anterior es una técnica de persuasión que funciona con seres humanos, pero es completamente inútil con una máquina, es decir, con una herramienta de inteligencia artificial que, además, puede ser programada para identificar esos marcos mentales y no entrar en ellos¹⁵, exactamente igual que ahora algunas de ellas —ChatGPT en el momento actual— por motivos técnicos¹⁶ tienden a ser complacientes y a darnos la razón¹⁷, lo que es positivo a efectos comerciales porque tiene la potencialidad de hacernos adictos a la aplicación¹⁸. De hecho, se trata del mismo marco

¹³ Druckman, J. N., Political Preference Formation: Competition, Deliberation, and the (Ir)relevance of Framing Effects, *American Political Science Review*, nov. 2004, vol. 98, 4, pp. 671 y ss.

¹⁴ Descartes, R., *Discurso del método*, Madrid 1986, p. 35.

¹⁵ Kim, T., Song, H., The Effect of Message Framing and Timing on the Acceptance of Artificial Intelligence's Suggestion, *Association for Computing Machinery*, 2020.

¹⁶ Laizure, S. C., Caution: ChatGPT Doesn't Know What You Are Asking and Doesn't Know What It Is Saying, *The Journal of Pediatric Pharmacology and Therapeutics*, 14-10-2024, <https://jppt.kglmeridian.com/view/journals/jppt/29/5/article-p558.xml>.

¹⁷ “These models aren't designed to transmit information, so we shouldn't be too surprised when their assertions turn out to be false”. Hicks, M.T., Humphries, J., Slater, J., ChatGPT is bullshit, *Ethics and Information Technology*, 2024, p. 38.

¹⁸ Vid. Golden, W., ChatGPT: A trusted source?, *Irish Journal of Technology Enhanced Learning*, vol. 7, 2, pp. 113 y ss.

mental, pero observado ahora de manera inversa. Por puro cálculo estadístico de palabras que suelen ir acompañadas de otras¹⁹, la herramienta parece identificar lo que nos parece más razonable, lo que se traduce, a efectos prácticos, en que nos sigue la corriente. Y esto quiere decir que también podría estar programada —aunque es más complejo²⁰— para llevarnos la contraria, utilidad que puede tener bastante recorrido en materia de prevención de conflictos y evitación del proceso²¹, por cierto, pero esa es otra cuestión.

En el estado actual de la técnica, sólo es viable la automatización de procesos repetitivos, en los cuales, aun sabiéndose perfectamente que, por decirlo así, su resultado son habas contadas, el abogado no suele perder la oportunidad de intentar descalificar el comportamiento de la parte contraria para que el juez identifique a una persona aviesa, es decir, para generar un marco mental desfavorable a su oponente. Pues bien, nada de esa literatura, tantas veces de pésima calidad, tendrá sentido si el interlocutor es una máquina puesto que, muy probablemente, en el futuro esos procesos estarán dirigidos por herramientas orientadas al dato, es decir, aplicativos en los que haya que aludir a los datos principales del proceso, dejando escaso espacio a los relatos persuasivos que, por entretenidos que sean, no alteren la realidad principal.

En consecuencia, la aplicación preguntará si el inquilino pagó y cuántos meses se retrasó, no si prometió pago varias veces y luego se escondía del acreedor cuando lo veía por la finca. La herramienta pedirá el dato de si el vuelo se retrasó o si se perdió la maleta, y no se hará referencia al estado de ánimo o agitación del pasaje o peor o mejor educación del personal de la compañía. La aplicación preguntará si se desea el divorcio, y no si el cónyuge es una buena o mala persona. Pedirá la descripción de la realidad de la ocupación del bien inmueble, y no serán relevantes para la herramienta las intenciones sociológicas del ocupante. Sólo si esas circunstancias sociales tienen importancia porque la ley las señala expresamente, se pedirá una respuesta clara y directa y una prueba, pudiendo llegar a indicarse incluso de qué pruebas puede tratarse, aunque en este sentido la herramienta debe ser sólo ejemplificativa²² y muy generosa, a fin de no recaer en una versión posmoderna del sistema medieval de valoración legal de la prueba²³, que ya casi, por fortuna, es historia.

Con todo, puede que quede un resquicio donde se puedan introducir marcos mentales. Al fin y al cabo, la I.A. sólo es una tecnología basada en una enorme base de datos que elabora sus resultados basándose en dichos elementos, es decir, en los citados datos. No es descabellado que si son co-

¹⁹ Hicks, M.T., Humphries, J., Slater, J., ChatGPT is bullshit, cit. p. 37.

²⁰ Gregorcic, B., Pendrill, A.-M., ChatGPT and the frustrated Socrates, *Physical Education*, 2023, pp. 1 y ss.

²¹ Vid. Larson, D. A., Artificial Intelligence: Robots, Avatars, and the Demise of the Human Mediator, *Ohio State Journal on Dispute Resolution*, Vol. 25, 2010, pp. 117-118.

²² En la línea de asistir, y no sustituir, establecida por el Reglamento (UE) 2024/1689 de 13 de junio, punto 61.

²³ Vid. Azón, *Summa Azonis*, Lib. IV, 19 (de probationibus). punto 21, p. 318. Degli Ubaldi, Baldo, *Practica Baldi*, 1521, folio 39 vuelto. Jaumar Carrera, *Práctica forense*, 1840, pp. 39-45.

nocidos esos elementos que la máquina tiene en cuenta para elaborar sus patrones, la estrategia del abogado vaya enfocada precisamente a reflejarlos y entrar en el terreno de la resolución favorable ya prevista por la aplicación. Se analizará esta cuestión en el siguiente epígrafe.

