E.M. nº 73 Enero-Abril 2023

¿PUEDEN LOS ALGORITMOS SER EVALUADOS CON RIGOR?

Juan Antonio Garde Roca.
Economista. Expresidente Agencia Estatal de Evaluación (AEVAL). Presidente de ALGOVERIT.

RESUMEN

Cada vez es más frecuente usar técnicas de inteligencia artificial –y de aprendizaje automático ("machine learning") en particular-, para crear algoritmos que resuelvan distinto tipos de problemas relacionados con la clasificación, la predicción, el análisis de riesgos, la toma de decisiones o la elaboración de recomendaciones, a partir del análisis de datos preexistentes.

Ello se debe a varios factores, entre ellos: 1. la existencia de una ingente cantidad de datos que arrojan luz sobre el comportamiento de cada vez más personas, entidades y sistemas; 2. la difusión de herramientas informáticas cada vez más fáciles de utilizar y más accesibles que simplifican el procedimiento de análisis de los datos existentes y la generación de los algoritmos de resolución de problemas;3. la práctica universalización de las infraestructuras y servicios digitales necesarios para realizar la toma de datos y su procesamiento.

Todos estos factores han permitido que hoy sea relativamente fácil aplicar estas técnicas de inteligencia artificial a la resolución de problemas cotidianos, cuestiones éstas que se abordan, entre otras, en el presente artículo.

1. INTRODUCCIÓN

La aparente simplicidad de las técnicas de inteligencia artificial no debe hacer olvidar la complejidad que acompaña un uso adecuado de estas técnicas.

En particular, se deben adoptar las precauciones necesarias para:

1. Evitar que los algoritmos resultantes presenten sesgos. Para ello, es necesario garantizar que el conjunto de datos que se ha utilizado para entrenar el modelo esté equilibrado, que la selección de las variables sea correcta, que la técnica utilizada sea adecuada y que su resultado sea relevante.

2. Garantizar que se respeten los derechos de las personas sobre las que se adoptan las decisiones o recomendaciones, lo que tiene implicaciones sobre las variables que puedan usarse para analizar los datos de partida y en el momento de la toma de decisiones.

3. Considerar que los resultados sean auditables, por la existencia de mecanismos que permitan conocer cómo se han adoptado las decisiones o ser posible analizarlas mediante el cruce con evidencias adecuadas.

En relación con este último punto, existe la posibilidad de una gran variedad de tipos de auditorías y también de objetos de evaluación en relación con los algoritmos.

Es cierto que la complejidad que puede alcanzar esta técnica, así como su innovación constante, llevan a preguntarse por la viabilidad de efectuar procesos integrales de análisis. También a considerar positivamente el uso de la Inteligencia Artificial (I.A.) como herramienta idónea para una auditoria viable, a partir del uso de algoritmos de aprendizaje automático o "machine learning".

Todo ello, no anula la posibilidad de que los algoritmos puedan y necesiten ser auditados con rigor, sino que refuerza la necesidad de evaluación y transparencia, siempre considerando con realismo lo que puede y no puede auditarse y el cómo se audita.

En este artículo se pretende mostrar las posibilidades que ofrece la auditoría algorítmica para evaluar los usos de la IA, a través de las diversas herramientas habituales de evaluación ya existentes.

En primer lugar, es necesario estudiar el sistema algorítmico en su contexto y en el de la organización a la que sirven. De esa forma, será posible considerar el perfil inicial de los riesgos a los que pueden enfrentarse sus resultados.

Dicha información resulta fundamental para la elaboración del plan de auditoría. Adicionalmente dicho plan deberá seleccionar el paquete de herramientas y pruebas de análisis precisos para su despliegue.

Ese despliegue puede integrar especificidades propias del funcionamiento de los algoritmos y la tecnología de la IA, también de la experiencia aprendida de los distintos métodos testing de software empleados en los últimos años, así como la larga experiencia en técnicas de auditoría y también los numerosos trabajos y guías existentes de evaluación.

La evaluación del cumplimiento del marco regulatorio, será un componente fundamental de todo tipo de auditorías algorítmicas que se pretendan realizar, antes de pasar a los referentes técnicos.

El conocimiento de las bases de datos utilizados y la información empleada en el desarrollo, entrenamiento y evaluación del sistema algorítmico, adquiere también una importancia singular.

Finalmente, se considera la integración en el marco de control interno de las organizaciones y siguiendo la referencia de los Informes COSO, cómo incorporar la auditoria algorítmica al proceso de gobernanza y gestión del negocio y las empresas.

Los algoritmos, al igual que las personas, podemos producir sesgos. Los primeros a partir de los datos y la información que utilizamos los humanos al entrenarlos. Estos sesgos pueden provenir de orígenes muy distintos**:** de la propia construcción de la herramienta y su objeto, de una incorrecta selección de la información que lo alimenta, de la falta de calidad de los datos que utiliza y, en ocasiones, puede ser relativamente fácil corregirlos con técnicas similares a las que vienen utilizándose en evaluación.

