Autoritat Catalana de Protecció de Dades

Inteligencia Artificial:
Decisiones Automatizadas en Cataluña

Preámbulo

Hacia una sociedad inteligente

Por M. Àngels Barbarà,
Directora de la Autoridad Catalana de Protección de Datos

En los últimos 10 años se ha producido una evolución tecnológica tan amplia, profunda y a tal velocidad que los seres humanos no somos capaces de prever cómo estas tecnologías se incorporarán en nuestras vidas[1].

La innovación y el desarrollo tecnológico fundamentan su crecimiento en los datos, personales y no personales. La convergencia en el tiempo de varias tecnologías-integradas para utilizar de forma masiva los datos y generar valor- amplifica el impacto en la vida de las personas. Además, la explotación de la información no agota su valor en un uso, sino que pueden existir múltiples, generándose así una cadena de valor que otorga un poder inmenso a aquellos que gobiernan la tecnología y los datos.

Me refiero a tecnologías como la computación en la nube, que permite acceso desde diferentes lugares y gran capacidad de almacenamiento a bajo coste; el Internet de las Cosas (IoT, en inglés) que recoge información de todos los entornos, en cualquier formato y momento; el análisis de datos masivos (Big Data Analytics) que permite generar nuevo conocimiento en salud, transporte, educación, medio ambiente, etc. Y, finalmente, la inteligencia artificial (IA), que engloba aquellos sistemas que manifiestan un comportamiento inteligente, capaces de analizar su entorno y actuar -con cierto grado de autonomía- con el fin de alcanzar objetivos específicos[2].

La IA entendida en un sentido amplio, la podemos ver como una herramienta de apoyo a la toma de decisiones en cualquier ámbito. Este apoyo en la toma de decisiones afecta, y de forma muy relevante, a las personas, ya que los datos y la tecnología dibujan perfiles y patrones que sirven para decidir e influir directamente.

Cambios invisibles

Nos adentramos en un contexto digital que nos expone a nuevas formas de vulnerabilidad a través del tratamiento de nuestros datos personales. Estos datos son utilizados por las tecnologías emergentes para generar valor y modifican, en gran medida, la forma en que las personas y organizaciones entienden la realidad que los rodea.

[1] "Hemos perdido la capacidad de censurar, de gobernar o de siquiera negociar con la tecnología, que es vista como una consecuencia lógica de nuestro progreso intelectual" (Joaquín David Rodríguez Álvarez, 2016, "La civilización ausente", p.11).

[2] Comunicación de la comisión al Parlamento Europeo, al Consejo Europeo, al consejo, al Comité Económico y Social Europeo y al Comité de las regiones: Inteligencia artificial para Europa. COM/2018/237 final.

El cambio se está produciendo de forma casi invisible, entrando en nuestra vida personal, profesional, social, política con apariencia de normalidad, de comodidad, de mejora de nuestras vidas sin impacto, sin consecuencias.

Este es uno de los riesgos más importantes que se derivan de las tecnologías emergentes: su intrusión silenciosa en nuestros derechos y libertades, transformando la sociedad en la que vivimos, de una forma que parece que no cause ningún impacto. Esta sociedad modulada por las tecnologías, la que llamamos sociedad digital, será diferente a la que hemos conocido hasta ahora.

En la sociedad digital -que se desarrolla dentro de la cuarta revolución industrial- las tecnologías se fusionan mediante los mundos físico, digital y biológico[3]. La tecnología se integra en nuestro cuerpo y se desarrolla a partir de los nuevos datos que generamos. Está diseñada para profundizar en el conocimiento de nuestros impulsos, con el objetivo de incidir de forma directa, invisible y en tiempo real- en la toma de decisiones de las personas y sobre las personas.

Su invisibilidad y opacidad hace que no seamos conscientes del riesgo que generan los valores y principios que han sustentado nuestra sociedad hasta ahora: la dignidad, la libertad, el libre desarrollo de la personalidad, la igualdad, reformulando aspectos básicos de la democracia y del Estado de Derecho.

Nuevo modelo de sociedad

Se está gestando un nuevo modelo de sociedad sin que aquellos que la integran se den cuenta. Es por este motivo, que desde la Autoridad de Protección de Datos -con la misma orientación que el conjunto de autoridades de control europeas y de las principales instituciones de garantía de los derechos humanos- queremos aportar claridad y comprensión en el contexto de la sociedad catalana. Igualmente, queremos ayudar a las personas a defender de forma más activa sus derechos y libertades desde el conocimiento y la información.

En esta ocasión, nos hemos centrado en la inteligencia artificial (IA) en Cataluña y, en concreto, en los algoritmos de decisión automatizada (ADA) por su potencial de impacto en la privacidad y en el conjunto de derechos y libertades de las personas.

Como no puede ser de otra manera, lo hacemos desde la regulación en materia de protección de datos. En mayo del 2018 fue de plena aplicación el Reglamento General de Protección Europea (RGPD). Esta norma -de aplicación directa al conjunto de los Estados miembros de la Unión Europea- ha venido a dar forma jurídica a la voluntad de proteger los derechos y libertades de las personas en el contexto del cambio de modelo de sociedad hacia la sociedad digital. Y es un impulso europeo a las tecnologías emergentes como herramienta para salir de la crisis coyuntural en la que se encontraba inmersa Europa.

El RGPD establece un nuevo modelo de garantía de los derechos y libertades de las personas basado en la responsabilidad y compromiso de las entidades con los

[3] Klaus Schwab (2016) “La cuarta revolución industrial”, Ed. Debate.

derechos de las personas, y en un enfoque en el riesgo donde cada entidad tiene que valorar como, el concreto tratamiento de datos que realiza, afecta a sus usuarios.

Este nuevo marco europeo se basa en los principios tradicionales de protección de los datos (lealtad, licitud, transparencia, limitación de la finalidad, minimización, exactitud, limitación de los plazos de conservación, seguridad), y en unos requisitos de legitimación de los tratamientos muy similares los existentes. Pero, cambia radicalmente la forma de afrontar el cumplimiento, justamente, en base a los criterios antes mencionados, de responsabilidad proactiva y enfoque en el riesgo.

El nuevo modelo obliga a las organizaciones a plantearse cómo hacer los tratamientos de datos para que la forma de hacerlo sea la más adecuada para reducir el impacto en los derechos de las personas, al tiempo que permite adaptarse de forma integrada con la idiosincrasia de cada organización.

Dimensión ética y social

La dimensión ética y social ha tomado, en los últimos años, una especial relevancia en el debate en torno a la evolución del derecho a la protección de datos, y cómo esta dimensión debe regir la innovación tecnológica, protegiendo los valores, principios y derechos que son inherentes a nuestra sociedad y que definen la tradición europea.

En el contexto de la IA y los ADA -de innovación y disrupción continua- se exige, a las entidades que quieren utilizar estas herramientas, un paso más en el cumplimiento. Ya no son suficientes las obligaciones mecánicas y formales. Hay que diseñar, desarrollar y utilizar estas herramientas incorporando una dimensión ética y social.

El uso intensivo de los datos -mayoritariamente datos personales- obliga, desde una perspectiva jurídica y ética. La IA no puede ser un fin en sí mismo, sino un medio para lograr una transformación de la sociedad poniendo a la persona en el centro. Pero cumplir esta premisa no es fácil, ya que la IA se incrusta en las tecnologías que utilizamos de forma rutinaria e invisible.

La IA sirve a finalidades muy diversas y tiene el potencial de mejorar el bienestar de la gente. Como indica la OCDE–en su reciente Recomendación[4]– la IA tiene el potencial de contribuir a una actividad económica global y sostenible, de aumentar la innovación y la productividad.

En este entorno, el derecho a la protección de datos personales se ha convertido en un elemento clave para proteger a las personas frente al potencial uso abusivo de la información. Vivimos en una sociedad donde la identidad digital de cada individuo, responda o no a la realidad, es la que se toma como base para la personalización de productos y servicios, para influir en las decisiones de las personas y para decidir sobre ellas. Pero, como recuerda el Grupo de expertos en ética creado por el Supervisor Europeo: la protección de datos no es un tema técnico o legal, es una cuestión profundamente humana.

[4] OECD, Recommendation of the Council on Artificial Intelligence (OECD/Legal/0449), adoptada el 21 de Mayo de 2019.

La tecnología mejora nuestra calidad de vida, nos la hace más cómoda, pero no puede ser a cambio de que nuestros derechos se vean limitados sin medida; por ejemplo, mediante la manipulación y distorsión de nuestra percepción de la realidad o "vigilando nuestras emociones"[5]. Como expone el Informe hecho por AlgorithmWatch: "Automating Society" hay que hacerse otras preguntas, como quién decide cuándo se debe utilizar IA, con qué fines y / o quién y cómo la desarrolla.

Son cuestiones básicas, y el Parlamento Europeo así lo indicaba cuando decía que en IA hay que invertir no sólo en tecnología e innovación, sino también en los ámbitos sociales, éticos y de responsabilidad. Hay elementos que hay que integrar en el diseño, desarrollo y aplicación de las tecnologías. Estos elementos deben ir más allá de las obligaciones reguladas en las normas, deben tener una proyección universal. Es decir, deben ser comprensibles y aplicables de manera global.

IA segura y fiable

Ante estas enormes capacidades de las tecnologías emergentes, debemos reflexionar sobre si el camino que ha tomado la innovación, su diseño y su uso, se pueden hacer de una manera menos intrusiva para la vida de las personas.

Debemos promover un diseño que no limite la libertad del individuo para tomar decisiones y donde su comportamiento no sea controlado ni dirigido por aquellos que ostentan el poder sobre la tecnología. Y eso sólo lo conseguiremos si la sociedad entiende cómo funcionan los algoritmos y puede hacer planteamientos críticos respecto de las tecnologías que se utilizan. Hay que diseñar una IA segura y confiable, donde se integren los principios de la transparencia, la explicabilidad, la seguridad, la auditabilidad y la responsabilidad.

En octubre de 2018, en la 40ª Conferencia Internacional de Autoridades de Protección de Datos, se aprobó la Declaración sobre Ética y Protección de Datos en Inteligencia Artificial. En abril de 2019, el Grupo independiente de expertos de alto nivel sobre Inteligencia artificial de la Comisión Europea hizo pública sus Directrices Éticas para una IA fiable. Y en mayo pasado, la OCDE publicaba sus Recomendaciones sobre IA. Finalmente, en el transcurso de la 41ª Conferencia internacional de protección de datos, el Consejo de Europa anunciaba el inicio de los trabajos para la elaboración de un marco legal en IA.

En este contexto, desde la Autoridad Catalana de Protección de Datos, conscientes de que Cataluña es un importante motor de desarrollo tecnológico, y partiendo del trabajo que hemos realizado en estos últimos años en materia de ética y protección de datos, hemos iniciado un proyecto específico destinado a identificar cómo se está utilizando la IA en Cataluña, y cómo se incorpora la vertiente ética en estos entornos.

En este Proyecto, nos hemos enfocado en un uso concreto de la IA: los algoritmos de decisión automatizada (ADA) y cómo afectan a los valores que definen la tradición europea. Forma parte de este informe una recopilación de aplicaciones de los ADA realizado por la periodista especializada en Tecnologías de la Información y la

[5] Resolución 2018/2088(INI) del Parlamento Europeo, de 12 de febrero de 2019, sobre una política industrial global europea en materia de intel·ligència artificial i robòtica.

Comunicación, Karma Peiró. Quiero señalar que, aunque todo el informe tiene un carácter divulgativo, este carácter está especialmente marcado en la descripción de las aplicaciones de los ADA. Con una voluntad de hacer estas tecnologías más comprensibles, huimos de los detalles más técnicos de su diseño o aplicación.

Es necesario, también, identificar las posibles desviaciones y discriminaciones derivadas del uso de las tecnologías emergentes para evitar que se perpetúen discriminaciones "históricas" en base a la raza o al sexo, o que se generen nuevas, no perceptibles, a partir de la creación de grupos en base a patrones y correlaciones algorítmicas.

El enfoque ético de Europa respecto a la IA refuerza la confianza de las personas y permitirá generar una ventaja competitiva para las empresas europeas de IA[6]. Desde la Autoridad queremos contribuir con la elaboración de los principios que deben gobernar el diseño, desarrollo y uso de la IA. Este informe que se presenta sólo es el comienzo del largo camino que nos queda por recorrer en la garantía de los derechos y libertades de las personas en un mundo en constante evolución.

[6] Comunicación de la comisión al Parlamento Europeo, al Consejo Europeo, al consejo, al Comité Económico y Social Europeo y al Comité de las regiones: Generar confianza en la inteligencia artificial centrada en el ser humano. COM/2019/168 final.

Índice de contenidos

1 Protección de datos y ética en las decisiones automatizadas

Nuestras vidas están cada vez más condicionadas por la interacción que tenemos con la tecnología. En el centro de este ecosistema están nuestros datos. Instagram tiene nuestras fotos, Google nuestro correo electrónico y documentos, Facebook conoce nuestro círculo social y Twitter nuestros intereses. En estos casos somos nosotros mismos que generamos los datos y los ponemos a disposición de multitud de organizaciones. Otras veces, los datos se generan y recogen sin que seamos conscientes. Por ejemplo, el rastro que dejamos con los dispositivos móviles (con sensores de todo tipo), la navegación en Internet o las compras con tarjeta.

Esta gran cantidad de datos personales disponibles acentúa la incidencia de la tecnología sobre la vida de las personas. En otras palabras, permite un altísimo grado de personalización en los servicios que nos ofrecen. Esto se ve claramente en el caso de las noticias. Antes, podíamos elegir un periódico o un telediario en función de su línea editorial. Ahora, hay aplicaciones de móvil, recomendadores de noticias o redes sociales (Google News, Facebook, Twitter) que adaptan la información al perfil de cada persona. El objetivo no es otro que mejorar la experiencia de los usuarios, mostrar sólo aquella información que es de su interés para fidelizarlo.

Este altísimo grado de personalización puede tener muchas ventajas para las personas, pero también riesgos para sus derechos y libertades. Los riesgos directamente relacionados con el tratamiento de datos personales son bastante obvios: un incidente de seguridad puede poner en peligro la confidencialidad de nuestros datos, se puede hacer un uso inadecuado de la información, etc. Otro tipo de riesgos, no tan obvios, son los que provienen de la aplicación concreta que se hace de los datos. En el caso de las noticias personalizadas, existe el riesgo de que la información que nos muestran esté demasiado restringida, lo que limita nuestro campo de visión de la realidad. Es el llamado *filtro burbuja*. Hay que tener presente que aquel que controla lo que ves, puede controlar lo que piensas, sientes y haces.

En este contexto, la capacidad de los algoritmos de tomar decisiones de forma automatizada es esencial. Un portal de noticias puede personalizar manualmente el contenido para un pequeño número de personas, pero no para miles. Hoy en día son los algoritmos que analizan el perfil de un usuario y deciden cuáles son los contenidos más adecuados.

En la discusión anterior nos hemos centrado en el caso de un portal de noticias, pero este es sólo un ejemplo de entre los miles de aplicaciones que algoritmos de decisión automatizada tienen en multitud de campos. Como parte de este informe, se incluye un estudio, realizado por la periodista Karma Peiró, sobre el uso de los algoritmos de decisión automatizada en Cataluña. En este estudio se explican multitud de aplicaciones en sanidad, asuntos sociales, trabajo, banca, etc. De estas aplicaciones resulta claro que los algoritmos de decisión automatizada pueden tener un gran impacto sobre la ciudadanía: diagnosis incorrecta de una enfermedad, denegación de una ayuda social, ser rechazado en un proceso de selección de personal, denegación de un crédito por parte de un banco, etc.

Este informe

El objetivo de este informe es concienciar a la ciudadanía y dotarla del conocimiento necesario para que haga un uso responsable de sus datos personales. Esto se consigue en dos fases.

La primera fase se materializa en un estudio de investigación sobre el uso de los algoritmos de decisión automatizada en Cataluña, realizado por la periodista especializada en Tecnologías de la Información y la Comunicación, Karma Peiró. El objetivo es informar a la ciudadanía, en tono divulgativo, sobre las ventajas y riesgos que conllevan.

En la segunda fase, entendiendo que la mayoría de estas aplicaciones hacen uso de datos personales, reflexionamos sobre las principales fricciones entre los algoritmos de decisión automatizada (y, por extensión, de la inteligencia artificial) además de los principios de protección de datos personales. Fruto de las reflexiones anteriores, proponemos una serie de recomendaciones finales.

Algoritmos de decisión automatizada en el RGPD

Desde mayo de 2018, el Reglamento General de Protección de Datos (RGPD) es la norma europea básica que regula el tratamiento de datos personales. La anterior norma tenía un carácter demasiado formalista y no pudo dar respuesta a los importantes retos que han acompañado al desarrollo tecnológico. Un enfoque basado en el riesgo y el principio de responsabilidad proactiva son dos novedades importantes del RGPD. El enfoque basado en el riesgo busca que las medidas que toma el responsable del tratamiento para proteger los derechos y las libertades de las personas sean proporcionales a los riesgos del tratamiento. La responsabilidad proactiva exige que el responsable del tratamiento sea capaz de demostrar que se hace de acuerdo con el RGPD. Es decir, no basta cumplir con el RGPD, hay que poder demostrarlo.

El RGPD refuerza los derechos que tenemos las personas respecto al tratamiento de nuestros datos; la llamada autodeterminación informativa. Ahora bien, para que las personas podamos ejercer los derechos que tenemos sobre nuestros datos de una manera apropiada, es necesario primero que conozcamos cuáles son estos derechos y, después, que seamos conscientes de la necesidad de protegerlos. Por ejemplo, uno de los principales errores es que damos el consentimiento para el uso de nuestros datos de manera demasiado generosa o, incluso, despreocupada. Dando siempre el consentimiento para el uso de nuestros datos reducimos sensiblemente la protección que el RGPD nos proporciona.

La redacción del RGPD busca ser lo más general posible. Es por ello que el RGPD no menciona tratamientos de datos concretos: se basa en unos principios generales. La única referencia a un tipo concreto de tratamiento está en el artículo 22 que habla de la toma de decisiones automatizadas. La gran relevancia que ha adquirido este tipo de algoritmos y las fuertes implicaciones que pueden tener sobre las personas hacen que sea necesario dar protecciones específicas.

Algoritmos de decisión automatizada e inteligencia artificial

Los algoritmos de decisión automatizada y la inteligencia artificial son conceptos fuertemente relacionados. Es cierto que es posible diseñar un algoritmo que toma decisiones automáticas de manera poco (o nada inteligente). Por ejemplo, un algoritmo para elegir el candidato más adecuado para un puesto de trabajo podría decidir de forma aleatoria (sin evaluar a ninguno de los candidatos). Ahora bien, con gran probabilidad, la decisión tomada por un algoritmo de estas características no será óptima. Para que los algoritmos de decisión automatizada sean realmente útiles es necesario que tengan un comportamiento inteligente.

La IA engloba aquellos algoritmos que buscan dotar a los ordenadores de un comportamiento inteligente. La definición de 'comportamiento inteligente' no es del todo clara. Hoy, la inteligencia artificial ha superado las capacidades humanas en muchas tareas concretas, pero todavía hay muchas tareas que no están a su alcance: es capaz de ganar al mejor jugador de ajedrez del mundo pero sería incapaz de realizar tareas ordinarias de una persona de mediana edad, como tener una reunión con la maestra de los hijos. Las personas nos movemos en diferentes contextos, ambientes y decidimos sobre multitud de temas cada día porque hemos acumulado un conocimiento previo que las máquinas todavía no tienen. En resumen, los humanos sabemos hacer muchas cosas con poca precisión; las máquinas, muy pocas con mucha precisión.

Hay varios niveles de comportamiento inteligente: desde aquellos sistemas más básicos que se limitan a aplicar un conjunto de reglas que han sido prefijadas, hasta complejos sistemas de aprendizaje profundo (que están inspirados en las redes neuronales biológicas). Fijar cuál es el límite a partir del cual un algoritmo es inteligente es una cuestión controvertida y que no tiene una relevancia especial para el fin de los algoritmos de decisión automatizada (tomar decisiones sin intervención humana). Por tanto, en este informe usaremos los términos inteligencia artificial y algoritmos de decisión automatizada de manera intercambiable. La única exigencia que ponemos para calificar como inteligencia artificial es que el algoritmo analice el escenario actual y tome la decisión en función de este análisis.

Decidir los límites y fijar qué es ético

Dado el impacto que los algoritmos de decisión automatizada y la inteligencia artificial tienen sobre las personas, hay que fijar el límite de lo razonable. Obviamente, la ley (particularmente el RGPD como principal regulación europea en protección de datos) es un instrumento básico. Ahora bien, no es el único instrumento. La ética tiene también un papel esencial. Por ejemplo, a la hora de definir la ley y de guiar su aplicación.

La determinación de lo que es y no es ético no es una cuestión que pueda resolver la APDCAT o un grupo de académicos. La ética es un contrato social y es la sociedad quien debe fijar que es ético. Para que este debate se produzca de forma satisfactoria, es necesario dotar a la sociedad de los conocimientos básicos necesarios. En esta tarea

este informe puede ser de gran utilidad: dando a conocer las aplicaciones de los algoritmos de decisión automatizada en Cataluña y analizando sus implicaciones.

Actualmente, la sociedad tiene una doble visión respecto la IA y los ADA. Por un lado, estamos acostumbrados a las mejoras que proporcionan los avances tecnológicos y que rápidamente adoptamos. Muchas veces el avance es incontestable. Por ejemplo, no hacer uso de las capacidades de la inteligencia artificial en la detección de enfermedades, no sería ético. Por otra parte, no debemos ser ingenuos y decir que todo vale. Ser demasiado permisivo con el uso de datos personales puede dar lugar a usos que interfieran de forma demasiado acentuada con los derechos y las libertades de las personas.

El gran potencial de la inteligencia artificial es un condicionante a la hora de fijar los límites. Se considera una tecnología disruptiva capaz de cambiar las relaciones de poder entre los estados y hace que nadie quiera quedar atrás. Ya se habla de la nueva carrera espacial entre superpotencias, esta vez centrada en el desarrollo de la IA. Los Estados Unidos -haciendo valer su posición como líder en desarrollo tecnológico- todavía tiene una posición dominante. China se ha incorporado más recientemente con un plan de inversiones multimillonario que, en una década, quiere, primero, igualar en conocimiento a los Estados Unidos y, después, superarlo. Japón es líder en robótica. Algunos países europeos buscan tener su parcela en el mundo de la inteligencia artificial, pero con presupuestos más modestos. Es en el desarrollo de una IA ética (también desde el punto de vista de la protección de datos) donde Europa tiene la delantera respecto de los otros actores.

Protección de Datos y IA

La gran cantidad de datos disponibles ha sido uno de los pilares fundamentales para el desarrollo de la IA que hemos vivido en la última década. Es por ello, que el efecto de la protección de datos sobre la inteligencia artificial es un tema controvertido.

Algunos sectores critican las garantías que ofrece el RGPD; diciendo que son un freno para el desarrollo de la IA en Europa. Estos actores abogan por una regulación más laxa, que facilite la reutilización de datos.

También existe la visión contraria, que dice que las limitaciones que impone el RGPD son un incentivo para el desarrollo de una inteligencia artificial que haga un uso más razonable y efectivo de los datos disponibles. Por ejemplo, fomentando el desarrollo de algoritmos más sofisticados que aprenden a partir de conjuntos limitados de datos, como lo hacen las personas. El ajedrez (aunque no trata con datos personales) es un claro ejemplo del funcionamiento de la IA actual. Los algoritmos más sofisticados se han entrenado jugando miles de millones de partidas, mientras que los grandes maestros humanos han jugado a lo sumo unos pocos miles.

El punto de vista de la APDCAT, como no puede ser de otra manera, es que el uso de la inteligencia artificial debe ser compatible con el RGPD.

La sociedad que queremos construir

Sin restar importancia a las críticas al RGPD, mirar al pasado nos puede ayudar a reflexionar a la hora de decidir el tipo de sociedad que queremos construir. La

primera mención al derecho a la privacidad es de 1980[7]y buscaba proteger a las personas de las intromisiones de los avances tecnológicos de la época:

> *"Las fotografías instantáneas y la empresa periodística han invadido los lugares sagrados de la vida privada y doméstica; y numerosos dispositivos mecánicos amenazan con hacer buena la predicción de que lo que se dice al oído en la habitación más retirada, lo pregonarán desde las azoteas."*

Salvando las distancias, estamos ante un caso similar: decidir si dejamos que la tecnología haga un uso indiscriminado de datos personales o si ponemos límites. Indudablemente, las nuevas tecnologías son mucho más intrusivas que una fotografía; en el sentido de que tienen la capacidad de generar un perfil de la persona mucho más extenso. Si hacer una fotografía de una persona en su ámbito privado puede violar el derecho a la intimidad, no deberíamos también limitar el uso de los datos que generamos en ese ámbito. Es la misma Declaración Universal de Derechos Humanos, en su artículo 12[8], que reconoce el derecho a la privacidad.

En el mundo actual podemos ver diferentes sensibilidades respecto a la relación entre inteligencia artificial y el respeto de los derechos y las libertades de las personas. En particular, respecto a la protección de datos personales tenemos estos tres escenarios:

Europa

Dispone de la legislación de protección de datos personales más avanzada. Salvo algunas excepciones (como obligaciones legales, interés público, interés legítimo), la legislación europea promueve la autodeterminación informativa. Es decir, el derecho de cada persona a decidir que el uso que debe hacerse de sus datos.

Como hemos comentado anteriormente, la normativa de protección de datos europea presenta algunas fricciones con la inteligencia artificial (*Véase el capítulo "Inteligencia artificial y protección de datos"*).

[7] Warren i Brandeis. "The Right to Privacy". Harvard Law Review. 1890.

[8]Declaración Universal de Derechos Humanos (https://www.ohchr.org/en/udhr/documents/udhr_translations/cln.pdf).

China

La protección de los datos personales es un tema poco desarrollado en China. A pesar de disponer de un estándar de protección de datos personales, este no es de obligado cumplimiento.

China hace un uso de la inteligencia artificial muy invasivo. Por ejemplo, el sistema de créditos sociales, actualmente en implantación, utiliza vigilancia electrónica masiva para controlar el comportamiento de los ciudadanos, asignar una puntuación a cada individuo, y premiarlos o penalizarlos. Algunas de las implicaciones que puede tener este sistema por los ciudadanos chinos son la prohibición de viajar, la exclusión de ciertas escuelas, la represión a minorías religiosas y la humillación pública.

