E.M. nº 82 Enero-Abril 2026

JUECES O ALGORITMOS: ¿SUSTITUIRÁ LA *INTELIGENCIA ARTIFICIAL* A LOS JUECES?

Manuel Alfonseca Moreno
Profesor Honorario de la UAM (antes catedrático de Lenguajes y Sistemas Informáticos).

RESUMEN

Un algoritmo de 2017 que predice la decisión de los jueces sobre enviar o no a un acusado a prisión preventiva obtiene buenos resultados y podría convertirse en una buena herramienta de ayuda a la decisión. Además, con el auge de los LLM a partir de 2022, la cuestión de su utilización como ayuda a los jueces ha pasado al primer plano. Es de esperar que en los próximos años todas estas herramientas irán mejorando y que serán cada vez más utilizadas como ayuda a la toma de decisiones, pero no se espera que los jueces humanos sean sustituidos por ellas, al menos en el corto plazo. Como toda tecnología, siempre se verá con reticencia.

1. INTRODUCCIÓN

En publicaciones recientes se suele decir que la así llamada "Inteligencia Artificial" (IA) va a provocar grandes cambios en el panorama profesional de la humanidad, eliminando algunas profesiones y haciendo aparecer otras nuevas.

Entre las nuevas profesiones se citan estas: Automatización e IoT industrial; diseño de herramientas de IA; ciberseguridad; arquitecto de *Big Data*; desarrollo del metaverso; Blockchain y criptomonedas; experto en ética profesional.

Entre las profesiones obsoletas se citan estas: a) Profesiones administrativas (operador de centralita telefónica, cajero bancario, agente de viajes, bibliotecario, recepcionista). b) Profesiones industriales (operario de línea de montaje, operador de maquinaria, empaquetador, transportista de materiales interno). c) Análisis y proceso básico (contable, analista financiero, corrector de pruebas). d) Profesiones rutinarias (taxista, cajero de supermercado, repartidor de prensa, archivero). e) Intermediación (corredor de bolsa, agente inmobiliario, vendedor de seguros). Y en general, cualquier profesión de bajo nivel de especialización.

¿Deberá añadirse a los jueces a la lista de las profesiones obsoletas? ¿Será posible que los jueces sean sustituidos algún día por herramientas de IA?

2. UN ALGORITMO DE AYUDA A LA DECISIÓN

La referencia 1, anterior en algunos años al auge del último grito de la IA, los Modelos Grandes de Lenguaje (LLM por sus siglas en inglés), compara la eficacia de un programa de toma de decisiones que utiliza algoritmos de *aprendizaje por ordenador*, con las mismas decisiones cuando las toman seres humanos. El problema que se abordó fue la toma de decisiones por los jueces del Estado de Nueva York respecto a dejar en libertad (con o sin fianza) al acusado de un delito, o bien enviarle a prisión preventiva hasta que se celebre el juicio, que puede tardar varios meses en llevarse a cabo. El

problema es importante, pues hay más de diez millones de arrestos anuales en todo el país, y las decisiones que se tomen tienen repercusiones económicas y sociales: las personas en prisión preventiva consumen recursos públicos, mientras que un acusado liberado puede cometer nuevos delitos.

Los criterios que utilizan los jueces no son los mismos en todo el país. En el Estado de Nueva York, el juez sólo debe considerar la posibilidad de que el acusado no se presente al juicio cuando llegue el momento (riesgo de fuga). En otros Estados se debe tener en cuenta la prevención de delitos adicionales realizados por el acusado si se encuentra en libertad bajo fianza. El trabajo comentado en este artículo no considera la cuantificación de la fianza, por lo que la decisión a tomar queda reducida al caso más simple: libertad provisional o prisión preventiva.

2.1 Descripción del algoritmo

El programa de ayuda a la decisión que se utilizó en este trabajo consta de dos partes:

- Un algoritmo que trata de resolver el siguiente problema: *¿debe el juez enviar a este acusado a prisión preventiva o no?* Este algoritmo está programado. Su funcionamiento depende de una serie de variables (varios cientos) cuyo valor exacto se desconoce inicialmente.
- Otro algoritmo de aprendizaje, que ejecuta el algoritmo anterior para una serie de casos concretos de resultado conocido y trata de asignar valores óptimos a cada una de las variables para que el otro algoritmo prediga correctamente lo que ocurrió.

Al algoritmo de aprendizaje se le proporcionaron 443.751 casos reales de la ciudad de Nueva York, junto con información sobre si las personas que los jueces dejaron libres se saltaron los controles judiciales. Una vez ajustados los valores de las variables para que el primer algoritmo obtenga las mejores predicciones posibles, el trabajo del algoritmo de aprendizaje ha terminado y el primer algoritmo puede utilizarse solo.