Los sesgos de información se producen por deficiencias o disparidades, dada la dificultad que implican comparaciones homogéneas y las posibles confusiones en la construcción y diseño de las bases de datos.

Existe la posibilidad de una gran variedad de tipos de auditorías y también de elegir múltiples objetivos de evaluación en relación con los algoritmos. Es cierto que la complejidad que puede alcanzar esta técnica y su proceso de innovación constante han desanimado a algunos, que se preguntan por la viabilidad de efectuar procesos integrales de análisis. También, a considerar positivamente el uso de la Inteligencia Artificial (IA) como herramienta idónea para una auditoria viable, a partir del uso de algoritmos de aprendizaje automático o *machine learning*, y conforme nos adentramos en el aprendizaje profundo.

Las cajas negras que incorporan estos últimos algoritmos adquieren un carácter mucho más intenso y desconocido, con mayores capas ocultas, no solo para los expertos sino también para sus propios programadores. Las técnicas al uso para analizar estas cajas negras pueden hacerse más difíciles en estos entornos.

Por otra parte, los costes de una auditoría de estas características son muy elevados y los mercados y los inversores, en ocasiones, resultan poco propicios a hacer transparentes y públicos sus limitaciones o vulnerabilidades.

Pero la auditoría y la evaluación pueden y deben tener objetos más acotados**,** centrarse en algunos procesos críticos singulares**.** Debemos también partir del conocimiento de lo que puede esperarse de la auditoría de un sistema de I.A. y qué no es posible esperar.

Es preciso propiciar el análisis de la composición y usos de los algoritmos para generar mejoras de su eficiencia que eviten riesgos y sesgos detectables, que sean asumibles con el acervo técnico e instrumental existente y con costes asequibles. La transparencia lo exige también, cuando el algoritmo se aplica sobre acciones de los seres humanos, ya que precisa cierto concepto respecto de la justicia, y eso implica la existencia de valores y su explicitación.

La existencia de regulaciones, modelos y prácticas corporativas regladas de control interno y externo, permiten incorporar la auditoria algorítmica al entorno corporativo de las organizaciones e integrar en sus procesos, actividades y negocios el análisis de los usos de la I.A., que no pueden ser ajenos al modelo de gobierno corporativo y su despliegue.

En estas líneas pretendo mostrar las posibilidades que ofrece la auditoría algorítmica para evaluar los usos de la IA, a través de las diversas herramientas habituales de evaluación ya existentes.

En primer lugar, es necesario estudiar el sistema algorítmico en su contexto y en el de la organización a la que sirven. De esa forma, será posible considerar el perfil inicial de los riesgos a los que pueden enfrentarse sus resultados y el impacto de sus errores**.**

Dicha información resulta fundamental para la elaboración del plan de auditoría. Adicionalmente, dicho plan precisará elegir el paquete de herramientas y pruebas de análisis específicos para su despliegue, que puede integrar especificidades propias del funcionamiento de los algoritmos y la tecnología de la IA, también de la experiencia aprendida de los distintos métodos *testing* de software empleados en los últimos años, así como la larga experiencia en técnicas de auditoría, y también los numerosos trabajos y guías de evaluación existentes.

La evaluación del cumplimiento del marco regulatorio será un componente fundamental de todo tipo de auditorías algorítmicas que se pretendan realizar, antes de pasar a los referentes técnicos.

El conocimiento de las bases de datos utilizados y la información empleada en el desarrollo, entrenamiento y evaluación del sistema algorítmico, adquiere también una importancia singular.

2. PROGRAMACIÓN REGULAR FRENTE AL APRENDIZAJE PROFUNDO.

Los algoritmos, al igual que la digitalización y las técnicas y usos de la Inteligencia Artificial, son motivo de presentación confusa, en algunas ocasiones, tanto en términos técnicos como científicos, en ciertos medios de comunicación.

No suele reseñarse suficientemente que la calidad y el origen de los datos empleados, su contexto, los atributos que se pretenden destacar, el propio entrenamiento, así como el perfil y la inspiración del programador, pueden llegar a tener un peso decisivo en sus resultados.

La ausencia en origen de una auditoría interna especializada adecuada, y también las diferencias entre las bases de información del entrenamiento y su práctica posterior una vez implementado , agudizan la incertidumbre.

Los algoritmos utilizan datos. Si estos se conciben como un activo valioso de la organización deben ser gestionados. En los últimos años ha tomado relevancia la figurea del Chief Datta Officer (CDO), que se ocupa de aspectos, como la calidad, la trazabilidad y el lineage. La cooperación entre el auditor, los científicos de datos y el CDO son hoy de la mayor importancia.

Las dificultades y distorsiones pueden ser previas a la calidad del dato empleado o a su entrenamiento, y vincularse al objeto o finalidad que se pretende alcanzar con su uso. Cuestiones que pueden llegar a ser relevantes en la evaluación de este ámbito tecnológico y para la regulación actual, como por ejemplo la determinación de la propiedad de los datos y los responsables de sus usos, así como del propio mecanismo del algoritmo, precisarían de una mayor atención.