Estados Unidos

Estados Unidos carece de una ley de protección de datos. Su marco legal está formado por cientos de leyes (estatales y federales) que tratan aspectos específicos.

Una diferencia que remarcar entre Europa y Estados Unidos tiene que ver con el consentimiento para el uso de datos personales opt-in contra opt-out. Esto quiere decir que en Europa el consentimiento se debe recoger de forma explícita, informada y para una finalidad determinada. En Estados Unidos el consentimiento es implícito: si una persona no quiere que una organización trate sus datos, lo debe pedir explícitamente.

Podemos decir que los Estados Unidos hacen un uso pragmático de la inteligencia artificial, aprovechando toda la información que está disponible.

Por lo tanto, la relación entre protección de datos personales e inteligencia artificial es un tema que puede enfocarse de diversas maneras. Corresponde a cada sociedad decidir los límites y fijar qué es ético. Ahora bien, esto no es posible si la sociedad desconoce los usos que se hacen y las implicaciones que tienen sobre las personas. El contenido de los siguientes capítulos busca dar respuesta a estas cuestiones.

2 Ada en acción: trabajo de investigación en Catalunya

Por Karma Peiró

Karma Peiró es periodista especializada en las Tecnologías de la Información y la Comunicación (TIC) desde 1995. Conferenciante en seminarios y debates relacionados con los datos, la transparencia de la información y la ética en la inteligencia artificial. Colaboradora en el Informe de AlgorithmWatch: Automating Society, una radiografía sobre la aplicación de los algoritmos de decisión automatizada en Europa.

Miembro del Consejo Asesor de Transparencia del Ayuntamiento de Barcelona; del Consejo Asesor de la Iniciativa Barcelona Open Data; y del Consejo para la Gobernanza de la Información y los Archivos. Más información a: http://karmapeiro.cat

2.1 Cómo la inteligencia artificial está cambiando el mundo

En 2011, el popular concurso de televisión estadounidense Jeopardy -basado en preguntas y respuestas de cultura general- acumulaba un bote de cinco millones de dólares. Al gran premio optaban dos de los mejores jugadores en décadas: Ken Jennings (74 veces ganador) y Brad Rutter (con más de tres millones de dólares ganados). Ambos aceptaron el reto de luchar, esta vez, con un oponente muy atípico que se encontraría en una habitación contigua al plató, debido al ruido del sistema de ventilación que lo hacía funcionar. Se trataba del sistema inteligente de IBM Watson[9], una máquina capaz de entender, responder a preguntas complejas en tiempo real y aprender de cada interacción. ¿Se imaginan quién ganó?

De manera muy simple, se podría decir que la inteligencia artificial (IA) es una parte de la informática que potencia que las máquinas funcionen y reaccionen como los humanos. Es decir, que puedan razonar, aprender y actuar de manera inteligente. Para conseguir este objetivo, la IA desarrolla algoritmos que predicen y toman decisiones de manera automatizada.

La idea no es nueva. El filósofo y escritor Ramon Llull[10] (1232-1316) dedicó su vida al *Ars Machina*, una máquina que servía para hacer pruebas lógicas y facilitar el razonamiento. Más recientemente, el matemático Alan Turing creó el *Test de Turing*[11] (1950) que planteaba la posibilidad de que un ordenador pudiera "pensar". Y se demostraba en el momento que una persona no podía distinguir si quien le hablaba era una máquina o un humano. A Turing también se le concede el mérito de haber salvado millones de vidas gracias a romper los códigos matemáticos de la máquina de encriptación de mensajes Enigma[12] de los nazis. Y así acortar la 2ª Guerra Mundial en dos años.

Hay muchos otros hitos históricos que nos acercan la inteligencia artificial en el presente, aunque hasta la fecha sólo eran casos excepcionales. Es ahora, que sin darnos cuenta la tenemos más cerca que nunca.

Algoritmos que nos acompañan

Los algoritmos son el conjunto de comandos a seguir para resolver un problema o una tarea. Fueron aplicados ya en las primeras civilizaciones de la humanidad. Y nosotros mismos, en nuestro cerebro, cuando reaccionamos a una situación imprevista empleamos algoritmos mentalmente para resolverla. Lo mismo hacemos cuando cocinamos un plato, siguiendo los minutos de cocción de cada ingrediente que marca la receta.

Los algoritmos que encontramos en los ordenadores nos hacen la vida más fácil y sencilla. Cuando buscamos la mejor ruta para llegar a un lugar; cuando pedimos un taxi desde una app móvil; cuando recibimos otras recomendaciones después de comprar un libro; cuando vemos nuestra serie preferida en una plataforma online; cuando

[9] Watson (https://ca.wikipedia.org/wiki/Watson_(intel%C2%B7lig%C3%A8ncia_artificial)

[10] Ramon Llull (https://www.cccb.org/es/exposiciones/ficha/la-maquina-de-pensar/223672)

[11] *Test de Turing (*https://ca.wikipedia.org/wiki/Test_de_Turing)

[12] Enigma (https://ca.wikipedia.org/wiki/M%C3%A0quina_Enigma)

buscamos vivienda en un barrio nuevo; cuando accedemos a una oferta de trabajo; cuando pedimos un crédito al banco; cuando entramos en una lista de espera para una operación quirúrgica; cuando encargamos una pizza a domicilio o cuando contratamos el seguro del hogar los algoritmos deciden y nos dan las mejores soluciones u opciones. ¿Qué haríamos sin ellos? ¡Sería una enorme pérdida de tiempo y de ineficiencia prescindir de ellos!

Los algoritmos más sofisticados utilizan aprendizaje automático (*machine learning*), una rama de la inteligencia artificial que consigue que las máquinas mejoren con la experiencia. Es bueno para establecer patrones y relaciones ingentes, así como para agilizar procesos. Dentro del *machine learning* está el aprendizaje profundo (deep learning), una técnica de procesamiento de datos basada en redes neuronales artificiales con muchas capas. La técnica está inspirada en el funcionamiento básico de las neuronas del cerebro. Existe hace más de 50 años pero ahora tenemos suficiente volumen de datos y capacidad de computación para aplicarla en una multitud de casos prácticos.

Fabricantes incansables de datos

Los 4 mil millones y medio de personas conectadas hoy en Internet (58% de la población mundial) generamos un número gigantesco de datos. Cada movimiento que hacemos es ya, prácticamente, un dato. Fabricamos datos las 24 horas al día, los 365 días del año. Incluso cuando dormimos estamos creando datos: los del descanso, la no actividad. Sin datos, el mundo hiperconectado se ralentiza tanto que parecería parado.

Nuestras acciones en los ordenadores y los móviles -que llevamos encajados permanentemente en las manos- producen datos que describen nuestros gustos, estado de ánimo, reacciones a estímulos, miedos y ambiciones. Los datos más valiosos para las empresas, partidos políticos o gobiernos son los que definen el comportamiento de grupos de gente. Después de analizarlos y clasificarlos, configuran perfiles o patrones de consumo para hacernos llegar ofertas de productos, servicios o mensajes tan personalizados (y tanto a nuestro gusto) que casi no podemos rechazarlos. La consultora Cambridge Analytica utilizó estas técnicas de perfilado para lanzar desinformación (fake news) a través de Facebook en campañas políticas como la del presidente Donald Trump o la del Brexit.

Otros datos, como las huellas digitales, la cara, el iris o el genoma nos identifican como seres únicos e irrepetibles. También hay datos acumulados, derivados de nuestras acciones pasadas que nos clasifican como buenos o mal pagadores, o que nos abren y cierran puertas a nuevos puestos de trabajo en función del lugar de residencia, género, edad o color de piel.

Los datos lo deciden todo sobre nuestras vidas en el presente, y en el futuro. Es como si de repente nos hubiéramos puesto unas gafas de aumento y hubiéramos descubierto un sinfín de matices de nuestro entorno, más apreciado antes. El editor tecnológico de la revista The Economist Kenneth Cukier[13],explica -en su libro *Big Data: A Revolution that will transform how we work, live and think*- que "el valor de la

[13] Kenneth Cukier (https://www.ted.com/speakers/kenneth_cukier)

información hoy reside en la manera de correlacionar los datos para descubrir patrones que ni siquiera se habían imaginado".

Confiaremos para evolucionar

Ya hemos normalizado que una máquina -o asistente virtual- nos desee buenos días cada mañana, tener un aparato que da vueltas por el suelo de casa limpiando o que un sistema automatizado nos sugiera la serie que nos apetece ver según el momento del día. Pronto aceptaremos que la nevera haga la compra de la semana, que la lavadora se ocupe de la colada de manera autónoma o que la calefacción de casa llame al técnico cuando falle la caldera. Todos estos objetos producen (o producirán) datos relacionados con el consumo, las interacciones y el uso. Se conoce como la Internet de las Cosas (IOT, en inglés), es decir, la interconexión digital de objetos para comunicarse e interactuar entre ellos-. Se calcula que, en 2025, habrá en el mundo 21.500 millones de objetos conectados[14].También tendremos robots domésticos ocupándose de los cuidados de mayores y pequeños en el hogar, los alumnos en clase estudiarán las materias más aptas para su personalidad y confiaremos en los coches autónomos como los vehículos más seguros para nuestros desplazamientos.

Confiaremos en la tecnología porque sabemos que, de lo contrario, no evolucionaremos como sociedad. Confiaremos porque la historia nos ha demostrado que la tecnología siempre nos ha hecho avanzar como seres humanos.

La complejidad se entiende con ejemplos

Pero para confiar debemos conocer los beneficios y riesgos que implica la tecnología que nos rodea y la que se nos promete. Sólo así podemos ser críticos y valorar en cada momento qué dosis de tecnología queremos en nuestras vidas.

El objetivo de esta investigación es sacudir la curiosidad individual, explicar -de una manera sencilla- para qué sirven los algoritmos de decisión automatizada (ADA) que se aplican ya en Cataluña. Para explicar la complejidad, hemos recopilado más de cincuenta ejemplos en el ámbito social, laboral, de la salud, de la educación, la banca, la comunicación o la ciberseguridad, entre otros.

Para elaborar este trabajo hemos contado con la colaboración desinteresada de una treintena de expertos de nuestro país -máxima referentes de la investigación académica en inteligencia artificial-, y de emprendedores y profesionales relacionados con la IA, que han aportado conocimiento y reflexión sobre los avances actuales y los riesgos a tener en cuenta. Seguro que nos hemos dejado de mencionar muchas investigaciones e iniciativas interesantísimas en los sectores mencionados o en otros no mencionados. Esta es sólo una muestra para entender y decidir.

Porque de eso se trata: de decidir a título individual qué hacer si un algoritmo de decisión automatizada nos afecta de manera negativa o discriminatoria. De saber qué organismos o legislación nos amparan. De conocer los procedimientos a seguir para

[14] Informe La Internet de les Coses a Catalunya. ACCIÓ. Generalitat de Catalunya https://www.accio.gencat.cat/web/.content/bancconeixement/documents/pindoles/iot-cat.pdf

que se revisen los criterios utilizado para el diseño del algoritmo, así como para detectar y mitigar los sesgos de los datos que han provocado una desviación desfavorable para una persona o un colectivo.

2.2 Riesgos de los ADA

Imagínese que busca trabajo en una empresa de gran prestigio, con condiciones de trabajo y sueldo más que aceptables. La compañía a la que aspira recibe muchas demandas, y no hace entrevistas personalizadas ni tests porque considera que las habilidades que destacan los currículums son a menudo exageradas. A cambio, pide la contraseña del correo electrónico del candidato para que un algoritmo rastree los mensajes personales y decida si es el candidato que busca la empresa. ¿Daría usted el consentimiento para que un sistema de inteligencia artificial (IA) revisara su buzón a cambio de optar al trabajo de su vida?

El ejemplo es real y se explica en el informe elaborado por la ONG *Algorithm Watch: Automating society: Taking stock of Automated Decision-Making in the E*U[15]. La agencia de recolocación de trabajo finlandesa Digital Minds tiene una veintena de grandes corporaciones como clientes. Al recibir miles de currículos cada mes, sólo puede hacer la selección con algoritmos de decisiones automatizadas (ADA). Digital Minds también los utiliza para escudriñar el Facebook y Twitter personal de los aspirantes. El sistema analiza la actividad del candidato y cómo reacciona. Con los resultados se puede saber si una persona es introvertida u otros aspectos de la personalidad. De nuevo la pregunta: ¿daría usted su consentimiento? Como la tecnología ya existe, no sería descartable que pronto se empiecen a aplicar métodos similares en Cataluña.

Toman decisiones por los humanos

Los algoritmos hoy ya son capaces de asumir algunas funciones de los humanos y esto los hace más imprescindibles cada día. Pueden hacer reconocimiento facial o de imágenes; interactuar (asistentes virtuales); decidir automáticamente (recomendadores de series, libros u otros productos); tener un impacto social (robots que se ocupan de enfermos en hospitales); y aprender de muchas situaciones en tiempo real (los coches autónomos que pronto tendremos).

"Hay tres tipos de algoritmos de decisión automatizada" -explica Jordi Vitrià, catedrático de Lenguajes y Sistemas Informáticos de la Universidad de Barcelona (UB) -. "Los que predicen un número (fabricaré 33 coches), los que predicen una clase (cierto/falso, si/no, enfermo/sano); y los recomendadores -que entre un conjunto muy grande de posibilidades- aseguran que comprarás un determinado producto. Estos tres tipos previamente hacen la predicción y, según el resultado, toman la decisión (automatizada o no)".

Elisabeth Golobardes, Doctora en Ingeniería Informática, explica lo siguiente: "Los algoritmos son de predicción y de decisión automatizada. Pueden predecir potenciales clientes en un ámbito comercial o un posible accidente en una central nuclear. Y también están los que toman decisiones como decidir evacuar a la población en riesgo 100km fuera del núcleo de la central nuclear ante una posible explosión. O decidir echar a 100 trabajadores de una empresa por posible quiebra. Son decisiones muy complicadas y confío que siempre haya un profesional detrás, que valore la situación, y

[15] Informe elaborado por AlgorithmWatch, *Automating society: Taking stock of Automated Decision-Making in the EU.* Karma Peiró es autora del capítulo 'Spain'. *(*https://algorithmwatch.org/en/automating-society/)

se haga responsable en caso de aplicarse la decisión de la máquina de manera automatizada".

Tradicionalmente, los bancos han empleado muchos algoritmos antes de conceder o no un crédito. La única diferencia respecto al momento actual es que antes era el director de la oficina bancaria quien lo decidía, en función de unas variables marcadas por la experiencia de años, y ahora lo hacen las máquinas de manera automatizada. Lo mismo ocurre con los seguros de coches o de vivienda: lo que antes hacía un profesional, el comercial de toda la vida, ahora lo hace una máquina. Todo es tener una gran base de datos con miles o millones de casos del pasado y entrenar el algoritmo. "Pero hay que tener cuidado, ¡por los sesgos!", Alerta Jordi Vitrià. "Si cojo una base de datos de un banco y durante 30 años ha habido una visión machista de la sociedad, por la que no concedían créditos a las mujeres, el algoritmo perpetuará este sesgo y hoy tampoco lo dará". Los algoritmos no son ni buenos ni malos, pero tampoco son neutros. Se alimentan de datos y éstos tienen sesgos. "Siempre han existido. Lo que hay que hacer es detectarlos y mitigarlos ", añade Vitrià.

¡Malditos sesgos!

En los años 80, el vicedecano de la Escuela de Medicina del Hospital St.George de Londres, Geoffrey Franglen, debía evaluar unas 2.500 solicitudes cada año. Para automatizar el proceso escribió un algoritmo que le ayudara a revisar, basándose en el comportamiento de evaluación de solicitudes anteriores. Aquel año, los candidatos se sometieron a una doble prueba antes de ser admitidos: la del algoritmo y la de los profesores. Franglen se dio cuenta de que las calificaciones coincidían en un 90-95%, lo que demostraba que la IA podía reemplazar a los humanos en esta fase tan tediosa. Pero cuatro años más tarde, la dirección del centro se dio cuenta de la poca diversidad de los candidatos. Y la Comisión para la Igualdad Racial del Reino Unido denunció a la Escuela por discriminación xenófoba y de género. Resultó que cada año, el algoritmo se había dejado fuera de la selección unas 60 personas: los discriminaba porque tenían apellidos no europeos, o eran mujeres. Se estaban perpetuando sesgos.

"Hay tres tipos de sesgos clásicos: el estadístico, el cultural y el cognitivo", explica Ricardo Baeza-Yates, Catedrático de Informática en la Universidad Pompeu Fabra y en la Northeastern University[16]. "El estadístico procede de cómo obtenemos los datos, de errores de medición o similares. Por ejemplo, si la policía está más presente en algunos barrios que en otros, no será raro que la tasa de criminalidad sea más alta donde tenga mayor presencia. El sesgo cultural es aquel que deriva de la sociedad, del lenguaje que hablamos o de todo lo que hemos aprendido a lo largo de la vida. Los estereotipos de las personas de un país son un ejemplo claro. Por último, el sesgo cognitivo es aquel que nos identifica y que depende de nuestra personalidad, los gustos y de los miedos. Si leemos una noticia que está alineada con lo que pensamos, nuestra tendencia será validarla, aunque sea falsa".

Esta última desviación se llama también 'sesgo de confirmación'. Buena parte de las noticias falsas (fake news) se alimentan de este razonamiento para difundirse más rápidamente. Por este motivo, si no nos cuestionamos lo que leemos o vemos,

[16]Ricardo Baeza-Yates i Karma Peiró (2019). *És possible acabar amb els biaixos dels algorismes (1a i 2a part)*, (https://www.karmapeiro.com/2019/06/17/es-possible-acabar-amb-els-biaixos-dels-algoritmes-1a-part/)

corremos el riesgo de avanzar hacia una involución humana. El historiador Yuval Noah Harari[17]alerta en su último libro *21 Lecciones para el siglo XXI de* que "con la tecnología actual, es muy fácil manipular a las masas". Y si seguimos lo que piensa la mayoría de la gente, ¿qué pasa cuando moralmente la masa está equivocada?

Y aún más sesgos...

Ricardo Baeza-Yates[18] explica que existe el sesgo de orden (ranking) cuando hacemos una búsqueda en la Web, ya que las personas tienden a hacer clic en las primeras posiciones y el buscador podría interpretar que estas respuestas son mejores que las siguientes.

Los sesgos de presentación son los que encontramos en las recomendaciones en el ámbito del comercio electrónico. Sólo lo que el buscador muestra al usuario podrá tener clics. Todo lo que no salga en la página de resultados queda fuera de consulta. Es un círculo vicioso, como el del huevo y la gallina. Y la única manera de romperlo es mostrar el universo total de resultados. "Esto se conoce como *filtro burbuja*: el sistema muestra únicamente lo que te gusta. Como se basa en las acciones del pasado, no es posible ver lo desconocido", añade Ricardo Baeza-Yates.

En caso de que el mundo siga funcionando de esta manera, llegará un momento en que nos sentiremos como el personaje principal de la película El show de Truman[19], que un buen día se dio cuenta de que todo su mundo era un engaño. Y que se había perdido lo que hay más allá del horizonte. Las redes sociales funcionan así y a las multinacionales ya les va bien. Potencian lo que se conoce con el nombre de Economía de la Dopamina. La dopamina es un neurotransmisor que tenemos en el cerebro y que proporciona sensación de placer, sentimientos de gozo o refuerzo para motivar a las personas a realizar acciones. Estar en el filtro burbuja de una red social puede llegar a crear una adición, que nos incita a pasar muchas horas interactuando.

Los que discriminan a las mujeres

En 2014, Amazon estrenó un algoritmo para reclutar nuevos trabajadores en sus almacenes[20]. La herramienta puntuaba de una a cinco estrellas a los mejores candidatos. Todo parecía ideal: la IA ahorraría horas en el departamento de recursos humanos. Pero un año más tarde, la multinacional se dio cuenta de que, en los puestos técnicos, como el de desarrollador de software, no se había contratado a ninguna mujer. ¿Podría ser que no hubiera ninguna candidata con aptitudes para ese trabajo?

El ejemplo es un clásico ya de los errores que ha provocado recientemente la IA. Los datos masivos que sirvieron para alimentar el algoritmo de selección de personal se

[17]Yuval Noah Harari (https://www.ynharari.com/)

[18] Ricardo Baeza-Yates i Karma Peiró (2019). *És possible acabar amb els biaixos dels algorismes (1a i 2a part)*, (https://www.karmapeiro.com/2019/06/17/es-possible-acabar-amb-els-biaixos-dels-algoritmes-1a-part/)

[19] Film: *El show de Truman* (https://ca.wikipedia.org/wiki/The_Truman_Show)

[20] Ricardo Baeza-Yates i Karma Peiró (2019). ¿Por qué la inteligencia artificial discrimina a las mujeres? https://medium.com/think-by-shifta/por-qu%C3%A9-la-inteligencia-artificial-discrimina-a-las-mujeres-18b123ecca4c

basaron en currículos recibidos en la década anterior, donde la mayoría de los programadores eran hombres. Cuando el sistema automático detectaba la palabra 'mujer' -o un sinónimo-, directamente el penalizaba poniéndole menos puntuación.

Otro ejemplo de discriminación de género pasó en el 2016. Investigadores de Microsoft Research y de la Universidad de Boston usaron una colección masiva de noticias de Google para entrenar algoritmos sobre los estereotipos femenino/masculino que salían en la prensa. Los resultados daban que los hombres eran programadores informáticos y las mujeres amas de casa. Ellos eran médicos y ellas enfermeras. Y tiene su lógica, porque en Estados Unidos la mayoría de los periodistas que redactaban las noticias, sobre las que se basó el entrenamiento del algoritmo, eran hombres. Así que Google sólo reflejaba el sesgo de género existente en la realidad.

Sesgos hay para dar y regalar. Hay decenas de culturales y aún más de cognitivos. Se han llegado a clasificar hasta un centenar, pero serían unos 25 los más importantes. En el artículo *The Ultimate List of Cognitive Biases: Why Humans Make Irrational Decisions*[21]se enumeran unos 49 sesgos cognitivos. Estos son los más peligrosos porque están arraigados a cada persona. La única manera de resolverlos es cambiar a cada persona, lo que de entrada ya parece una proeza imposible. Hay que dar la razón al historiador Harari cuando dice que "es muy fácil manipular las personas y muy complicado eliminar los sesgos".

Pero los sistemas inteligentes no son autónomos, ni actúan por sí mismos. Y lo explica muy bien la investigadora Elisabeth Golobardes: "El algoritmo como tal no tiene ningún sesgo. Son los datos que se introducen y el objetivo por el que ha sido diseñado el que discrimina".

¿Se puede ser justo con todos?

Tampoco todos los sesgos son perjudiciales. "Por ejemplo, que haya más enfermeras que enfermeros puede ser positivo por sus cualidades empáticas en el trato de los pacientes. Pero que los políticos sean mayoría no lo es porque un punto de vista de la población (el femenino) no está igualmente representado", dice Baeza-Yates. Los resultados de los algoritmos pueden discriminar por razones de género, raza, edad o clase social, por mencionar los más importantes.

Sabiendo que los algoritmos tienen sesgos y que estos pueden discriminar... ¿por qué se aplican? "Una respuesta podría ser que el beneficio o acierto de los resultados es considerablemente superior (más de un 90% en la mayoría de los casos) que el perjuicio o error. ¿Es esto justo por los que salen perjudicados? En este punto se podría iniciar una larga discusión sobre qué es o no es justo en la vida ", continúa Baeza-Yates.

Resulta muy difícil ser justo con todos. Un algoritmo puede serlo con un colectivo de mujeres, pero discriminar a un hombre. En este sentido, el investigador Andrew

[21]Tomer Hochma. *The Ultimate List of Cognitive Biases: Why Humans Make Irrational Decisions (*https://humanhow.com/en/list-of-cognitive-biases-with-examples/)

Selbst [22]—del Data & Society Research Institute[23]— explica que decidir la discriminación en inteligencia artificial es bien complicado. "Es un proceso en constante evolución, al igual que cualquier aspecto de la sociedad".

Para ser rigurosos también hay que decir que los algoritmos bien diseñados -aunque tengan sesgos- son justos de acuerdo con los parámetros que se les han dado. A diferencia de los humanos -que pueden variar su decisión en función de un estado de ánimo o de cansancio físico y mental- los algoritmos siempre funcionan igual.

Los prejuicios

Los sesgos son similares a los prejuicios: todos tenemos, en mayor o menor grado. Muchos los heredamos de nuestro entorno social o familiar sin darnos cuenta. El sesgo mayor es creer que no tenemos ningún prejuicio. Pero... ¡atención! Si los sesgos no se corrigen, existe el riesgo de habitar un futuro donde cada vez sea más difícil el progreso social porque se perpetúen.

¿Cómo estar seguros, pues, que todos los datos que introducimos en el algoritmo representan el universo que queremos predecir y que no discriminarán a nadie? No podemos estar seguros. Y es aquí donde entran los aspectos éticos de la inteligencia artificial que se deben tener muy presentes (véase capítulo 6. Aspectos éticos de la IA con las opiniones de investigadores).

[22] Andrew Selbst (https://andrewselbst.com/)

[23]Data & Society Institute (https://datasociety.net/)

2.3 ¿De decisión automatizada o de soporte a la decisión?

La matemática Cathy O'Neal[24] –autora del libro Armas de Destrucción Matemática, dedicado a poner en evidencia a los algoritmos de decisión automatizada (ADA) en Estados Unidos- en una de sus visitas a Barcelona, me planteó el siguiente dilema: "¿Serías capaz de decidir si un menor corre peligro en casa porque está a punto de ser maltratado o abusado sexualmente por alguno de sus progenitores?". Ante mi cara de preocupación, ella añadió:" en EEUU lo están decidiendo las máquinas. El algoritmo AURA[25] fue creado para detectar los menores que podían ser víctimas potenciales de abusos antes de producirse los hechos, con la buena intención de evitar un trauma al niño ". Pero... ¿cómo puede saber a ciencia cierta una máquina que se ha de alejar al niño de sus padres antes de que pase nada?, repregunté perpleja y bastante preocupada. Y la respuesta de O'Neal fue: "No lo puede saber". El algoritmo Aura finalmente fue una prueba piloto que no se llegó a aplicar. Pero otros sistemas automatizados, como el COMPASS[26] –que predicen la criminalidad en las cárceles– sí han sido muy polémico. El medio independiente Pro Publica realizó una investigación periodística[27]donde demostraba los sesgos que tenía sobre las personas negras.