3. SÓLO RELATOS DE HECHO COMPATIBLES CON EL PASADO

La I.A., como cualquier otra base de datos, solamente es depositaria del pasado, y puede dar la sensación de que predice el futuro solamente porque su “marco mental”, por decirlo así, es que el pasado siempre se repite. Dicho con otras palabras, solamente puede “predecir” situaciones que ya se hayan producido en un tiempo anterior, pero es absolutamente incapaz de generar una nueva situación, salvo que construya falsariamente una de esas situaciones combinando elementos del pasado, como por otra parte hacen tantas personas cuando recurren a su “experiencia”²⁴. Se trata del llamado heurístico de la representatividad²⁵, según el cual el ser humano toma sus decisiones de acuerdo con aquello que recuerda como exitoso, es decir, a aquello que ya conoce y que, en consecuencia, pertenece a su pasado.

El hecho es que las herramientas de IA aplicadas al proceso, tarde o temprano, van a recoger los indicios que el juez vaya estableciendo como probados y elaborará un relato de hechos que será propuesto al juez para integrar la redacción de su sentencia²⁶. Este proceso se podrá automatizar aún más cuando la herramienta sea capaz de confirmar por sí sola la existencia de los indicios, teniéndolos por probados, como sucederá con la prueba pericial²⁷ sobre todo, y hasta con la prueba documental²⁸. Más dificultad plantearán los interrogatorios²⁹, aunque la herramienta podrá descubrir contradicciones del testigo o lagunas en su relato que le lleven a proponer que su declaración fue poco creíble³⁰, todo ello después de haber sido utilizada tal vez para detectar con celeridad las preguntas impertinentes porque contengan información que se busque que el declarante asuma³¹, o porque le infundan emociones que alteren la objetividad que se busca en sus respuestas, entre

²⁴ Bentham, J., *Traité des preuves judiciaires*, Paris 1823, pp. 33 y ss. Allen, R.J.; Pardo, M.S., Relative Plausibility and Its Critics, *International Journal of Evidence & Proof*, vol. 23, no. 1-2, April 2019, p. 21.

²⁵ Tversky, A.; Kahneman, D. (1982). Judgments of and by Representativeness, en Kahneman; Slovic; Tversky (ed.). *Judgment under Uncertainty: Heuristics and Biases*, pp. 84 y ss.

²⁶ Nissan, E., “Digital technologies and artificial intelligence’s present and foreseeable impact on lawyering, judging, policing and law enforcement”, *AI & Society*, 2015, p. 17. Prakken, H., “Modelling Reasoning about Evidence in Legal Procedure”, *ICAIL* 2001, pp. 119 y ss.

²⁷ Metallo, V.N.A., The Impact of Artificial Intelligence on Forensic Accounting and Testimony - Congress Should Amend “the Daubert Rule” to Include a New Standard, *Emory Law Journal Online*, 69, 2019-2020, pp. 2039 y ss. Fournier, L. R., “The Daubert Guidelines: Usefulness, Utilization, and Suggestions for Improving Quality Control”, *Journal of Applied Research in Memory and Cognition*, 5, 2016, p. 308.

²⁸ Vid. Nieva Fenoll, J., *Inteligencia artificial y proceso judicial*, cit. pp. 90 y ss.

²⁹ Nieva Fenoll, J., *Inteligencia artificial y proceso judicial*, cit. pp. 80 y ss.

³⁰ Vid. Manzanero, A. L., *Psicología del testimonio*, Madrid 2008, pp. 27 y ss.

³¹ Manzanero, *Psicología del testimonio*, cit. p. 137.

otros casos que la herramienta podría detectar simplemente evaluando el uso del lenguaje.

Pero todo lo anterior, de momento, no ha llegado todavía y son simplemente ideas para desarrollar herramientas en este sentido en el futuro. Lo que sí que está aquí ya son los relatos de hechos de procesos sencillos que suelen ser idénticos y que dependen de la existencia de algunos indicios. Es más que conocido que la IAGen puede producir relatos, que genera precisamente escogiendo palabras de pasajes de otros relatos que ya tiene en su base de datos, alterando la gramática y vocabulario para que parezca un texto original. Esa misma funcionalidad podría emplearse —si se desea— en los procesos.

Tal vez el ejemplo más sencillo de ello es el desahucio por falta de pago³². En ese proceso se busca simplemente acreditar que el inquilino no ha pagado la renta. Es, de hecho, el único dato relevante si no existen situaciones de vulnerabilidad. Pues bien, la demanda exitosa será aquella que identifique ese dato del impago de manera confirmada, por ejemplo, con un documento bancario del que se desprenda incluso la cuantía total de la deuda no satisfecha en la cuenta corriente que se dispuso. De hecho, no hará falta hacer absolutamente nada más. La defensa, en estos casos, sólo tiene como salida intentar obstaculizar el desahucio, y no es nada fácil cuando lo que se enfrenta no es a un juez que puede ser sensible, sino, como vamos a ver, a una máquina sin emociones. En caso de que se utilice una herramienta de inteligencia artificial, el demandado puede intentar que el aplicativo, de manera automática, detecte un caso de *mora creditoris* si logra introducir en la aplicación un dato que confirme que el banco se ha negado a recibir la transferencia, siguiendo para ello la orientación que ofrezcan relatos de hecho de la jurisprudencia en que se haya declarado probada la existencia de esa mora del acreedor. Habitualmente será imposible en caso de que no haya ocurrido algo así, pero lo importante es entender que sólo con que se consiga crear una apariencia de intento de pago, es posible que el sistema detecte una oposición que puede llegar a ser exitosa, al menos durante un cierto tiempo, hasta que el aplicativo sea corregido para detectar la estratagema.

