

FORREST, Katherine (2019): *When Machines Can Be Judge, Jury, and Executioner* (Singapur, World Scientific) 368 pp.

Muchas veces las publicaciones sobre derecho y tecnología parecen repetirse como *matryoshkas*; hay una primera obra atractiva y con nuevas ideas y luego una seguidilla de reproducciones similares que no aportan mucho a la discusión. Particularmente en el tema de “*jueces robot*” hay pocos trabajos que salgan de las mismas retóricas y ejemplos que tradicionalmente se han expuesto. Por eso es que es novedoso cuando se encuentra un trabajo que es capaz de diferenciarse del resto de las publicaciones y logrando añadir nuevos elementos y atraer visiones nuevas a un tema que ha sido ampliamente tratado, pero que está lejos de haber sido resuelto.

Katherine B. Forrest logra hacer la diferencia en su libro *When Machines Can be Judge, Jury, and Executioner*. En su vida previa al ejercicio privado del derecho, específicamente en el área de la tecnología, fue jueza federal del distrito sur de Nueva York. Desde esa posición le tocó resolver materias penales, impregnando en el centro de su práctica los principales lineamientos cuanto se trata de decidir sobre la libertad de un tercero y transmitiendo dicha experiencia en la obra. El libro está lleno de ejemplos y situaciones que le tocó vivir en su rol como jueza y cuenta, en su mayoría, con un estilo cercano al de la memoria.

El foco central de la obra está en el cuestionamiento de la implementación de herramientas de Inteligencia Artificial (en adelante IA) en los procedimientos criminales. En particular, en el problema que significa externalizar completa o parcialmente la labor de decidir sobre la libertad de un individuo en una máquina.

El trabajo empieza bajo la premisa que parte del trabajo del juez está siendo externalizado, y en virtud de aquello, vale la pena hacerse la pregunta filosófica sobre las diferencias existentes en la forma de tomar decisiones. La autora hace un interesante contraste entre la forma en que los humanos toman decisiones en situaciones de justicia penal frente a cómo lo hacen las máquinas. Las personas tomamos decisiones basadas en una combinación de elementos: conjeturas al azar, intuición, sistemas de creencias basadas en reglas, marcos éticos, normas culturales y sociales, asociaciones, sesgos y datos. Esta forma de tomar decisiones trae aparejada al menos dos grandes problemas: por un lado, si queremos automatizar dicha función, debemos ser capaces de incorporar el listado de componentes antes indicados al proceso de una máquina. Como conclusión, no declarada, la autora intuye que la complejidad y extensión de elementos que componen la toma de decisiones humana parece imposible de ser transmitida a una máquina. El otro problema que trae consigo, es reconocer que el pensamiento humano no ha resultado ser necesariamente el más eficiente. Independientemente de la plétora de elementos que alimentan nuestras decisiones, muchas veces éstas no son las mejores y son susceptibles de error. Un excelente ejemplo de ello son los juicios que hacemos contra alguien cuando recién lo conocemos.

Un problema que sigue al anterior, es establecer cuál es el objetivo de la toma de decisiones al interior de un proceso. Pensar en automatizar nos haría creer que estamos buscando simplemente eficientizar los juicios. Sin embargo, el punto de optimización en materia procesal debería ser el de justicia; buscar procesos más justos, no necesariamente más

eficientes (midiendo la eficiencia términos de tiempo y uso de recursos). Sobre este punto la autora se detiene en analizar los fines de la justicia penal.

Bajo la premisa que el derecho busca la justicia más que la eficiencia, el derecho procesal ha creado contrapesos para reducir los riesgos de error. Un ejemplo de estos se encuentra en la necesidad de que las decisiones judiciales sean realizadas a través de una sentencia. Una condena sin sentencia viola el debido proceso en su esencia, ya que no le permite a las partes saber los fundamentos de la misma. Decisiones no fundamentadas no permiten a los afectados ejercer su derecho a la defensa o constatar que se hayan tomado en consideración todos los elementos relevantes expuestos (argumentales y probatorios), ni tampoco permite conocer la argumentación o lógica detrás de la decisión. La racionalidad reglada que permite el procedimiento a través de la sentencia es una garantía para la sociedad. La sentencia y su fundamentación se encuentra en el corazón del debido proceso.