En Cataluña no hemos llegado a estos extremos. Estamos en una fase inicial de la inteligencia artificial y muchos de los ejemplos que contamos todavía forman parte de una investigación o prueba piloto. Por otra parte, todos los expertos consultados por este informe dejan claro que las decisiones automatizadas se aplican únicamente en aquellos sectores que no conllevan un riesgo para la vida de una persona. Por tanto, en el ámbito de la salud, en el judicial (y en ciertos casos de enseñanza), la decisión del algoritmo es un apoyo al profesional y no se hace de manera automatizada. La última palabra siempre la tendrá el médico, el juez o el maestro responsable.

Cuestión de escala

En sectores como el de reclutamiento de puestos de trabajo, donde se ha de atender una demanda de cientos o miles de personas, el algoritmo hará una primera selección. Para dar una ayuda social, con muchos solicitantes en las mismas condiciones, el sistema automatizado decidirá entre miles. Y de alguna manera, también puede conllevar cambios importantes y repercusiones desfavorables para la vida de las personas.

"Automatizar la selección de personal en departamentos de recursos humanos es un ejemplo muy común. Si recibes mil currículums cada mes, o cada semana, no te los puedes mirar todos y decidir. Los algoritmos hacen una preselección. Y los

[24]Cathy O'Neal (https://mathbabe.org/about/)

[25]Christine Zhang (2016) "Q&A: Cathy O'Neil, author of 'Weapons of Math Destruction,' on the dark side of big data". Los Angeles Times. https://www.latimes.com/books/jacketcopy/la-ca-jc-cathy-oneil-20161229-story.html

[26]Correctional Offender Management Profiling for Alternative Sanctions (COMPASS) https://en.wikipedia.org/wiki/COMPAS_(software)

[27] «Machine Bias», Pro Publica https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

responsables se miran sólo aquellos que han pasado el filtro de los ADA. Pero las máquinas no ven cosas que los humanos sí. Y entonces es cuando vienen las discriminaciones", explica Ramón López de Mántaras, Profesor Investigador del Consejo Superior de Investigaciones Científicas (CSIC).

"En medicina es diferente. Los médicos utilizan los ADA sólo como apoyo a la decisión. La Unidad de Cuidados Intensivos está llena de algoritmos, y claro que alguno puede disparar una alarma, pero un médico (o personal sanitario) mira qué pasa y toma la decisión final. No hay ningún sistema automatizado que te administre un medicamento sin la supervisión de un profesional y espero que siempre sea así", añade. "Porque el médico ve cosas que el algoritmo no ve. Y el algoritmo encuentra patrones que el ojo humano no ve. El algoritmo mira los árboles, y el médico el bosque. Por eso, cuando trabajan juntos disminuye mucho el error. En la detección del cáncer de mama, hay estudios que demuestran que el mejor médico tiene un error del 5 o 6% con las mamografías, y la ADA un 6-7%. Pero que trabajando juntos el error es sólo del 0,5%".

Las 'cajas negras'

Recapitulemos: la inteligencia artificial utiliza algoritmos que aprenden de datos masivos. Estos datos tienen sesgos que pueden discriminar. Se deben detectar y mitigar. Los algoritmos predicen y luego, algunos, toman decisiones automatizadas. Buena parte de estos sistemas son de aprendizaje automático, es decir, que aprenden antes de usarlos. Y algunos de ellos de aprendizaje profundo (deep learning).

Cualquier sistema que se considere inteligente debe tener la habilidad de aprender, es decir, de mejorar con la experiencia. Lo que no se ha conseguido aún es explicar cómo el algoritmo ha llegado a la predicción o decisión. Son lo que se conoce como 'cajas negras'. Las aplicaciones dedicadas al procesamiento del lenguaje natural como los traductores, el diagnóstico médico, la bioinformática o la detección del fraude financiero son cajas negras.

Por lo tanto, si un sistema automatizado predice alguna acción que me discrimina, ¿puedo preguntar cómo ha llegado a esa decisión? Puedes preguntar, pero no obtendrás, por ahora, una respuesta. Y este es, actualmente, uno de los grandes dilemas de la IA.

La explicabilidad o transparencia algorítmica

Para entender cómo hizo la predicción o tomó la decisión el algoritmo se habla de explicabilidad (Explainable Artificial Intelligent)[28]. "En los últimos dos años se ha trabajado mucho en este problema de los algoritmos de aprendizaje automático. En particular, en los más opacos que son los de aprendizaje profundo", explica Ricardo Baeza-Yates. "Pero aún faltan respuestas".

La investigadora del Oxford Institute Sandra Wachter[29]considera que deberíamos tener el derecho legal de saber por qué los algoritmos toman las decisiones que nos afectan. Explica que los propietarios de los algoritmos -multinacionales, empresas de todos los

[28]*Explainable Artificial Intelligent* (https://en.wikipedia.org/wiki/Explainable_artificial_intelligence

[29]Amit Katwala (2018) *"How to make algorithms fair when you don't know what they're doing"*, Wired. https://www.wired.co.uk/article/ai-bias-black-box-sandra-wachter

tamaños, bancos, pero también gobiernos, administraciones públicas o la policía- pondrán muchos inconvenientes en hacer transparentes las fórmulas de sus sistemas para defender la propiedad intelectual o la seguridad ciudadana. Por ello propone que se den "explicaciones contra-fácticas". Es decir, si te han denegado una hipoteca, poder preguntar al banco: "Si ganara 10 mil euros más al año, ¿me la hubieran concedido?". Si no has pasado la selección de un puesto de trabajo, poder preguntar: "Si hubiera tenido un máster, ¿me lo habrían dado?".

La comunidad científica trabaja para que los algoritmos den autoexplicaciones de sus decisiones. Y podemos pensar que muy pronto se encontrará la manera. Estamos en los inicios de un cambio social que debe magnificarse en los próximos años. Y la ciudadanía tiene un rol muy importante. La tecnología tiene sus partes brillantes y oscuras. Por partes oscuras consideramos errores involuntarios, pero también el mal uso intencionado que se puede hacer a una persona o a un colectivo en función de los sesgos de los datos, o de los criterios con los que se haya creado el sistema inteligente. Hay que estar alertas a las anomalías y discriminaciones para conseguir -con la legislación vigente y los organismos públicos responsables- que se siga garantizando la privacidad, la confidencialidad y la libertad individual.

2.4 ¿Dónde se aplican los ADA en Cataluña? Más de 50 ejemplos para entender…

La mejor manera de entender conceptos tecnológicos es con ejemplos. Cuando nos ponemos en una situación determinada, podemos imaginar con más claridad qué haría la máquina en cuestión o cómo actuaría el sistema inteligente. Sólo así podemos valorar la importancia, los riesgos y las consecuencias de su aplicación.

Por este motivo se ha considerado importante incluir en este informe un capítulo dedicado a explicar donde se encuentran los algoritmos de decisión automatizada (ADA) en diferentes sectores: salud, sistema judicial, educación, ámbito social, comercio, ciberseguridad, banca, ámbito laboral, comunicación, etc. No es, ni mucho menos, el recopilatorio más exhaustivo. Somos conscientes de que nos dejamos infinidad de ejemplos. Pero sí consideramos que es una muestra para hacerse una idea como de invisibles son ya los ADA, y para entender tanto el valor de su aplicación como los riesgos.

De manera generalizada, se puede decir que mientras las empresas y sector privado está muy avanzado en su implementación, a la administración pública todavía le queda un largo camino por recorrer. Es destacable también el sector salud y judicial que, a partir de colaboraciones entre hospitales locales e internacionales, universidades, entes privados, start-ups o con desarrollos propios, hace años que exploran las posibilidades que ofrece la inteligencia artificial para avanzar en un mundo más eficiente y justo.

2.4.1 Salud

La inteligencia artificial es un gran apoyo en el ámbito de la salud. Los algoritmos ayudan a entender o extraer conclusiones sobre enfermedades en mucho menos tiempo, a sugerir un diagnóstico o medicación pertinente, a hacer una mejor gestión de los centros hospitalarios o a leer historias clínicas a gran escala.

En este caso, las decisiones automatizadas a diferencia de otros sectores -no se aplican sin la supervisión de un doctor o especialista en la materia. Y este es un hecho importante. ¿Qué pasaría si el algoritmo no detectara un cáncer en un paciente y el doctor no diera un tratamiento? En Cataluña -al menos cuando se redacta este informe- la IA siempre es un apoyo a la decisión del médico especialista.

La inteligencia artificial tiene un potencial enorme pero siempre en colaboración, y como complemento del potencial y conocimiento que tienen los profesionales de la salud. Como se ha dicho en el apartado anterior, cuando la decisión automatizada implica un riesgo para la vida de las personas, la última palabra la tiene el profesional especialista.

"El cerebro humano es muy lento, comparado con lo que hoy pueden hacer las máquinas", apunta la investigadora Itziar de Lecuona, Profesora del Departamento de Medicina y Subdirectora del Observatorio de Bioética y Derecho de la Universidad de Barcelona (UB). "Pero no por ello debemos confiar ciegamente en la tecnología. Todos los modelos se deben validar, se deben hacer pruebas de concepto y preguntarse si el algoritmo funciona para el objetivo inicial", añade.

"Los algoritmos siempre han existido en medicina, pero ahora hemos dado un paso más: vamos hacia la predicción, hacia la experiencia más personalizada. Con la tecnología actual (y sus capacidades) sería muy poco eficiente no aplicarla ". Lecuona insiste que en el ámbito de la salud todo el mundo está pensando en hacer el bien. "Pero también se debe tener en cuenta la ética. Porque los algoritmos se nutren de datos que forman parte de la intimidad de las personas".

Por otra parte, la experta en seguridad y resiliencia de la Agencia Europea de Seguridad de Redes e Información (ENISA), Dimitra Liveri, habló en el Barcelona Cybersecurity Congress[30] de los hospitales inteligentes, que muy pronto dejarán de ser excepcionales. "Un hospital inteligente es un ecosistema de máquinas interconectadas que toman decisiones automatizadas", explicó. ¿Qué pasaría si una bomba de insulina inteligente comete un error? El paciente podría morir. La experta de ENISA recordó en su presentación que aún quedan retos por resolver. "1. La disponibilidad del aparato: si estoy enfermo quiero que mi dispositivo funcione todo el tiempo, sin detenerse. 2. La integridad: quiero que me inyecte las dosis correctas de glucosa y que nunca se equivoque".

- Una píldora que graba el intestino

[30]Barcelona Cybersecurity Congresshttps://www.barcelonacybersecuritycongress.com/es/front-page/

El cáncer de colon, en Cataluña, es una enfermedad que afecta mayoritariamente a población de más de 50 años. Se calcula que se diagnostica cada año a unas cinco mil personas y está ligado al estilo de vida y la alimentación de cada uno. Desde hace años, la Generalidad de Cataluña se ocupa de hacer llegar a domicilio unos tests para detectarlo. A partir de una pequeña prueba, se puede hallar a tiempo en caso de anomalía. Cuando el test da positivo se realiza una prueba endoscópica o colonoscopia, que requiere - en algunos casos - el ingreso hospitalario, sedación y resulta costosa para la sanidad pública.

Desde hace unos años, en EEUU y en algunos países de Europa se está practicando una nueva técnica que se podría implementar en Cataluña pronto. Investigadores del Hospital Vall d'Hebrón junto con el Departamento de Matemáticas e Informática de la Universidad de Barcelona (UB) experimentan con una cápsula endoscópica[31]para estudiar el estado de los intestinos. Los pacientes deben tragarse una pequeña píldora que va dotada de una cámara, cuatro LEDs[32]y una batería. La píldora viaja por el cuerpo del paciente y graba imágenes que pueden ser de gran ayuda para el doctor. Las imágenes se envían por wifi en conexión con un dispositivo que lleva la persona, sujeto a un cinturón. La grabación dura entre 8 y 12 horas.

La píldora no tiene ninguna contraindicación y si da alguna alerta se procede a realizar la colonoscopia. ¿El invento parece perfecto, pero entonces surge la pregunta de si los médicos tienen tiempo de visionarse tantas horas de grabación de cada paciente? "Imposible", responde Jordi Vitrià, director del departamento de Matemáticas e Informática de la UB. "O haces un sistema automatizado, que analice y detecte si ve alguna anomalía o no tiene sentido. Con lo que el algoritmo detecte, el doctor tomará la decisión de si se hace la endoscopia u otro tipo de pruebas ", añade Vitrià.

- Detector del dolor exagerado

Imaginemos que un trabajador ha sufrido un accidente o lesión en un brazo, en una pierna o en la espalda. Y pide la baja de media/larga duración para detener su actividad profesional. La duración y cuantía de la prestación va en función del tiempo que tenga que estar de baja. Por norma, estos casos los gestionan las aseguradoras laborales con un reconocimiento físico, donde hacen pasar al trabajador lesionado por diferentes máquinas que ponen a prueba las articulaciones. Pero siempre hay un porcentaje de casos en que las aseguradoras no pueden detectar realmente la magnitud del dolor.

"Se ha constatado que aproximadamente un 50% de las personas son magnificadores del dolor", explica Jordi Vitrià. Es decir, que esa parte del cuerpo no les duele tanto como dicen. Cuando esto se detecta, la Seguridad Social deniega la baja solicitada. "Hay mucho dinero público en juego. Por ello, pidieron a la UB crear un modelo algorítmico, de acuerdo con todo el historial de pruebas y diagnósticos acumulados durante años. Ahora vemos que, en muchos casos, los pacientes tienen razón, pero en otros no. Se puede detectar la exageración del dolor de manera automatizada", puntualiza el investigador de la UB. Tras implementarse en Barcelona, se ha

[31] Cápsula endoscópica (https://hospital.vallhebron.com/en/diagnostic-tests/endoscopy-endoscopic-capsule)
[32]LEDs, dispositivos miniaturizados que emiten luz (https://ca.wikipedia.org/wiki/D%C3%ADode_emissor_de_llum)

implementado en Sevilla, con los mismos criterios, por lo que los niveles de permisividad del dolor y decisiones sean similares en ambas ciudades.

- Teléfonos inteligentes para los ancianos

La población de más de 80 años ha aumentado en Europa en las últimas décadas. Una mayor esperanza de vida y el descenso de la natalidad debería servir para plantearse cómo se invierten los recursos de manera más eficiente. Una vida más larga no necesariamente significa mejor salud. ¿Cómo hacer para mantener el bienestar de los ancianos? ¿Qué ajustes se necesitan a las políticas sociales, económicas y de salud pública?

El proyecto europeo NESTORE[33] se presenta como un entrenador virtual que acompaña a las personas mayores, con buena salud, y les da consejos personalizados sobre los hábitos alimentarios o los ejercicios diarios que hacer. El seguimiento es a través de objetos conectados y sensores repartidos por la casa, así como una aplicación de móvil que monitoriza todos sus movimientos.

"Se les entrega un teléfono inteligente y deben fotografiar lo que tienen en la nevera, que comen, etc. Con algoritmos de visión por computación se hace la predicción de lo que deben ingerir de acuerdo con su estado de salud", explica Itziar de Lecuona, una de las investigadoras del NESTORE. "En la vivienda también tienen un aparato (asistente virtual) que les recuerda que deben salir a caminar, los pasos mínimos u otro tipo de ejercicio físico. Se acaba convirtiendo en un acompañante muy íntimo porque lo sabe todo de ti: las enfermedades y la medicación, la comida, las horas de descanso, si te has movido o no, si has estado en contacto con la familia, si ha recibido visitas, etc.", añade.

- Acertar la medicación después de un trasplante

El Hospital del Mar de Barcelona tiene una de las unidades de trasplante de riñón más grandes de toda España. Comenzó en 1973, y hoy en día ya ha realizado más de 1.400[34].En Cataluña, casi la mitad de los nuevos pacientes sometidos a terapia renal sustitutiva tienen más de 70 años, y buena parte de los donantes renales superan los 60.

Uno de los problemas principales después de un trasplante es acertar la dosis de medicación que se da de por vida al receptor del riñón. Por eso este paso es muy sensible. "Si te pasas, puede tener efectos secundarios y si la medicación es poca, el órgano puede ser rechazado por el cuerpo. Es muy complicado", explica Fernando Cucchietti, Director de Analítica de Datos y Visualización en el Barcelona Supercomputing Center. "Propusimos al Hospital del Mar utilizar algoritmos para decidir si una persona recibe o no un riñón con todos los condicionantes que actualmente se tienen en cuenta entre el donante y receptor, pero también para

[33] Proyecto europeo NESTORE (https://nestore-coach.eu/home)

[34] 40 años de trasplantes renales al Hospital del Mar https://www.parcdesalutmar.cat/es/noticies/view.php?ID=1003

ayudarles a precisar el protocolo de medicación exacto", añade. Por ahora, el proyecto aún no se ha puesto en marcha.

- Detectar a tiempo la cirrosis

Cada vez que el hígado sufre una lesión, ya sea por enfermedad, consumo excesivo de alcohol u otra causa, el órgano intenta repararse a sí mismo. En el proceso, se forma un tejido de cicatrización. A medida que la cirrosis avanza, se forman más tejidos de cicatrización y hacen que el hígado funcione con dificultad (cirrosis descompensada). La cirrosis avanzada es potencialmente mortal. Pero si se diagnostica con tiempo y se trata la causa, se puede limitar el daño y -a veces- revertirse.

El Hospital Clínico de Barcelona tiene una de las mejores unidades de tratamiento del hígado, pero cuando los enfermos acuden ya es demasiado tarde. Por eso es tan importante desarrollar tecnologías y sistemas de detección temprana, explica el Director de Analítica de Datos y Visualización en el Barcelona Supercomputing Center. "Ahora formamos parte de un proyecto europeo con el Clínico, que nos permitirá tener los síntomas preliminares a una cirrosis de hígado. La prueba se hará con 40 mil personas de toda Europa, durante cinco años. Y se coordina desde Barcelona ", añade Fernando Cucchietti. Según explica, la idea es instalar un aparato en centros de atención primaria y hospitales de manera que los algoritmos hagan el diagnóstico y los médicos decidan si el paciente debe hacer una biopsia o trasplante. "Se quieren detectar los síntomas 5 o 10 años antes de que sea demasiado tarde, porque en esta fase temprana hace falta un cambio de estilo de vida y muy poca medicación", concluye.

- Diagnóstico con electrocardiogramas digitales

Una empresa del sector de la salud vascular fabrica aparatos que hacen electrocardiogramas en formato digital. Esto permite procesarlos y asociarlos a patologías. "Tener todo este conocimiento encapsulado es como tener un comité de médicos que, a lo largo del tiempo ha diagnosticado una serie de enfermedades, con muchos cientos de miles de electros", explica Jordi Navarro, experto en datos y CEO de una empresa catalana dedicada a la IA.

"El algoritmo hace un diagnóstico similar al del médico, a partir de encontrar patrones. Con la diferencia de que si había veinte cardiólogos y uno se hubiera equivocado en el diagnóstico, pero 19 hubieran acertado, el algoritmo siempre coge el resultado de la mayoría. En este caso, el bueno. En este sentido podemos decir que los algoritmos son democráticos".

- Reconocimiento facial para detectar el TDAH

El Trastorno por Déficit de Atención con Hiperactividad (TDAH) es de origen neurobiológico y se caracteriza por la imposibilidad de mantener la atención, hiperactividad e impulsividad que influye considerablemente en la vida de un menor. Los síntomas del TDAH se manifiestan en el colegio o en cualquier otro ambiente social. Este trastorno se inicia en la infancia, pero puede continuar en la adolescencia y

la edad adulta. A nivel mundial, se estima que entre el 3 y el 7% de los niños pueden estar afectados. En Cataluña, el 1,3% de las niñas y el 3,7% de los niños de 6 a 17 años residentes en Cataluña consumieron algún fármaco para el TDAH durante el 2015. Esta proporción es de 1,9% y de 5,5% en las niñas y niños de 12 a 15 años, respectivamente, que es el grupo con más consumo[35].

Ahora con el reconocimiento facial se puede diagnosticar más acertadamente la enfermedad. El Centro de Visión por Computación de la Universidad Autónoma de Barcelona (UAB) ha puesto en marcha un sistema inteligente que analiza los gestos y las expresiones faciales del menor. Con los resultados se puede dar un grado de certeza sobre los síntomas. Este diagnóstico servirá al médico/psicólogo que trata al menor a hacer el protocolo de actuación. Con el mismo algoritmo intenta ahora diagnosticar otras enfermedades de tipo depresivo entre los jóvenes. "Apenas está empezando", explica Meritxell Bassolas, Directora de Conocimiento y Transferencia de Tecnología del CVC. "Además de los resultados del algoritmo, se analiza el comportamiento del menor en las redes sociales. Con todo, el profesional tiene más información para aplicar un tratamiento", añade. (Ver otros ejemplos con visión por computación en el apartado 2.4.10)

- El enfermo no va al hospital que le toca

El CatSalut -organismo que gestiona la sanidad pública de Cataluña- distribuye los recursos sanitarios entre las diferentes comarcas catalanas. Estos se reparten en 9 regiones sanitarias, y en subregiones de diversa categoría. Todos tienen asignado un centro de atención primaria, además de un hospital para las urgencias o las estancias más largas. Pero no siempre se acude a donde se está asignado, por lo que los recursos -previstos inicialmente para un número determinado de personas- pueden ser escasos o sobrantes. ¿Cuáles son los motivos que hacen que los enfermos no acudan allí donde les corresponde por zona de residencia? Las razones pueden ser múltiples: problemas logísticos, o porque faltan profesionales en el centro, o porque las esperas son largas o porque los horarios del transporte público para llegar son insuficientes.

Hasta la fecha esta pregunta se respondía con la minuciosa tarea de los técnicos que buscaban las anomalías después de mucha observación de funcionamiento. Ahora lo puede hacer un programa automáticamente[36]: el algoritmo coge las variables de comportamiento de los pacientes y las muestra en verde o rojo en función si están yendo al lugar que les corresponde o no. Con esta información extrae patrones anómalos de cada centro de salud. De modo que son más rápidos en organizar el territorio y minimizar un problema de recursos públicos.

- Predecir si un enfermo volverá al hospital

[35] Observatorio de la Salud de la Generalitat de Cataluña http://observatorisalut.gencat.cat/web/.content/minisite/observatorisalut/ossc_central_resultats/informes/fitxers_estatics/MONOGRAFIC_26_TDAH_CdR.pdf

[36] Amalfi Analytics (https://www.amalfianalytics.com/)

Las personas que padecen enfermedades crónicas del corazón o el riñón vuelven, cada cierto tiempo, al hospital. A veces por pequeños imprevistos que los desestabilizan: como un día de calor intenso o correr para coger el bus. El indicador de readmisión hospitalaria - reconocido internacionalmente como prueba de calidad[37]– marca en 30 días la tasa óptima para que un paciente regrese al hospital. "Si el paciente vuelve antes de un mes, algo falla", explica Ricard Gavaldà, profesor y coordinador del laboratorio de investigación en la Universidad Politécnica de Cataluña (UPC). "Quizás el hospital necesitaba camas y ha enviado al paciente a casa antes de lo previsto, a lo mejor no ha habido comunicación con el médico de cabecera en el seguimiento de la medicación administrada, o simplemente porque la persona necesitaba más atención", añade.

"En Cataluña, el CatSalut paga el 100% del gasto hospitalario. Pero si el paciente vuelve a los 29 días de haberse ido, el hospital sólo recibe el 40%. De este modo, se persigue que haya una mejor atención", explica Gavaldà. Con la información básica de miles de casos de un hospital, se ha hecho un programa que predice cuál es el riesgo de que el enfermo vuelva antes de 30 días. "Con un riesgo elevado, el hospital lo puede tener un par de días más en la cama. Con un riesgo bajo, se puede hacer un seguimiento diario en casa para saber si todo va bien, pero queda libre una plaza para otra urgencia", continúa Gavaldà. "Así se ahorran recursos públicos y se da bienestar al paciente si no es necesario que esté hospitalizado".

Sobre las decisiones automatizadas que los algoritmos ya pueden tomar en el ámbito de la salud, para cualquier diagnóstico de enfermedades, el profesor y coordinador del laboratorio de investigación en la UPC lo tiene claro: "No se trata de si la máquina lo hace mejor o no que el médico. Sino que el médico con la máquina, lo haga mejor que sin ella", sentencia Gavaldà.

- Agilizar el cribaje en urgencias[38]

Cuando una persona acude de urgencias a un hospital -sobre todo en las grandes ciudades- soporta horas de espera, sea en la sala de entrada, un pasillo o en el box (reservado). Por norma, después de un largo rato es visitada por el médico, que envía hacer pruebas. Después de horas, a veces todo un día y una noche, el médico decide si le asigna una cama.

¿Qué pasaría si desde el primer momento, en el cribaje de entrada, se solicitara una cama para los casos más recurrentes? Con los miles de entradas acumuladas en urgencias y su resolución, un algoritmo puede decidir de manera automatizada qué personas lo necesitarán. Entrarían en lista de espera de cama desde el momento de la selección, y no después de horas de espera acumuladas.

Si la máquina acierta, el enfermo sólo espera en urgencias el tiempo que tardan las pruebas encargadas para el diagnóstico. Y el hospital gana en eficiencia. Si el algoritmo se equivoca y el paciente no necesita cama, rápidamente se dará a otro enfermo.

---

[37]Re admission rate as an indicator of hospital performance. The case of Spain https://www.researchgate.net/publication/8267981_Readmission_rate_as_an_indicator_of_hospital_performance_The_case_of_Spain

[38]Amalfi Analytics (https://www.amalfianalytics.com/)

- Traducir las historias clínicas

Cuando un investigador quiere hacer una búsqueda se encuentra con el gran problema de que el 80-85% de los historiales están en lo que se llama 'formato libre'. Es decir, el paciente explica donde tiene la dolencia y el médico lo añade a su historial. El diagnóstico de la enfermedad, tratamiento, medicamentos suministrados, etc. todo queda en un formato incomprensible para las máquinas. Y con la información así no se puede hacer ninguna investigación a gran escala.

El Hospital de la Vall d'Hebron utiliza un programa[39]que automatiza y convierte en datos estructurados las historias clínicas para un análisis posterior. Esta tarea a mano es muy lenta, y siempre lleva años de retraso. "Utilizamos procesamiento de lenguaje natural. Es decir, hemos enseñado al ordenador a leer historias clínicas, a partir de las variables solicitadas por el investigador. El algoritmo hace una lectura automatizada y vuelve los datos estructurados en una hoja de cálculo", explica Gabriel Maeztu, Doctor y Científico de Datos del proyecto. Maeztu quiere dejar muy claro que "en ningún caso, los datos de los pacientes salen del entorno virtual del hospital. Siempre se respeta toda la cadena de la privacidad".