Imagínese ahora todo lo anterior en el marco de un proceso cuyos hechos sean más complejos. La labor del abogado consistirá en presentar los datos de manera que lo que se esté aportando coincida con un relato que se halle ya en la jurisprudencia anterior y, por tanto, en la base de datos del sistema. Exactamente igual que cuando se genera un *prompt*, existen reglas³³ para conseguir que sea eficaz en la búsqueda, es decir, que la herramienta capte lo que se le está preguntando, en este caso lo adecuado será presentar la realidad de manera que el aplicativo perciba como estadísticamente existente

³² Vid. Julià Pijoan, M., *La computerización del derecho, a partir del proceso y de los procedimientos judiciales*, cit. pp. 146 y ss.

³³ Entre otros muchos, Liu, P., Yuan, W., Fu, Y., Jiang, Z., Hayashi, H., Neubig, G., Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing, *ACM Computing Surveys*, Volume 55, 9, n. 195, pp. 1 y ss.

el relato que interese a cada parte. Y para ello será imprescindible conocer muy bien la jurisprudencia, pues con sus retazos, fundamentalmente, estará construida la base de datos.

Desde luego, la labor del creador del sistema será evitar que se le engañe, pero conseguir eso hasta las últimas consecuencias es muy complicado, y se trata de algo que habrá que ir actualizando de manera constante y periódica. Ya existen, por ejemplo, aplicativos que tratan de identificar los datos de un *curriculum*³⁴, resultando, sin embargo, muy difícil detectar la auténtica realidad que se esconde detrás de esos datos, lo que es aprovechado por no pocos pícaros. Esta realidad no será diferente con los procesos judiciales, lo que, como ya se ha dicho, obligará a ajustar y perfeccionar la herramienta constantemente con la ayuda de abogados y jueces que vayan detectando los errores del sistema. Es muy probable que ello provoque una extensión de la eficacia de los recursos de reposición para permitir que el propio juez pueda corregir esas disfunciones. Y, desde luego, puede que asistamos en el futuro a recursos de apelación cuya única finalidad, de hecho, sea la corrección, no de un error del juez, sino de un fallo del sistema.

Además, los abogados contarán con la ventaja sustancial de que el funcionamiento de la herramienta deberá ser en todo caso transparente³⁵, como se explicará al final, puesto que de lo contrario será imposible ejercer el derecho de defensa en esta época con estas nuevas herramientas. Esa transparencia también permitirá adecuar mejor, de manera preventiva, la introducción de datos por parte del abogado a aquello que “espere” la aplicación, es decir, a lo que sea coherente con todo lo que figure en sus bases de datos. Por tanto, la labor del abogado para lograr la “convicción” de la herramienta consistirá en obtener y preparar los datos de la forma que sea más adecuada para que la aplicación actúe correctamente, de manera que su funcionamiento pueda ser favorable al cliente. En el fondo, la estrategia es parecida a la elaboración de un *prompt*, pero puede tornarse más compleja en función del contenido del asunto.

Con todo y con eso, las perspectivas de defensa en ese escenario son bastante más escasas que en la actualidad, precisamente porque la máquina no se deja impresionar en absoluto por nuestra representación escénica ni por la retórica empleada en los diversos escritos. Como veremos a continuación, la máquina carece de emociones, lo que no quiere decir que sean del todo irrelevantes.

³⁴ Haddad, R., Mercier-Laurent, E., Curriculum Vitae (CVs) Evaluation Using Machine Learning Approach, en Mercier-Laurent, E., Kayalica, M.Ö., Owoc, M.L. (eds), *Artificial Intelligence for Knowledge Management. AI4KM 2021. IFIP Advances in Information and Communication Technology*, vol. 614. Springer 2021.

³⁵ Vid. STS (Sala 3ª) 1119/2025, 11-9-2025.

4. EL FIN DE LAS EMOCIONES

Aunque la *IA de la teoría de la mente*³⁶ dibuja un futuro realmente distópico, lo cierto es que un aplicativo de IA carece de emociones. Las emociones³⁷ son solamente una herramienta del cerebro para generar reacciones rápidas, muy instintivas, alejando al sujeto de un peligro inminente o aproximándolo a situaciones, lugares, personas, alimentos, etc., que favorecen su existencia y que, por ello, son percibidas como agradables. Para todo eso sirve el miedo y el afecto, igual que la tristeza, la alegría, el odio, el asco, o la sorpresa, aunque el recuento de cuáles sean las emociones básicas no parece ser pacífico entre los psicólogos³⁸. En todo caso, su concepto y virtualidad son evidentes.

Una herramienta de IA no necesita esas emociones, como no las precisa ninguna otra máquina, porque carecen de autoconsciencia y, por tanto, de instinto de supervivencia. Un aplicativo puede ser creado para que se proteja de amenazas externas, pero no reacciona con miedo, sino simplemente de manera mecánica cuando detecta aquello que su programador identificó en el listado de las amenazas. No aumenta su presión arterial ni su frecuencia cardíaca porque no tiene que combatir o huir corriendo porque no es un ser vivo³⁹. Para una aplicación de IA, responder a una amenaza es exactamente lo mismo que responder a una pregunta. Simplemente ejecuta la tarea para la que está programada. Tampoco se alegra la máquina cuando obtiene la información que busca su aplicativo. Simplemente la procesa y puede emitir sonidos o imágenes que a un ser humano nos evoquen alegría, como han hecho tantísimos videojuegos a lo largo de la historia.