Para validar el proceso, se sometió el primer algoritmo a 110.938 casos de la ciudad de Nueva York, distintos de los anteriores, y se compararon sus predicciones con lo que ocurrió en realidad. Si el algoritmo logró predecir los resultados reales con cierta aproximación, puede utilizarse para maximizar o minimizar alguna función: el coste económico de un proceso, o el número de delitos que tienen lugar en determinadas circunstancias.

El programa no produce un resultado binario (sí o no), sino un nivel de riesgo, la probabilidad de que el acusado viole las condiciones de la libertad provisional, lo que permite clasificar a los acusados en grupos de riesgo y comparar los resultados del programa con las decisiones tomadas por los jueces.

El número total de casos reales considerados por el algoritmo fue de 554.689, suma de los dos grupos de casos indicados en los párrafos anteriores, que corresponden a delitos cometidos durante cinco años en la ciudad de Nueva York. El 80% de estos casos se utilizó para entrenar el algoritmo, mientras que el 20% restante (casos de validación) sirvió para comprobar la eficacia del algoritmo entrenado y compararla con la de los jueces que tomaron decisiones sobre los mismos casos.

El trabajo se repitió con 151.461 casos de otras ciudades en los que se aplicaron otros criterios, con resultados parecidos.

2.2 Validación del algoritmo

La eficacia del programa se comprobó correlacionando el riesgo predicho por el algoritmo para los casos de validación con el riesgo real observado en los mismos casos. El artículo de la referencia 1

concluye que la actuación de los jueces fue defectuosa, pues aunque se suele tener en cuenta los riesgos posibles de dejar en libertad a un acusado, algunas de sus decisiones parecen aleatorias. Se ha intentado discutir esta conclusión, clasificando a los jueces en grupos según su grado de indulgencia o dureza y analizando las diferencias entre los distintos grupos, o suponiendo que los jueces hacen uso de información no cuantificable (como el aspecto físico del detenido), pero ninguna de estas comprobaciones permite afirmar que la conclusión anterior sea falsa.

3. CONSIDERACIONES GENERALES

A esta descripción del proceso de aprendizaje por ordenador se deben añadir tres consideraciones adicionales:

1. Si el primer algoritmo no está bien construido o no se adapta bien al problema que trata de resolver, el proceso de aprendizaje no podrá tener éxito. El algoritmo de aprendizaje no es tan crítico: aunque pueda tardar más o menos en conseguir su objetivo, si el algoritmo al que se aplica está bien diseñado, acabará ajustando las variables a un conjunto de valores óptimo. El algoritmo descrito en el artículo debe de ser bueno, pues obtiene sistemáticamente mejores resultados que los jueces. Además, es fácil modificar el criterio de aprendizaje para adaptarlo a situaciones diferentes, como simular el comportamiento de jueces concretos, en lugar de cuantificar el riesgo de liberar a los acusados.

2. Los datos que se proporcionan al algoritmo durante la fase de aprendizaje deben estar libres de sesgos, porque si no lo están los resultados del algoritmo también serán sesgados. Por ejemplo, en el caso de una herramienta parecida llamada COMPAS (véase la referencia 2), desarrollada por la empresa Northpointe (cuyo nombre actual es Equivant), que se ha utilizado en la práctica en tribunales de los Estados Unidos, se detectó un sesgo racial en los resultados proporcionados por COMPAS, aunque la empresa ha negado que dicho análisis fuera correcto. En cualquier caso, el Tribunal Supremo de Wisconsin decidió en 2016 que los jueces podían utilizar la herramienta como apoyo a sus decisiones, pero que las puntuaciones ofrecidas por esta debían ir acompañadas de advertencias y precauciones.

3. Los programas de toma de decisiones no deben sustituir a los seres humanos. Los autores del artículo en la referencia 1 no proponen que se sustituya a los jueces por programas de ordenador, aunque la conclusión de su análisis no sea favorable a la actuación de dichos jueces. Por el contrario, programas como este podrían utilizarse como sistemas de ayuda a la decisión, que los jueces podrían consultar antes de tomar la suya, aunque no debería ser obligatorio que obedecieran sus consejos.

El objetivo de ese trabajo no es, como suele ocurrir en investigaciones sobre aprendizaje por ordenador, la mejora de los algoritmos o del proceso de aprendizaje automático, sino comprender mejor qué hacen los jueces cuando toman este tipo de decisiones y proponer procedimientos para mejorar su actuación de dos maneras: a) minimizando el número de personas enviadas a prisión preventiva sin aumentar el riesgo de que los que quedan libres delincan, lo que reduciría el gasto público; b) minimizando el número de delitos cometidos sin aumentar el número de presos preventivos. Para ello se proponen dos procedimientos:

1. Contracción: un sistema que avise al juez cuando está a punto de soltar a un acusado de alto riesgo. Esto permitiría reducir la tasa de delitos cometidos, a costa de aumentar ligeramente el número de acusados enviados a prisión preventiva.