Una vez la información está estructurada se puede predecir, por ejemplo, si un paciente tendrá o no complicaciones al pasar por el quirófano. "Por ejemplo, el departamento de traumatología de Valle Hebrón utiliza IA para predecir qué pasará en una operación de túnel carpiano. Esto se puede saber a partir de los datos estructurados de su historial clínico, pero también con la de otros pacientes que han pasado por la misma operación en los últimos años", añade Maeztu. Con toda esta información se hace un pronóstico que, finalmente, el médico deberá valorar y decidir si opera o no.

[39]IOMED i el Hospital Vall d'Hebrón desarrollan nueva tecnología para la gestión de historias clínicas https://www.europapress.es/catalunya/noticia-iomed-vall-dhebron-desarrollan-nueva-tecnologia-gestion-historiales-clinicos-20180211115250.html

2.4.2 Sistema Judicial

La aplicación de la inteligencia artificial en el ámbito judicial es un tema muy delicado porque -como en el ámbito de la salud- las decisiones que tome el algoritmo afectan directamente a la vida de las personas. El eco de experiencias similares en Estados Unidos -donde el programa *Correctional Offender Management Profiling for Alternative Sanctions (COMPASS)* decidía la reincidencia criminal- tampoco ayudan a tener confianza en los resultados de la máquina. El medio independiente ProPublica delató[40]que COMPASS -elaborado por la empresa Equivant[41]- tenía sesgos que marcaban una probabilidad más alta de cometer crímenes por los acusados negros que por los blancos.

Sin embargo, los sistemas automatizados han evolucionado mucho en los últimos años. Y a pesar de tener presente los sesgos, también se debería pensar en los prejuicios de los jueces (por motivos de raza, género, religión, etc.) y cómo estos pueden influenciar a la hora de tomar una decisión. Por lo tanto, sería de rigor preguntarnos: ¿quién puede ser más justo, un juez o una máquina?

Humans decisions and machine predictions[42]es un interesante estudio estadounidense que demuestra cómo -en determinar una fianza en un proceso judicial- el aprendizaje automático puede funcionar mejor que las decisiones de un juez. Los resultados dieron que, cuando era muy evidente que el preso tenía muy bajo riesgo en reincidir, tanto los jueces como el algoritmo coincidían en liberarlo bajo fianza antes del juicio. Pero que la máquina era más justa que el juez en predecir casos de mayor riesgo de reincidencia criminal. Y esto es porque son sistemáticas, incluso cuando son tan racistas como los jueces[43].

En Cataluña, hace unos diez años que se aplican programas similares al COMPASS para detectar la reincidencia criminal en adultos y en jóvenes. Hasta la fecha, ninguna investigación ha demostrado que haya sesgos perjudiciales para los internos. El investigador Carlos Castillo -Director del grupo Web Science and Social Computing de la Universidad Pompeu Fabra (UPF) - ha realizado diversas investigaciones sobre estos sistemas inteligentes y -a criterio suyo- funcionan bastante bien. "Los técnicos que hacen uso de ellos en última instancia, valoran individualmente los resultados que ha dado la máquina y deciden la medida a aplicar", explica Castillo.

- Predecir la reincidencia criminal

Los permisos de salida se utilizan para la reinserción y rehabilitación de los internos. El hecho de poder pronosticar, con la eficacia predictiva más grande posible, la probabilidad de rotura futura de un permiso representa una gran ayuda para los

[40]«How we analyzed the Compass recidivisim algorithm», Pro Publica (2016) https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
[41]Equivant https://www.equivant.com/
[42]*Human decisions and machine predictions*. https://academic.oup.com/qje/article-abstract/133/1/237/4095198?redirectedFrom=fulltext
[43] Ricardo Baeza-Yates i Karma Peiró , "És possible acabar amb els biaixos dels algorismes?"

técnicos penitenciarios. El Riscanvi[44]es un protocolo (o herramienta de valoración del riesgo) que se puso en marcha en 2009 en todas las cárceles de Cataluña, para estimar las posibilidades de que una persona vuelva a delinquir una vez haya salido de la cárcel. El Departamento de Justicia de Cataluña hizo un encargo al Grupo de Estudios Avanzados en Violencia[45] (GEAV) de la Universidad de Barcelona (UB) y, en el tiempo que lleva en funcionamiento, ya se ha aplicado a unos 20 mil presos.

La predicción de la criminalidad que hacen los algoritmos es individualizada y personalizada. "En las cárceles -en un momento u otro- los técnicos, directores, psicólogos o juristas han de tomar la decisión sobre qué puede pasar ante un permiso de un interno o cuando éste está a punto de salir a la calle para siempre", explica Antonio Andrés Pueyo[46], Investigador Principal del GEAV y Catedrático de Psicología de la UB. "La preocupación es si volverá a cometer un delito".

Pueyo explica que tradicionalmente se hacía un análisis, a partir de unos parámetros y, si el interno cumplía ciertos requisitos, decidían la actuación en función de los resultados. "Hace diez años que se hace de manera automatizada con inteligencia artificial. A partir de 43 variables -que un sistema matemático combina a la perfección- el técnico puede tomar la decisión más acertada", añade Pueyo. "Si un preso pide un permiso para visitar a la familia, y la respuesta del RisCanvi es de alta probabilidad de delinquir, alerta al técnico con una señal roja. El preso sale igualmente, pero se activa un seguimiento diario, una pulsera electrónica o el contacto con un familiar", explica el director del GEAV.

El programa también acumula el comportamiento violento del interno en prisión: si ha agredido a otro, si ha intentado autolesionarse, suicidarse, etc. Es una tarea individual para cada interno y habitual en el sistema penitenciario. Pero el técnico ya no tiene que retener toda esta información, lo hace el sistema automatizado, aportándole mucho conocimiento cuando tiene que tomar una decisión.

El Riscanvi ya va por su versión 3.0. Cada tres años se actualiza y se incorporan mejoras para que sea más preciso. Ante posibles dilemas éticos, Pueyo puntualiza que "la respuesta del algoritmo es validada siempre por la Junta de Tratamiento. Y esta puede: a) mantener el mismo nivel de riesgo (bajo, alto) del algoritmo, b) aumentarlo; y/o c) disminuirlo. En los casos b y c, debe justificar este cambio con las evidencias que lo sustentan".

- Algoritmos que detectan la reincidencia juvenil

El programa *Structured Assessment of Violence Risk in Youth* (SAVRY[47]) funciona con la misma lógica que el RisCanvi, pero este fue elaborado en 2003 en Estados Unidos y no por el GEAV. "Se utiliza en muchos países del mundo: Canadá, EEUU y muchos otros de

[44]Riscanvi http://cejfe.gencat.cat/es/recerca/cataleg/crono/2017/eficacia-del-riscanvi-2017/
[45] GEAV http://www.ub.edu/geav/
[46]Antonio Andrés Pueyo http://www.ub.edu/personal/docencia/profes99_2000/pueyoficha.htm
[47] SAVRY https://treballiaferssocials.gencat.cat/web/.content/03ambits_tematics/07infanciaiadolescencia/temes_relacionats/jornades_treball_dgaia_2012/docs_3_maig/valoracio_risc_reincidencia.pdf

Europa como Holanda", explica Antonio Andrés Pueyo. "Tiene menos factores de valoración, sólo 26, e igualmente se aplica de manera muy individualizada.

Es más manual que el RisCanvi y la valoración final depende más del técnico. Porque en el caso de los jóvenes, los cambios de comportamiento son muy bruscos o van más deprisa comparados con los adultos. Por lo tanto, ya está bien que sea así ", añade el director del GEAV. "Automatizar mucho más el SAVRY no sería una buena estrategia".

Cuando la valoración de riesgo de reincidencia es alta -tanto en el RisCanvi como en el Savry- los equipos de tratamiento deciden las actuaciones o medidas que deben tomar para evitar que pase lo que indica la máquina.

- Estadística predictiva para abogados

Hay herramientas[48]jurisprudenciales que ayudan también a los abogados en la ardua tarea de leerse sentencias y combinarlas para extraer nuevas conclusiones. Los modelos matemáticos pueden operar con información cualitativa del análisis de millones de casos, de los jueces que han dictado sentencia, de la aplicación de artículos y leyes, fecha, lugar, etc. Los algoritmos combinan toda la información procedente de órdenes e instancias jurisdiccionales de todo el estado español y definen la estrategia procesal más idónea para cada caso.

- Orientación para la extradición (o no) de migrantes

Cuando detienen a un migrante ilegal y lo llevan ante el juez hay ciertos criterios por los que este puede decidir la no devolución a su país: por riesgo de muerte, riesgo de contraer enfermedad grave, por una epidemia, o para ser sometidos a tortura o castigo cruel, inhumano o degradante[49]. El juez pone una puntuación a estos criterios. Y a partir de ahí decide.

El departamento de Matemáticas e Informática de la Universidad de Barcelona cogió todos los casos que habían pasado por el Tribunal Supremo, porque se entiende que una vez han pasado por aquí sienten jurisprudencia y crearon un modelo matemático. Jordi Vitrià, investigador de la UB explica que la herramienta da al abogado una información muy valiosa a las asociaciones de ayuda a los migrantes. "Los datos son todos de casos reales y la herramienta facilita una referencia sobre por dónde podría ir la extradición", añade Vitrià. "Es una referencia para el abogado y podría serlo también para el juez, porque el modelo está creado en función de los casos del Supremo. Pero está claro que el juez tendrá la última decisión". Por ahora no se ha puesto en práctica en Cataluña.

[48]Jurimetria (https://jurimetria.wolterskluwer.es/content/Inicio.aspx)
[49]Migration HR and Governance (https://www.ohchr.org/Documents/Publications/MigrationHR_and_Governance_HR_PUB_15_3_SP.pdf)

2.4.3 Educación

La inteligencia artificial ha revolucionado todas las industrias y la educación es un sector que no se queda al margen. El pasado mes de mayo, la UNESCO reunió a 50 ministros en Beijing[50] para consensuar el documento: *Artificial Intelligence in Education. Challenges and Opportunities for Sustanaible Development*[51]. Entre los puntos principales está el mensaje que la IA tiene el potencial de transformar profundamente la educación, siempre respetando los derechos humanos y los valores sociales. Y que las políticas educativas se deberían planificar con este enfoque, de manera que se pueda empoderar a la juventud para su futuro. Se ha acabado estudiar sólo de pequeño. El aprendizaje es un hecho que se alargará toda la vida.

La IA ayudará muchísimo en el ámbito de la educación porque -según comentan los expertos- se pasará del aprendizaje en tareas al aprendizaje basado en la colaboración. Por otra parte, el tratamiento y análisis que se puede hacer hoy en día de las evaluaciones, de plataformas de aprendizaje a distancia, de la interacción en clase con el móvil o de la navegación por webs educativas permite imaginar nuevos métodos de estudio. Los algoritmos ya pueden establecer relaciones y adaptarse a cada estudiante prediciendo la mejor manera de adquirir conocimiento.

En las aulas de estudios primarios y secundarios, el sistema automatizado puede ser una guía para el profesor. Y en el ámbito universitario puede servir para gestionar la investigación, así como hacer más eficientes los recursos de la Facultad.

- Aprendizaje activo

El programa *Assessment and Learning in Knowledge Spaces* (ALEKS)[52] se ha probado en Estados Unidos durante más de 12 años con millones de estudiantes, y ahora llega a las escuelas catalanas. "No hay que tener miedo a introducir la IA en las aulas, puede convertirse en una herramienta que mejore el proceso de aprendizaje", explica Carles Sierra, Director del Instituto de Investigación en Inteligencia Artificial (IIIA).

ALEKS utiliza preguntas adaptativas para determinar, de una manera rápida y precisa, lo que un estudiante sabe o no sobre una materia. También aconseja al estudiante sobre los temas para los que está más preparado. Durante todo el curso, el programa va reevaluando periódicamente al estudiante. Sirve para planificar un aprendizaje personalizado a cada alumno, o para aquellos que tengan más problemas en alcanzar una determinada materia.

"El sistema hace un modelo de cada estudiante, y sabe su estado de conocimiento", añade Sierra. El sistema aporta un gráfico al profesor para que haga seguimiento de la consecución de los conceptos que adquiere el alumno. "Hay que tener en cuenta

[50] International Conference on Artificial Intelligence and Education (https://en.unesco.org/events/international-conference-artificial-intelligence-and-education)
[51] UNESCO *Artificial Intelligence in Education. Challenges and Opportunities for Sustainable Development (*https://unesdoc.unesco.org/ark:/48223/pf0000366994)
[52] ALEKS https://www.aleks.com/

siempre que es un apoyo al profesor en el aula. Estudios de evaluación han demostrado que, con esta tecnología, el grado de abandono es menor en la universidad y que las notas en la secundaria mejoran", concluye Sierra.

- El algoritmo hace los grupos en clase

En una clase con 30 alumnos que deben realizar una tarea por equipos, un sistema automatizado agrupa a los estudiantes en función de sus personalidades. "El profesor define el número de estudiantes por equipo y las competencias necesarias para la tarea. La IA agrupa y asigna una responsabilidad a cada estudiante", explica el Director del Instituto de Investigación en Inteligencia Artificial (IIIA), Carlos Sierra.

¿Pero cómo puede una máquina hacer los equipos para realizar una tarea sin equivocarse en la agrupación? "Primero se evalúa las personalidades a partir de 20 preguntas, y el sistema inteligente los clasifica: introvertido-extrovertido; sensato-intuitivo; analítico-emotivo; resolutivo-reflexivo", responde Carlos Sierra. "En función del resultado, el sistema automatizado los agrupará de la manera más diversa y eficiente para resolver el ejercicio propuesto por el profesor. El resultado del algoritmo podría servir para reforzar (o no) la opinión del profesor, dado que él también tiene mucho conocimiento de los alumnos". Lo que se intenta, según el director del IIIA, es buscar equipos equilibrados y que todos ellos sean lo suficientemente buenos. "Esta es la parte difícil. Encontrar un equipo bueno es fácil, pero que todos sean mínimamente buenos es muy complicado. Y eso es lo que hace el algoritmo".

Varios institutos de Cataluña ya han probado esta tecnología, y la mejora del rendimiento osciló entre un 25 y un 30%, según explica Carlos Sierra. Éticamente, podrían surgir dudas de estas aplicaciones. ¿No existe el riesgo de que la máquina se equivoque y que un alumno salga perjudicado? "Los resultados que tenemos ahora es que en promedio funciona bien. Esto no quiere decir que dé error en algún caso en concreto. Pero hay que tener en cuenta que los profesores tampoco están libres de sesgos y se pueden equivocar en la agrupación de los alumnos", responde el investigador.

- Corrección de exámenes

Otra técnica que el Instituto de Investigación en Inteligencia Artificial (IIIA) ha llevado a la práctica -con alumnos de secundaria- es lo que se conoce como evaluación entre pares.

Una herramienta tiene definida una rúbrica de cómo se debe evaluar a los alumnos. El profesor sólo evalúa varios estudiantes, pero finalmente toda la clase tiene su nota personalizada. ¿Cómo puede ser?

"Cada alumno tiene un grupo de compañeros a evaluar. ¿Qué hacemos? Comparamos la manera de evaluar de los alumnos. Y eso da un nivel de similaridad en evaluar", continúa explicando Carlos Sierra[53]. ¿Pero cómo puede ser que un alumno tenga suficiente criterio para evaluar a su compañero, como hace un profesor? "El sistema

[53] Vídeo: Inteligencia Artificial y Educación, con Carles Sierra https://www.youtube.com/watch?v=Xd5ZoKdl33A

inteligente crea similitudes entre la forma de evaluar del profesor y la de los alumnos. Los algoritmos crean un nivel de confianza entre las valoraciones del profesor y de los estudiantes evaluadores".

Este experimento se ha hecho en una clase de inglés y los resultados no se alejaban más de un 10% de los resultados que habría dado el profesor.

- Superar la dislexia

Desde hace un par de años, unas 70 escuelas de Cataluña disponen de un sistema de inteligencia artificial para ayudar a alumnos con dislexia o dificultades de lectura. En total son unos 3.000 niños y niñas -entre 5 y 12 años- los que se benefician y unos 200 más que, a título particular, hacen seguimiento intensivo.

UBinding[54]es una plataforma de aprendizaje basada en el desarrollo cognitivo de cada niño, la familia y el entorno escolar. Un grupo de investigadores de la Universidad de Barcelona (UB) la desarrolló en 2007. Según explican sus promotores, el método Ubinding consigue ayudar al 90% de los alumnos a mejorar su fluidez lectora en unos 7-8 meses de media.

"Se trata de un algoritmo de apoyo a la decisión. Un equipo formado por psicólogos, logopedas y matemáticos especializados en los trastornos del aprendizaje hacen seguimiento diario", explica Jorge López, Responsable de Área de Investigación y Desarrollo. Igualmente, se podría aplicar a niños con déficit de atención porque fallan en funciones ejecutivas, o para reforzar el repaso de algunas asignaturas.

"El entrenamiento permite adaptarse a cada niño o niña en función de la velocidad a la que asimila la lectura", añade la matemática Adina Nedelea. "Teniendo miles de alumnos, resulta muy complicado hacer el seguimiento individualizado sin la ayuda de la inteligencia artificial. El algoritmo nunca decide, pero sugiere cómo será la siguiente sesión del estudiante y el profesional valora si la hace o no".

El equipo de UBinding cuenta con el feedback de los padres/madres, y del propio niño o niña que ha hecho los ejercicios de lectura. "Sin los algoritmos no habríamos podido llegar a miles de alumnos", concluye Nedelea.

- Asistentes virtuales en el aula

Los programas más avanzados en aprendizaje automático y profundo pueden crear para cada alumno un método personalizado de estudio, a partir de los datos generados por su participación en el aula o a distancia. El algoritmo recomienda el tramo de estudio de acuerdo con el ritmo, aptitudes y objetivos de cada alumno. También puede sugerir los mejores compañeros para hacer un trabajo, analizando las compatibilidades y habilidades de cada miembro del grupo. "En pocos años, el aula se transformará: el profesor dispondrá de un asistente virtual que le ayudará en sus tareas, los alumnos tendrán un programa personalizado y la asistencia a clase será

[54]Ubinding http://www.ubinding.cat/

presencial o virtual", explica Ivan Ostrowicz[55], experto en inteligencia artificial aplicada al sector educativo.

Que todos los alumnos estudien con el mismo libro de texto, ya es antiguo. Según Ostrowitcz, los algoritmos recomendarán contenidos para cada uno de acuerdo con su nivel; cuadros de mando que ayudarán al maestro a verificar el avance de cada alumno y asistentes virtuales para que los alumnos puedan resolver dudas en todo momento. Igualmente, tendrán un recomendador de los mejores estudiantes con los que se puede aprender y los temarios adaptados a cada uno. Pero lo más singular de todas estas promesas es que un algoritmo será capaz de alertar al alumno cuando esté a punto de olvidar el conocimiento aprendido. Con el sistema de retención adaptativa, se mide la velocidad del olvido y se recomienda una revisión de la lección justo antes de que se pierda lo que se había memorizado.

- Reconocimiento facial para entrar en el instituto[56]

Este ejemplo no es una aplicación de la IA para mejorar el aprendizaje, sino un sistema de reconocimiento facial que controla la presencia de los alumnos en un centro educativo.

Desde hacía siete años que funcionaba en un instituto de secundaria de Barcelona y la mayoría de las familias lo veían con buenos ojos. "El primer año me chocó pero va muy bien porque si tu hijo no pasa por el sistema de reconocimiento facial, envían un SMS a los padres", explica una de las madres en una entrevista de TV3. A través de diferentes cámaras instaladas en el centro, los estudiantes debían fichar al llegar a la escuela, sino se advertía de la falta del menor a los padres. El sistema no controlaba la asistencia a clase, aunque si el alumno estaba en el instituto, seguramente, iría a clase.

Según la profesora de Derecho y Ciencia Política, Mònica Vilasau Solana[57], los datos biométricos y los datos de menores están marcados como especialmente sensibles. Por este motivo, desde la Autoridad Catalana de Protección de Datos se le abrió una investigación. "No es suficiente con que los padres den su consentimiento, sino que se ha de evaluar el impacto que puede tener usarlos y mirar si no hay otra alternativa disponible", explica Vilasau.

Otros centros educativos catalanes habían implementado el reconocimiento facial de los alumnos pero, con el Reglamento Europeo de Protección de Datos, lo deshabilitaron. Este instituto de Badalona también lo terminó retirando, a raíz de la investigación abierta por la APDCAT. El ejemplo choca con la vertiente más ética de la IA: se utiliza la tecnología para controlar y no para mejorar el rendimiento en la formación de los jóvenes.

---

[55] IvanOstrowitczhttps://www.linkedin.com/in/ivanostrowicz/?locale=fr_FR

[56] Noticia publicada al 324.cat (5/10/19) https://www.ccma.cat/324/reconeixement-facial-per-passar-llista-en-un-institut-de-badalona/noticia/2952712/

[57] Mònica Vilasau https://www.uoc.edu/portal/es/news/kit-premsa/guia-experts/directori/monica-vilasau.html

2.4.4 Banca

Los bancos siempre han sido los primeros en aplicar los algoritmos de decisión automatizada. Ya desde los años 70, con operaciones mucho más simples que las actuales. Ahora son redes neuronales avanzadas que ofrecen una variedad enorme de 'productos financieros' en función de los datos que se les introducen: promocionar o no un servicio a través del buzón de correo electrónico de un cliente o recomendarle productos cuando visita la web de la entidad financiera son acciones habituales. La inteligencia artificial también se utiliza para tareas internas para ganar eficiencia y rendimiento en una entidad bancaria, para gestionar las llamadas de los clientes, para conseguir unos tiempos de espera telefónica aceptables, para diseñar el trato preferencial de los clientes Premium o para decidir en qué parte de la ciudad ubicar nuevas oficinas.

Los bancos están muy controlados respecto a la privacidad y anonimización de los datos que nutren sus algoritmos. Por lo tanto, de entrada, impera la lógica de no perjudicar a ningún cliente con posibles discriminaciones generadas por los sesgos. "Asimismo se debe entender que un banco vive de discriminar, aunque el verbo tiene una connotación negativa", explica Marco Bressan, ex Jefe Científico de datos de una entidad bancaria de ámbito estatal. "La entidad debe saber si recuperará el dinero de una hipoteca o de un préstamo. La utilización de algoritmos para decidir a quien los concede o no es básica. Por su parte, el banco ha de detectar y mitigar los posibles sesgos (de género, nivel de renta, origen del cliente, etc.) que puedan haber".

Bressan está convencido de que los algoritmos -si están bien diseñados- son más precisos que las personas. "Imagina un director de un banco misógino: seguramente tomará decisiones muy equivocadas por los prejuicios personales que puede tener. Los sistemas automatizados responden a diferentes variables y la decisión que toman es más justa ", añade. "Los datos con los que se alimentan los algoritmos es lo más importante. Si de entrada hay una discriminación, la perpetuará en el tiempo y afectará a mucha más gente".

Por otra parte, los humanos -según Bressan- ya no llegamos a calcular la cantidad de operaciones en tiempo real que pueden hacer las máquinas con datos masivos. "Sin embargo, como actúan a gran escala, sólo que el algoritmo falle una vez perjudica miles o millones de personas de golpe. En cambio, el director del banco cuando se equivocaba afectaba como mucho a 100 personas.

- El crédito, para quien tenga mejor puntuación

Las acciones más reguladas de una entidad financiera son las de acceso a crédito. El algoritmo que decide a qué tipo de clientes le concede o no - a priori- está validado y regulado por el Banco Central de cada país. Pero el cliente desconoce esta valoración.

¿Cómo decide un algoritmo conceder o no un crédito? Cuando un banco presta dinero se asegura de que lo recuperará. Antes se recopilaba información del cliente y el director de la entidad decidía. Con tecnologías más potentes, la predicción del dinero que se recuperará o no es más acertada. Se sabe la cantidad de dinero que la persona gana, y el ahorro que puede realizar cada mes después de pagar los gastos. Se clasifican los clientes en grupos (o clusters), y se asigna una puntuación (o scoring) que

no está necesariamente relacionada con lo que gana. Una persona con un salario de unos 1.500 euros mensuales puede ahorrar 100 y tener una puntuación más alta que uno que gane 4.500 euros y no sea capaz ni de ahorrar 50.

¿Cómo se entrenan estos algoritmos para que sean eficaces? Pongamos que tenemos los datos de miles de clientes a los que se les ha concedido créditos en el pasado. En función de unas variables, se programa el algoritmo para que prediga si estas personas son lo suficientemente solventes para devolverlo o no. Como son datos del pasado son fácilmente comparables para saber si el sistema automatizado está funcionando de manera correcta. Ya ha quedado demostrado que la precisión siempre es mayor con un sistema automatizado que haciéndolo manualmente.

- Contratación de seguros

Los sistemas de *scoring* (puntuación) son frecuentes en la contratación de seguros (de vehículos, salud, casa, empresa, etc). Por lógica, la probabilidad de que un joven tenga un accidente de tráfico es más alta que la de una persona de 55 años, y por tanto el coste del seguro será más elevado.

Podríamos preguntar si estos sistemas automatizados ¿no se equivocan nunca? y la respuesta es que el error o acierto dependerá de la información previa (datos masivos) que se introduzca en el algoritmo. Cuando las predicciones en grupos se equivocan sale más gente perjudicada. Puede darse el caso de que un joven sea más prudente que un hombre de más edad, y que los jóvenes siempre terminen pagando más. Los algoritmos de las aseguradoras tienen en cuenta otras variables para decidir la prima, como la renta. Se ha demostrado que la gente de renta más baja -por norma- tiene más siniestralidad que uno de renta alta. La orografía también se tiene en cuenta en el precio a pagar porque condiciona mucho los accidentes.

- Predictores de hipotecas

A medida que la inteligencia artificial es más potente, los algoritmos son más matemáticos (y menos estadísticos). "Esto ha hecho que pasemos de la analítica descriptiva a la predictiva", explica Pier Paolo Rossi, Director de Analítica y Marketing de una entidad bancaria catalana. "Se estudia el pasado para predecir el futuro. El algoritmo analiza, por ejemplo, todos los clientes que han pedido hipoteca en los últimos cinco años y, a partir de unas variables, predice los que potencialmente pedirán una hipoteca pronto".

Las variables pueden ser de tipo sociodemográfico (edad, género, hijos, etc.), de productos financieros contratados (cuenta corriente, hipotecas, préstamos, etc.), de cómo interactúa con el banco (si tiene o no tarjeta de crédito, si la utiliza mucho o poco, si siempre saca dinero del cajero o va a la oficina, etc). La relación con los bancos en los pueblos es más cercana que en las ciudades, y esta es otra variable. Con los datos obtenidos se entrenan los algoritmos que predecirán, con una probabilidad muy alta de acierto.

