

CONGRESO NACIONAL DE DERECHO PROCESAL

Comisión de Derecho Procesal Penal I

Inteligencia artificial: proceso penal y derechos fundamentales
(arts. 75, inc. 19 y 22, CN)

“Inteligencia Artificial y valoración de la prueba: problemas y desafíos en relación al derecho a la defensa.”

Autor: Ian Silberberg

Síntesis: Este trabajo analiza la introducción de sistemas de IA en el proceso penal, puntualmente en la valoración probatoria. Se argumenta que esto no necesariamente implicaría un perjuicio para la garantía penal del derecho a la defensa.

Postula para participar del concurso por el premio de AADP.

Sumario: Este trabajo analiza y defiende la introducción de sistemas de Inteligencia Artificial en el proceso penal (puntualmente en la valoración probatoria), y explora cómo impacta esto en las garantías procesales. Se examina el recorrido histórico de la valoración probatoria y nuestro procedimiento para entender cómo interactúan las garantías con estos nuevos agentes analíticos y, finalmente, se abordan los problemas y desafíos actuales de la IA en torno a este tipo de funciones.

I – Introducción:

En este momento, mientras estamos aquí reunidos, hay decenas de miles de máquinas formadas por aleaciones de carbono, impulsadas por señales eléctricas entre distintos nódulos intercomunicados, tomando decisiones judiciales y cambiándole la vida -para bien o para mal- a millones de personas. Estas máquinas, que se valen de sensores que responden a estímulos como pueden ser los lumínicos o sonoros, tienen sobre ellas el peso de la última herramienta coercitiva del Estado de derecho. Valoran prueba, e imponen penas y sanciones bajo su propia lógica, cometiendo errores, a veces, por exceso en la carga de información, y otras, por una tan conocida obsolescencia programada. Hablamos, claro está, del juez humano.

Las siguientes páginas no intentarán abordar soluciones mágicas, ni hacer futurología sin sentido respecto de las limitaciones mencionadas. La propuesta ofrecida a continuación apunta a un ejercicio de reflexión con la mirada puesta en el cada vez más cercano -y a veces ya presente- choque entre el juez descrito y un nuevo tipo de magistrado inorgánico al que se le atribuye el nombre de “juez máquina” ¹. Un choque que, en incontables escenarios tanto presentes como futuros, afecta al proceso penal y las garantías del acusado dentro del mismo. Caso clave y ejemplificativo de esto es lo atinente a una nueva valoración probatoria realizada por inteligencia artificial. Me centraré en este punto justamente porque de él deviene la mayor tarea de procesamiento de datos presente en nuestro proceso, y si

¹ Dentro de esta categoría no entiendo únicamente a los programas informáticos que actúan tomando decisiones judiciales de forma “autónoma”, tal cual lo harían sus pares humanos, sino también a aquellas herramientas que asisten e influyen significativamente al juez de carne y hueso en su rol.

bien desde el campo de las neurociencias, día a día, entendemos lo extraordinarios que son nuestros cerebros, no pueden competir contra los nuevos agentes digitales que procesan cantidades enormes de información en un tiempo ínfimo y con un margen de error mucho más pequeño². Casos de valoración digital ya existen en diversos países del mundo -tales como Alemania y Estados Unidos- y no representan una obra de ciencia ficción sino una discusión obligatoria por un fenómeno que avanza con una velocidad mucho mayor que con la que se mueven nuestros debates académicos.

Antoine Garapon y Jean Lassègue afirman, intentando dar sentido a esta nueva revolución tecnológica, social, cultural, y gráfica, que “Lo digital actúa como un reorganizador simbólico que afecta todos los sectores de la vida.”³. En consecuencia de ello, el derecho -y puntualmente el derecho penal- sufrirá, inevitablemente, la misma suerte que tantos otros fenómenos culturales creados por nuestra especie. Desde la gran -y tardía- digitalización aplicada a nuestra justicia nacional, hasta los más avanzados softwares que emplean inteligencia artificial para predecir las probabilidades de reinserción de los detenidos en Estados Unidos ⁴, todo parecería indicar que estamos iniciando nuestro recorrido por una nueva era, y el derecho penal no podrá escapar -ni, tal vez, debería- al inminente cambio paradigmático que se avecina. Ante este viraje, la estructura procesal del derecho penal, junto con las garantías que los distintos modelos occidentales procuran proteger, parecería ser una de las aristas más comprometidas.