En el entendido que hemos creado contrapesos y pasos procesales que permiten garantizar la racionalidad y, con ello, la justicia para las partes dentro del proceso, tenemos que cuestionar el funcionamiento y efectos de la implementación de los sistemas de IA en el sistema procesal. Para la autora automatizar los elementos que un juez toma en consideración al sentenciar conlleva una dificultad importante, ella asume que cualquier implementación de IA busca incrementar únicamente la eficiencia procesal. Dejando de lado las consideraciones propias de la Justicia.

Es en ese entendido que el libro se hace cargo de encasillar a qué corriente filosófica puede ajustarse la implementación de IA. Eficientar el procedimiento, afirma la autora, siempre responde a una visión filosófica del proceso judicial. Para fundamentar el punto, hace una presentación de las corrientes que podrían explicar el fenómeno. El desarrollo lo hace desde dos ejemplos de herramientas de automatización: por un lado desde los sistemas de evaluación de infractores (*assessments for offenders*), y por el otro, desde las armas letales autónomas.

Una teoría utilitarista de la justicia, en la que los grandes números y la tendencia parecen más importantes que el caso particular, aceptaría sin mayor complicación la automatización de la toma de decisiones judiciales. Sin embargo, facilitaría y fomentaría que incrementaran los prejuicios sobre ciertos grupos determinados, en el entendido de que estos podrían ser criminalizados *a priori*. Esto sucede actualmente con los sesgos que pueden encontrarse en ciertos sistemas autónomos respecto de ciertas minorías. Para extremar el argumento, la autora dice que si como sociedad siguiéramos las teorías que responden a lógicas utilitaristas, deberíamos encarcelar a toda la población joven, ya que son estos quienes en su mayoría delinquen, salvaguardando al resto de la población.

Luego de un barrido general sobre algunas tendencias filosóficas que pudieran ayudarnos a entender la implementación de IA, la autora concluye que el uso de IA en justicia es utilitarista tanto desde su diseño como por sus objetivos. En mi impresión, habría sido un aporte si la autora no se refiriera únicamente a experiencias de implementación en cortes estadounidenses, ya que llega a una conclusión general –que la implementación de IA sigue siempre una lógica utilitarista– basada en una experiencia particular –lo que se ha implementado en Estados Unidos–.

Tal vez uno de los aportes más interesantes del libro es el análisis detallado al sistema COMPAS, un software de determinación de riesgo potencial de reincidencia de imputados utilizado en cortes estadounidenses. Existen muchos trabajos que han documentado sobre casos de injusticia y discriminación que han surgido desde COMPAS, pero pocas publicaciones —o al menos ninguna que yo haya tenido la oportunidad de ver— que estén desarrollados desde el punto de vista de alguien que haya sido interviniente o usuario. Forrest hace un análisis bastante interesante del Manual del Usuario del Sistema; revisa las explicaciones que se dieron por la empresa en la actualizaciones del Manual luego que se hicieran públicas los distintos errores que se cometían. En ese sentido, sigue el análisis tratando de identificar las teorías sociales de comportamiento criminal que subyacen y que se declaran por los desarrolladores de este tipo de sistemas. Finalmente, termina por discutir sobre las distintas categorías usadas en la base de datos para hacer la predicción o atribución de puntaje a los sujetos analizados. Esta sección del libro resulta ser de gran utilidad para quienes trabajan en la materia, y que necesitan contrastar la visión académica con la experiencia real de los practicantes, en este caso jueces.

Como es de esperar, la conclusión del análisis que realiza la autora es que si bien estos sistemas significan un riesgo para los sujetos pasivos de su utilización, no podemos renunciar a su uso. Sin embargo, es necesario para su implementación, más que mejoras o soluciones parche, un rediseño desde el inicio del desarrollo.

Si bien el libro carece de un estudio exhaustivo de criminología o computación, la visión de la autora es enriquecedora para el lector y valiosa para la discusión. Es un libro basado en experiencias y apreciaciones personales, que lo hacen un texto cercano e interesante por las anécdotas de la jueza Forrest y que logra hacerse cuestionamientos fundamentales de la aplicación de IA en la aplicación de justicia. Deja, intencionalmente, preguntas abiertas y logra dejar al lector con conocimientos nuevos y un mejor entendimiento de los desafíos de implementar más tecnológicas las cortes.

MATÍAS ARÁNGUIZ V.
Shanghai Jiaotong University