Evaluar el riesgo de cumplimiento regulatorio, respecto del uso de la IA, en un proceso de innovación tan inacabado y acelerado, es el primer eslabón a tener en cuenta para una auditoría instrumental algorítmica.

Todavía existe al respecto una escasa regulación, pero el diseñador y el usuario deben situarse en el "framework" del Reglamento en ciernes de Inteligencia Artificial de la U.E, y la propia Declaración conjunta de la Comisión, el Parlamento y el Consejo de la UE acerca de los derechos digitales ya en vigor, que ofrecen un marco para el cumplimiento, en el contexto de las recomendaciones de organismos internacionales como la OCDE y La Unesco.

El desarrollo científico actual no puede ser ajeno al rigor metodológico, tanto en auditoria como en evaluación, para lograr alcanzar las necesarias evidencias. La causalidad matemática y también la correlación, contribución y/ o atribución, tan en uso en la evaluación, siguen aportando un peso significativo. La física cuántica nos proyecta una realidad en que la naturaleza puede desenvolverse dentro de un contexto de probabilidades menos previsibles y más coyunturales. La investigación, la economía, la política, la sociedad, los intereses geoestratégicos, nos lo están proclamando continuamente. En la teoría de la negociación, cuando los intereses son sustituidos por las emociones, los desenlaces pueden perjudicar a todos los negociadores y a sus resultados.

Los primeros algoritmos usados por los robots, y en la toma de decisiones en procesos organizativos, mostraban y muestran una tendencia causal netamente "ingenieril". Son deterministas en su implantación para alcanzar un resultado, se basan en datos de probabilidad de contextos acotados y estables, se configuran a partir de técnicas de decisión de arriba-abajo, y promueven la traslación en la programación del lenguaje humano al de las máquinas, a partir de las bases de datos y la información existentes.

El reino de la probabilidad que se deriva de la magnitud de las bases de datos ha resultado fundamental, pero lógicamente no puede servir per se. Los cambios de contexto o la pretensión de su uso en otros contextos generan grandes imperfecciones y puntos ciegos continuos, cuando pretenden aplicarse en distintas organizaciones o en diversos sectores y realidades.

Los algoritmos de caja blanca, de carácter principalmente determinista, que inicialmente podemos denominar regulares, seguirán teniendo en el futuro un gran desarrollo. Tanto en el sector público como en el privado, precisan de seguridad-riesgos, supervisión humana, regulaciones adecuadas y auditorías.

Lo anterior resulta todavía más imprescindible, dada su evolución hacia otra realidad de acelerada innovación: la del del algoritmo de aprendizaje automático o *machine learning*, que usa técnicas de caja negra. Generando otro proceso cualitativamente distinto que ya no es sólo

predeterminado por el programa inicial. Entre las distintas modalidades de aprendizaje automático, a través de combinaciones de algoritmos, se encuentran aquellos denominados de aprendizaje supervisado, aprendizaje no supervisado, aprendizaje por refuerzo, y la caracterizada por el uso de redes neuronales. Todos ellos precisan de análisis adaptados a sus peculiaridades.

Gracias al proceso que incorpora el aprendizaje automático, los nuevos programas pretenden, de forma dinámica, que el ordenador no solo se limite a reproducir con sus resultados el funcionamiento deseado, sino que aprenda también con la propia información contenida, superior en varios ordenes de magnitud a la que puede manipular el ser humano. Analizar esos millones de datos y extraerles un sentido a través de un funcionamiento similar a nuestras redes neuronales, ampliadas en su programación, les permite barajar en segundos un número ingente de intenciones probabilísticas que no son perceptibles para nuestro cerebro. Se configuran en consecuencia a través de técnicas de decisión de abajo a arriba, incorporando también progresivamente un proceso desde el lenguaje del ordenador a lenguajes humanos de salida.

En este caso, las cajas negras adquieren un carácter mucho más intenso y desconocido, con mayores capas ocultas, no solo para los expertos sino también para los propios programadores, que los utilizados con anterioridad a la IA.

Los métodos se tornan más y más complejos, con este extraordinario despliegue fundamentado en el aprendizaje automático, y de un subcampo dentro éste que viene denominándose *deep learning* o aprendizaje profundo.

En la programación regular, el algoritmo convierte la entrada en resultados, a través de reglas lógicas y matemáticas. En el aprendizaje automático, y como caso particular en el aprendizaje profundo, se parte de un juego de casos o valores de entrada, a los que se dota de un cierto peso de importancia en cada caso. Sirven para que la red neuronal juegue contra sí misma**,** usando para ello ecuaciones agrupadas en múltiples capas, sin conocimiento de las reglas lógicas y matemáticas en el que se produce el aprendizaje que precisa para obtener las salidas. Las neuronas de cada capa se conectan con las neuronas de la capa siguiente, en ciertos casos ocultas, y así sucesivamente hasta una capa de salida.

Este aprendizaje de los algoritmos permite, a partir de procesos de prueba- error, investigar en un instante las decisiones, sin que resulte siempre viable una supervisión humana. Toman decisiones vitales que pueden plantear, ya en la actualidad, dilemas en materias transcendentes, no sólo de la vida cotidiana, sino que influyen así mismo de forma creciente y decisiva en nuestro modelo cultural y civilizatorio.