Nuestra forma de concebir el proceso penal, junto con los consensos que damos por sentados, serán tomados por sorpresa por una serie de modificaciones que llevarán a procedimientos automáticos, jueces-máquina y herramientas digitales de resolución de conflictos, que tendrán como consecuencia directa la transformación del proceso desde sus cimientos y

² OpenAI. "GPT-4 Technical Report". OpenAI, 2023, pp. 1-10.

³ Antoine Garapon y Jean Lassègue, “Justicia Digital. Revolución gráfica y ruptura antropológica”, Editorial Jusbares, Poder Judicial de la Ciudad de Buenos Aires, Buenos Aires, 2024.

⁴ COMPAS, por ejemplo “...es un programa hecho para ayudar a los jueces a determinar si un acusado debería continuar bajo el régimen de prisión preventiva mientras aguarda su juicio.” (Karen Hao, "AI is more fair than a judge in criminal risk assessments", MIT Technology Review, October 17, 2019, <https://www.technologyreview.com/2019/10/17/75285/ai-fairer-than-judge-criminal-risk-assessment-algorithm/>)

pilares fundamentales. En el centro de esta mutación podemos encontrar al derecho a la defensa en relación con la valoración probatoria que, de quedar en manos de una máquina incapaz de dar explicaciones que transparenten la “caja negra” -que resulta de la mayoría de los desarrollos lógicos que se expresan a partir de estos sistemas-, podríamos encontrarnos ante un gran argumento para desenchufar toda incursión hacia el futuro.

Y es que muchas de las nuevas herramientas que van surgiendo en esta época de cambios, como afirma Daniel Pastor, analizando el cercano caso del proceso digital por infracciones detectadas a partir de la máquina de foto-multas, no hacen mejor lo que ya se viene haciendo en términos de justicia, sino que hacen justicia de otro modo, una *justicia mejor*⁵. En esta frase creo que se encuentra concentrado el campo más rico para explorar dentro de lo que nos compete: todo indicaría que, la meta final, probablemente no sea un mejoramiento en el proceso como lo conocemos, sino la instauración de uno nuevo, al que, con un poco de suerte y rapidez, tendremos el privilegio de estudiar y, en alguna medida, hasta encauzar.

Pastor menciona también que en este sistema automático, varios de los instrumentos más sagrados de nuestro orden jurídico quedan de lado, ello debido a que “...los derechos que esos instrumentos debían garantizar están incluso mejor asegurados por la propia máquina de multar.”⁶. Muchos colegas, al día de hoy, se verán aterrorizados o espantados por la idea de un mundo en el que prescindamos de garantías fundamentales en virtud de un nuevo proceso que tiende a aproximarse a la perfección y a extinguir el margen de error, pero permítanme encender la luz y mostrar que el monstruo en el rincón de la habitación es en realidad el perchero -aunque tal vez la preocupación esté simplemente mal direccionada-, y que lo señalado por el autor, lejos de ser un motivo para escandalizarse, no es nada más que la evolución lógica del régimen del proceso penal cuando se le aplica este tipo de presiones tecnológicas. Para ver dicha progresión, encuentro útil dar un paso atrás y observar el devenir acumulado de nuestro proceso, haciendo foco en los cambios sufridos por los sistemas de valoración probatoria.

⁵ Pastor, Daniel. "Inteligencia artificial y poder judicial: hacia otra idea de justicia". En Juan Gustavo Corvalán (dir.), Tratado de Inteligencia Artificial y Derecho, Tomo III. Thomson Reuters La Ley, Ciudad Autónoma de Buenos Aires, 2021, pp. 63-66.

⁶ *Ibid*

II – Un poco de historia:

Julio Maier realizó un atento análisis acerca de la forma en la que se debería abordar la historia del proceso penal en base a la no-linealidad de su desarrollo a lo largo del tiempo⁷. Y es que la evolución de las garantías penales y los modelos de enjuiciamiento se pueden entender como un constante y paulatino cambio paradigmático que no siempre responde al efecto del paso del tiempo como flecha que apunta en una única dirección. Aquí, la distinción entre modelos teóricos (es decir, los modelos de enjuiciamiento en su estado más “puro”) y modelos históricos (el modelo aplicado en determinado período, y en determinado lugar) es clave para entablar nuestro diálogo con el pasado, y nos ayudará no solamente a entender el presente, sino también a pensar -y tal vez moldear- el futuro. Me enfocaré, por una cuestión práctica, en los modelos históricos que se manifestaron en cuatro etapas diferentes de la humanidad.