La analítica prescriptiva también comienza a implementarse. Los algoritmos identifican las necesidades de los clientes en función de su ciclo de vida. No es lo mismo una

pareja de 25 años, que vive en un piso de alquiler, sin hijos; que la misma pareja diez años más tarde. Si todo les ha ido bien, seguramente tendrán un sueldo más alto, querrán un piso de compra, tendrán hijos y las necesidades serán diferentes. "Con esta analítica gana el banco y el cliente, porque antes se ofrecían hipotecas a todos y ahora sólo a aquellas personas que la necesitan", añade Pier Paolo Rossi. Los datos no fallan.

- Recomendadores financieros

Como ya ha quedado claro en los anteriores ejemplos, los bancos juegan con ventaja respecto a otras empresas porque tienen muchos datos de sus clientes. Los datos son dinero. Para saber qué productos financieros les interesarán sólo tienen que hacer un análisis descriptivo'. Es decir, estudian cómo son, cuántos de ellos llegan económicamente bien a fin de mes, cuántos son de la zona alta de la ciudad, de qué edades son, qué ritmo de vida llevan, etc. Los algoritmos ayudan a los clientes a tomar decisiones financieras, a comprar a crédito, a gestionar los gastos del hogar o la vida familiar, contratar seguros, etc.

El negocio de los bancos ha cambiado considerablemente en los últimos años y ya no consiste sólo en prestar y guardar dinero, sino en ofrecer servicios que los diferencien. ¿De qué manera? Con la información relacionada con los hábitos de vida de miles de personas: desde la zona de la ciudad donde vive, hasta el tipo de trabajo, el sueldo, los miembros de la familia, los menores a su cargo, etc. Pero también con datos de consumo: de la luz, agua, gas o cualquier otro gasto domiciliado, las compras diarias, gastos de viajes, estudios, actividades extraescolares, vacaciones, reservas de aviones y hoteles de trabajo, o de ocio. Con todos estos datos los algoritmos hacen una radiografía integral de la vida de cada persona, y ofrecen estos servicios que 'les pueden interesar'. De este modo, consiguen fidelizar al máximo a cada cliente.

- *Coaching* financiero

Como los datos de consumo son los más valiosos, con ellos se hacen 'perfilados' anonimizados, por lo que se clasifica al cliente en un grupo determinado. Así se detectan productos que todos ellos comprarían. En el ejemplo anterior ya hemos visto las variables de nivel económico y gasto mensual que se pueden tener en cuenta.

Los bancos envían y/o hacen ofertas muy personalizadas de compra que, con muy poca probabilidad, el cliente rechazará porque se ahorrará dinero en un servicio que utiliza o será una ventaja familiar en poco tiempo. Es decir, si una persona de la misma edad que yo, con el mismo sueldo y con el mismo gasto tiene un wifi contratado en casa mucho más económico que yo, el banco me ofrecerá que cambie de compañía y que contrate uno nuevo mediante la entidad financiera. Estas ofertas personalizadas van destinadas a cientos o miles de clientes.

El recomendador financiero automatizado sustituye al gestor tradicional del banco, aquella persona en la que se confiaba para abrir un plan de pensiones o hacer un ahorro a largo plazo. Pongamos que una persona tiene 10 mil euros que quiere guardarlos para su jubilación o invertirlos para sacarle más rendimiento. El algoritmo lo clasificará según si tiene un perfil de riesgo agresivo '(arriesgará mucho) o es más bien' conservador '(poco riesgo), y conseguirá que la propuesta del banco se adapte a

sus preferencias. Estos tipos de asistentes sin intervención humana ayudan tanto al personal del banco como al cliente. Le alertan si tiene riesgo de perder su dinero, aunque en esta operación el banco no gane nada, pero fideliza.

Son acciones destinadas a recuperar la confianza en los bancos, muy dañada tras la crisis económica. Otra operación que fideliza es avisar al cliente de que se quedará en números rojos en muy poco tiempo. Como el sistema automático monitoriza las cuentas, le puede alertar de una situación incómoda que todo el mundo se quisiera ahorrar. Y aconseja frenar ciertos gastos de restaurantes, ocio, compras secundarias, etc.

- Fraude en tarjeta de crédito

¿Alguna vez le ha pasado que han rechazado su tarjeta de crédito en una tienda en la que no había comprado nunca o se la han bloqueado porque había comprado más de lo habitual?

Aunque no somos conscientes, esto es más habitual de lo que podríamos imaginar: cada día se bloquean millones de tarjetas en todo el mundo.

El uso del aprendizaje automático para detectar fraudes financieros se utiliza desde los años 90 del siglo pasado y hoy ya está muy avanzado, pero todavía no es del todo perfecto. Los algoritmos están entrenados para bloquear tarjetas de crédito cuando detectan un consumo extraño o exagerado. Y lo hacen monitorizando millones de transacciones diarias.

Pero ¿qué pasa si el sistema automatizado se equivoca al considerar 'extraña' una operación que no lo es? Esto se conoce como 'falso positivo', es decir, errar en la predicción y en la decisión de bloquear la tarjeta. Un estudio realizado en 2015 por la consultora Javelin Strategy Research[58] estimó que sólo una de cada cinco predicciones de fraude es correcta, y que los errores en el banco le podían costar miles de millones en ingresos perdidos porque las personas que han visto su tarjeta bloqueada ya no la utilizan de nuevo. Aunque la detección del fraude automatizado está muy avanzada, todavía presenta algunas limitaciones. Porque para detectar casos fraudulentos reales, es necesario que los algoritmos se equivoquen muchas veces.

Ahora para evitar más frustraciones a los clientes y dolores de cabeza a los bancos, investigadores del Massachussets Institute Technology (MIT) utilizan una técnica[59]—-que ya se aplica también en bancos ubicados en Cataluña- para reducir casi a la mitad la posibilidad de error. Esta técnica se fija en otras 'características' de la compra que hasta ahora no se habían tenido en cuenta, como la distancia entre dos comercios y la hora en que se produjo la compra, además de si se han hecho las dos de forma presencial u online.

[58]2015 Data Breach Fraud Impact Report https://www.javelinstrategy.com/coverage-area/2015-data-breach-fraud-impact-report

[59]Solving the false positives problem in fraud prediction using automated feature engineering http://www.ecmlpkdd2018.org/wp-content/uploads/2018/09/567.pdf

- Algoritmos que interactúan con el cliente

Los asistentes virtuales online o *xatbots* son algoritmos que interactúan directamente con los clientes, simulando un gestor de banco. Se recoge la conversación escrita en forma de chat, como los movimientos que hace para las diferentes opciones de la página web. Las cookies (programas que siguen el rastro) permiten al banco saber por dónde se ha desplazado una persona en el tiempo que ha visitado la web.

"En función del movimiento, de las preguntas que ha hecho al *xatbot*, de las respuestas que ha dado, el algoritmo va aprendiendo y puede personalizar la pantalla del cliente, en ese momento, lo que necesita. Así sabemos si está interesado en un préstamo, si es cliente o no, qué cantidad necesita, para qué uso (viaje, casa, estudios, coche), y el algoritmo personaliza la respuesta al momento ", explica Pier Paolo Rossi. Este sistema es el que utiliza también Amazon o cualquier tienda online. "Si conoces a tu cliente, tienes una ventaja respecto a los demás. Y generas una confianza que antes no tenías", concluye.

2.4.5 Comercio

Sin darnos cuenta estamos rodeados de algoritmos que nos sugieren productos, recomiendan una serie, nos alquilan un piso y nos ofrecen un coche de segunda mano, etc. Y siempre aciertan nuestros gustos.

De igual manera, hemos introducido en nuestras vidas pequeños aparatos que nos saludan a primera hora de la mañana, o con los que interactuamos para pedir música, encender la televisión mientras cocinamos, o que nos diga la previsión del tiempo. Son los asistentes virtuales, micro inteligencias que entienden el lenguaje natural. Los encontramos también en formato *xatbot*, a menudo en páginas webs e interpretan lo que un cliente o ciudadano les pregunta. Con la información que reciben, deciden la respuesta. Son sistemas que ahorran tiempo al servicio de atención al cliente de una empresa, pero también sustituyen a trabajadores que, en un pasado no tan lejano, atendían llamadas y resolvían dudas por teléfono o correo electrónico. Pero lo que tenemos hoy son inteligencias artificiales muy prematuras. No hemos visto nada de lo que está por llegar.

En Cataluña, las multinacionales tecnológicas copan el mercado y ser competitivo en inteligencia artificial (IA) no es fácil, explica Jordi Navarro[60]– CEO de una empresa dedicada a la analítica predictiva y al *machine learning* desde hace 4 años-. Él considera que el sector turístico aún tiene camino por recorrer. "A partir de información histórica, podemos entender el pasado y extraer patrones. Esto nos permite operar en el sector hotelero, en las cancelaciones de habitaciones de hotel, por ejemplo. Con la IA, de alguna manera, es como si tuviéramos una bola de cristal, que nos permite acertar con un grado bastante alto. El comportamiento humano es muy previsible. Si no ¿por qué crees que triunfa el líder mundial del comercio online?", pregunta.

A Jordi Mas -miembro de Softcatalà y pionero de la Internet catalana- le preocupa la parte ética de los sistemas de aprendizaje profundo que se utilizan en ámbitos comerciales, sobre todo porque no se puede explicar cómo una máquina toma ciertas decisiones. "Me preocupa que no haya un código deontológico respetado por todos. Por ejemplo, ningún profesional debería trabajar en un proyecto comercial que no respete las convenciones de derechos humanos internacionales. También me preocupa que la regulación vaya por detrás de la tecnología, y que seamos muy reactivos. Como lo que está pasando ahora con el reconocimiento facial, que no somos conscientes de que los datos biométricos tienen mucho valor. Los algoritmos eligen en cada momento el contenido que leemos, miramos o escuchamos, los productos que acabamos comprando, filtran lo que no nos interesa, y mil cosas más. Existe el peligro de crear burbujas reduccionistas, en un mundo que es complejo y con muchas opciones. Me preocupan las empresas poco éticas pero también (y mucho) los gobiernos. Sólo hay que pensar que mucho de lo que tenemos nos llega de EEUU o de China".

[60] Jordi Navarro https://www.linkedin.com/in/jordi-navarro-perez/

- Cancelación de habitaciones

En 2017, un pasajero de la aerolínea United Airlines[61] acaparó todas las miradas al ser expulsado, por la fuerza, de un avión. Había *overbooking* (se habían vendido más pasajes de la capacidad del aparato) y alguien tenía que quedarse en tierra. Le tocó al cliente que menos millas (o vuelos) acumulaba. Otro pasajero grabó el incidente, y el vídeo dio la vuelta al mundo. La compañía reembolsó el precio de los billetes a todos los que viajaban en el desafortunado vuelo 3411.

Esta situación no es nueva. El overbooking siempre ha existido. Antes por intuición o conocimiento del sector se hacía a ojo y ahora se hace con las matemáticas y la IA. Jordi Navarro explica que la misma estrategia se emplea en el sector hotelero, con las reservas y cancelación de habitaciones. "Ahora nos equivocamos menos. Está todo basado en estadística a partir del perfil de la persona que ha hecho la reserva, el comportamiento en anteriores ocasiones, y otras variables.

- Un clic no quiere decir un cliente

A menudo cuando buscamos un alojamiento en un destino turístico accedemos a los hoteles desde un portal o web de precios intermediaria. En el buscador se introduce la ciudad, la cantidad de noches y te da un resultado. Estas webs intermediarias son mayoristas que ofrecen todo lo que los establecimientos tienen, como las agencias de viajes de antes. ¿Cuál es el negocio del mayorista? Los anuncios de los hoteles. Por ejemplo, si alguien busca pasar dos noches en Menorca, la web mayorista ofrecerá un listado de alojamientos con los diferentes precios. Sólo para clicar en alguna de las ofertas, el propietario del establecimiento ya paga al mayorista. Si después no se reserva la habitación, ha perdido dinero. Y esto ocurre miles de veces cada día. Por lo tanto, el hotelero debe estar muy convencido de que lo que ofrece al mayorista es lo suficientemente atractivo para que los clics supongan reservas de habitaciones.

Ahora con la IA se puede acertar mucho más -en función de la ciudad, de la hora, de si piden 1 o 2 plazas, de para cuando las piden, etc.-, y asignar una probabilidad alta de reserva. Así, el hotelero sólo ofrece al mayorista una habitación en Menorca cuando sabe casi a ciencia cierta que el cliente se la quedará. ¿Se podría hacer sin algoritmos? "No" responde Jordi Vitrià, Investigador del Departamento de Matemáticas e Informática de la UB, desde donde se ha colaborado en este proyecto. "Es una cuestión de escala. Antes la agencia de viajes de toda la vida tenía contactos con hoteles, empresas de alquiler de coche, aviones, etc. Y se quedaba un margen para todo lo que ofrecía y vendía. Este sería el papel del mayorista. Ahora en tiempo real, las 24 horas del día, es imposible hacerlo personalmente".

¿Es ético que si te conectas desde el barrio más rico de la ciudad, el precio de aquella habitación te cueste el doble? "*Business is business*", añade Vitrià. "Aunque no juegas con unas reglas claras, porque nadie te ha contado que te han tomado los datos de geolocalización, y que esto puede influir en el precio final que pagas".

[61] «United Airlines reembolsará los billetes a todos los clientes del vuelo del que expulsó un pasajero», 20 Minutos https://www.20minutos.es/noticia/3011324/0/united-airlines-devolvera-billetes-pasajeros/

- La ITV más asequible

En Francia el precio que se paga para pasar la Inspección Técnica de Vehículos está liberalizado. Como hay mucha competencia entre las franquicias que revisan los coches, la Universidad de Barcelona colaboró en un proyecto para establecer precios automáticos en función del lugar donde estuviera ubicado el centro de la revisión. "No es lo mismo si tienes siete centros de ITV de la competencia, que si eres el único en la zona", explica Jordi Vitrià de la UB. "Hicimos un algoritmo que -con todos los datos de inspecciones de vehículos de Francia de los últimos años- automáticamente proponía al gestor de la franquicia de ITV el precio que le podía hacer competitivo, en función de las circunstancias que le rodean, o del día del mes, etc.".

- Factores externos para acertar en la producción

Los grandes concesionarios de coches necesitan tener un control muy estricto sobre el aprovisionamiento de materiales. Algunas piezas deben solicitarse con tiempo porque proceden de países asiáticos. La central de la marca de vehículos hace previsiones, pero a menudo no se ajustan a la realidad porque no tienen en cuenta otros factores externos a la fabricación. Esto provoca que haya grandes stocks en los almacenes. Lo que quiere decir, dinero parado.

Las predicciones de producción se pueden mejorar con un modelo matemático, dice el investigador de la UB, Jordi Vitrià. "Con el histórico de aprovisionamientos, pero también con la previsión meteorológica, los días festivos y de vacaciones del lugar de cada almacén, y con datos de factores socioeconómicos los resultados son muy diferentes. Es información que nunca habrías tenido en cuenta para la fabricación de coches, pero que influyen mucho".

- Cómo explotar al máximo las bodas

El *customer journey* es el seguimiento que se hace al cliente, desde el momento que entra en contacto con una empresa (para solicitar información, por ejemplo) hasta que la abandona (se borra de un servicio). A partir de las interacciones se puede predecir cuándo se perderá ese cliente. Si nos damos de alta en un servicio de alquiler de motos por la ciudad, por ejemplo, y durante un tiempo no contratamos ninguna, la empresa nos enviará correos y mensajes al móvil recordando ofertas, descuentos, oportunidades, rutas a hacer hasta que alquilamos de nuevo una moto. Y si no reaccionamos, nos habrá dado por perdidos.

Con esta misma estrategia operan la mayoría de las empresas hoy. En Sant Cugat hay una que ofrece todo tipo de servicios y productos para una boda. Es el gran buscador de las bodas: desde maquilladores, hasta restaurantes, alquileres de coches especiales, vestidos, etc. Trabajan más de un centenar de personas que responden dudas y consultas en cualquier idioma. Está ubicada en Cataluña pero el mercado es global.

El negocio funciona igual que el de las reservas de habitaciones de hotel. La empresa gana de cada clic que hacen en un servicio, y le paga el restaurante que se anuncia. Éste paga por salir de los primeros en su buscador. Si sales de los primeros, tienes

muchas probabilidades de que hagan reservas de boda. "Ni los nacimientos, a pesar de ser un momento importante de la mayoría de personas, causa tanta expectación, ni mueve tanto dinero como las bodas", explica Jordi Vitrià de la UB. "Los algoritmos ayudan a mejorar el *costumer journey* de esta empresa y alertan cuando hay peligro de que se vaya el cliente, cuando se debería hacer una acción presencial o por correo-e", añade.

- Algoritmos para ser más rápidos

Las empresas de repartidores en bicicleta o moto también utilizan algoritmos. En el reparto de última milla (distancias cortas) se piden cosas pequeñas y muy diversas. Estas empresas confían a los algoritmos la eficiencia de la gestión de los pedidos y las entregas. Por ejemplo, asignan al repartidor que cumplirá antes con la petición del cliente y le indican dónde comprar lo que pide, en función de la distancia a recorrer y el tiempo establecido. Si la distancia es muy grande, la aplicación traspasará el pedido a un repartidor que vaya en moto. Esto también supone un beneficio para el repartidor, que hace esperar menos tiempo al cliente, y puede recibir una mejor calificación al terminar el servicio.

Los algoritmos asignan una puntuación a los repartidores en función de la reputación que tengan, basada en criterios como las horas dedicadas al reparto, si ha estado en fines de semana, si ha estado en las horas de mayor demanda, si no tienen quejas de los clientes, etc. Con aprendizaje automático, el sistema también puede entender qué piden los clientes y cómo lo piden. "Por ejemplo, si se piden tres aguas, ¿qué significa?, ¿Tres garrafas de cinco litros de agua o de un litro? La plataforma dispone de una base de datos con toda la información acumulada y procesa para que la entrega sea lo más precisa posible", explica el responsable de comunicación de una empresa de reparto a domicilio.

- Los datos de comportamiento, los más valiosos

Todos visitamos portales con anuncios de compra-alquiler de vivienda, de coches de segunda mano, o de ofertas de trabajo, para ver una serie. Pero ¿cómo funcionan los algoritmos que tienen detrás? El objetivo principal es tener al usuario el máximo tiempo enganchado (*engagement* de usuario) en la plataforma o web. Este es el negocio: a más tiempo invertido en estos portales, más posibilidad de hacer dinero con los anuncios, productos o servicios que se ofrecen. Con aprendizaje automático se hacen recomendaciones que aciertan casi siempre, además de detectar si alguien pone anuncios fraudulentos (que los hay y muchos), y pueden interactuar con el usuario con respuestas inteligentes (*smart replies*).

"Los algoritmos de recomendación se inventaron hace veinte años, no son nada nuevo. Aprenden del comportamiento del usuario", explica el responsable de un *marketplace*, empresa internacional ubicada en Barcelona y propietaria de varios portales de anuncios ubicada en Barcelona. "Sin embargo, tenemos dos problemas: 1) Al introducir nuevos productos o servicios. Si tienes un portfolio de 7 mil productos, que ya funcionan solos pero a uno no lo conoce nadie porque es nuevo, ¿cómo haces para que se vea? Se entrena el algoritmo para que lo recomiende hasta que consigue la

atención del usuario. 2) Segundo problema: lo que se conoce como *cold start* (usuario nuevo). No sabemos nada de él, no tenemos datos de comportamiento. Tal vez ha navegado por veinte páginas, pero no ha comprado nada. El algoritmo no sabe qué le gusta, por lo tanto, va a ciegas en las recomendaciones. Si consigue que compre o meta algún producto en la cesta, ¡esto ya es muchísima información! Normalmente, una persona está 20 minutos en el portal y le muestra cosas muy genéricas", explica este profesional.

Todas las empresas de comercio electrónico funcionan igual. Cuando una persona llega a la web de un portal de anuncios, se le inyecta una cookie (o pequeño archivo que lo identifica con una numeración única). Las cookies perduran en el tiempo, y durante meses o años ofrecen mucha información. ¿Qué datos se recogen? Las categorías visitadas, las búsquedas, las páginas vistas, el tiempo invertido en cada página, las compras, etc. "Si tengo un usuario que cada vez que entra en el portal va a la sección de libros, hace búsquedas de Java, Python, HTML, me ha visitado mil veces pero no se ha dado de alta y no me ha comprado nada, igualmente ¡tengo muchos datos!", añade el responsable del *marketplace*. "Analizamos toda esta información, y cada vez que un usuario es identificado, se ponen en marcha los recomendadores. Hacemos millones de recomendaciones diarias para todos los portales que tenemos. Son del tipo: si has visto esta bicicleta, seguro que te interesa este casco. Los anónimos también cuentan. Quizás entran hoy y no vuelven hasta dentro de seis meses, pero todavía tiene la cookie porque no expira", concluye.

2.4.6 Social

La inteligencia artificial en el sector social resulta muy útil sobre todo para procesos a gran escala. Los algoritmos pueden conceder (o no) ayudas económicas, asistencia de algún tipo y derivar a otros servicios especializados en función de la problemática personal. Hoy ya son de gran ayuda, sobre todo en grandes ciudades cuando hay miles de ciudadanos solicitando servicios.

Los algoritmos -como ya se ha explicado- se alimentan o entrenan con datos masivos del pasado. Y estos, es muy probable que tengan sesgos porque las sociedades evolucionan y los hábitos, costumbres, modos de vivir, educarse o de acceder al mundo laboral han variado. No es lo mismo la sociedad catalana de hoy que la de hace cincuenta años.

Por ejemplo, las mujeres no accedíamos tanto a puestos de trabajo, prácticamente no había familias monoparentales, tampoco se adoptaba tanto como hoy, etc. Podemos tener problemas similares (económicos, de inmigración, de violencia, etc.) pero las realidades son diferentes. Por eso, hablando de la aplicación de la IA en sector social es tan importante detectar y mitigar- cuanto antes mejor- las posibles discriminaciones que vayan adosadas a los datos masivos que se utilizan.

En Cataluña, la Administración Pública apenas empieza a ponerse las pilas en utilizar la IA para redistribuir ayudas sociales, pero sobre todo se está contemplando -por ahora- para ser más eficientes en los trámites y gestionar mejor los recursos.

- Inteligencia colectiva para las ayudas sociales

El Área de Derechos Sociales del Ayuntamiento de Barcelona puede atender un promedio de 50 mil primeras visitas al año. Las personas que acuden a los 40 centros de servicios sociales repartidos por la ciudad tienen problemas económicos, o de dependencia, enfermedad mental, de alcoholismo, puede necesitar ayuda psicológica, de adaptación, puede acudir por un caso de violencia de género, etc. Problemáticas muy diversas que son atendidas por una plantilla de más de 700 profesionales, entre trabajadores sociales, psicólogos y educadores sociales.

Cuando la persona llega al centro, se le atiende en unas cabinas privadas. El trabajador social registra la conversación y al terminar transcribe la problemática, así como la ayuda o servicio a la que ha sido derivada. En el sistema interno se describe con tres letras: Demanda (D), Problema (P), Recurso (R). Actualmente, el Ayuntamiento dispone de cientos de miles de entrevistas, muchas de las cuales acaban siendo repetitivas porque los problemas se parecen.

"Entramos en un repositorio 300 mil entrevistas y lo dotamos de técnicas de aprendizaje automático", explica Lluís Torrens, Director de Innovación Social del Área de Derechos Sociales, Justicia Global, Feminismos y LGTBI en el Ayuntamiento de Barcelona. "La máquina leyó todos los comentarios anotados por los trabajadores sociales para las D, las P y las R. Ahora sugiere los recursos a partir de lo aprendido. Clasifica las demandas y las respuestas posibles".

Los resultados se han aplicado ya en tres centros. Según explica Torrens, las recomendaciones son más precisas que la de los profesionales porque evitan la dispersión. Y el nivel de satisfacción es alto. "Si tienes 700 profesionales es muy fácil que no todos destinen los recursos de la misma manera. Bien porque el profesor universitario le contaba las materias a su manera, bien porque el profesional ha investigado más en un tipo de problemática, etc. La máquina homogeniza las respuestas, dando la libertad al profesional que acabe decidiendo. Me gusta decir que no es un sistema de inteligencia artificial sino colectiva", apunta el Director de Innovación Social del Área de Derechos Sociales del Ayuntamiento de Barcelona.

- Potenciar las evaluaciones

Se calcula que unas 25 mil familias de Barcelona viven en situación de pobreza cronificada. Por extraño que pueda parecer, nunca hasta ahora se han hecho evaluaciones de las ayudas sociales que el propio ayuntamiento destina. No se sabe si les fue bien o no a las personas o a familias a las que se les ha concedido una prestación o un servicio de recuperación. "Un médico cuando te da una pastilla sabe en qué porcentaje te curará la enfermedad. Los profesionales sociales no saben si funcionará o no el tratamiento que aplican porque en las personas influyen un montón de factores", explica Lluís Torrens. "De entrada si los que han pedido la ayuda ya no vienen más al servicio, puede que se les haya resuelto la situación, pero también que hayan desistido de nuestra ayuda o se hayan marchado de la ciudad", añade.

"La IA nos permite hacer evaluaciones precisas y comparar lo que habría decidido el profesional, con el que decide la máquina. Ahora queremos crear un sistema integral de datos masivos, que consiste en coger toda la información que podemos obtener de registros administrativos de una familia demandante de ayudas económicas, y saber sus ingresos a través de la Agencia Tributaria, si cobra una pensión, si ha pedido otras ayudas a otras administraciones, para conseguir una evolución temporal y saber si está mejorando o no ", continúa el Director de Innovación Social del Área de Derechos Sociales del Ayuntamiento de Barcelona.

Las evaluaciones se han puesto en práctica en el proyecto europeo BMincome[62] , donde durante dos años se ha hecho un seguimiento exhaustivo a 950 familias de los diez barrios del Besós. "Les hemos inyectado una renta, con políticas activas de reinserción. Todo un proyecto muy controlado, con la agencia pública Ivàlua detrás. Este proyecto (que ahora se acaba) ha permitido tener mucho conocimiento de cómo una acción muy evaluada está afectando a las familias. Estamos intentando entrar la filosofía de que las intervenciones deben evaluarse y los algoritmos nos ayudan a hacerlo", añade Torrens.