Los abogados en particular —y no pocos litigantes— se han pasado milenios intentando captar las emociones del juez, como atestigua uno de los primeros cuentos escritos de la historia: “el campesino elocuente”, de hace más de 4.000 años⁴⁰. Las emociones son, de hecho, aquello que más fácilmente puede persuadir, porque se apela a la empatía del interlocutor, que es otra manera de decir que se intenta conectar con lo mejor de su persona, en principio. Que un juez vea en el litigante alguien que podría ser él mismo, es tal vez la cima del éxito en la defensa, aunque suele bastar con que la persona del litigante o su caso le suene a una historia conocida o, mejor aún, un caso similar en el que tal vez no se hizo justicia, que ese juzgador sí va a aprovechar para hacer ahora. En el primer caso —la historia conocida—, hará uso

³⁶ Cuzzolin F, Morelli A, Cirstea B, Sahakian BJ. Knowing me, knowing you: theory of mind in AI, *Psychological Medicine*. 2020, 50, 7, pp. 1057 y ss. Langley, C., Cirstea, B.I., Cuzzolin, F., Sahakian, B., Theory of Mind and Preference Learning at the Interface of Cognitive Science, Neuroscience, and AI: A Review, *Frontiers in Artificial Intelligence*, Vol. 5, 2022.

³⁷ Hülshoff, T., *Emotionen*, München 2006.

³⁸ Dubois, J., Adolphs, R., Neuropsychology: How Many Emotions Are There?, *Current Biology. A Cell Press journal*, 2015, vol. 25, 15..

³⁹ Guimarães, S., Moura, D., Vascular adrenoceptors: an update, *Pharmacological Reviews*, vol. 53, n. 2, pp. 319 y ss. Hoffman, B.B., *Adrenaline*, London 2013, p. 1.

⁴⁰ Assmann, J., *Ma'at. Gerechtigkeit und Unsterblichkeit im Alten Ägypten*, München 1995.

del heurístico de representatividad, como ya se dijo. En el segundo, del heurístico de accesibilidad⁴¹, que le llevará a tomar su decisión precisamente de acuerdo con aquello que recuerda con más facilidad.

El hecho de que alguien, en su defensa, aproveche todo lo anterior no es más que una manipulación, incluso cuando esa activación de las emociones es espontánea. Puede que la emoción refleje la realidad de lo que ha sucedido y debe ser enjuiciado, pero un juez debe cumplir con lo que ordena el ordenamiento jurídico sin que esas emociones le lleven a alterar su contenido, porque de esa forma estaría perdiendo su imparcialidad y, a la postre, tal vez prevaricando. El estudio de independencia y de la imparcialidad, aunque pocas veces se haya explicado así⁴², es un estudio de emociones⁴³. Las leyes han solido tener presente la realidad —o su apariencia— del afecto y el odio por las partes o por lo juzgado⁴⁴, pues constituye la base de todas las causas de recusación que existen. No está carente de consideración el miedo cuando se estudia la independencia⁴⁵, puesto que las presiones de otros órganos del Estado, o de poderes fácticos, lo que intenta infundir tantas veces es miedo. En la época de las monarquías absolutas era el miedo a ser despedidos por el rey, y justamente por ese devenir histórico, que provocó sonados casos de *lawfare* sobre todo en Inglaterra⁴⁶, nació el concepto de independencia judicial. Independencia del rey⁴⁷, naturalmente, en un primer momento. Pero como se ve, de nuevo lo que se tenía en cuenta era simplemente una emoción del juez⁴⁸.

Nada de eso sucede cuando se trata con una máquina, pues ni siente ni padece. Por ello, la estrategia de defensa ya no puede estar enfocada hacia las emociones, salvo que el ordenamiento recoja su consideración, como sucede con las circunstancias de vulnerabilidad en los desahucios, situación en la que en España se codificó⁴⁹, a mi modo de ver de manera desacerta-

⁴¹ Tversky, A., Kahneman, D., Availability: A heuristic for judging frequency and probability, en Kahneman, Slovic, Tversky (ed.), *Judgment under Uncertainty: Heuristics and Biases*, cit., pp. 163 y ss.

⁴² Vid. Nieva Fenoll, Trasfondo psicológico de la actividad judicial, en *Hacia el proceso judicial del siglo XXII*, Valencia 2025, pp. 213 y ss.

⁴³ Forza, Menegon, Rumiat, *Il giudice emotivo*, Bologna 2017, pp. 18-19.

⁴⁴ Vid. Ya desde hace muchos años la STEDH 8692/79 Piersack v. Bélgica, 1-10-1982, 30: “Whilst impartiality normally denotes absence of prejudice or bias, its existence or otherwise can, notably under Article 6 § 1 (art. 6-1) of the Convention, be tested in various ways. A distinction can be drawn in this context between a subjective approach, that is endeavouring to ascertain the personal conviction of a given judge in a given case, and an objective approach, that is determining whether he offered guarantees sufficient to exclude any legitimate doubt in this respect.”

⁴⁵ Tarr, G. A., *Without fear or favor: Judicial independence and judicial accountability in the States*, Stanford 2012, p. 91.