Recorriendo los pasillos históricos tal y como lo propone Maier, nos debemos una primera parada en la Europa previa al siglo VII, donde el Derecho Germano había instaurado en distintos pueblos (bárbaros, como los llamarían sus vecinos latino-parlantes) un sistema acusatorio privado que tenía entre sus principales características la gran connotación ritual y espiritual que es le atribuía al proceso. El autor propone en sus páginas que *“La prueba no se dirigía a demostrar un hecho con objetividad, es decir, a establecer la verdad sobre la hipótesis objeto del proceso, ni tan siquiera a formar la convicción del tribunal sobre la responsabilidad o inocencia del acusado, sino que representaba un medio de lucha entre las partes, a través del cual, por actos sacramentales, cada una de ellas daba mayor fuerza a sus afirmaciones subjetivas sobre el derecho que invocaba.”*⁸ Entre las pruebas más comunes se encontraban el duelo, el juramento y las ordalías. Podríamos decir que, en este tipo de modelos, la valoración probatoria era ejercida, en muchos casos, por dioses o entes sobrenaturales, a los que

⁷ Julio B. J. Maier, “Derecho Procesal Penal. Tomo I: Fundamentos”, Editorial Ad-Hoc, Buenos Aires, 2023.

⁸ Julio B. J. Maier, “Derecho Procesal Penal. Tomo I: Fundamentos”, Editorial Ad-Hoc, Buenos Aires, 2023

pedirle algún nivel de “explicabilidad”⁹ del sistema sería, cuanto poco, infructífero -esto es, claro, si esperamos una respuesta de algún tipo que no tenga que ver con la interpretación de un relámpago o la manifestación de espíritus en una fogata-.

Como paso siguiente y continuando con esta “evolución racional” del proceso penal, cabe analizar el procedimiento griego -anterior en términos cronológicos-, en el que, si bien la tortura y los “juicios de dios” continuaban siendo utilizados como medios de prueba, la tarea de valoración se encontraba en manos de un juez -perteneciente a la misma comunidad que el acusado- que la ejercía a través del método de la íntima convicción, votando finalmente -sin necesidad de explicar su decisión- para determinar si el acusado era considerado culpable o no del hecho imputado. Podemos empezar a hablar ahora sí, de un proceso mental con cierto grado de lógica en el que -valorando prueba que, a nuestros ojos, lejos está de tener cualquier condimento lógico- humanos valoraban los elementos presentados (muchas veces de índole divina), debiendo realizar algún tipo de trabajo mental o pasos ordenados de razonamiento para llegar a una conclusión sobre el caso.

Podemos diferenciar este último sistema del romano -sin ahondar demasiado en un análisis que podría hacerse sumamente extensivo sobre las tres etapas de régimen político y procesal penal de dicha civilización- principalmente en el aspecto mítico de la prueba. Los romanos abandonan la idea de la prueba divina, las ordalías y las pruebas para pasar a un sistema de conocimiento histórico y objetivo para la reconstrucción de un hecho acaecido anteriormente. Así, el derecho deja -sólo de momento- a dios afuera del recinto, atando tanto la producción como la valoración probatoria al plano terrenal.

Habiéndose plantado sus semillas en el régimen de la *cognitio extra ordinem* romana, la inquisición volvió a dar otro giro en términos de régimen procesal penal, y nuevamente, la valoración de la prueba no estuvo exenta de ello. El condenado perdió todos los rasgos que lo acercaban a una

⁹ La “explicabilidad” de una IA es la capacidad que tiene un modelo de estas características de ayudar a los creadores y usuarios de esta a entender cómo arribó a una decisión en particular. (“Expleinable AI”, IBM, <https://www.ibm.com/topics/explainable-ai>)

aproximativa igualdad con su contraparte acusadora, y los oficiales del estado inquisitorial tomaron las riendas de una investigación que intentaba culminar con la confesión por parte del reo. Tanto la habilitación al juzgador para aplicar torturas al acusado, como para condenar al mismo, estaban dadas por un sistema de prueba tasada que, a partir de una enumeración puntualizada de las condiciones necesarias para tomar ciertas decisiones, permitían al juez proceder de una u otra manera.