Evitaré explicaciones complejas respecto de un uso más especializado de las cajas negras y las propias formas de control de sus resultados. Me centraré en la descripción más general de los instrumentos de auditoría y evaluación, así como en señalar algunas de las pistas y pruebas para alcanzar evidencias más empleadas habitualmente.

No obstante, no debemos descartar un uso adecuado de la propia IA para evaluar otros algoritmos mediante programas específicos de auditoría, lo que constituye un camino de gran recorrido en el futuro.

Podemos diferenciar entre varios tipos de auditorías: a) la de los datos con que se entrena un algoritmo; b) la del propósito y finalidad del algoritmo, c) la del algoritmo y d) la de los resultados.

No siempre es posible conocer el "código fuente" y solo en ocasiones es útil poder interpretarlo, al menos en una primera fase, pero si es posible, considerando su objeto y la finalidad para la que sirve, discernir su calidad y abordar su análisis.

No se trata de reproducir debates tecnológicos difícilmente comprensibles para la mayoría de la sociedad y difíciles de dilucidar en el marco judicial, sino de los efectos tangibles más fácilmente verificables como evidencias, que sirvan para mejorar su legítimo uso y su eficiencia real.

Todo lo descrito anteriormente, no sólo no anula la posibilidad de que los algoritmos puedan y necesiten ser auditados con rigor, sino que refuerza su necesidad de evaluación y transparencia y siempre, considerando con realismo, lo que puede auditarse y el cómo se audita.

Existen elementos y consideraciones que hacen más complejo el proceso y precisan de una visión más pragmática: un ex ante regulador que no penalice innovación, una auditoria tanto de su diseño como para su implementación, y el análisis riguroso de los efectos de su uso, a través de la formulación de una ética y unos valores explícitos institucionales y de gobernanza.

Crear el entorno adecuado, implica una responsabilidad específica de los diversos sectores económicos, sociales e institucionales, suficientemente conscientes e involucrados en los beneficios que a la humanidad puede aportar la Inteligencia Artificial, pero que también lo deben estarlo, de los peligros, sesgos y riesgos de un uso no responsable. Así mismo, resulta fundamental una sociedad civil que actúe de forma proactiva con las instituciones, empresas y el sector público, en este proceso de transición, para posibilitar el mejor uso humano de estas tecnologías.

Las auditorías precisan contemplarse desde un enfoque muy pragmático, de ítems que resulten fundamentales, que analicen y evalúen los algoritmos para comprobar cómo funcionan y si están cumpliendo sus objetivos declarados, o produciendo resultados sesgados y generando nuevas vulnerabilidades sociales, al margen del marco jurídico y regulatorio existente en cada caso.

El Instituto de Auditores Internos de España, a través de La Fábrica de Pensamiento y del trabajo “Auditoría Interna de la Inteligencia Artificial aplicada a procesos de negocio”, las auditorías y consultoras privadas, la Agencia de Protección de datos, los trabajos efectuados en diversas universidades como la Pompeu Fabra, Pablo Olavide, Complutense, Carlos III o Politécnicas, junto a otros expertos y una amplia experiencia internacional en centros de excelencia, promueven trabajos y guías a través de análisis y estudios relevantes. Cada vez más empresas recurren a otras especializadas para encargar o revisar sus algoritmos, cuando se enfrentan a críticas y pérdidas de mercado y reputación, por resultados defectuosos y sesgados.

3. CÓMO ANALIZAR EL SISTEMA ALGORÍTMICO.

En primer lugar, es necesario estudiar el sistema algorítmico en su contexto y en el de la organización a la que sirven, establecer su tipología y conocer el objetivo y la finalidad que pretende, recabando información de los actores e intereses implicados en su diseño, financiación e implementación.

De esa forma, será posible considerar el perfil inicial de los riesgos a los que puede enfrentarse. Dicha información resulta fundamental para la elaboración de un plan de auditoría para el algoritmo, que no puede desvincularse de la gobernanza de la organización que lo integra, así como de los modelos de control interno y riesgos establecidos en su contexto.

Adicionalmente, dicho plan precisará elegir el paquete de herramientas y pruebas de análisis precisos para su despliegue, que puede integrar especificidades propias del funcionamiento de los algoritmos y la tecnología de la IA, también de la experiencia aprendida de los distintos métodos *testing* de software empleados en los últimos años, así como la larga experiencia en técnicas de auditoría, y también los numerosos trabajos y guías existentes de evaluación.

3.1 Calidad de la información y bases de datos.

El conocimiento de las bases de datos utilizados y la información empleada en el desarrollo, entrenamiento y evaluación del sistema algorítmico, es un apartado fundamental.

Tanto la validez de las muestras, como los grupos relevantes elegidos en las bases de datos, así como las llamadas variables proxi, que son aquellas que no parecen influir de forma aislada, pero que mediante inferencias o correlaciones con otras variables pueden repercutir en el desarrollo de sesgos y en las propias decisiones algorítmicas, deben ser analizados.