2. Ordenación: un sistema que ordene a los acusados en función de su nivel de riesgo y sugiera intercambiar un acusado de alto riesgo a quien el juez se disponía a dejar libre por un acusado de menos riesgo que el juez pensaba enviar a prisión preventiva. Este sistema haría posible reducir la tasa de delitos manteniendo constante el número de personas encarceladas, o bien reduciría este número, manteniendo estable la tasa de delitos prevista.

Esta es una cita del artículo en cuestión (mi traducción):

> *Reducir la población de las cárceles sin aumentar los delitos es una prioridad clave. La investigación económica empírica suele enfocar cuestiones causales... Nuestro análisis aborda otra posibilidad: mejorar las predicciones.*

En cualquier caso, los algoritmos de este tipo no deberían sustituir a los jueces, sino ayudarles a tomar decisiones con mejores elementos de juicio, proporcionados por herramientas que no están sometidas a variaciones aleatorias de estados de ánimo, enfermedades y otros efectos que afectan a los seres humanos.

4. LA SITUACIÓN ACTUAL

Con la llegada de los LLM en 2022, la cuestión de su utilización como ayuda a los jueces ha pasado al primer plano. El problema es que el objetivo de los jueces es (o debería ser) descubrir la verdad sobre el caso que deben juzgar, mientras que herramientas como CHATGPT, GEMINI, DEEPSEEK y otras que van apareciendo, no utilizan el criterio de lo que es verdadero o falso para construir sus contestaciones, sino que se basan en la predicción de la palabra siguiente más probable, utilizando como fuente de datos la información contenida en Internet (véase la referencia 3).

A menudo se han señalado multitud de casos en que la contestación de estas herramientas era absurda o se inventaba las contestaciones (estas respuestas se llaman *alucinaciones*), por lo que su utilización en casos jurídicos reales debería realizarse (si se realiza) con las mayores precauciones. Por esa razón, algunas sentencias judiciales recientes han sido apeladas (véase la referencia 4), porque el juez validó información alucinatoria proporcionada por ChatGPT, que se inventó jurisprudencia falsa del Tribunal Supremo, o incluso han sido anuladas por el simple hecho de haber utilizado herramientas de IA para escribir la sentencia (véase la referencia 5).

Una herramienta de ayuda a los despachos de abogados es un objetivo menos ambicioso que una de ayuda a los jueces, y de hecho ya existe una, basada en IA Generativa y asociada a ChatGPT, llamada Harvey (véase la referencia 6).

Es de esperar que en los próximos años todas estas herramientas irán mejorando y que serán cada vez más utilizadas como ayuda a la toma de decisiones, pero no se espera que los jueces humanos sean sustituidos por ellas, al menos en el corto plazo. A este respecto, la referencia 7 dice lo siguiente:

> *En el sistema judicial se debate sobre su inclusión no solo porque podría ser una herramienta muy útil, sino porque para muchos es un hecho que su inclusión se irá dando paulatinamente y por ello presenta un desafío enorme para no rechazarla del todo…*
>
> *Como toda tecnología, siempre se verá con reticencia, especialmente por la vulneración que podría suponer el hecho de dejar en manos de un programa… la predicción de un resultado judicial y, ya que se le considere en su totalidad o se le valide o bien que sea una herramienta en las funciones públicas, siempre habrá que considerar factores éticos que no son fáciles de prevenir, limitar y señalar, como también ciertos factores legales que son los más difíciles de considerar.*

5. REFERENCIAS

1. Jon Kleinberg; Himabindu Lakkaraju; Jure Leskovec; Jens Ludwig; Sendhil Mullainathan. THE NATIONAL BUREAU OF ECONOMIC RESEARCH, Working Paper 23180, Feb. 2017, http://www.nber.org/papers/w23180
2. COMPAS Field Guide, https://njoselson.github.io/pdfs/FieldGuide2_081412.pdf. Véase también https://en.wikipedia.org/wiki/COMPAS_(software).

3. Alfonseca, M. “Un modelo de ChatGPT”, https://divulciencia.blogspot.com/2023/06/un-modelo-de-chatgpt.html
4. https://diariosabemos.com/sociedad/tribunales/escandalo-judicial-ia-inventa-jurisprudencia-supremo-juez-usa_514285_102.html
5. https://www.lanacion.com.ar/seguridad/copiar-y-pegar-la-frase-que-delato-al-juez-que-uso-ia-provoco-la-nulidad-de-un-fallo-penal-y-ahora-nid16102025/
6. “Harvey: la IA al alcance de los abogados”, https://www.pwc.es/es/newlaw-pulse/legaltech/havery-ia-alcance-abogados.html. Véase también https://openai.com/es-ES/index/harvey/.
7. García Sánchez, M.I. “La facultad de juzgar vs inteligencia artificial: más allá de la automatización del sistema de justicia”, Ius et Scientia, 11:2, 2025, DOI: https://dx.doi.org/10.12795/IETSCIENTIA.2025.i02.11