Como respuesta a los regímenes absolutistas que controlaron el escenario europeo durante más de diez siglos, los movimientos culturales e intelectuales del iluminismo idearon una estructura procesal penal que devolvía al individuo su dignidad como parte en el entramado judicial. Se establecieron límites al Estado persecutor y se consolidaron las bases para dar paso a un sistema mixto que justamente buscaba reformar el modelo anterior, para dar paso a una estructura de corte más racional, que introduciría tanto enjuiciamientos por jurados como por cueros jurisdiccionales unipersonales. Es en este período histórico en el que surge el método de valoración probatoria de la sana crítica racional (aplicado principalmente en los casos resueltos por jueces profesionales) en el que los jueces y tribunales basan sus decisiones en una evaluación lógica y razonada de los elementos probatorios, siguiendo las reglas de la lógica, la experiencia y los conocimientos científicos aplicables al caso concreto. Un modelo en el que la “fundamentación” -podríamos llamarla también “explicabilidad”- de la sentencia emana directamente de ese sistema de valoración, que a su vez se propone como garantía de alto calibre para el acusado, quien debe poder acceder a esa explicación lógica que derivó en el decisorio judicial sobre su persona, y que puede ser atacado por la defensa para conseguir se abra paso a una segunda instancia revisora.

En estas últimas líneas se puede ver cómo el período iluminista, junto con su modelo estructural del régimen procesal, funcionaron como “estructuras madres” para las garantías que hoy vemos reflejadas en, por ejemplo, nuestra Constitución Nacional. Derecho a la defensa, derecho al recurso, garantía de imparcialidad, y muchísimos otros pilares fundamentales se alzan a partir de esta etapa. Aunque me veo inclinado a reforzar la idea de que lo que intento mostrar con este desarrollo histórico,

es que hay una tendencia hacia la racionalización y la implementación de procesos lógicos más eficientes para la valoración de las pruebas que se producen para cada caso. Los jueces están, en un sentido genérico del término, *ejecutando una serie de pasos lógicos* -también llamados algoritmos- para analizar *una gran cantidad de datos* que se presentan desperdigados en el plano real y que debe *estructurar para devolver una respuesta -o “output”* en forma de sentencia fundamentada.

III – Garantía de futuro:

Nuestro sistema penal, hijo de los procesos iluministas y de la racionalización de las estructuras políticas que se inició en el S.XVII, llega a nosotros con un aprendizaje tan acelerado como intenso en términos de desconfianza del poder, que conlleva un intento extremadamente alto y costoso -un sacrificio que el pasado nos insiste en pagar- por impedir que el aspecto persecutorio de ese poder se desbande, incurriendo así en prácticas oscuras a las que nuestra especie ha acordado no retroceder.

En ese marco, las garantías funcionan como sistema de protección ante la arbitrariedad del poder humano concentrado y con el fin de proteger la dignidad del individuo. Sin ellas, no solo nos expondríamos a las prácticas de un Estado que en determinadas circunstancias ha sabido convertirse en un monstruo incontenible, sino que, además, el proceso sería muchísimo más permeable a resultados disvaliosos o arbitrarios por culpa del error humano.

Es esperable que nuestros cerebros, con su limitada velocidad de procesamiento de datos, y acotada memoria, brinden resultados decepcionantes o, directamente, fallen en analizar grandes cantidades de datos. Esto se presenta como una cuestión de notable importancia a la hora de hablar de valoración probatoria. Es en este momento de nuestro proceso penal en el que la efectividad para analizar grandes cantidades de información se vuelve clave y hasta necesaria para que resulten de esto la menor cantidad de resultados equivocados. Ahí es donde entran las herramientas de Inteligencia Artificial como un aliado indispensable para poder alcanzar las capacidades de procesamiento requeridas para analizar toda la información que pueda incorporarse como prueba dentro del proceso.

Imaginemos a modo de ejemplo que un juez debe valorar pruebas en contra de un acusado por el delito de homicidio. Entre las pruebas se encuentran 5 videos de las cámaras de seguridad del edificio en donde ocurrió el delito, 6 informes periciales -de unas 7 u 8 carillas cada uno-, 100 fotografías del lugar de los hechos, y 18 testimonios. Ya nos resulta intuitivo pensar que la velocidad de respuesta de la IA será extremadamente mayor a la del humano a la hora de analizar todo este material. Pero no solamente eso, sino que, a medida que pasa el tiempo (día a día nos encontramos, cada vez más, con avances y mejoras en este campo tecnológico) los LLM (Large Language Models)¹⁰ se vuelven más efectivos a la hora de analizar estructuras complejas de datos, teniendo, estadísticamente, mejores resultados que los humanos incluso en exámenes académicos de renombre internacional¹¹.