Se puede generar un tratamiento desigual de los distintos colectivos, bien por la falta de calidad de los datos elegidos como por el hecho de que sus usos en el entrenamiento algorítmico vengan ya sesgados en las bases de datos, o debido a que su aplicación en el algoritmo produzca ese efecto.

Para verificar la necesaria representatividad de las variables elegidas, y que estas se contemplan de forma adecuada, se viene utilizando todo un conjunto de tablas de naturaleza estadística referidas entre otros aspectos a materias como la edad, el género, la raza, y una distribución por grupos específicos de los colectivos incorporados.

La calidad, fiabilidad y validación del origen de estos datos es en consecuencia esencial para la propia evaluación del algoritmo. Junto a ello, el análisis del código fuente, cuando sea viable su conocimiento e interpretación, fundamentalmente a través del uso de nuevas pruebas de entrada y salida y la experimentación con nuevas órdenes de trabajo.

A través de ratios de impacto, tasas de falsos positivos y falsos negativos y otras pruebas, estas aplicaciones métricas, conjuntadas con los *testing de software*, permitirán un análisis de rigor acerca de la confianza que ofrece el algoritmo.

3.2. Métodos de *testing de software*.

Como referencias auditoras de algoritmos pueden utilizarse también modelos más tradicionales de *testing de software*. Específicamente:

El llamado enfoque funcional o de caja negra, que pretende considerar los riesgos potenciales de un modelo, abstrayéndose del código fuente, las rutas de tipo interno existente y la información que desconoce. Utilizando pruebas del software, se posibilita validar o no los requisitos funcionales del propio sistema. Su finalidad es descubrir en qué instancias el resultado no se comporta o se atiene a lo que se espera de él, o a sus pretendidas especificaciones. No se preocupa de lo que el software puede estar generando con sus mecanismos de aprendizaje, lo hace enfocándose en su entrada y salida y, tomando como base las especificaciones, selecciona métricas y resultados, a través del comportamiento del sistema y su comunicación. Es ciertamente habitual su utilización dentro de los propios sistemas o en aquellos módulos que van a actuar como interfaz con los usuarios

En su plan de trabajo, se suele primero examinar los requisitos y especificaciones del sistema. Después, se seleccionan pruebas de entradas válidas (caso de prueba positivo), para verificar si las procesa correctamente. También se seleccionan algunas entradas no válidas (caso de prueba negativo), para verificar que el sistema puede detectarlas. A través del medidor de salida esperada, determina las expectativas de todas esas entradas. Finalmente, se ejecutan los casos de prueba y, con un probador de software, se comparan los resultados reales y esperados. Los defectos, si los hay, se corrigen y se vuelven a probar.

Existen diversos tipos de pruebas de caja negra. Entre otras se encuentran las siguientes:

- *Pruebas funcionales.* Relacionado con los requisitos funcionales del sistema; lo hacen mediante probadores de software.

- *Pruebas no funcionales.* Analizando rendimientos, escalabilidad y usabilidad.
- *Test y pruebas de regresión.* Tras actualizaciones o mantenimientos del sistema.

En los algoritmos de caja negra no es posible expresar de manera explícita las reglas utilizadas para las decisiones que genera. Se trata por tanto de patrones no explicitados.

La prueba se limita a analizar los datos que componen las funciones de entradas y salidas, con la pretensión de generar cierta amplitud de cobertura y valorar la eficiencia respecto del resultado esperado, minimizando el número de casos de análisis sobre los que puede actuar. Pretenden alcanzar evidencias acerca de la presencia de aciertos o errores asociados, tan solo a la prueba efectuada que es la que resulta disponible. Tiene un carácter esencialmente funcional, detectando fallos en las respuestas cuando la operación depende de respuestas interdependientes o provenientes de otros módulos de los sistemas.

En la auditoría interna resulta conveniente establecer algunos mecanismos de monitorización continua para la identificación de procesos ineficaces de los sistemas de IA, en el caso de que se produzcan incidentes importantes o las soluciones hayan evolucionado/aprendido de manera inapropiada. Ciertas ineficiencias de los sistemas de IA pueden corregirse a través de mecanismos de reversión para la corrección de algoritmos y acceso a datos limpios, con revisiones en el tiempo.

En el método de pruebas estructurales o de caja blanca, se analiza la estructura interna del algoritmo y el código fuente. Se vincula al análisis de los valores de entrada, para examinar flujos y detalles procedimentales en la ejecución del programa, para cerciorarse de que se desenvuelven con los valores de salida adecuados. Pueden desplegarse por unidades, entre flujos, y en la integración de subsistemas. Para realizar las pruebas de los requerimientos funcionales existen varias opciones, la más utilizada es por medio de *check-lists.*

Se pueden vincular también a requerimientos, que no se refieren directamente a las funciones específicas que incorpora el sistema sino a propiedades emergentes de éste como la fiabilidad, la respuesta en el tiempo, la capacidad de almacenamiento, y la interoperabilidad con otros sistemas de software o hardware. Así mismo, a factores externos como los reglamentos de seguridad y las políticas de privacidad.