- Detectar los sesgos

Después de poner en marcha algunas pruebas piloto con inteligencia artificial, la máxima preocupación del Director de Innovación Social del Área de Derechos Sociales

[62] Projecte BMincome http://ajuntament.barcelona.cat/bmincome/ca/pressupost-ajudes-barcelona

del Ayuntamiento de Barcelona, Lluís Torrens, es que los algoritmos con los que opera tengan el mínimo de sesgos posibles. Por ello, han encargado una auditoría a una empresa externa. "Es muy probable que, por el origen de las familias, por género o por edad -de manera involuntaria tengamos resultados que no corresponden. Y que no haya ninguna discriminación por el propio algoritmo que no haya generado alguna sin que nosotros seamos conscientes. Es lo más normal, ya que las respuestas sociales han ido cambiando a lo largo del tiempo. Pero debemos estar alertas a los sesgos y mitigarlos", explica Torrens.

"Igualmente nos gustaría detectar con el algoritmo una problemática que el profesional social no haya sido capaz de captar. Como un hecho posible de violencia en un hogar (sea hacia la mujer, los hijos o los ancianos). Esto se realizaría correlacionando información de otros servicios sociales, por ejemplo, los servicios de acompañamiento a las mujeres con problemática de violencia machista. Conectando los servicios se podría hacer un programa de alerta precoz y detectar una situación complicada en casa. Aún lo estamos trabajando y lo exploraremos más. Pero el profesional asistencial podría trabajar coordinadamente con los Mossos de Escuadra, con la Guardia Urbana y con otros para ayudar a las familias", concluye el responsable del consistorio de la capital catalana.

- Estudiar el envejecimiento activo

La directora del Grupo Consolidado de Investigación *Machine Learning and Computer Vision* de la Universidad de Barcelona (UB), Petia Radeva, lidera el proyecto *Lifelogging*[63], que consiste en capturar imágenes de una persona a lo largo de su vida con una cámara portátil.

Fue financiado por La Marató de TV3 y se desarrolla junto con el equipo de la Doctora Maite Garolera del Consorcio Sanitario de Terrassa. El interés es monitorizar la vida diaria para estudiar el envejecimiento activo de las personas mayores. Las cámaras de resolución temporal larga (LTR) son excelentes para este propósito.

Comer, moverse, dormir, ir de compras, cocinar, limpiar, socializar, etc. de las personas mayores están relacionadas con el deterioro cognitivo. Y se sabe que los que mantienen habilidades altas para hacer diferentes actividades cada día son menos frágiles física y cognitivamente. Con técnicas de visión por computación y aprendizaje profundo se están analizando estas acciones a partir de las imágenes capturadas con cámaras LTR. También sirven como álbum de recuerdo para mejorar la memoria.

- Asistente inteligente de dietas personalizadas

La Universidad Politécnica de Cataluña (UPC) participa en el proyecto, "*Diet for you*" junto con el Instituto Hospital del Mar de Investigaciones Médicas (IMIM) para potenciar un estilo de vida saludable y la adherencia a la dieta. "El algoritmo toma todos los datos del paciente -de salud, de estilo de vida, la dieta que ha hecho o está haciendo, el ejercicio que hace, eventualmente se podría tomar información

[63]Lifelogging http://www.ub.edu/cvub/egocentric-vision/

genómica, etc.- y hace un perfil de la persona ", explica Karina Gibert, Investigadora Principal del Proyecto y Miembro del *Intelligent Data Science and Artificial Intelligent Research Center* (IDEAI-UPC). Así el sistema tiene conocimiento para saber sus patrones de dieta", añade.

Automáticamente crea los menús para 3 o 5 meses, con todos los platos preparados, que se adaptan a la prescripción nutricional más estándar. "Con la gracia", continúa Gibert, "que el sistema se configura con contexto de qué gusta y no gusta a la persona, a qué ingredientes no puede acceder porque son demasiado caros, cómo es el estilo de vida donde vive (si el almuerzo es muy largo o corto, si se come deprisa un sandwich o hace 3 comidas fuertes, si toma té o café, etc.) además de saber todas las restricciones de tipo alérgico o alimentario, como por ejemplo, que la persona sea diabética", explica. Sobre esta base de conocimiento se crean los menús descompuestos nutricionalmente. Por último, el nutricionista-dietista revisará la recomendación automática y activará las restricciones que correspondan o ajustará la sugerencia del algoritmo. El proyecto está financiado por el Ministerio de Ciencia, Innovación y Universidades dentro del Programa Retos.

- El robot que hace compañía

La robótica es todavía muy experimental en nuestro país, pero cada día nos acostumbraremos más a tenerla bien visible en la asistencia sanitaria, así como en residencias de ancianos, guarderías y escuelas.

En Cataluña, el robot Pepper se paseó por salas de centros hospitalarios en 2018. Lo incluimos en este apartado social (y no en el de salud) porque las principales funciones de la máquina eran informar o acompañar a pacientes. Se desplaza silenciosamente y tiene aspecto de humanoide, además de interactuar con personas en 21 idiomas. Aún está en fase de desarrollo, pero ya se ha pensado para que explique a las personas mayores como se deben hacer las curas o cómo tratar su enfermedad. Sería también de ayuda para el acompañamiento a menores inmunodeprimidos después de una operación quirúrgica, que residen aislados.

El robot utiliza técnicas de aprendizaje automático, y puede reconocer a personas de manera muy precisa memorizando los rasgos faciales. También puede identificar el estado de ánimo de los pacientes, a partir de la expresión de la cara y del tono de voz. En el desarrollo del proyecto se han unido diferentes centros médicos como el Sant Joan de Déu o el Clínic, el departamento de robótica de la Universidad La Salle y la empresa YASYT[64].

- Gavius, para ser más sociales

Un proyecto similar a MySocialGov, se inicia en 2020 en el Ayuntamiento de Gavà. Gavius[65]es un asistente virtual, que comunicará a la ciudadanía qué ayudas sociales tienen a su alcance, cómo se tramitan, cómo se otorgan y cómo se pueden percibir de forma cómoda, rápida, sencilla y a través del móvil.

[64]YASIT https://yasyt.com/ca

[65]Gaviushttps://www.aoc.cat/2019/1000265639/laoc-collabora-amb-gavius-un-projecte-innovador-en-lambit-dels-ajuts-socials/

Con financiación del fondo europeo Urban InnovationActions -dotado de 4,3 millones de euros- es un proyecto en combinación entre administración (Consorcio AOC, Ay. de Gavà y Mataró), empresa (GFI y EY), ciudadanía (Xnet) e investigación (UPC y CIMNE).

2.4.7 Trabajo

Las innovaciones tecnológicas en el sector del reclutamiento de personal no son ninguna novedad: los primeros formularios en formato test aparecieron en la década de los años 40 del siglo pasado; en los 90 ya se usaban técnicas digitales y, ahora, es el turno de la inteligencia artificial. Y se puede llegar hasta límites no imaginados antes. Por ejemplo, la agencia de recolocación de trabajo finlandesa Digital Minds tiene una veintena de grandes corporaciones como clientes. Al recibir cientos de demandas de candidatos, quiere ahorrar tiempo para seleccionar al mejor trabajador para sus clientes. Así que no hace entrevistas personalizadas: pide la contraseña de correo electrónico (y perfiles en las redes sociales) del aspirante, y un algoritmo de decisión automatizada escudriña toda su información personal (mensajes enviados y recibidos, interacciones, etc.), además de decidir si se merece el puesto de trabajo.

En Cataluña también encontramos aplicaciones de IA en el sector laboral. Afortunadamente estas no son tan intrusivas como el caso finlandés. Ya se utiliza para la automatización de la selección de personal o para predecir las probabilidades que una persona desempleada tiene de encontrar un nuevo trabajo.

- Selección de personal por los gestos de la cara

El Centro de Visión por Computación presentó una demo - en el Mobile World Congress 2019- un recomendador de recursos humanos. Hay un grupo de investigación que a partir del análisis comportamental y gestual de las personas, dicen qué tipo de perfiles ocupacionales, y qué aptitudes tienen más o menos enfatizadas los candidatos.

Nos podemos situar en la escena de una persona que acude a una entrevista de trabajo. El director de recursos humanos la entrevista, pero hay una cámara que está grabando y un software con algoritmos que está captando los rasgos de la personalidad[66]. "El responsable de recursos humanos se basará en su impresión personal pero también en el análisis del sistema inteligente para decidir si es el candidato buscado o no", explica Meritxell Bassolas, Directora de Conocimiento y Transferencia de Tecnología en el CVC. "La máquina puede detectar que tenía una actitud nerviosa, confiada, menos artística, más reflexiva, etc. Y lo hace tanto por los gestos como por las emociones y las microexpresiones faciales", añade Bassolas. La investigadora considera, sin embargo, que nunca tendremos entrevistas únicamente con una cámara y las valoraciones de los algoritmos. "Recordemos que estos sistemas deben ser de apoyo a la decisión del profesional siempre. Los expertos en recursos humanos tienen mucho conocimiento que también deben poner sobre la mesa al reclutar a personas. Saben qué parámetros deben evaluar en función del cargo que buscan o las competencias necesarias. La máquina no llega hasta aquí".

- Los sesgos de las plataformas profesionales

[66]CVC at Mobile World Congress 2019 http://www.cvc.uab.es/outreach/?p=1780

El Grupo de Investigación de Ciencia de la Web y Computación Social de la UPF, junto con la Universidad Técnica de Berlín y el Centro Tecnológico Eurecat, creó un algoritmo que detecta y mitiga sesgos de otros algoritmos. Al sistema lo han llamado FA*IR.[67]"Estudiamos datos de ofertas de trabajo, de reincidencia de presos y rankings de admisión a las universidades para detectar patrones de discriminación en directorios que puedan favorecer o relegar ciertos colectivos, por género, edad o raza", Carlos Castillo, director del Grupo de Investigación de Ciencia de la Web y Computación Social de la UPF.

Una de sus alumnas de doctorado, Meike Zehlike, investigó cómo se clasificaban los hombres y las mujeres en las plataformas de profesionales LinkedIn y Viadeo. "Si hay 100 perfiles de hombres y mujeres igualmente cualificados y en los primeros resultados del buscador sólo aparecen hombres, tenemos un problema".

FA*IR detecta este tipo de discriminación y la corrige incorporando un mecanismo de acción positiva para reorganizar los resultados, y evitar la discriminación sin afectar la validez del resultado.

- Predecir las bajas laborales

A partir de las asistencias o ausencias del personal de un centro hospitalario, un algoritmo puede predecir en cada servicio, en cada perfil laboral, cuántas bajas pueden esperarse cada día. El sistema da una estimación de cuántas personas no se presentarán a su puesto de trabajo.

"No perfilamos trabajadores nunca, lo hacemos por servicio", explica el coordinador del laboratorio de investigación en la Universidad Politécnica de Cataluña (UPC), Ricard Gavaldà. Ahora se hacen contrataciones de personal de salud por días, o fines de semana, y van a urgencias o a planta, pero no tienen la experiencia de haber trabajado regularmente en ese centro. Esto provoca distorsiones, no hay tiempo de formar a los profesionales, no conocen el contexto del hospital, etc. "No se puede predecir el motivo por el cual se tomará la baja pero sí que el servicio se quedará cojo de personal durante ciertos días", argumenta Gavaldà.

- Automatizar para gestionar en tiempo récord

Una de las tareas del Servicio Público de Empleo Catalán (SOC) es la formación de las personas trabajadoras o desocupadas. Cuando gestionan las subvenciones destinadas a esta formación, esto les genera un problema de tiempo y recursos enorme, porque son muchas las entidades públicas y privadas al servicio público que optan a hacer la formación y recibir dinero de la administración.

"Hablamos de cantidades de dinero muy importantes, que pueden llegar a 50 millones de euros, y a los que optan unas 300 o 400 entidades de formación. Cada expediente es un proyecto, y cada proyecto representa 3 o 4 cursos ", explican responsables de la Secretaría Técnica del SOC. "Antes de otorgar las subvenciones, el personal interno debía introducir todas las variables manualmente para evaluar, de acuerdo con unos

[67] M. Zehlike, F. Bonchi, C. Castillo, S. Hajian, M. Megahed y R. Baeza-Yates: FA*IR: A Fair Top-k Ranking Algorithm: https://arxiv.org/pdf/1706.06368.pdf

criterios establecidos previamente. Y así, expediente por expediente ", añaden. Como era un volumen enorme, el proceso tardaba meses antes de que se supiera el resultado de las entidades que harían la formación.

En 2014, se decidió automatizar este trámite, y ahora se hace en sólo unas horas. En este caso, no hay aprendizaje automático pero sí algoritmos de órdenes lógicas que se aplican a una máquina. Hay una secuencia lógica - más o menos compleja, porque afectan muchos parámetros- que resuelve una tarea pesada y complicada.

- Automatizar para planificar mejor

El Centro Tecnológico Eurecat ha diseñado un algoritmo que analiza las ofertas de los portales Infojobs y Feina Activa[68](los más consultados por los empleadores). El objetivo de este proyecto piloto es evidenciar que la inteligencia artificial puede ser muy útil para la clasificación de la oferta formativa, tanto en ámbito territorial como en el sectorial. "Podemos saber las carencias, las coincidencias y también adjuntar las competencias que piden los empresarios con la formación de los departamentos competentes, el Servicio Público de Empleo de Cataluña y el Departamento de Educación", explican desde la Secretaría Técnica del SOC.

Una de las potencialidades es apoyar al orientador. En función de las características de la persona y de su experiencia profesional el algoritmo podría predecir qué probabilidad de inserción tiene. "Por ejemplo, si eres una mujer con estudios universitarios, que ha pasado por diferentes puestos de responsabilidad, se te clasifica en un clúster determinado. Con los datos masivos de toda la población catalana -previamente introducidos-, el algoritmo podrá predecir el porcentaje de probabilidad de encontrar trabajo". A continuación, el orientador del SOC ¬-con otros recursos de orientación profesional- dará las opciones que mejor se adecuen a su perfil para mejorar su empleabilidad. "Es una herramienta más que tiene el orientador, nunca se hará una decisión automatizada para conceder o no formación, o cualquier otro servicio porque se podría caer en discriminaciones", añaden. Apenas se ha estrenado este 2020.

[68] Portal Feina Activa de la Generalitat https://feinaactiva.gencat.cat/web/guest/home

2.4.8 Ciberseguridad

Los ataques informáticos a las empresas -grandes y pequeñas-, así como a gobiernos y organizaciones que manejan grandes volúmenes de información privada o financiera obligan a tener una ciberseguridad cada vez más sofisticada. "Las empresas dedicadas a la ciberseguridad utilizan algoritmos para caracterizar el funcionamiento normal de todos los ordenadores e interacciones que se producen en el entorno de la organización, a partir de información encontrada en diferentes fuentes de datos, los correlacionan y así detectan desviaciones de la normalidad, indicadoras de alguna posible anomalía", explica Manel Medina[69], Catedrático en la Universidad Politécnica de Cataluña y fundador y director del esCERT-UPC, el equipo español de seguridad en red.

Pero el cibercrimen está en expansión y tal como reconocieron la mayoría de los principales ponentes en el Barcelona *Cybersecurity Congress* - celebrado el pasado mes de octubre-, los que quieren hacer daño lo tienen más fácil que los que luchan por la defensa y la protección. "Las ciberamenazas son una tendencia en aumento en todo el mundo, que afecta a todas las industrias", explica el Informe **La ciberseguridad en Cataluña**[70]-elaborado por la Generalitat–.

El motivo principal es la transformación digital que se ha experimentado en todos los sectores de la sociedad en la última década. "Los mismos avances tecnológicos que han impulsado la productividad y la eficiencia de los negocios son los que han hecho a las organizaciones más vulnerables a los ciberataques", añade el informe. Y señala que cada vez que una empresa sufre un ataque se ve obligada "a interrumpir sus operaciones una media de 17 horas al año. Otros efectos negativos pueden ser: el paro completo de las operaciones, la disminución de la facturación; la revelación de información confidencial; las implicaciones legales, la pérdida de calidad de los productos, los daños a la propiedad física e incluso a la vida humana".

Lo peor es que la complejidad de las amenazas aumenta rápidamente y de manera constante. Esto obliga a las empresas a estar en permanente alerta. Manel Medina explica que los sistemas automatizados actuales pueden detectar si te conectas a direcciones IP de países inusuales para las actividades de la organización, o si en algún ordenador de la empresa hay un tráfico de datos que no se corresponde con el habitual, para detectar posibles fugas de información. "El problema viene con los ataques permanentes avanzados (*Advanced Permanente Threats*- APT[71]) –programas espías que se auto instalan en los ordenadores de empresas y quedan latentes durante largos períodos de tiempo. Durante este tiempo, van generando un goteo de datos, con muy poca intensidad, para que los programas de detección de anomalías no perciban desviaciones significativas del comportamiento normal".

Los APT son un conjunto de procesos sigilosos y continuos con la intención de saltarse la seguridad informática de una empresa u organización, normalmente por motivos de

---

[69]Manel Medina (https://inlab.fib.upc.edu/es/persones/manel-medina) https://www.linkedin.com/in/manelmedina/?originalSubdomain=es

[70]*La ciberseguretat a Catalunya* (https://www.accio.gencat.cat/web/.content/bancconeixement/documents/informes_sectorials/ciberseguretat-informe-tecnologic.pdf)

[71] APT (https://en.wikipedia.org/wiki/Advanced_persistent_threat)

negocio o políticos. Y se programan para permanecer activos un periodo largo de tiempo. "Hace 4 años hubo un robo de datos personales en la oficina de gestión de personal del gobierno estadounidense (OPM de USA), que afectó incluso a los que estaban jubilados. Tres meses antes el gobierno se había protegido con un programa de detección y prevención de intrusiones[72]. No se explicaban cómo podía haber pasado, y el motivo era que la fuga de información ya se estaba produciendo cuando el programa de detección "aprendió" lo que se podía considerar un comportamiento normal. "Son sistemas de aprendizaje automático (*machine learning*[73])que aprenden los parámetros normales de un conjunto de datos supuestamente normales y cuando le das una muestra que no se ajusta a esta normalidad, te avisan. Pero si la muestra 'normal' ya contiene tráfico generado por el programa espía, éste es considerado parte de la normalidad, y no generará ningún aviso".

- El nuevo desafío: robo de datos biométricos

Hoy las identificaciones humanas se hacen habitualmente con huellas digitales, escaneo de iris, reconocimiento facial, de voz o ADN. Cada vez es más común que las empresas obliguen a los trabajadores a elegir uno de estos sistemas para asegurarse de que unos no fichan en nombre de otros.

Los datos biométricos son únicos e intransferibles. No hay dos caras iguales, ni dos huellas digitales iguales. Un artículo de la revista Forbes[74]recordaba hace unos meses que "el uso de la tecnología biométrica y autenticación humana supondrá un gran impacto en la sociedad. Si bien son una gran solución para la identificación, plantea problemas serios de seguridad. Los países no están preparados para asegurar los patrones de detección biométricos. Por otra parte, los riesgos para el rendimiento laboral, la precisión, la privacidad, la interoperabilidad y los posibles riesgos para la salud -problemas de visión por los escáneres de retina- han de ser gestionados de manera efectiva", comentaba el periodista.

El director del esCERT-UPC también alerta de estos riesgos de seguridad, que a priori se venden como un avance tecnológico. "Habría que concienciar a las empresas que si recogen datos biométricos, ¡los protejan mucho!".

- Algoritmos para proteger a los clientes

Internet ha cambiado la forma en que se compran bienes y servicios y esto ha provocado una rápida transformación también en la organización interna de las empresas. Si no se invierte en seguridad, el costo ante un ataque informático de gran tamaño puede repercutir en graves pérdidas económicas y de reputación.

---

[72]Intrusion Prevention System (IPS) (https://es.wikipedia.org/wiki/Sistema_de_prevenci%C3%B3n_de_intrusos) https://en.wikipedia.org/wiki/Office_of_Personnel_Management_data_breach
https://www.csoonline.com/article/3130682/the-opm-breach-report-a-long-time-coming.html

[73] Aprendizaje automático (https://ca.wikipedia.org/wiki/Aprenentatge_autom%C3%A0tic)

[74] Hacking OurIdentity: TheEmergingThreatsFromBiometricTechnology (https://www.forbes.com/sites/cognitiveworld/2019/03/09/hacking-our-identity-the-emerging-threats-from-biometric-technology/#353ed3505682)

El riesgo de recibir ataques cibernéticos -cada vez más sofisticados- es muy alto. Desgraciadamente, los pequeños negocios y comercios siempre llegan tarde en la transformación digital y la ciberseguridad sólo la tienen en cuenta después de haber sufrido un ataque. "Los cibercriminales tienen tácticas, técnicas y procedimientos para comprometer a las empresas y monetizar los datos robados", exponen en el informe *Cyberthreat Intelligence for Retail & E-Commerce*[75], publicado por Blueliv (una de las start-ups de ciberseguridad creada en Barcelona, y que ya cuenta con sedes en San Francisco y Londres). "El riesgo nunca ha sido tan alto como ahora: desde campañas de suplantación de identidad (*phishing)*[76]que engañan a los usuarios para compartir información personal y financiera, secuestran cuentas para realizar un fraude, desarrollan y despliegan *crimeware* y *malware* (programas maliciosos que se auto instalan en los ordenadores), sistemas de pago digital y bases de datos", continúa el informe. Los algoritmos de decisión automatizada proporcionan notificaciones en tiempo real de detección de fraude para tarjetas de crédito robadas, pero también para prevenirlo, interceptando tarjetas antes de que se revendan en el mercado negro.

- Aprendizaje automático para detectar ataques

La inteligencia artificial se está utilizando tanto para la defensa como para el ataque. "Es una carrera para ver quién se defiende o ataca mejor", explica Josep Domingo Ferrer[77], director del Centro de Investigación en Ciberseguridad de Cataluña (CYBERCAT) y Profesor Catedrático de la Universidad Rovira i Virgili. "Para detectar posibles amenazas se aplican métodos de aprendizaje profundo, a partir del entrenamiento con datos del pasado. Por ejemplo, a partir de ataques previos, se analizan las características de quien los provocó y se alerta si se dan las mismas condiciones. Lo mismo para encontrar el virus que ha causado un ataque: se analizan casos anteriores. Los algoritmos funcionan muy bien cuando se dispone de grandes bases de datos históricos."

[75]*CyberthreatIntelligenceforRetail&E-Commerce (*https://www.blueliv.com/thanks-ecommerce-retail-whitepaper/)
[76] *Phishing* (https://ca.wikipedia.org/wiki/Pesca_(inform%C3%A0tica)
[77]Josep Domingo Ferrer (http://www.urv.cat/es/universidad/conocer/personas/profesorado-destacado/2/josep-domingo-ferrer)

2.4.9 Comunicación

En el ámbito de la comunicación, hace más de diez años que se aplican los algoritmos, sobre todo con técnicas de visión por computación, por ejemplo, para interpretar la lengua de los signos[78]. Las personas con sordera pueden, así, mantener una conversación y comunicarse con quien no entiende la lengua de los signos, ya que realiza una traducción de signo a palabra en tiempo real.

Más recientemente la IA se ha puesto en práctica en la Wikipedia -para detectar el vandalismo o las entradas incorrectas-, así como para generar noticias en medios o para valorar de qué manera se muestran series y películas en las plataformas de vídeo bajo demanda, tales como Netflix.

Surgen algunos dilemas éticos, líneas rojas que no se deberían traspasar y riesgos de vivir en filtros burbuja para intereses comerciales. También preguntas que no querrían ni plantear: ¿Las máquinas sustituirán a los periodistas? Tecnológicamente ya se puede hacer. O esta otra: ¿Podríamos acabar perdiendo de vista la producción audiovisual local porque un algoritmo sólo me enseña la estadounidense? Tecnológicamente ya se hace.

- El algoritmo periodista

"Una de las áreas más interesantes y con futuro de la inteligencia artificial (IA) es el aprendizaje automático" -explica David Llorente, fundador de una empresa que genera noticias con IA-. "La máquina es súper exacta en la creación de contenidos. No dirá que otro jugador ha marcado el gol, ni se equivocará en los resultados electorales de tal partido, ni en la previsión del tiempo porque bebe de los datos".

La empresa de Llorente trabaja actualmente para 25 medios de comunicación y agencias de noticias españolas. Otro tema -comenta Llorente- es la percepción social. "Es normal que los periodistas se sienten amenazados profesionalmente. Pero aún queda camino por recorrer. No se aceptará que una máquina haga una noticia donde se delaten a ciertas personas de fraude".

El algoritmo de esta empresa redacta noticias sólo de aquellas temáticas que pueden tener datos objetivos, es decir, resultados de electorales, deportivos, datos económicos, loterías, previsión del tiempo, tráfico, etc. "El medio que nos contrata ve como doblamos o triplicamos el volumen de noticias que hacen diariamente, y con mucha precisión. Esto les puede ayudar a vender más suscripciones o a conseguir más publicidad".

David Llorente explica que el sistema inteligente elabora la noticia en tres fases: en la primera, la máquina decide los datos más relevantes para la construcción de la noticia. "En resultados electorales, lo más importante es que se explique quién ha ganado. Pero en un partido de fútbol narrar un gol, cuando han marcado ocho, tal vez no es tan

[78] Creado un sistema visual para interpretar lenguas de signos

https://www.uab.cat/web/noticies/detall-d-una-noticia/creat-un-sistema-visual-per-interpretar-llengues-de-signes-1090226434100.html?noticiaid=1275458325318

importante ", apunta Llorente. En la segunda fase se decide cómo escribir la información, a partir de un volumen de noticias similares la máquina aprende la estructura de la noticia y la forma. Aquí hay una revisión final de periodistas que trabajan en la empresa para validar la narrativa. Cuando ya está entrenado el sistema, toma los datos definitivos, selecciona las frases y redacta la noticia final. La máquina siempre se entrena con la hemeroteca del medio para el que está en funcionamiento, para coger el tono y el estilo de escritura. No sólo se gana en velocidad, volumen de información sino también en precisión y corrección ortográfica.

David Llorente cree que actualmente los periodistas hacen muchas tareas rutinarias, que no son propiamente creativas, que podrían confiarse a los sistemas inteligentes. "Lo que se intenta es que los periodistas dispongan de más tiempo para hacer información mejor elaborada, de contexto, reportajes en profundidad, etc". Sin embargo -según Llorente- se debería evitar que la máquina exprese opinión, aunque tecnológicamente ya se puede hacer. "El tema de la opinión es más delicado. La opinión la generamos los humanos. Se debe tener mucho cuidado. Son líneas rojas para nosotros. Porque enseguida podríamos acusarnos de crear *fakenews*, con titulares engañosos de grandes temas como el cambio climático o temas políticos, que podrían cambiar el sentido de la realidad. El tema ético lo tenemos muy presente. Hemos recibido ofertas para hacer este tipo de trabajos y nos hemos negado rotundamente", concluye David Llorente.