Cabe preguntarse qué pasa con las garantías procesales cuando el error humano deja de existir y el proceso lógico pasa a estar en manos de una máquina. Es aquí donde debemos plantar los cimientos de los debates venideros y entender que esos derechos que la máquina de multas cuida mejor que los instrumentos garantistas ya existentes no serán desechados y olvidados, sino modificados, mutados y trasladados a lo largo del proceso¹². ¿A dónde se van las garantías en un proceso creado para cuidar -mejor que nosotros- que no se avance sobre los derechos del acusado? ¿Qué pasa puntualmente con el derecho a la defensa y una valoración probatoria manejada enteramente por la máquina? Enriquecer el debate sobre estos

¹⁰ "Los LLM son modelos que contienen miles de millones de parámetros y están entrenados en vastos corpus que bordean el tamaño de Internet. Debido a su increíble tamaño, los LLM pueden considerarse tecnologías de propósito general, con la aparente capacidad de comprender y generar texto similar al humano en una amplia gama de temas, incluyendo medicina, finanzas, educación, comercio y derecho." Matthew Dahl, Varun Magesh, Mirac Suzgun, and Daniel E. Ho. "Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models." Journal of Legal Analysis (próximamente), Yale Law School and Stanford Law School, March 8, 2024.

¹¹ "GPT-4 exhibe un rendimiento a nivel humano en la mayoría de estos exámenes profesionales y académicos. Notablemente, aprueba una versión simulada del 'Uniform Bar Examination' con una puntuación en el top 10% de los examinados" (OpenAI. "GPT-4 Technical Report." OpenAI, 2024, pp. 1-10).

¹² No descarto que en el futuro los modelos digitales y la imposición de un nuevo sistema procesal, hagan que las garantías se vuelvan una reliquia de la práctica acusatoria que hoy en día se mantiene como un norte -al que nos hemos acercado insatisfactoriamente-. Sin embargo, creo que lo más probable es que los humanos inevitablemente sigamos en alguna medida involucrados en la tarea de juzgar a nuestros pares ya sea desde el control, la revisión, o entrenando a estos nuevos "jueces máquina" para que innoven en el camino que como sociedad consideremos adecuado.

temas será de vital importancia para moldear el proceso penal del futuro cercano que ya empezó a gestarse y que, conforme las investigaciones que surgen en distintos medios académicos desde hace ya más de un año, continuará avanzando de forma exponencial¹³.

Es de suma importancia entender que la implementación de sistemas informáticos con inteligencia artificial no lleva necesariamente ni a la eliminación del juicio como etapa central del procedimiento penal, ni a socavar los principios englobados en el sistema acusatorio del régimen planteado por nuestra constitución. Es decir, que el juez máquina nos lleve a hacer un nuevo tipo de justicia, no implica que el humano deba desaparecer por completo del mismo, ni que, en un futuro cercano, nuestra idea modelo de proceso haya quedado obsoleta. Es cierto que si estos sistemas no van a hacer “lo que ya se viene haciendo” de una manera más eficiente, sino que vienen a “hacer algo nuevo y mejor”, es muy probable que tengamos que replantearnos muchos de los lejanos objetivos que hoy fija como norte nuestra constitución. Sin embargo, creo que la única manera de avanzar en forma coherente y armónica para con los procesos histórico-culturales que nos trajeron hasta aquí es idear un modelo que, habiéndose convertido en este “algo nuevo y mejor”, conserve las guías, pautas y principios fundamentales que hoy se resguardan detrás de las garantías existentes.

IV – La valoración digital:

Poniendo la lupa puntualmente en el momento de la valoración probatoria (que destaco nuevamente por ser el momento más activo en cuanto a “procesamiento de datos” dentro del procedimiento) nos encontramos con un escenario que -si nos centramos únicamente en la implementación de IA de dicho punto- no generaría grandes estragos para la configuración estructural del resto del procedimiento. Supongamos que continuamos teniendo una primera etapa de investigación en donde se recolecta prueba, una etapa intermedia crítica de su anterior, y luego un juicio oral y público en el que el tribunal se encuentra con los elementos probatorios, para luego “alimentar” a una IA con todos estos (podría ser de

¹³ “2024 AI Index Report”, Stanford University, 2024

una cuantía similar a la ejemplificada anteriormente) y que luego de pocos momentos arrojará un porcentaje probabilístico y matemático sobre las chances de que el acusado sea efectivamente el autor del hecho. Este tipo de herramientas pueden procesar las imágenes del lugar del hecho, cruzar esa información con pericias y videos, calcular cuestiones como la altura del autor, la distancia entre el arma homicida y la víctima, y demás cuestiones importantísimas en cuestión de segundos. Estos sistemas, a su vez, están preparados para comprender, interpretar e interactuar con conceptos emocionales y demás características humanas que suelen considerarse -erróneamente- como fuera del alcance de las IA¹⁴. Después de todo, los LLM están hechos con el propósito de interactuar con nosotros, aprender, y comprender nuestros comportamientos.