Estos requerimientos no funcionales generales suelen estar enmarcados en aspectos como la escalabilidad del sistema, el uso óptimo de las conexiones a las bases de datos, y las posibles necesidades de crecimiento. También la disponibilidad continua del servicio, contemplando requerimientos de confiabilidad y la consistencia de los componentes de la actividad ante recuperaciones, y si cuenta con alarmas precisas para contingencias, evitando las pérdidas de información y contemplando la interrupción de transacciones para que estas finalicen de forma correcta.

Planes de seguridad acordes con la delicadeza y la sensibilidad de la información que maneja, de acuerdo a las especificaciones funcionales y las políticas, normas y estándares de seguridad requeridas.

El mantenimiento implica estructurar el código fuente de una manera consistente y predecible, e implementar las interfaces asegurando una fácil implementación en el sistema, generalmente con modelos estandarizados.

Los requerimientos no funcionales son básicos dentro de un desarrollo de software. Al ser requerimientos implícitos internos, se perciben menos por los usuarios, pero resultan muy importantes para las pruebas porque sin estos no se puede asegurar la funcionalidad y un aceptable nivel de calidad del sistema.

• *Otras principales pruebas y técnicas de uso* se vinculan al flujo de control, al flujo de datos, a pruebas de bifurcación, pruebas de caminos básicos, y alcanzan a las pruebas de integración. Pueden también efectuarse con particiones en clases de equivalencia y el análisis de casos límites y fronteras.

4. EL PAPEL DE LAS TÉCNICAS CUALITATIVAS

Las técnicas cualitativas pueden tener también un papel importante en el proceso de la auditoría algorítmica. La auditoría y la evaluación nos muestran herramientas y usos habituales que pueden incorporarse fácilmente al análisis algorítmico. Los usuarios y destinatarios muy diversos de los algoritmos, al incorporar diversos sectores: empresas, sector público, técnicos, tecnólogos, y diversos grupos de ciudadanos, integran una amplia gama de intereses.

También los medios de comunicación, las universidades y centros de investigación, las asociaciones ciudadanas y las propias asociaciones de auditores son actores interesados en el uso y la mejora de los algoritmos, tanto en su vertiente interna más técnica como en su faceta social, regulatoria y ética.

El análisis cruzado de información entre los distintos actores puede ser una potente herramienta de auditoria en las diversas vertientes. Puede reforzar el papel de arbitraje o integrador que algunas escuelas de evaluación proponen para las políticas públicas y, en este caso, puede servir también para validar la realidad compleja del despliegue técnico que integra. La triangulación de focos de análisis y pruebas, también de equipos plurales de auditores, parece ser ciertamente recomendable.

La revisión de la creciente literatura académica en este campo, y los observatorios algorítmicos y repositorios de experiencias, están llamados a cumplir un papel creciente en la auditoría algorítmica en el futuro. La consulta a expertos y validadores también se abre camino para reforzar sendas de evidencias.

Cualquiera de las técnicas de las ciencias sociales puede ser empleada, siempre con el necesario rigor, y utilizada en el análisis algorítmico.

5. LAS CONCLUSIONES Y RECOMENDACIONES DE LAS AUDITORIAS

El objetivo de la auditoría es adquirir una buena comprensión de cómo y a quien afecta la creación y uso de los algoritmos. Con el uso de métricas, las pruebas e información obtenida y las validaciones, es posible alcanzar las evidencias y los análisis que se precisan sobre sus efectos y los cambios que necesitan incorporar, en su caso, para un uso eficaz por parte de empresas y del propio sector público.

Pero todavía es necesario incorporar a la auditoría el marco de cumplimiento de legalidad, llamado a formar parte del paquete principal específico del contenido del informe.

La auditoría de legalidad deberá considerarse un factor esencial, conforme el nuevo marco regulador vaya entrando en vigor, y es un objeto de preocupación creciente.

Quisiera referirme de forma especial, dentro de los criterios habituales en evaluación, a la importancia singular que en la auditoría algorítmica adquieren los principios de relevancia y pertinencia. Las características de complejidad y dificultad hacen más importante aún si cabe, una buena elección del objeto auditado y el refuerzo de la "auditabilidad" de ambos principios.

Algunas de las preguntas que podrían resultar útiles en la evaluación de los algoritmos, podrían adoptar la forma siguiente:

¿Qué objetivo(s) tiene el algoritmo?
¿Qué se espera de la utilización de dicho algoritmo?

¿Es relevante dicho objetivo?
¿Se ha alcanzado ese objetivo?
¿Ha tenido el algoritmo efectos inesperados?
¿Cuál es el su coste?
¿Se encuentra su coste en el mismo nivel que el de algoritmos similares?
¿Hay capacidades en la organización que utiliza el algoritmo para adaptarlo cuando es necesario? (sostenibilidad)
¿Qué impacto ha tenido la utilización del algoritmo?