- El maldito filtro burbuja

El Consejo del Audiovisual de Cataluña (CAC) encargó al investigador Carlos Castillo[79]un estudio para detectar cómo el algoritmo presentaba vídeos en plataformas como Netflix, OrangeTV y otros. Europa obliga a que haya un 30% de producción audiovisual europea en cada plataforma. "Pero una cosa es que exista esta producción audiovisual en el catálogo y, otra, que el algoritmo la acabe mostrando al usuario", explica Castillo. "La nueva revisión de la Directiva de Servicios de Comunicación Audiovisual (DSCAV) explicita que garanticen esa cantidad de producción europea a los usuarios en sus catálogos. Si el algoritmo te lo ha mostrado los primeros días y no la has seleccionado, puede que ya no te la muestra nunca más porque interpreta que no te interesa".

Lo primero de que tenemos que ser conscientes -explica Castillo- es que los usuarios tenemos la sensación de que en una plataforma de vídeo bajo demanda estamos eligiendo la serie o película que queremos ver, "pero no es así en absoluto. Quizás tiene delante 40 títulos pero esta es una parte muy ínfima del catálogo total. Y el usuario no decide el criterio por el que se han seleccionado aquellos primeros 40 títulos".

En el informe que encargó el CAC se habla de los riesgos a la autonomía y a la diversidad, la burbuja informativa, vivir en un filtro burbuja. "Si sólo ves películas de

[79]Carlos Castillo, Profesor Investigador Distinguido en el Departamento de Tecnologías de la Información y Comunicación de la Universidad Pompeu Fabra. Informe para el CAC: "La oferta y la demanda del contenido audiovisual en la Era de los Datos Masivos" (2018" (2018) https://chato.cl/papers/castillo_2018_oferta_disponiblidad_contenido_audiovisual-ES.pdf

acción, no te muestra más que eso. Por lo tanto, aquí no hay diversidad. Pero también hay riesgo de autonomía porque deberías poder explorar el catálogo entero. Ahora sólo hay una cajita de buscar, que es imposible que tú accedas al volumen total. Podrían haber otros diseños de interfaz que, por ejemplo, hablando pudieras realizar la búsqueda. De igual manera, podrían dar la opción de decidir que el catálogo esté ordenado o desordenado".

- IA para la Wikipedia en catalán

Una de las preocupaciones más importantes de los proyectos abiertos Wikimedia es la revisión de contribuciones potencialmente perjudiciales ("ediciones"). Así como las contribuciones que, sin querer, contienen errores. Desde el 2018, Wikipedia utiliza aprendizaje automático para hacer las revisiones de artículos a través de ORES[80], una herramienta que "marca las acciones que son potencialmente vandalismos, por ejemplo. De esta manera, los voluntarios que revisan las modificaciones recientes en Wikipedia pueden hacer su trabajo más fácilmente", exponen en su página web. "Para poder habilitar ORES, la comunidad de editores ha de enseñar al sistema, etiquetando ediciones ya realizadas, indicando si son correcciones ortográficas, gramaticales, actualizaciones de contenidos o -dado el caso- vandalismo".

[80] ORES- https://ca.m.wikipedia.org/wiki/Viquiprojecte:Viquirepte/ORES

2.4.10 Visión por Computador

La Visión por Computador (VC) es una rama de la inteligencia artificial que hace que los algoritmos entiendan una imagen igual que lo hacemos las personas. Ya pueden distinguir objetos y personas, así como describirlos o decir qué hacen estas personas.

La VC no es un terreno nuevo. "Hace más de 40 años que se empezó a investigar. En los años 70 del siglo pasado, el científico y padre de la inteligencia artificial Marvin Minsky[81]experimentó con ella", explica Petia Radeva, la directora del Grupo de Investigación *Machine Learning and Computer Vision* de la Universidad de Barcelona (UB). "Hace unos 30 años que la VC entró en las fábricas de automoción para ensamblar coches y otras funciones en entornos controlados. La novedad es el aprendizaje automático porque es capaz de resolver problemas más complicados, como podría ser la conducción de coches autónomos", añade Radeva. "Ahora la investigación va veinte veces más rápida que hace una década".

Hemos decidido dedicar una sección a esta tecnología porque -aunque los ejemplos que se exponen podrían haberse ubicado dentro del ámbito de la salud, del social o del comercial- es bueno diferenciar el potencial del aprendizaje profundo en la VC.

¿Dónde encontramos Visión por Computador?

En todas partes ya. En los parkings, cuando una máquina lee la matrícula del coche y se abre la barrera de salida. En el reconocimiento facial, el del iris, o el de la huella digital que utilizan los sistemas de seguridad de aeropuertos y de la policía, pero también nuestros móviles en sustitución de una contraseña.

La VC está en los porteros automáticos de edificios, de empresas y de algunas casas. En algunas escuelas e institutos han comenzado a aplicarlo para el control de los alumnos en el centro. Los bancos lo ofrecen a sus clientes para acceder a la cuenta corriente desde un cajero. En los hospitales la VC se encuentra en el análisis de todo tipo de imágenes del cuerpo: tomografías, radiografías, densitometrías, resonancias, ecografías, mamografías, etc. para detectar y predecir enfermedades.

Se utiliza también para contabilizar personas -como el número de manifestantes en un espacio público-, o en acciones de deportistas, como en las veces que cada jugador tiene posesión del balón en un partido, los ataques, las faltas, etc. Estos algoritmos no son propiamente de decisión automatizada, sino que calculan, estiman y cuantifican. "Los humanos somos mejores en cualificar, las máquinas en cuantificar", explica la investigadora de la UB.

En los móviles es útil para añadir información turística cuando se fotografía un monumento, sin importar el ángulo o el tipo de luz del momento. También se utiliza para el reconocimiento de documentos digitalizados de siglos pasados o para decir cuál es el objeto encontrado en un yacimiento arqueológico. Y la VC está presente en los robots de la planta automatizada que Amazon tiene en Cataluña, y en los que pasean por pasillos de hospitales (todavía en pruebas) para llevar comida a los

[81]Marvin Minsky https://ca.wikipedia.org/wiki/Marvin_Minsky

enfermos o hacerles compañía. Y en la edición de 2019 del Mobile World Congress[82]de Barcelona se entraba -literalmente- 'por la cara'. En lugar de mostrar la acreditación, las barreras se podían abrir también con el reconocimiento facial.

Un último ejemplo: gracias al reconocimiento del iris con VC, el fotógrafo Steve McCurry localizó Sharbat Gula, la mujer afgana fotografiada en 1985, cuando sólo tenía 12 años y huía de su país en guerra. Su cara se hizo mundialmente famosa al ser publicada en la portada del National Geographic[83].

A continuación detallamos algunos ejemplos que se encuentran en las fábricas o algunas industrias catalanas y que son fruto de la colaboración de diferentes centros de investigación.

- Detección de pizzas defectuosas

El Centro de Visión por Computación (CVC) de la Universidad Autónoma de Barcelona (UAB) hace más de 15 años que ayuda a las empresas catalanas a desarrollar proyectos digitales y procesos industriales. "Cuando hablamos de inteligencia artificial pensamos en la movilidad autónoma y una toma de decisión de los algoritmos a gran escala. Pero existen muchos otros proyectos interesantes, a una escala menor que están impulsando grandes avances", expone Mertixell Bassolas, la Directora de Conocimiento y Transferencia de Tecnología en el CVC.

En el ámbito de la alimentación - y con aprendizaje profundo- podemos detectar si la estética de las pizzas es correcta no, si el reparto de los ingredientes en la base de pizza es el establecido o detectar materiales intrusivos que hayan podido entrar en el envasado", comenta Bassolas. "Para lograr esto, el algoritmo necesita entrenarse con muchas imágenes, millones si puede ser. Cuantas más imágenes, más acertado será el resultado. Y a poder ser de pizzas en buen estado y en mal estado, para que aprenda a diferenciarlas. Y necesita anotaciones para distinguir qué es cada ingrediente".

Según Bassolas, el reto es conseguir lo que queremos pero con el mínimo esfuerzo inicial por nuestra parte", expone la directora de Conocimiento y Transferencia de Tecnología en el CVC. Con videojuegos se generan imágenes sintéticas de pizzas virtuales, o de coches autónomos, tantas como necesita el algoritmo para ser entrenado. "Pero esto también da mucho trabajo", añade Bassolas. "Y al final, conseguimos que un sistema inteligente con muy poca información sea capaz de resolver lo que queremos. Y además, que no se olvide de lo que ha aprendido, para que pueda ser funcional en otro caso sin empezar de cero".

- Seguimiento del engorde del cerdo

Sectores como el agroalimentario se están digitalizando poco a poco. Procesos que hasta ahora se hacían de manera muy manual son automatizados para extraer nuevas

[82]«Comença el Mobile WorldCongressamb la mirada posada en el 5G». Febrer, 2019. Betevé. https://beteve.cat/economia/comenca-mobile-world-congress-2019-5g/

1 [83]Sharbat Gula https://www.nationalgeographic.com/news/2017/12/afghan-girl-home-afghanistan/

conclusiones que ayuden a rentabilizar más los recursos o mejorar los productos. En el caso de las granjas de animales ya se monitoriza todo el proceso: desde el engorde hasta la llegada del animal en el matadero. "Controlamos el cerdo durante toda la cadena de valor y esto permite comparar qué tipo de comida es la mejor, sea para acortar tiempo o para incrementar la calidad del producto al final", dice Meritxell Bassolas, Directora de Conocimiento y Transferencia de Tecnología en el CVC .

El Centro de Visión por Computación hizo una herramienta para detectar los cerdos en el matadero. "Estamos aplicando redes neuronales que hoy ya son capaces de detectar cualquier movimiento en una escena". En las granjas de engorde también se implementa el mismo sistema inteligente para seguir la evolución del animal, los pesos, los volúmenes, el estado del cerdo y valorar cómo impacta en la calidad del producto.

Anteriormente este sistema se había utilizado para clasificar residuos de manera selectiva en una planta de basura, o para leer los dígitos de los contadores del agua.

- La complejidad de los coches autónomos

Hemos oído hablar mucho de las maravillas que harán los coches autónomos, pero también hay mucha incertidumbre respecto a las capacidades de las máquinas en cualquier imprevisto. ¿De quién será la responsabilidad en caso de un accidente?

Entrenar a los algoritmos para que se apliquen en la movilidad es uno de los retos más complicados que se han planteado los investigadores de Inteligencia artificial hasta ahora. Porque la máquina debe saber todo lo que hay en la escena mientras circula. Lo debe tener previamente identificado, y reconocerlo. Desde objetos estáticos (edificios o mobiliario urbano) hasta todo lo que se mueve: personas y animales, pero también la hierba, una hoja de papel, una estrella o cualquier objeto que salga volando o lo tire alguien, como una pelota. Y, por último, el coche también debe saber cómo reaccionar en cada situación. "Es inviable entrenar a los algoritmos en un entorno real", puntualiza Bassolas. "Por lo tanto, generamos videojuegos y los sistemas inteligentes se entrenan a partir de situaciones virtuales".

Pero después de haber sido entrenada en un entorno virtual, ¿se adaptará a las calles de una ciudad o una carretera real? Hasta hace muy poco no había respuesta a esta pregunta, expone Bassolas. Ahora ya se habla del *Domain adaptation*[84], es decir, la capacidad que el algoritmo tiene de cambiar de entorno y adaptarse. "En un proyecto internacional, hemos creado un súper entorno virtual en código abierto, en colaboración con Intel. Empresas como Toyota se han mostrado interesadas. Ahora hay muchas otras marcas que aprovechan lo que hemos hecho para adaptar diferentes partes del proceso".

- Reconocimiento facial

Uno de los grandes problemas que ha dado hasta ahora el reconocimiento facial es que los algoritmos han sido entrenados con pocos datos de un determinado tipo de

[84] *DomainAdaptation* https://en.wikipedia.org/wiki/Domain_adaptation

personas. Por ejemplo, si se han puesto en práctica en Europa, tal vez les han dado menos imágenes de cara o de cuerpo de personas negras, asiáticas o hispanas. Si después este sistema inteligente debe adaptarse a una situación concreta, puede dar problemas como el del secador de manos de un lavabo público que sólo secaba las de piel blanca porque las negras no las reconocía.

En los últimos años se han detectado y denunciado ampliamente los sesgos de los algoritmos de reconocimiento facial porque discriminan, especialmente las caras no blancas y las femeninas. En 2015, Google Photos clasificó a algunas personas negras como 'gorilas[85]'. La solución que dio es eliminar ciertas etiquetas, como el término 'gorila', en lugar de arreglar el algoritmo, con el que continúa sesgado. También, en 2018, el programa Rekognition de Amazon identificó erróneamente a 28 de los 435 miembros del congreso de Estados Unidos como delincuentes[86].

En Cataluña, en el ámbito judicial se aplica el reconocimiento facial tanto para reconocer un posible criminal como para saber si el acusado está mintiendo, en situaciones de custodia de menores. "Si la persona miente, el sistema lo detecta a partir de las microexpresiones faciales", comenta Bassolas. La pregunta es... ¿el algoritmo de decisión automatizada puede fallar? Y la directora de Conocimiento y Transferencia de Tecnología del Centro de Visión por Computación niega con la cabeza. "Hace unos diez años que se utiliza y se perfecciona esta técnica".

- Fotografías satelitales

Las imágenes satelitales pueden ayudar a resolver un montón de problemas socioeconómicos de un país o población. Desde ahorrar costes y recursos en la producción y distribución de alimentos, hasta cuestiones de energía, para entender qué se ha cultivado en un país en un año, temas de polución, desastres naturales, inundaciones, etc. "El problema es que la tecnología empleada hasta ahora tiene un coste muy elevado. Si quiero tener una foto satelital de alta resolución de toda Cataluña, actualizada cada mes, me podría costar unos 400 millones de dólares por foto", explica Marco Bressan, científico de datos y experto en imagen satelital.

"Para que realmente sean útiles debe haber una frecuencia de captura, en tiempo real. Por ejemplo, si hay un tifón en Bangladesh, quiero entender el impacto en los cultivos de arroz, tener una información prácticamente diaria del efecto, cuántas hectáreas han sido anegadas, qué tipo de medidas debo tomar para ayudar a la población, etc. Pero para obtener todos estos datos, debo enviar no un satélite, sino cientos. Pero a 400 millones de dólares el satélite es inviable", continúa Bressan. Después de ciertas investigaciones, la empresa que lidera ha conseguido mantener la resolución de la imagen por menos de un millón de dólares.

---

[85]«Google apologizes after Photos app tags twoblackpeople as a gorillas» https://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas

[86]«MIT researchers: Amazon's Rekognition shows gender and ethnic bias (updated)» https://venturebeat.com/2019/01/24/amazon-rekognition-bias-mit/

El objetivo de la empresa es conseguir una imagen de la Tierra cada semana. Y, más adelante, una al día. Actualmente, estas fotografías son útiles para el sector agrícola, forestal, de seguros o para el sector energético. Y se aplica sobre todo para gestionar infraestructuras. "Si tienes un gasoducto que cruza Siberia o La Patagonia, verificar el estado de la instalación por temas regulatorios, de seguridad, o porque se ha desbordado un río, han caído árboles, etc. es muy complicado y supone un coste enorme. Pero tener una suscripción a imágenes satelitales de la zona de interés, puede reducir mucho el gasto ", explica Marco Bressan.

"Los algoritmos analizan en tiempo real todo lo que dicen las imágenes. Si hay coches afectados por una inundación, ¿cuántos? En tiempo real, se puede saber cuántas hectáreas hay sembradas en un país. Por ejemplo, si es un productor de alimentos y depende de esta materia prima para la producción, puede saber el inventario. Esto era imposible antes. Antes si había una plaga que mataba el cereal, te quedabas pagando una barbaridad. Y ahora se puede predecir, etc.".

Los gobiernos ya utilizan este tipo de imágenes para el seguimiento de movimiento y estimación de los campos de refugiados. O en defensa, para planificación de ataques, ataques con drones.

2.5 La ética de la inteligencia artificial

El interés actual por la inteligencia artificial (IA) no tiene precedentes. Muy pronto será imprescindible para la sociedad porque aporta mucha eficiencia, conocimiento y creatividad. El aprendizaje automático (machine *learning*) -y, en particular, el aprendizaje profundo (*deep learning*) - han permitido avanzar como no se podría haber imaginado una década atrás.

Pero también hay que pensar en el impacto que tendrá sobre las personas. Es el momento de hacerse preguntas sobre el significado de lo que está bien o mal, sobre la relación entre poder y abuso, o entre sesgo y distorsión. La Guía Ética de la Comisión Europea para una IA fiable[87], la Declaración de Barcelona [88] y la Estrategia IA para Cataluña [89] –presentada este 2019 y que incorpora la propuesta de creación de un Observatorio Ético- son algunos ejemplos de la preocupación por asegurar los derechos básicos ante este avance tecnológico.

Los expertos consultados por este informe han sido preguntados por las cuestiones éticas. Todos insisten en encontrar la manera de explicar cómo un algoritmo ha tomado una decisión o predicho una situación, además de tener la posibilidad de cuestionar el razonamiento del sistema automatizado, y plantean el debate de la 'responsabilidad' en el caso de un mal funcionamiento o discriminación del algoritmo. Igualmente señalan la urgencia de formar a la ciudadanía -desde la clase política hasta los educadores, pasando por los padres y madres y la opinión pública en general. Formarlos y o concienciarlos para que sepan cuando deben reclamar sus derechos en caso de que un sistema automatizado reduzca (de manera parcial o total) su libertad individual.

[87] EU artificial intelligence ethics checklist ready for testing as new policy recommendations are published https://ec.europa.eu/digital-single-market/en/news/eu-artificial-intelligence-ethics-checklist-ready-testing-new-policy-recommendations-are

2 [88] Barcelona Declaration for the proper development and usage of artificial intelligence in Europe (2017) https://www.iiia.csic.es/barcelonadeclaration/

[89] Estrategia IA per Catalunya. Conselleria de Polítiques Digitals i Administració Pública, de la Generalitat de Cataluña (2019). (https://participa.gencat.cat/uploads/decidim/attachment/file/818/Document-Bases-Estrategia-IA-Catalunya.pdf)

Catedrática de ética y filosofía del derecho moral y político de la Universidad Autónoma de Barcelona (UAB).

VICTÒRIA CAMPS[90]: " Con la inteligencia artificial se ha perdido la privacidad "

¿Qué es la ética? y ¿Por qué es tan importante ocuparse de ella en la aplicación de la inteligencia artificial?

"La ética se podría definir como un conjunto de principios, normas o valores que orientan nuestra conducta y que no pueden ser anulados por los demás. Por ejemplo, el respeto a la dignidad es un valor que no puede ser eliminado para que sea beneficioso para otro", explica Victoria Camps, Catedrática de Ética y Filosofía del Derecho Moral y Político de la UAB.

"Las religiones no nos sirven para definir la ética porque es cierto que tenemos los Diez Mandamientos en la cristiana, que nos dicen: 'No matarás' o 'No robarás', pero se podría cuestionar el de: 'No desearás a la mujer del otro'. La ética va más allá, incluye una exigencia de universalidad -como la no discriminación o el respeto hacia las otras personas-".

"La ética tiene obligatoriedad, pero no tiene la coacción del derecho -que obliga y si no cumples, te penaliza, comenta Camp." La norma ética obliga en conciencia", añade." Con la inteligencia artificial hay muy poca conciencia de los peligros. Se ha perdido lo que antes considerábamos un bien preciado: la privacidad. Los jóvenes (pero también personas de todas las edades) no tienen ningún tipo de pudor y lo dan todo: fotos, datos, reconocimiento facial y luego salen las violaciones filmadas, acciones de *bulling* en las escuelas, etc.".

La equidad no la resuelve sólo el algoritmo

"Lo más importante en ética es la equidad", continúa Victoria Camps. "Un sistema equitativo significa favorecer a los que están en peor situación. El Estado tiene la obligación de proteger a las personas y su equidad. ¿Esto lo puede hacer un algoritmo? Sí, podría favorecer a quien han desahuciado, a los que no tienen trabajo y viven con hijos menores a su cargo. Esto podría acreditar que se le concediera una ayuda social. Pero cada caso es individual, y esto también se debe tener en cuenta cuando se programa el algoritmo. Porque podría haber quien se quejara, y la administración pública debería tener respuestas para el uso de la inteligencia artificial. La equidad no la resuelve sólo el algoritmo".

"La IA no puede obviar las preguntas básicas de la ética que hacen referencia a la privacidad, a la confidencialidad. Al informático que programa el algoritmo no le puedes pedir que piense si se utilizará bien o mal. La pregunta se debe hacer a quien tiene la responsabilidad máxima, a quien encarga y decide lo que debe hacer el algoritmo. Formo parte del Comité de Bioética de Cataluña [91]. Hace unos años se discutió mucho en el Parlamento sobre si comercializar los datos de salud de las personas con el programa Visc++. Finalmente, se aprobó con reservas y enmiendas. La

[90] Victòria Camps (https://www.uab.cat/web/el-departament/victoria-camps-cervera-1260171817458.html)
[91] Comité Bioética de Cataluña (http://canalsalut.gencat.cat/ca/sistema-de-salut/comite-de-bioetica-de-catalunya/)

pregunta a hacerse es: ¿tener compartidos los historiales clínicos es un avance? Obviamente que sí. Pero sin perder el control del uso que se pueda hacer de estos datos". Y ¿qué podría ser hacer un mal uso? "Pues que se supiera que un político tiene una enfermedad no explicada. La persona tiene derecho a que sus datos de salud sean privados y si da el consentimiento de usarlos, sólo sea para aquello que dé servicio a la ciencia y la medicina ".

La responsabilidad del algoritmo es un tema recurrente, sobre todo imaginando posibles accidentes en los coches autónomos (sin conductor) que tienen que llegar. "Los problemas éticos en este caso, son de manual. Olvídate de la IA. Piensa en el conductor de un tren: detecta que hay un suicida en la vía. ¿Qué hace? ¿Una maniobra para esquivarlo poniendo en peligro la vida de muchos más pasajeros? Desde el punto de vista ético, no hay una respuesta única. Siempre hay que procurar hacer el mal menor. Y con la inteligencia artificial se debe hacer lo mismo".

Diferencia entre el interés comercial y la ética

Victoria Camps considera que hay que marcar una línea entre lo que puede ser interés comercial y lo que puede ser ético. ¿Es ético que una compañía, siga nuestro rastro de navegación y nos bombardee con publicidad en el buzón de correo, en el móvil, en las webs que consultamos? ¿Es ético que los bancos sepan todo de nosotros, hábitos de consumo, gastos y preferencias de ocio y nos quieran fidelizar con ofertas de productos, en función del perfil que han hecho de nosotros? ¿Es ético que la habitación de un hotel me cueste más dinero a mí que a mi vecino porque saben en qué barrio vivo y pueden interpretar mi renta?

"Vivimos en una sociedad de consumo, basada en la oferta y la demanda", responde la Catedrática de Ética de la UAB. "Y si compras una aspiradora, te puede incomodar toda la información o publicidad que te enviarán en los próximos meses, pero es el mundo de los negocios. No lo aceptes. Otra cosa es que con tus datos te quieran perjudicar. Como, por ejemplo, que el banco sepa que eres alcohólico y que se lo diga a la aseguradora del coche para que te discrimine en el pago".

La confianza no debe perderse

"La vida en común necesita la confianza hacia el otro. No hay que perder nunca", comenta Camps. "Tú renuncias a parte de tus libertades a cambio de que unos representantes públicos que has elegido te protejan. Pero si confías tu dinero a unos políticos y luego se destapan casos de corrupción, pierdes la confianza".

Sobre la confianza que ponemos en redes sociales como Facebook, y que nos fallan constantemente porque el propietario vende los datos personales a consultoras (Cambridge Analytica) que aprovecha para cambiar la orientación de voto en unas elecciones (EE. UU., Brexit, Brasil, etc.), la Catedrática de Ética y Filosofía argumenta que "estas plataformas son totalmente prescindibles. No es el algoritmo de Twitter que te muestra la vida en una burbuja, eres tú, que aceptas esta situación. Lo que es una lástima es que los políticos se comuniquen con la población a través de Twitter porque es un deterioro de la política". Insiste en que antes de los algoritmos nos pasaba lo mismo, eligiendo el medio de comunicación con el que nos informamos.

"Pero esto no tiene mucho que ver con la ética. Lo que éticamente importa es preservar la libertad individual, que nadie te acabe diciendo como debes vivir".

Relator de Inteligencia Artificial y Privacidad del Consejo de Europa. Profesor de derecho de la Universidad de Turín.

ALESSANDRO MANTELERO[92]: " No hay una ética global que se pueda aplicar a todos los países "

"La ética y la inteligencia artificial (IA) es un debate muy vigente en Europa ", explica el autor de informes sobre regulación del big data y la IA[93]. Alessandro Mantelero quiere puntualizar que, cuando hablamos de ética, se debe dejar claro a que nos referimos. "La ética se puede considerar como un nivel adicional y complementario del nivel de protección de la ley. Ética y derecho están conectados, pero son diferentes. Si la ética se transforma en ley, tiene sanciones y ya no es ética", puntualiza. Y continúa: "Cuando hablamos de protección de datos, no hablamos de riesgos y beneficios. Si no de derechos fundamentales. Y hay que tener en cuenta que la ética en España no es el mismo que en Rusia". Y entonces, ¿qué prevalece? ¿Es posible una ética global?

No hay una respuesta concreta: "La ética nunca puede ser global, corresponde a un entorno, a una comunidad. Por lo tanto, es muy difícil que se llegue a una ética europea. Tampoco será el mismo en función de las aplicaciones de la inteligencia artificial. La ética por un algoritmo que actúe sobre enfermos no será la misma que para un robot asistencial de personas mayores. Ni por aquel robot que cuide de un niño y que deberá transmitirle ciertos valores de vida. Además, en cada familia predomina valores diferentes ".

Ante este panorama complejo, Mantelero concluye que todos los documentos e informes que se están redactando actualmente no definirán un marco único de la ética. "Estamos al principio. Se habla mucho de ética, pero poco de sociedad. Si digo que invertiré mucho dinero para predecir y controlar el crimen, a partir de una aplicación policial con IA, la pregunta ética y social podría ser: ¿Y por qué no invertir ese dinero para crear escuelas, con un buen sistema educativo, que intente reducir la criminalidad y la diferencia social? Así se ataca el origen del problema ".