No se ve a simple vista un problema con que el juicio de culpabilidad sea realizado enteramente en base a los procesos matemáticos -precisos e incalculablemente mejores que los nuestros- de este agente informático. Si podemos creer en que el criterio de íntima convicción para el juicio por jurados tiene firmes cimientos de legitimidad dentro de nuestro sistema, por qué no confiar en una máquina que es objetivamente mejor que nosotros para calcular, interpretar, comparar y sistematizar conocimientos. Sin embargo, tenemos la obligación de pensar, como mencioné antes, en este tipo de “nuevos modelos procesales” dentro un proceder histórico que avance sobre los campos que permitan mejoras y cuide las conquistas del iluminismo. Y el derecho a la defensa frente a una valoración probatoria automática e instantánea necesita un nuevo hogar dentro del procedimiento.

El derecho a la defensa del imputado comprende la facultad de intervenir en el procedimiento penal abierto para decidir acerca de una posible reacción contra él y la de llevar a cabo en él todas las actividades necesarias para poner en evidencia la falta de fundamento de la potestad penal del Estado o cualquier circunstancia que la excluya o atenúe. Se entiende sintéticamente que puede entenderse como la facultad de ser oído, la de controlar la prueba de cargo que podrá utilizarse válidamente en la

¹⁴ Tyler Cowen, "GPT-4 Beats Psychologists on a New Test of Social Intelligence," *Marginal Revolution*, May 2024, <https://marginalrevolution.com/marginalrevolution/2024/05/gpt-4-beats-psychologists-on-a-new-test-of-social-intelligence.html>

sentencia, la de probar los hechos que él mismo invoca para excluir o atenuar la reacción penal, la de valorar la prueba producida y exponer las razones, fácticas y jurídicas para obtener del tribunal una sentencia favorable según su posición, que excluya o atenúe la aplicación del poder penal estatal¹⁵. Queda en evidencia ante la descripción formulada por Maier, que tenemos un problema. No me adelantaría a descartar una forma tan superadora de razonamiento y análisis valorativo de la prueba -en el contexto del surgimiento de un nuevo esquema procesal con sistemas informáticos del calibre mencionado- cuando podemos utilizarla de forma harmónica con el derecho a la defensa. La comparación con el juicio por jurados parece darnos una rápida respuesta sobre cómo se podría lidiar con esta situación: el juicio de culpabilidad podría ser definido por el velocísimo análisis hecho por la IA, mientras que lo concerniente al monto de la pena quedaría en manos humanas¹⁶.

Adicionalmente, el punto más relevante para la estructura defensiva, se dará al momento del control de los “input” o, traducido, al momento en el que se define qué pruebas se ingresarán al sistema para que el mismo haga los análisis pertinentes. De esta manera, y sin aplicar grandes cambios al sistema procesal actual, podemos ver cómo, en la medida que la máquina no proteja mejor que nosotros a las garantías, podemos conservar gran parte del proceso, ganando muchísimo tiempo, celeridad y exactitud a partir de la implementación de sistemas de análisis probatorio con IA.

Finalmente, dentro del plano revisorio del derecho de la defensa, y en diálogo constante con el derecho del imputado al recurso contra la sentencia, es importante comprender que no por acercarnos a un modelo en el que la herramienta arroje un resultado probabilístico¹⁷ se pierde capacidad

¹⁵ Julio B. J. Maier, “Derecho Procesal Penal. Tomo I: Fundamentos”, Editorial Ad-Hoc, Buenos Aires, 2023.

¹⁶ “Human In The Loop” o “HITL” es un modelo de actividad con Inteligencia Artificial en el que se propone no expulsar a los humanos de las actividades ya realizadas por nuestra especie, sino, por el contrario, mantenernos dentro del “loop” o ciclo de desarrollo de forma activa supervisando, entrenando y operando a los sistemas de Inteligencia Artificial.