La protección de los datos, junto con la garantía de los derechos y principios digitales, la garantía de privacidad y todo aquello que se deriva de la propiedad del dato, como es el consentimiento vinculado a la cesión sólo para fines específicos, repercuten en el análisis auditor. También la problemática de la responsabilidad jurídica en los usos indebidos de los algoritmos adquiere una importancia singular.

Por lo demás, dicho informe puede tener unas características similares a las que figuran en las numerosas guías y manuales de auditoría y evaluación existentes, incluyendo la necesidad de incorporar un informe ejecutivo y unas conclusiones y recomendaciones coherentes con el trabajo realizado.

En las conclusiones y recomendaciones, y dentro del objeto de mejora de los algoritmos y su funcionamiento, suelen incorporarse como hallazgos los aspectos vinculados al cumplimiento o incumplimiento del marco regulador existente, la gobernanza de la IA establecida, el contexto de su aplicación, los riesgos relevantes en las bases y el despliegue metodológico de los algoritmos, así como las principales variables de entrenamiento y sus efectos.

Respecto a las bases de datos empleadas, resulta inexcusable el chequeo de datos, la veracidad, fiabilidad y/o necesidad de actualización de las fuentes y los resultados de las muestras de variables efectuadas. También en la selección de datos, la idoneidad de los atributos elegidos para obtener causalidad o atribución para que garanticen efectividad o eficiencia del sistema. Así mismo, las recomendaciones para modificar la entrada de datos, cuando no son acordes con objeto pretendido, y la restructuración o limpieza de las bases de datos y variables empleadas. Especial mención precisa el resultado del análisis respecto a los atributos identificativos de grupos vulnerables o grupos sociales específicos afectados.

6. LA AUDITORÍA ALGORITMICA Y EL ENTORNO DE LA IA EN LA EMPRESA.

El Instituto de Auditores Internos de España acaba de publicar un interesante trabajo aplicado al uso de la IA en los procesos de negocio.

A partir de la metodología del modelo COSO 2013 para el control interno, considera cinco actividades clave para integrar en estos procesos:

1. La identificación de los riesgos según la complejidad de los algoritmos, objetivos y el entorno. Vinculando también el apetito de riesgos como referencia.
2. La supervisión de la arquitectura de datos empleados. Incluyendo protección de la privacidad, actualización, seguridad y atención a ciber ataques.
3. El análisis del perfil y relevancia de los modelos empleados.
4. La supervisión de la implementación de los algoritmos, su entrenamiento, garantía de la existencia de pruebas suficientes y otros *testings*.
5. La revisión periódica, actualización de métricas e indicadores de desempeño.

Específicamente, en relación con la primera etapa COSO *Entorno de Control,* que agrupa los conceptos de gobierno, cultura, ética y valores, señala el trabajo comentado la necesidad de una política de gobierno de los datos, dado el volumen ingente de su uso en IA, y considerar las amenazas y oportunidades para su actividad.

También se ocupa, en este ámbito específico de la IA, de la necesaria garantía del cumplimento regulatorio, una cultura ligada a los componentes éticos, el despliegue de la evaluación de riesgos derivados, así como de contar con personal con competencias adecuadas.

Propugna el aseguramiento o evaluación en la gestión de riesgos, relacionados con la confiabilidad de los algoritmos subyacentes y de los datos en los que se basan.

Considera el Instituto de Auditores Internos que entre las funciones de la Auditoría Interna se encuentra la evaluación el cumplimiento de sus objetivos estratégicos y, entre ellos, los vinculados al despliegue del sistema de IA. Deberá por tanto incluir la IA en la evaluación de gestión de sus riesgos relevantes y considerarlos en su plan de auditoría.

También, la Auditoría Interna debe ayudar a la compañía a evaluar, comprender y comunicar el grado en que los algoritmos de la IA tienen un efecto positivo o negativo para la creación de valor en el corto, medio y largo plazo.

7. DE LA LIMITACIÓN A LA OPORTUNIDAD

La complejidad, y quizás un academicismo excesivo, limitan el avance en este campo. La asunción de una práctica corporativa más adecuada se dificulta al no contar, aunque ya por poco tiempo, con una regulación institucional europea de Inteligencia Artificial que, aprobada en el parlamento, sirva de soporte jurídico global más allá de la problemática de protección de datos existentes.

Para quienes desean contratar auditores externos, y para la propia auditoría interna, no existen estándares suficientemente claros sobre lo que deberían implicar estas auditorías. Incluso propuestas institucionales que requieran auditorías anuales de los algoritmos de contratación, un sello de aprobación de un auditor, en estas condiciones, puede significar un escrutinio mucho mayor que el de otros y, adicionalmente, los informes de auditoría pueden también estar sujetos a acuerdos de no divulgación.

No obstante, el desarrollo de análisis algorítmicos es una necesidad social cada vez más urgente y relevante. Resulta imprescindible para que esta tecnología sea más explicable, más transparente, más predecible y controlable por la ciudadanía y las instituciones públicas reguladoras. También para su uso adecuado en las empresas y en los programas y políticas públicas.