⁴⁶ Bill of Rights, 1689: “Whereas the late King James the Second, by the assistance of divers evil counsellors, judges and ministers employed by him, did endeavour to subvert and extirpate the Protestant religion and the laws and liberties of this kingdom;”

⁴⁷ La Declaración de independencia de los EEUU de 1776 se refirió particularmente a este punto: “He has obstructed the Administration of Justice, by refusing his Assent to Laws for establishing Judiciary powers”, “He has made Judges dependent on his Will alone, for the tenure of their offices, and the amount and payment of their salaries”.

⁴⁸ Vid. las reflexiones de Andrés Ibáñez, P., *El tercero en discordia*, Madrid 2015, p. 158.

⁴⁹ Art. 441.5 a 7 de la Ley de Enjuiciamiento Civil.

da, algunos aspectos de la desolación que podía sentir un juez a la hora de decidir sobre la ejecución de un desalojo. No es un caso inédito. Lo mismo sucede cuando se concede la justicia gratuita para personas con escasos recursos, lo que es más fácil de observar si se hace un estudio histórico de esta institución⁵⁰.

En consecuencia, es absurdo que el abogado realice un relato emocional cuando trate con una máquina, porque no es esa la forma de persuadirla. Pero como se señaló, eso no quiere decir que las emociones vayan a estar del todo ausentes. No es solamente que habrá que estar atentos a las circunstancias emocionales que el ordenamiento haya recogido como relevantes, como se acaba de decir, sino que, además, habrá que estar pendientes de esos destellos de humanidad que puedan encontrarse en la jurisprudencia, porque pueden formar parte de un relato de hechos que cuadre con el presentado en el proceso y que, a la postre, puedan otorgar la victoria a uno de los litigantes. Veamos cómo.

5. SÓLO INTERPRETACIONES JURÍDICAS COHERENTES CON LÍNEAS JURISPRUDENCIALES

Los juristas estamos ya muy acostumbrados a la elaboración de *prompts* para la búsqueda de jurisprudencia. Explicitar los términos precisos de la búsqueda ahorra muchísimo trabajo cuando se interactúa con esta herramienta de inteligencia artificial tan básica. Es importante ser conscientes de ello porque no será tan distinto el trabajo jurisprudencial que deberá hacerse en casos más complejos en los que se interactúe con una de estas herramientas de IAGen que sirva a la justicia.

Si queremos obtener el éxito en el caso, ya no se tratará de buscar abusivamente jurisprudencia que nos dé la razón, o lo parezca al menos, ni tampoco aislar los párrafos de la jurisprudencia que nos interesen. Lo más útil será encontrar las grandes líneas jurisprudenciales que se parezcan lo más posible a nuestro caso concreto, puesto que la máquina trabajará con ellas y aplicará las mismas decisiones que ya se hayan emitido anteriormente. Al menos en un primer momento, será muy útil identificar las palabras que más repitan los tribunales al tratar con estos asuntos, para que la herramienta incluya el caso que se proponga como un guante, lo que tiene una importancia decisiva sobre todo en la elaboración del relato de hechos.

Sin embargo, más adelante, y puede que ese momento no tarde tantísimo, no será ya tan importante identificar las palabras concretas a incluir en el relato. Más bien se tratará de ser capaz de condensar la idea principal, lo que obligará a un análisis de la esencia de lo que diga la jurisprudencia. Con todo, será difícil que una herramienta trabaje con esas variaciones del lenguaje con toda perfección. Detrás de toda la tecnología que comporta la IA, se encuen-

⁵⁰ Vid. Partida III, Tít. VI, Ley 6.

tra la minería de datos⁵¹. Se trata de un trabajo muchas veces automatizado de elaboración de patrones, pero que requiere irremediablemente de muchos seres humanos para que la clasificación de los datos sea correcta. Ello precisa un esfuerzo inmenso que consume recursos energéticos extraordinarios y que en no pocas ocasiones provoca explotación laboral⁵². Por ello, esa labor debe ser lenta y actualizarse constantemente. Requerirá de muchísimos juristas que trabajen en ese sector y clasifiquen la información de manera óptima, anticipándose a errores y corrigiendo los que vayan surgiendo. Precisamente, el hecho de que la labor emplee tantos recursos humanos, puede provocar discrepancias inevitables que poco a poco irán siendo resueltas, aunque llevará su tiempo, naturalmente.

Con todo, existe un problema que es preciso prever ya desde ahora. Conforme avance el uso de estas herramientas, la jurisprudencia innovadora será cada vez menor; lo que, desde luego, aumentará el cumplimiento del derecho a la igualdad a los tribunales, pero al mismo tiempo puede provocar el anquilosamiento que tantas veces se observa en los sistemas del *Common Law* que trabajan con precedentes, así como las tremendas dificultades para que los tribunales superiores quieran operar un *overruling*, incluso pese a las más recientes tendencias al efecto⁵³. Además, debe tenerse en cuenta que la IA alcanzará a todos los niveles de la jurisdicción, lo que provocará que los tribunales supremos, sin ir más lejos, cada vez sientan una menor necesidad de modificar la jurisprudencia que ya estará contenida en los aplicativos, por pura comodidad. Es cierto que también es posible que tengan mucho menos trabajo; en la actualidad existe algunas veces la expectativa de que el Tribunal Supremo se separe en aquel caso concreto de su jurisprudencia, si es que ya no lo hicieron los tribunales inferiores. Pero en un contexto de IA, esa posibilidad puede ser marginal ante la enorme fuerza que tendrá la plena consciencia de cuál es la jurisprudencia aplicable al caso, ya prevista en la herramienta IA. Por consiguiente, al desaparecer la citada expectativa, es factible que se favorezca algo que viene siendo un sueño de los tribunales desde hace muchísimo tiempo: la descarga de asuntos.