¹⁷ Jordi Nieva Fenoll en “Inteligencia Artificial y Proceso Judicial” critica la implementación de IA en la valoración probatoria desde el rechazo a la respuesta probabilística de estos sistemas. Argumenta que en caso de ponerse en práctica, volveríamos a una nueva prueba tasada donde un porcentaje específico definiría la culpabilidad. Considero que esto no es acertado, sobre todo porque los humanos utilizamos el pensamiento probabilístico para tomar decisiones en entornos inciertos, lo

revisora. De acoger un sistema valorativo mediante inteligencia artificial, se volvería fundamental repreguntarse por un eventual recurso de casación aplicable a estos casos, que en consonancia con la modalidad que hoy adopta nuestro país a partir del fallo “Casal”, se encuentre en condiciones de aplicar un control sobre el “prompt” o instrucciones que el juez le dio a la IA, los “inputs” o elementos probatorios ingresados al sistema, y obviamente el juicio relativo al monto de la pena aplicable. Debería haber lugar también en estos sistemas revisorios para que la máquina indique qué antecedentes y métodos de análisis utilizó. No encuentro sin embargo asidero en el razonamiento que varios autores sostienen con respecto a que la programación o estructura lógica debería ser revisada, dado que ni los propios creadores de la mayoría de los softwares a los que me refiero saben realmente qué está sucediendo adentro de la caja.

V –Problemas del futuro, hoy:

La exploración del tópico desarrollado implica una parada obligatoria en los problemas más limitantes que tienen los sistemas impulsados por inteligencia artificial hoy en día. Y si bien la categorización conceptual puede variar, identificaré aquí tres circunstancias -interrelacionadas entre sí, conformando las afectaciones a los derechos y garantías advertidos por la discusión dogmática sobre este tema¹⁸- que presentan los mayores desafíos a la hora de pensar en conservar fundamentalmente el derecho a la defensa en juicio.

En primer lugar y tal vez uno de los aspectos de mayor polémica en los modelos más avanzados de IA es el “problema de la caja negra” que hace referencia a la imposibilidad por parte del usuario de acceder al procedimiento algorítmico mediante el cual el sistema llegó al resultado entregado. La inhabilidad de comprender completamente el proceso de toma de decisiones de la IA y la imposibilidad de predecir cómo ésta se va a comportar son aspectos que deberíamos tener en cuenta a la hora de ceder espacios de actividad humana para estas nuevas tecnologías. Esto está

que nos permite evaluar riesgos, estimar probabilidades y manejar la incertidumbre, procesos que la IA puede complementar analizando, comparando y contrastando datos de manera más precisa.

¹⁸ Al respecto, desarrolla Juan Luis Colomer Gómez en su libro, “El Juez-Robot y el respeto a los derechos fundamentales de los ciudadanos”, Editorial Tirant lo Blanch, Valencia, 2023.”

íntimamente relacionado con el concepto de “explicabilidad” del modelo. Cuanto más complejo sea el modelo, menor será la explicabilidad y mayor la opacidad de este.

Las redes neuronales a partir de las cuales se desarrollan los modelos más avanzados (como “GPT” o “Claude”) representan una verdadera imitación digital de cómo funciona el cerebro humano¹⁹, por lo que no poder leer la “mente” del agente que va a estar tomando la decisión judicial podría significar un problema. Yo creo que no. Si al agente digital (suponiendo que hablamos de un LLM como, por ejemplo, GPT) le pedimos en nuestro “prompt” que argumente por qué arribó a una decisión en particular, lo hará sin ningún problema. Podrá indicar fuentes, explicarnos qué elementos probatorios relacionó con determinados estudios estadísticos, mencionar qué parte de su inconmensurable base de datos se usó, etc. Si al juez humano (ya sea un juez profesional o miembro de un jurado) le pedimos que nos explique su decisión en particular, lo hará también sin ningún problema. Podríamos decir que pretender acceder al procedimiento algorítmico puntual que llevó a la IA al decisorio cuestionado -incluso si nosotros, desde el campo del derechouviésemos las herramientas para comprender alguna parte de lo que sucede tras bambalinas-, es equivalente a exigir que ingresemos al procedimiento lógico interno y subjetivo de cada magistrado (o miembro de un jurado) a la hora de tomar una decisión. Los jueces humanos, en efecto, también tienen cajas negras que encierran estructuras lógicas colmadas de errores y sesgos.

Lo cual nos lleva al segundo problema a señalar, que apunta a los sesgos que puede contener la máquina que valore. Independientemente de que estamos hablando de un modelo muy distinto en términos de funcionamiento a sistemas como COMPAS, no es menos cierto que todas las inteligencias artificiales tienen sesgos importantes que se vuelven especialmente problemáticos a la hora de ejercer juicios de valor sobre testimonios y culpabilidad.

¹⁹ “A neural network is a machine learning program, or model, that makes decisions in a manner similar to the human brain, by using processes that mimic the way biological neurons work together to identify phenomena, weigh options and arrive at conclusions.” (IBM, “What is a Neural Network?”, <https://www.ibm.com/topics/neural-networks>.)