Precisa de organismos reguladores y también certificadores independientes, con perfiles muy específicos, y no deben ser ajenos tampoco a la presencia de la sociedad civil.

La auditoría algorítmica, en la vertiente auditora interna y auditora externa, tienen cada cual un papel específico complementario fundamental y precisa equipos multidisciplinares con competencias adecuadas. La actividad necesita, por otra parte, de una información y una colaboración estrecha con la programación o la empresa suministradora del algoritmo, y no sólo con la gestora.

Las administraciones públicas deben dotarse de recursos con equipos competentes, inscritos en sus propios mecanismos de inspección, y control interno y externo. No puede mimetizarse sólo con el mercado existente, su responsabilidad ética y ciudadana no se lo permite, dados los riesgos de gobernanza, transparencia y reputación en los que puede incurrir.

Precisamente, la sociedad civil, en un modelo en el que se exige transparencia en la explicación del funcionamiento de los algoritmos, debe contar también con un marco de participación para incidir adecuadamente en el proceso verificador y auditor algorítmico.

Estaremos atentos a cómo evoluciona el mercado auditor, señalaba Gemma Galdón en una entrevista en Forbes de hace ya algún tiempo. Compartimos con ella que lo que necesitamos no es tanto la teoría sino la experiencia que supone su aplicación progresiva,también añadiría que es imprescindible culminar una regulación todavía demasiado inicial y una adecuación oportuna.

Algunos de los aspectos críticos que habitualmente acompañan al trabajo, según las características propias de cada auditoría, serían las siguientes:

- Un Plan de trabajo que tras contemplar la gobernanza del diseño del sistema con el gestor y/o el usuario, programe de manera consensuada o independiente el objeto a auditar, las técnicas de evaluación que se pretenden utilizar, la elección de los *testing* de software más adecuados, junto con la propuesta de un equipo plural de evaluadores experimentados que lo ejecuten.
- El análisis de los riesgos y sesgos que ofrecen mayor vulnerabilidad en función del tipo de algoritmo, la literatura de evidencias, los observatorios y el marco jurídico, económico, tecnológico y ético de actuación.
- El análisis del origen de la información y la evaluación de la correspondiente métrica de las bases de información incluidas en el entrenamiento y los usos del algoritmo.
- Las garantías establecidas respecto de la privacidad de los datos y el respeto a los límites mostrados del consentimiento de su uso.
- Una descripción general de las medidas técnicas y organizativas de seguridad implementadas con el algoritmo.
- Información sobre las responsabilidades de las partes implicadas respecto al funcionamiento del modelo.
- Hallazgos de posibles sesgos en los resultados obtenidos por las pruebas en el uso de los algoritmos.
- Funcionamiento de la interacción algoritmos-humanos y otros puntos críticos.
- Problemáticas específicas de sectores vulnerables o protegidos.

En consecuencia y como resumen, contestando a la pregunta del inicio de este artículo, más allá de la complejidad y dificultades objetivas se puede (y se debe) evaluar rigurosamente los algoritmos, dada la existencia de una gran variedad de tipos de auditorías y también de objetos de evaluación y siempre que se considere con realismo lo que puede y no puede auditarse y el cómo se audita.

8. BIBLIOGRAFÍA

Guías de Tecnología y Protección de datos, de adaptación al Registro de Protección de datos y Modelo de Informe de Evaluación de Impacto en la protección de datos en las Administraciones Públicas (2019, 2020,2021). Agencia Española de Protección de Datos.

Declaración conjunta de la Comisión, el Parlamento y el Consejo de la UE acerca de los derechos digitales en la década digital. (2023) Comisión Europea.

Benjamins Richard, Salazar Idoia y otros autores (2020). *El mito del algoritmo*. Ediciones Multimedia.

Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial" (2021,2022). Comisión Europea.

Du Sautoy Marcus (2020). *Programados para crear*. Acantilado.

Guía de Auditoria Algorítmica (2021). Éticas Consulting

Gómez Emilia (Editora). Castillo Carlos, Charisi Vicki, Dalh Verónica y otros numerosos participantes. Comisión Europea. Conferencia-Taller (2018). *Evaluación del Impacto de las máquinas inteligentes en el comportamiento humano: esfuerzo interdisciplinario*. https://arxiv.org/pdf/1806.03192.pdf

Lagares JA. Díaz Norberto. Barranco Carlos D. (2022) Aprendizaje profundo: una nueva vía para convertir el dato en conocimiento. *Revista de Economía Industrial, número 423. Monográfico de Economía del Dato (2022)*. Ministerio de Industria, Comercio y Turismo.

Ríos Insua David (2022). Economía del dato: Luces y Sombras. *Revista de Economía Industrial, número 423. Monográfico de Economía del Dato (2022*). Ministerio de Industria, Comercio y Turismo.

Tortosa Illana (Coordinador), Ausin Sánchez, Pablo. Corredera L.E. y otros (2023). La fábrica de Pensamiento. *Auditoría Interna de la Inteligencia Artificial aplicada a procesos de negocio (2023).* Instituto de Auditores Internos de España.