Y es que si la solución es previsible, es más que probable que acabe existiendo una herramienta que informe al ciudadano de cuál es la solución de su caso en muchos supuestos, lo que prevendrá, sin duda, conflictos. También hará que los abogados tengan menos trabajo, naturalmente, pero justamente en ese escenario serán mucho más que necesarios los letrados que sean más creativos, es decir, aquellos que sean capaces de visualizar una solución más

⁵¹ Carroll, M.W., Copyright and the Progress of Science: Why Text and Data Mining Is Lawful, *UC Davis Law Review*, 53, 2019-2020, pp. 899 y ss.

⁵² Rani, U., Dhir, R.K., The Artificial Intelligence illusion: How invisible workers fuel the “automated” economy, *International Labour Organisation*. 10-12-2024, <https://www.ilo.org/resource/article/artificial-intelligence-illusion-how-invisible-workers-fuel-automated>. Vid. también Bostanci, M., Yilmaz, Ö, Labour and exploitation processes in artificial intelligence: example of digital taylorism in data labelling, *Toplum ve Kültür, Journal of Social & Cultural Studies*, 15-05-2025.

⁵³ Es interesante el estudio del llamado *prospective overruling*. Vid. Passanante, L., *Il precedente impossibile*, Torino 2018, pp. 273 y ss.

imaginativa que la establecida por la jurisprudencia y que, por tanto, haga imprescindible la alteración del algoritmo. No sería mala idea que las Facultades de Derecho empezaran a modificar, no tanto sus planes de estudio, sino sus sistemas de evaluación, sobre todo, y la preparación de esas pruebas, a fin de responder a esa nueva realidad. En la actualidad se evalúa fundamentalmente la memoria, demasiadas veces acrítica. En el futuro, partiendo del conocimiento básico de la realidad del ordenamiento, lo que realmente se debiera valorar positivamente es la capacidad de ir más allá, porque solamente de ese modo podrá evolucionar el Derecho, contando con este tipo de juristas tanto entre la abogacía como, desde luego, en los tribunales.

6. LA CAJA NEGRA DEBE SER BLANCA

Ha existido, y aún existe, cierta incertidumbre sobre la transparencia del funcionamiento de las herramientas de IA, en el sentido de conocer la configuración de los algoritmos de cada aplicativo. Los fabricantes han solido negarse a desvelar el funcionamiento de sus ingenios, básicamente por temor a ser plagiados y dejar de ganarse la vida con ellos. Es por ello que nadie conoce todavía, a día de hoy, el funcionamiento real de COMPAS, sin ir más lejos⁵⁴.

Sin embargo, no es tal vez necesario conocer tanto, entre otras razones porque el funcionamiento interno de las herramientas no está al alcance de cualquier usuario, igual que tampoco tenemos conocimiento preciso hasta las últimas consecuencias de cómo funciona un automóvil, y sin embargo lo utilizamos. La clave, por tanto, no está exactamente ahí.

Lo esencial es que el funcionamiento de este tipo de sistemas sea público, porque de lo contrario sería imposible ejercer el derecho de defensa hasta las últimas consecuencias en caso de que la aplicación se emplee por los jueces. Con todo, lo que es más importante no es tanto conocer con precisión ese funcionamiento, cuestión habitualmente inasequible, sino cuáles son las líneas de comportamiento de la herramienta en el sentido de que sus decisiones sean previsibles.

El problema principal de COMPAS, y de otras herramientas similares, es que maneja una serie de circunstancias que son extraordinariamente sensibles, como el lugar de residencia, el nivel de estudios, la capacidad atencional, la raza o los consumos privados de una personas⁵⁵. Y todo ello para saber si alguien posee un riesgo de reincidencia en un delito violento... El funcionamiento interno de los algoritmos es muy complejo, pero lo que sí que sería imprescindible en este caso es saber en qué medida se aprecian las citadas circunstancias para establecer un pronóstico de riesgo, porque

⁵⁴ Ran. Xi, Systems approach to shedding sunlight on A.I. black boxes, *Hofstra Law Review*, 53, 2, 2025, pp. 451 y ss.

⁵⁵ Northpointe, *Practitioners Guide to COMPAS*, 17-8-2012, pp. 23 y ss.

ello sí es conocido por los que elaboraron la herramienta. En el fondo, este tipo de aplicativos intentan emular la labor de los psicólogos de la personalidad⁵⁶ —tantas veces responsables de estos aplicativos—, que con las entrevistas cognitivas no dejan de depender en una medida relativamente amplia del razonamiento experiencial, es decir, de la intuición⁵⁷. Lo que sería preciso saber es si la herramienta simplemente suma porcentajes en cuanto a la cantidad de personas agresivas de una determinada característica en las que luego se ha observado la comisión de un acto violento. Si es eso solamente, ciertamente es decepcionante, porque implicaría contemplar cómo desde las perspectivas científicas se está nuevamente llamando a la puerta del sistema de valoración legal de la prueba.

Sea como fuere, lo que necesitamos es una previsibilidad de las decisiones judiciales, y para ello sería utilísimo que sus fabricantes, además de publicar el funcionamiento de su sistema, elaboraran un manual completo de las respuestas del sistema ante los casos planteados, porque ello sería lo que orientaría la labor de los abogados, aunque sea para corregir el sistema. Y como ya se dijo antes, evitaría no pocos conflictos que actualmente se plantean ante la jurisdicción como si el juez fuera un oráculo, simplemente probando suerte.