Los modelos de lenguaje (así como la gran mayoría de los sistemas impulsados por IA) son entrenados con información a partir de producciones humanas. Si una IA tiene determinados sesgos contra determinados sectores de la población, es porque los datos incluidos en los documentos con los que se entrenó a esa IA presenten los mismos sesgos. Creo que es este tal vez el punto más conflictivo para analizar estos problemas desde el derecho de defensa, pero también creo que los modelos de lenguaje actuales (que se irán perfeccionando rápidamente con el tiempo) son suficientemente capaces de mostrar las fuentes utilizadas para elaborar respuestas a un prompt solicitado.

Creo que este problema tiene una solución similar al de la caja negra, y retomando lo antedicho al hablar de la etapa recursiva, sostengo que debería haber lugar a un análisis detallado de las fuentes y documentación utilizada por la IA para arribar a la decisión respecto de la culpabilidad del acusado.

Al respecto, Juan Luis Colomer Gómez ha escrito que “El respeto al derecho de defensa plantea cuestiones de suma relevancia. Se trata de que la IA sea accesible a las partes en las cuestiones básicas que afectan al conflicto. Por ejemplo, la información relativa al número de decisiones procesadas para obtener la escala (la posible predicción o una de las varias posibles predicciones), el origen de esas decisiones, la representatividad de las muestras seleccionadas, la distribución de decisiones entre diferentes criterios (como el contexto económico y social). Ello les permitirá comprender cómo se han construido las escalas, de cara a conocer los límites de sus alegaciones y poderlas debatir con mayor y mejor conocimiento ante un juez.”²⁰.

Por último, no quiero dejar de mencionar lo que creo que es el mayor obstáculo por sortear para que el resto de las soluciones propuestas surtan efecto y la IA pueda adaptarse como herramienta en un proceso acusatorio. Las llamadas “alucinaciones” consisten en respuestas erróneas²¹ que se

²⁰ Juan Luis Colomer Gómez, “El Juez-Robot y el respeto a los derechos fundamentales de los ciudadanos”, Editorial Tirant lo Blanch, Valencia, 2023, p. 40.

²¹ “Las alucinaciones en el campo de los LLM pueden subdividirse en tres tipos: el primero hace referencia a respuestas que no se corresponden o directamente contradicen al 'prompt' o a la consigna aportada por el usuario. La segunda está relacionada con las respuestas que contradicen el

presentan hoy en día como el principal motivo por el que, por ejemplo, un letrado no puede confiar en herramientas como GPT para que se lo asista en ejercer la profesión²². Si bien es cierto que este tipo de tecnologías avanzan rápida y exponencialmente, todavía tenemos -a mi entender, y al contrario de lo que la mayor parte de los autores considera- unos años antes de que se perfeccione al punto de poder evitar por completo las alucinaciones. De todas formas, las alucinaciones podrían ser fácilmente detectables y revisadas con ayuda de los mecanismos detallados anteriormente.

VI – Conclusión:

Armonizar las garantías fundamentales con la utilización de IA en el proceso penal es una obligación a la que se le prestará más atención en los próximos años. Pese al escepticismo y posiciones negativas por parte de una gran cantidad de autores, no parecería ser una tarea imposible. Estamos entrando en una etapa de aceleración técnica en la que nuestra sociedad y proceso penal posiblemente queden irreconocibles. Habilitar de forma progresiva pero decisiva la ayuda que la IA puede darnos en materia de valoración probatoria es una oportunidad que podría traer muchísimos más beneficios que costos si se hace en consonancia con las garantías y principios que pretendemos reforzar.

Debemos preguntarnos, en una época en la que se pretenden desarmar garantías y derechos fundamentales en el nombre de un “mejor funcionamiento” del sistema penal, si estas herramientas superinteligentes y con velocidades de análisis excesivamente altas, no son un principio de solución para agilizar los procedimientos sin retroceder al oscurantismo procesal del que salimos con el iluminismo.

'corpus' de entrenamiento de la máquina, y la tercera sigue a las respuestas que presentan errores con respecto a información del mundo que el sistema debería conocer." (Matthew Dahl, Varun Magesh, Mirac Suzgun, y Daniel E. Ho, "Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models." *Journal of Legal Analysis* (próximamente), Yale Law School y Stanford Law School, 2024).

²² Distintas actualizaciones de ChatGPT han permitido la creación de “Custom” GPT’s que permiten entrenar de forma personalizada a un modelo con información o documentación del usuario. Este tipo de prácticas achica considerablemente las alucinaciones, ya que el modelo de lenguaje cuenta con un corpus establecido por el usuario especialmente para la conversación elegida.