En conclusión, la transparencia en el funcionamiento de estas herramientas es fundamental. Como ya ha declarado muy acertadamente la jurisprudencia⁵⁸, la administración pública no puede utilizar herramientas cuyo funcionamiento sea opaco, lo que debe ser aplicado indudablemente a la Justicia.

7. BIBLIOGRAFÍA

- Allen, R.J.; Pardo, M.S., Relative Plausibility and Its Critics, *International Journal of Evidence & Proof*, vol. 23, no. 1-2, April 2019, p. 21.
- Andrés Ibáñez, P., *El tercero en discordia*, Madrid 2015.
- Andrés-Pueyo, A., Arbach-Lucioni, K., Redondo, S., The RisCanvi: A New Tool for Assessing Risk for Violence in Prison and Recidivism, en *Handbook of Recidivism Risk/Needs Assessment Tools*, Chichester 2018, pp. 255 y ss.
- Andrews, A.; Clawson, R.A.; Gramig, B.M., Finding the Right Value: Framing Effects on Domain Experts, *Political Psychology*, vol. 38, 2, Abril 2017, pp. 261 y ss.
- Assmann, J., *Ma'at. Gerechtigkeit und Unsterblichkeit im Alten Ägypten*, München 1995.
- Azón, *Summa Azonis*, Lib. IV, 19 (de probationibus). punto 21, p. 318.
- Bentham, J., *Traité des preuves judiciaires*, Paris 1823,
- Bornstein, A. M., Are algorithms building the new infrastructure of racism?, *Nautilus*, 21-12-2017, <http://nautil.us/issue/55/trust/are-algorithms-building-the-new-infrastructure-of-racism>.

⁵⁶ Andrés-Pueyo, A., Arbach-Lucioni, K., Redondo, S., The RisCanvi: A New Tool for Assessing Risk for Violence in Prison and Recidivism, en *Handbook of Recidivism Risk/Needs Assessment Tools*, Chichester 2018, pp. 255 y ss.

⁵⁷ Cfr. Julià Pijoan, M., RisCanvi. ¿Un protocolo de discriminación algorítmica?, en AAVV (coord.). Bueno de Mata), *Justicia y derecho en datos*, 2023, pp. 123 y ss.

⁵⁸ STS (Sala 3ª) 1119/2025, 11-9-2025.

- Bostanci, M., Yilmaz, Ö, Labour and exploitation processes in artificial intelligence: example of digital taylorism in data labelling, *Toplum ve Kültür, Journal of Social & Cultural Studies*, 15-05-2025.
- Brandom, R., Mechanical Turkers may have out-predicted the most popular crime-predicting algorithm, *The Verge*, 17-1-2018, <https://www.theverge.com/2018/1/17/16902016/compas-algorithm-sentencing-court-accuracy-problem>.
- Carroll, M.W., Copyright and the Progress of Science: Why Text and Data Mining Is Lawful, *UC Davis Law Review*, 53, 2019-2020, pp. 899 y ss.
- Cuzzolin F, Morelli A, Cîrstea B, Sahakian BJ. Knowing me, knowing you: theory of mind in AI, *Psychological Medicine*. 2020, 50, 7, pp. 1057 y ss.
- Degli Ubaldi, Baldo, *Practica Baldi*, 1521, folio 39 vuelto. Jaumar Carrera, *Práctica forense*, 1840, pp. 39-45.
- Descartes, R., *Discurso del método*, Madrid 1986, p. 35.
- Druckman, J. N., Political Preference Formation: Competition, Deliberation, and the (Ir)relevance of Framing Effects, *American Political Science Review*, nov. 2004, vol. 98, 4, pp. 671 y ss.
- Dubois, J., Adolphs, R., Neuropsychology: How Many Emotions Are There?, *Current Biology. A Cell Press journal*, 2015, vol. 25, 15,.
- Engel, C., Linhardt, L., Schubert, M., Code is law: how COMPAS affects the way the judiciary handles the risk of recidivism, *Artificial Intelligence and Law*, 2025, pp. 383 y ss.
- Forza, Menegon, Rumiati, *Il giudice emotivo*, Bologna 2017.
- Fournier, L. R., "The Daubert Guidelines: Usefulness, Utilization, and Suggestions for Improving Quality Control", *Journal of Applied Research in Memory and Cognition*, 5, 2016, p. 308.
- Furnham, A.; Chu Boo, H., A literature review of the anchoring effect, *The Journal of Socio-Economics*, vol 40, 1, Feb. 2011, pp. 35 y ss.
- Fussell, S., Study Finds Crime-Predicting Algorithm Is No Smarter Than Online Poll Takers, *Gizmodo*, 18-1-2018, <https://gizmodo.com/study-finds-crime-predicting-algorithm-is-no-smarter-th-1822173965>.
- Golden, W., ChatGPT: A trusted source?, *Irish Journal of Technology Enhanced Learning*, vol. 7, 2, pp. 113 y ss.
- Gregorcic, B., Pendrill, A.-M., ChatGPT and the frustrated Socrates, *Physical Education*, 2023, pp. 1 y ss.
- Guimarães, S., Moura, D., Vascular adrenoceptors: an update, *Pharmacological Reviews*, vol. 53, n. 2, pp. 319 y ss. Hoffman, B.B., *Adrenaline*, London 2013, p. 1.
- Haddad, R., Mercier-Laurent, E., Curriculum Vitae (CVs) Evaluation Using Machine Learning Approach, en Mercier-Laurent, E., Kayalica, M.Ö., Owoc, M.L. (eds), *Artificial Intelligence for Knowledge Management. AI4KM 2021. IFIP Advances in Information and Communication Technology*, vol. 614. Springer 2021.
- Hicks, M.T., Humphries, J., Slater, J., ChatGPT is bullshit, *Ethics and Information Technology*, 2024, p. 38.
- Hülshoff, T., *Emotionen*, München 2006.
- Julià Pijoan, M., *La computerización del derecho, a partir del proceso y de los procedimientos judiciales*, Madrid 2024.
- Julià Pijoan, M., RisCanvi. ¿Un protocolo de discriminación algorítmica?, en AAVV (coord.. Bueno de Mata), *Justicia y derecho en datos*, 2023, pp. 123 y ss.
- Khosrowi, D., Van Basshuysen, P., Making a Murderer: How Risk Assessment Tools May Produce Rather Than Predict Criminal Behavior, *American Philosophical Quarterly*, 2024, 61 (4), pp. 309 y ss. <https://scholarlypublishingcollective.org/uip/apq/article/61/4/309/391697>

- Kim, T., Song, H., The Effect of Message Framing and Timing on the Acceptance of Artificial Intelligence's Suggestion, *Association for Computing Machinery*, 2020.
- Laizure, S. C., Caution: ChatGPT Doesn't Know What You Are Asking and Doesn't Know What It Is Saying, *The Journal of Pediatric Pharmacology and Therapeutics*, 14-10-2024, <https://jppt.kglmeridian.com/view/journals/jppt/29/5/article-p558.xml>.
- Langley, C., Cirstea, B.I., Cuzzolin, F., Sahakian, B., Theory of Mind and Preference Learning at the Interface of Cognitive Science, Neuroscience, and AI: A Review, *Frontiers in Artificial Intelligence*, Vol. 5, 2022.
- Larson, D. A., Artificial Intelligence: Robots, Avatars, and the Demise of the Human Mediator, *Ohio State Journal on Dispute Resolution*, Vol. 25, 2010, pp. 117-118.
- Larson, J., Mattu, S., Kirchner, L., Angwin, J., How We Analyzed The Compas Recidivism Algorithm, *Propublica*, 23-5-2016, <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>.
- Liu, P., Yuan, W., Fu, Y., Jiang, Z., Hayashi, H., Neubig, G., Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing, *ACM Computing Surveys*, Volume 55, 9, n. 195, pp. 1 y ss.
- Manzanero, A. L., *Psicología del testimonio*, Madrid 2008, pp. 27 y ss.
- Metallo, V.N.A., The Impact of Artificial Intelligence on Forensic Accounting and Testimony - Congress Should Amend "the Daubert Rule" to Include a New Standard, *Emory Law Journal Online*, 69, 2019-2020, pp. 2039 y ss.
- Myers, D. G., *Intuición. El poder y el peligro del sexto sentido*, trad. de Guillermo Solana de *Intuition: its power and perils*, New Haven y Londres 2002, Barcelona 2003.
- Nieva Fenoll, J. *Inteligencia artificial y proceso*, Madrid 2018.
- Nieva Fenoll, J. Trasfondo psicológico de la actividad judicial, en *Hacia el proceso judicial del siglo XXII*, Valencia 2025, pp. 213 y ss.
- Nissan, E., "Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement", *AI & Society*, 2015, p. 17.
- Northpointe, *Practitioners Guide to COMPAS*, 17-8-2012.
- Passanante, L., *Il precedente impossibile*, Torino 2018, pp. 273 y ss.
- Plous, S., *The psychology of judgment and decision making*, 1993.
- Prakken, H., "Modelling Reasoning about Evidence in Legal Procedure", *ICAIL* 2001, pp. 119 y ss.
- Quattrocchio, S., *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, Springer 2020.
- Ran, Xi, Systems approach to shedding sunlight on A.I. black boxes, *Hofstra Law Review*, 53, 2, 2025, pp. 451 y ss.
- Rani, U., Dhir, R.K., The Artificial Intelligence illusion: How invisible workers fuel the "automated" economy, *International Labour Organisation*. 10-12-2024, <https://www.ilo.org/resource/article/artificial-intelligence-illusion-how-invisible-workers-fuel-automated>.
- Seymour, W., Detecting Bias: Does an Algorithm Have to Be Transparent in Order to Be Fair?, 2018, <https://ir.shef.ac.uk/bias/pdf/seymour2018detecting.pdf>
- Tarr, G. A., *Without fear or favor. Judicial independence and judicial accountability in the States*, Stanford 2012, p. 91.
- Tashea, J., Risk-Assessment Algorithms Challenged In Bail, Sentencing And Parole Decisions, *ABA Journal*, Marzo 2017, http://www.abajournal.com/magazine/article/algorithm_bail_sentencing_parole.
- Tversky, A., K., Daniel, The Framing of decisions and the psychology of choice, *Science*, 211, 1981, (4481), pp. 453 y ss.

- Tversky, A., Kahneman, D., Availability: A heuristic for judging frequency and probability, en Kahneman, Slovic, Tversky (ed.), *Judgment under Uncertainty: Heuristics and Biases*, cit., pp. 163 y ss.
- Tversky, A., Kahneman, D., Judgments of and by Representativeness, en Kahneman, Slovic, Tversky (ed.), *Judgment under Uncertainty: Heuristics and Biases*, pp. 84 y ss.
- Tversky, A., Kahneman, D., Judgment under Uncertainty: Heuristics and Biases, *Science*, New Series, Vol. 185, n. 4157, Sep. 27, 1974, p. 1128.

PREPRINT