---

[92] Alessandro Mantelero (http://staff.polito.it/alessandro.mantelero/)

[93] Regulating big data. The guidelines of the Council of Europe in the context of the European data protection framework
http://isiarticles.com/bundles/Article/pre/pdf/106203.pdf
Artificial Intelligence and Data Protection: Challenges and Possible Remedies
https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6
También el *High Level Expert Group on Artificial Intelligence* de la Comissió Europea avanza en estas cuestiones. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

Profesora del departamento de Medicina y subdirectora del Observatorio de Bioética y Derecho de la Universidad de Barcelona.

ITZIAR DE LECUONA[94]: " Hay que moderar el entusiasmo tecnológico que vivimos "

La experta en los aspectos éticos y legales de la biomedicina alerta que, en el ámbito de la investigación biomédica, buena parte de los proyectos de investigación que se evalúan están basados en la convergencia de tecnologías, entre las que se encuentran la inteligente inteligencia artificial y la aplicación de la analítica de datos masivos.

"La inteligencia artificial ya está aquí, y todos estamos contribuyendo con nuestros conjuntos de datos personales a su desarrollo", explica. "Hoy es posible explotar datos a escala masiva, para mejorar la toma de decisiones mediante el desarrollo de algoritmos, que permitan -correlacionando los datos- determinar patrones de comportamiento y predecir conductas. Y esto lo quieren hacer todos: desde la iniciativa privada hasta el sistema público de salud. Para tener sistemas de salud más eficientes o para que la ciudadanía se beneficie de una medicina personalizada. No hacerlo sería gravísimo desde el punto de vista ético ", añade De Lecuona.

El papel del ciudadano

Llevamos la inteligencia artificial integrada en los aparatos, como es el caso de los móviles. Emitimos datos continuamente desde los diferentes dispositivos digitales que utilizamos. También, existen sistemas más complejos en los que se explotan nuestros datos personales con varios objetivos.

"El problema es que se tomen decisiones sobre mí, mediante datos que son míos pero que yo nunca sabré quien las tiene, ni para qué se están utilizando. Si la ética trata de la felicidad y de sociedades más libres, la aplicación de la IA nos debe hacer reflexionar sobre el papel que juegan los individuos, como titulares de estos datos personales. ¿quién tiene el control y quién debería tenerlo? Se trata de repensar la capacidad de controlar nuestros datos y nuestra intimidad en la sociedad digital ", argumenta Itziar de Lecuona.

Hay que fomentar la alfabetización digital, para poder tomar decisiones libres e informadas como ciudadanos. En este sentido, la experta en aspectos éticos señala que son numerosos los profesionales y empresas que, en tanto que terceros, colaboran en la investigación científica. Por lo tanto, tienen responsabilidades sobre la programación de los algoritmos y el tratamiento de los datos confidenciales. Son, en la mayoría de los casos, profesionales que tienen un conocimiento experto, pero no han recibido formación en ética.

De Lecuona propone revisar qué perfiles -como los científicos de datos- se deberían incorporar en la evaluación de la investigación biomédica basada en inteligencia artificial por parte de los comités de ética de la investigación. Así como analizar qué conocimientos y formación en ética y protección de la confidencialidad de los datos personales sería necesario que recibieran todos los agentes que trabajan en IA, para

[94] Itziar de Lecuona (http://www.bioeticayderecho.ub.edu/es/itziar-de-lecuona)

asegurar que los principios de autonomía, beneficencia, justicia y explicabilidad[95] se cumplan.

Pereza tecnológica

La subdirectora del Observatorio de Bioética y Derecho de la UB destaca que en Cataluña tenemos unas bases de datos de salud de calidad, tales como la Historia Clínica Compartida [96](HC3) o el SIDIAP[97]. El reto ahora es que la información que contienen sea interoperable y se pueda reutilizar. Pero De Lecuona hace énfasis en que "hay que moderar el entusiasmo tecnológico que tenemos y evitar una confianza ciega en la IA en salud. La decisión final -especialmente en el ámbito biomédico- debe permanecer en el profesional, no en la máquina. los algoritmos per se no pueden tener el poder de decisión ".

Desde el punto de vista individual, Itziar de Lecuona propone quitarse la pereza tecnológica que arrastramos y que nos informamos sobre cómo actúan los sistemas automatizados que afectan a nuestra vida cotidiana. Hay que evitar que el algoritmo contribuya a fomentar las desigualdades y las discriminaciones, o perpetuar la foto. "No podemos seguir pensando que nosotros y nuestros conjuntos de datos no son importantes, y que nos da igual quien pueda acceder. Somos relevantes, de vez en cuando contribuimos a generar ontologías[98] para diseñar y perfeccionar la inteligencia artificial mediante nuestros conjuntos de datos. Debemos ser cuidadosos y promover una cultura de respeto de la intimidad. En la sociedad digital, proteger los datos personales es proteger a las personas ".

[95] Principios establecidos por el High Level Expert Group sobre Inteligencia Artificial de la Comisión Europea https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

[96] Historia Clínica Compartida http://salutweb.gencat.cat/ca/ambits_actuacio/linies_dactuacio/tecnologies_informacio_i_comunicacio/historia_clinica_compartida/

[97] SIDIAP https://www.sidiap.org/index.php/es

[98] Ontología https://ca.wikipedia.org/wiki/Ontologia_(tecnologia_de_la_informaci%C3%B3)

Profesor investigador del Consejo Superior de Investigaciones Científicas (CSCI) y ex director del Instituto de Investigación de Inteligencia Artificial (IIIA).

RAMÓN LÓPEZ DE MÁNTARAS[99]: " Se debería poder certificar que los algoritmos están limpios de sesgos "

Ramon López de Mántaras coordinó la elaboración de la Estrategia Española para la Inteligencia Artificial [100] – hecha pública este mes de marzo-, un paso importantísimo para definir las normas éticas que corresponden a su implementación. Igualmente, promovió la Declaración de Barcelona [101], redactada hace dos años por la comunidad científica y que marca muy bien las directrices éticas que deberían seguirse.

Uno de los puntos prioritarios que la Declaración de Barcelona menciona es el de la rendición de cuentas (*accountability*). Es decir, que cuando un algoritmo tome decisiones las personas afectadas puedan recibir - en términos que se entiendan- una explicación de por qué las ha tomado y permitan ser cuestionadas con argumentos razonados. Pero dos años más tarde, este es un aspecto pendiente de resolver que preocupa mucho a los investigadores.

" Como consumidor, ciudadano, tienes derecho a pedir explicaciones [102], con la Ley de Protección de Datos europea. Puedes preguntar por qué te han denegado una ayuda, o no te han dado un crédito. El código del algoritmo no te lo darán, pero debería poder exigir que una entidad neutral evaluara si el algoritmo es justo. No existe una entidad similar en Cataluña. Algorithm Watch [103]o Ethical Tech Society[104] pueden hacer esta función en ámbito internacional.

" Estaría bien que la Generalitat tuviera un organismo que certificara si el algoritmo tiene un sesgo o no. Esto ya se hace con los alimentos y los medicamentos, pero no con la IA. Se debería crear un sello de justicia (*fairness*). Tampoco habría que certificar todos los algoritmos, pero sí donde hay una decisión automatizada que puede perjudicar significativamente a las personas. Esto tiene un coste elevado para la administración. Pero mira, los coches bien que pasan la ITV. Si se hace por otras cosas...".

Perpetuar sesgos

"A menudo sucede que los datos masivos con las que entrenas el algoritmo están sesgados, porque son datos del pasado, y en utilizarlas de nuevo, perpetuas y amplifican sesgos del pasado. Pero los datos no te las puedes inventar. Las tienes que coger de algún sitio".

---

[99] Ramon López de Mántaras (http://www.iiia.csic.es/staff/ramon-l%C3%B3pez-de-m%C3%A1ntaras)

[100] Estrategia española de I+D+I en Inteligencia Artificial (http://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf)

[101] Barcelona Declaration for the proper development and usage of artificial intelligence in Europe (2017) https://www.iiia.csic.es/barcelonadeclaration/

[102] "La ética en la inteligencia artificial", de Ramón López de Mántaras https://www.investigacionyciencia.es/revistas/investigacion-y-ciencia/el-multiverso-cuntico-711/tica-en-la-inteligencia-artificial-15492

[103] Algorithm Watch (https://algorithmwatch.org/en/)

[104] Lorena Jaume-Palusí (https://twitter.com/lopalasi)

¿Como detectarlos? "Si al eliminar de un algoritmo la información de raza, la decisión cambia, ya tienes detectado un sesgo. El 100% de corrección no existe, pero sólo que se puedan sacar los sesgos más importantes ya sería un avance. Como el dispensador de jabón que no quita nada cuando se le acerca una mano negra, porque el sensor sólo funciona con la piel blanca. Fatal!. Lo mismo con el reconocimiento facial que confunde las caras negras con chimpancés ".

"El énfasis en la importancia de los aspectos éticos es la principal diferencia que Europa puede abanderar respecto a lo que se está haciendo en IA en Estados Unidos y China. Por ello, la UE exige a cada país que tenga una estrategia nacional para su aplicación y que defina muy bien estos criterios mínimos ".

Doctora en informática y profesora de investigación en el Instituto de Robótica e Informática Industrial (CSIC-UPC)

CARME TORRAS[105]: " Hay que formar en criterios éticos a quien desarrolla la tecnología "

"Hay actualmente una gran tendencia a trabajar con aprendizaje profundo (*deep learning*), un tipo de red neuronal con muchas capas que lo que hace es asociar entradas con salidas. Es un tipo de aprendizaje llamado de 'caja negra', porque no se sabe con exactitud qué pasa en este proceso de asociación. Si los datos tienen sesgos, el algoritmo aprende mal. Pero quien lo pone en marcha, sea un banco para darte un crédito o un organismo público para evaluar un currículum o asignar una subvención, no te podrá dar una explicación de cómo se ha llegado a esa decisión. Hay un proyecto muy ambicioso de la Unión Europea[106] de introducir el concepto de explicabilitat (*explainability*) para obtener explicaciones de cómo la máquina ha procesado los datos para llegar a una determinada conclusión. La comunidad informática está muy sensibilizada por encontrar una manera general de argumentar las decisiones que toma la máquina. Además, hay que expresar los argumentos en términos comprensibles para personas no-expertas en informática".

Carme Torras -autora de novelas de Ciencia ficción desde las que enfoca hacia dónde evolucionaremos los humanos y las máquinas- reclama dos aspectos: "1. Regulación para asegurar los criterios mínimos en la aplicación de la inteligencia artificial a la sociedad, en el momento actual. 2. Educación, empezando por formar en criterios éticos a quienes desarrollan la tecnología".

Formación en tecnoética

Según Torras, "por ser partícipe de este futuro, se debe conocer cómo se está creando". Ella es una defensora de formar a formadores, para que entre ellos vayan diseminando el concepto de 'tecnoética' o 'roboética', que sería la fusión del conocimiento de las tecnologías y / o robótica con una base de conceptos éticos.

"Quisiera llegar a secundaria, los institutos, los jóvenes de primaria. Habría que empezar ya, para que no sea demasiado tarde. Esta generación crecerá en un mundo totalmente tecnológico que estamos creando ahora. Debería ser una materia transversal, para asignaturas de lenguas, de matemáticas, de sociales, de filosofía, de tecnología. Y se podría decir: "Ética en la Sociedad Digital".

También propone la creación de un portal informativo, en permanente renovación y ampliación donde se encontrarán desde consejos prácticos para preservar la privacidad en el móvil, hasta regulaciones y normativas, derechos y dudas digitales. "Creo que hay una franja muy grande de la población, gente muy diversa, con curiosidad para saber cómo manejarse mejor en este momento tecnológico, pero no encuentra este servicio mínimo de información rigurosa. Aprovechamos esta inquietud y vehiculamos la", añade Carme Torras.

---

[105] Carme Torras (https://www.iri.upc.edu/staff/torras)

[106] Ethics Guidelines for Trustworthy AI https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1

Jefe de Tecnología de NCENT. Catedrático de informática en la Universidad Pompeu Fabra y en la Northeastern University.

RICARDO BAEZA-YATES[107]: " La revolución tecnológica actual necesita ética ".

"Hoy, las personas generan datos de muchas maneras "-explica Ricardo Baeza-Yates-." Y estas se recogen, analizan y tienen un valor mayor ahora que en el pasado. Pueden utilizarse con motivos legales, comerciales o para manipularnos electoralmente. Los datos ya permiten predecir comportamientos o prácticas de compra. Si tú me dijeras qué preguntas a un buscador web en un día, yo podría deducir si eres hombre / mujer, joven o mayor, localización, etc. Esta información que damos voluntariamente, junto con el rastro de nuestros hábitos, puede ser utilizada con finalidades positivas o negativas ", añade[108].

"La revolución actual necesita ética, sí ", enfatiza el catedrático de informática." Una parte de regulación- no demasiado estricta para no detener el desarrollo tecnológico- y unas mínimas normas éticas como la IA explicable. Es decir, si un algoritmo predice o decide automáticamente, debe poder darme una respuesta de cómo lo ha hecho. Para que esto ocurra, la gente debe saber que existen sesgos en los datos que se introducen los algoritmos ", explica el catedrático de informática de la UPF. Baeza-Yates dice que la tecnología siempre va más rápida que la parte social, y que sólo nos preocupamos de la ética cuando surgen problemas, como ahora mismo.

El 'Pepito Grillo' de los prejuicios

Utilizando los algoritmos sabiendo que tienen sesgos. ¿Por qué dejamos, pues, que tomen decisiones? Y Ricardo Baeza-Yates responde: "¿Por qué tenemos humanos que toman decisiones si también se equivocan? Todos tenemos prejuicios. Los sistemas automatizados son de gran ayuda en situaciones donde los sesgos no influyen mucho. Como por ejemplo en el control aéreo: tener personas muchas horas, en tensión, es más peligroso que entrenar máquinas para esta tarea. No se cansan, están programados, son más eficientes ".

" De sesgos, hay un montón, el problema es que no nos damos cuenta de que existen hasta que cometen un error. Lo mismo ocurre con nuestros prejuicios, que no somos consciente la mayoría de las veces. Pero las máquinas pueden ayudar a darnos cuenta en y crear un mundo más justo ". Sí, pero ¿cómo?" Se podría crear un 'Pepito Grillo'[109] (O asistente virtual) que alertara cuando se está cometiendo un prejuicio, al hablar, actuar, juzgar, etc. O que nos avisara cuando alguien nos está tratando de manipular ".

Y surge una nueva contra-pregunta: ¿Cuántos de nosotros aceptaríamos que una máquina (móvil u otro aparato) nos escuchara lo que decimos (en privado y público) y detectara nuestros prejuicios?

---

[107] Ricardo Baeza-Yates (http://www.baeza.cl/spanish.html)

[108] Participación en el documental chileno: "Por la razón y la ciencia" https://www.youtube.com/watch?time_continue=1253&v=7PCC7tRyM2I

[109] Participación en el podcast: "La inteligencia artificial tiene que ser nuestro Pepito Grillo", realizado por el BBVA https://www.bbva.com/es/podcast-la-inteligencia-artificial-tiene-que-ser-nuestro-pepito-grillo-ricardo-baeza-yates/

Mejorar la societad

"En los próximos años veremos como la detección de los sesgos de los algoritmos ayuda a mejorar el mundo ", concluye Baeza-Yates para dejar un mensaje optimista. Pero cuidado:" ¡Atención! En este nuevo ecosistema de los datos, la tendencia es tener cada vez menos privacidad. Cuando uno acepta una aplicación digital, hay que leer los términos de uso y comprobar si el servicio que obtendrá vale la pena respecto a la privacidad que perderá ", añade." Y saber si legalmente te pueden pedir los datos que te piden. Hay un cambio cultural enorme porque todo el mundo respete la privacidad de los demás ". Y recomienda más educación, saber que fácilmente se puede ser manipulado, entender la tecnología actual y defender los derechos que aseguren la libertad individual.

Director científico del grupo de Inteligencia Artificial de alto rendimiento en el Centro de Supercomputación de Barcelona. Catedrático de Inteligencia Artificial, Universidad Politécnica de Cataluña

ULISES CORTÉS[110]: " Hay que enseñar a los límites de la tecnología y pensamiento crítico "

"Antes quien decidía si darte un préstamo, hacerte una operación de corazón era el algoritmo que estaba al frente del banquero o del médico. Tú no sabías con qué criterios lo decidían. Ahora está pasando lo mismo, pero con máquinas. Sería bueno que una empresa pudiera asegurar que el préstamo es explicable si estás en desacuerdo. Pero hoy esto es imposible, porque nadie les está diseñando para que sean explicables".

"Hay una pequeña porción de la población que tiene miedo de los automatismos por los errores cometidos hasta ahora y, en concreto, del uno del aprendizaje automático (*machine learning*). Pero a la mayoría de gente les importa bien poco lo que se haga con los algoritmos. Y los debería importar porque sus vidas cambian constantemente en función de lo que deciden un montón de máquinas. Se deben producir casos muy flagrantes, como el de Cambridge Analytica, para empezar a tomar conciencia. No es malo usar Facebook, si se saben las consecuencias. Lo malo es dejar a un menor que entre sin advertirle".

"Tengo claro que, en una sociedad occidental, de pasado cristiano, hay un conjunto de valores comunes. De los que ya hablaban Platón y Aristóteles. En EEUU es de utilidad plantearse: ¿Cuál es el mal mínimo que puedo causar? Es muy evidente que lo más justo para nosotros no lo sea por los de Malawi. Pero nos juntamos y decidimos cuál es el consenso".

La ética es cultura

"Haría tres preguntas a todos: 1. ¿Conoce la constitución de su país? 2. ¿Sabe los mandamientos de su religión? 3. ¿Cuál es el último libro de ética que ha leído? ¿Si las respuestas son negativas, como le podemos pedir a los ingenieros que sean éticos? La ética lo has de poner dentro de una cultura. Y luego pedir a los programadores que hagan algoritmos que no perjudiquen los derechos de los demás".

"Ninguna tecnología debe dar miedo. Lo que da miedo es la gente que la utiliza. De diseño es inocua, pero puede tener unos objetivos ilícitos, buena parte de ellos para ganar dinero. Mark Zuckerberg tiene una frase buenísima: "Nosotros no necesariamente somos éticos, pero somos legales". Debemos enseñar a entender los límites de la tecnología. Pero también pensamiento crítico. A la gente no le interesa ser crítica, le interesa ser feliz. Vive en una distopía increíble y se pregunta: "¿Por qué debería preocuparme de mis datos? ¡Si no tengo nada que esconder!".

"El problema es que ni siquiera nos planteamos estas cuestiones. La única manera de revertir esta situación sería que los políticos se ocuparan de los problemas derivados de la tecnología; y educar a los educadores. Los padres han delegado la instrucción en

[110] Ulises Cortés (https://www.bsc.es/cortes-ulises)

el estado. Pero cuando se trata de la educación sobre la tecnología, no existen estos padres. Si quieres cambiar el mundo, tienes que conseguir que los profesores de escuelas, de institutos y de universidades sepan de tecnología. Y que los programadores aprendan más ética. Pero cada vez se ve más la televisión, se pasa mucho tiempo en las redes sociales, se hace menos deporte y se lee muy poco".

Doctora en ingeniería informática y profesora catedrática de La Salle Universidad Ramon Llull, de Inteligencia artificial y Ciencia de los datos. Miembro del Grupo de investigación en Data Science for the Digital Society (DS4DS)

ELISABET GOLOBARDES[111]: " Se están resolviendo problemas que no sabemos cómo se han resuelto"

"Todo cambia cuando los teléfonos inteligentes superan las ventas en móviles sin conexión a Internet. Democraticemos los datos y ahora las recibe y genera cualquier usuario. En África habrán pasado de no tener ninguna conexión a tener, en breve, 5G porque todo el mundo tendrá un móvil. Todo el mundo quiere saber. El poder ya no está en unos cuantos, lo tiene la gente".

"En mis clases del grado de informática y de robótica doy conceptos de IA-robótica-ética. Es una cuestión de educación. Hay mucha polémica también con el reconocimiento facial, por cómo entregamos -sin conciencia- nuestros rasgos de la cara, que son únicos. Les damos a cualquier empresa que nos ofrece un entretenimiento, y no sabemos de qué manera los tratará ni bajo qué legislación. Sería mucho más interesante que diéramos los datos a un servicio de salud público".

" La tecnología -por primera vez- ha superado las demostraciones matemáticas que hacen los humanos. Este es el único punto oscuro de la inteligencia artificial actual. Resolviendo problemas que no sabemos explicar cómo se han resuelto, especialmente en ciertas redes neuronales (*deep learning*)".

El algoritmo es neutro

"Pero me gustaría desvincular el algoritmo de la parte ética. El algoritmo es un tornillo de un coche, sólo una pieza. Es neutro. El problema está en los datos masivas que se le introducen por entrenarlo y en el objetivo de salida. Los programadores tampoco son los culpables: cumplen con lo que dice la empresa o quien encarga hacer ese algoritmo. Harían falta políticas públicas para obligar a que los datos para entrenarlos fueran representativos de la realidad, plurales, éticas e igualitarias. El problema es que quien manda crear el algoritmo -con una determinada fin- puede tener un determinado objetivo económico o comercial".

"Habría una etiqueta de "Datos éticos", un sello de calidad que asegurara que no hay discriminaciones ni sesgos. Igualmente habría que asegurar que el objetivo del algoritmo tiene un propósito ético".

"Sin embargo, me fío más de la inteligencia artificial que de algunos políticos. ¿Porque bajo qué intereses trabajan ellos y decidan lo que me repercute como ciudadana? No me lo explican tampoco. Me fiaría más de una máquina que intentara ser justa a partir de unos criterios definidos, que no de las decisiones no razonadas de muchos líderes del mundo. Se está poniendo demasiado el foco en la IA pero no sé si es para despistar".

---

[111] Elisabet Golobardes (https://www.salleurl.edu/es/elisabet-golobardes-ribe)

Catedrático de Lenguajes y Sistemas Informáticos de la Universidad de Barcelona (UB) y miembro del Departamento de Matemáticas e Informática de la UB.

JORDI VITRIÀ[112]: " Si no nos ponemos las pilas, estamos vendidos "

"A mí no me gusta la palabra ética, porque no se entiende. Prefiero hablar de transparencia, privacidad, diversidad y rendición de cuentas (*accountability*). Si un hotel me dice que compartirá mis datos con otros, puedo aceptar o no. Pero sé a qué jugamos. Que una compañía aérea me diga que las compartirá con el hotel para hacerme ofertas, puedo aceptar o no. Esta es la transparencia que quiero. Si un hospital o una administración me asegura que anonimizará los datos y que hará un uso responsable, yo puedo aceptar o no. Pero exijo esta privacidad. El criterio de diversidad también debe ser.

¿Qué empresa o gobierno me puede certificar que los datos no contienen sesgos que puedan provocar una discriminación? Hoy ninguno. Quizás necesitan asociaciones ciudadanas que lo defiendan. Pero antes hay curiosidad y entender el funcionamiento de la IA. Uno de los posibles negocios del futuro será el de auditor / certificador de algoritmos, para que los sesgos sean los mínimos. Por último, pienso la rendición de cuentas, debería ser también para las empresas. Si recibo una oferta de wifi poco después de decirle a mi hermano para Whatsapp que tenía un problema con la actual wifi. ¿Me puede demostrar Whatsapp que no ha hecho uso de mis datos para generar esta oferta? Porque a mí nadie me ha pedido consentimiento para ello. Deberíamos discutir sobre estos términos. La ética por definición es voluntaria, pero el cumplimiento de estos cuatro puntos debería ser obligatorio".

Los peligros indirectos

"Existen otros peligros para la sociedad ", continúa explicando Jordi Vitrià, miembro del Departamento de Matemáticas e Informática de la UB." Y para hacer frente sólo puedes concienciar y educar a la población. Uno de ellos es el de "la Economía de la dopamina" '. Y argumenta que uno de los bienes más preciados hoy en día es la atención, el tiempo que cada persona invierte en una red social, un juego, un medio, navegando por una web de compraventa de productos, etc. "Los modelos de negocio online intentan maximizar el tiempo de atención, ofreciéndote lo que te tendrá enganchado: una serie, los tuits de tus seguidores, las fotografías de tus amigos, los productos que te intentarán, etc.".

Por otro lado, nos alertamos mucho del reconocimiento facial que hacen países como China, o de experimentos como el que se hizo en Londres captando las imágenes de todos los peatones sin avisarles [113], "pero ... este año, en Barcelona, durante la celebración del Mobile World Congress muchos congresistas entraron por la cara. ¡Literalmente! Animaban a entrar con reconocimiento facial. Si me lo proponen, aunque sea opcional, es preciso saber qué uso hará la organización de todas estas

[112] Jordi Vitrià (http://www.ub.edu/dept_matinfo/professors/vitria-marca-jordi/)

[113] Karma Peiró y Ricardo Baeza-Yates. "Algoritmo, yo también existo", (https://www.karmapeiro.com/2019/09/24/algoritmo-yo-tambien-existo/)

caras. Porque no lo han explicado y espero transparencia ", añade Vitrià." Para mí educar a la gente en tecnología es esto".

"Desde el mundo tecnológico suele decir sólo parte de la verdad. Cualquier revolución tecnológica acaba generando más puestos de trabajo. Pero lo que no se explica es que hay un periodo más o menos largo- donde mucha gente se queda sin trabajo porque nunca será capaz de acceder a las nuevas profesiones. A un panadero o un taxista no lo puedes reconvertir en un analista de datos. ¿Cuánto tiempo les das a los taxistas? ¿Y los médicos de cabecera? ¿Los investigadores que hacen investigación médica siempre estarán, pero y los cirujanos? Desaparecerán, sin duda. Las enfermeras no tienen que sufrir, porque hablamos de empatía. Si queremos ser más inteligentes que en el pasado deberíamos diseñar una estrategia que permitiera pasar el período de transición, con mecanismos de regulación".

Alfabetización en datos (*Data Literacy)*

"El problema es que alegremente todo el mundo suba sus caras en Facebook. La plataforma de Mark Zuckerberg tiene una voluntad de manipulación impresionante. Nos lo demostró en las elecciones estadounidenses. Todavía nos hace falta mucha formación".

" Los comités de ética de los hospitales, tienen que entender de qué hablamos. ¿Los bancos ?, lo mismo. Y los políticos también deberían formarse en inteligencia artificial para que tomarán muchas decisiones al respecto. Lo ideal sería introducir un curso de alfabetización en datos a todas las carreras universitarias (de letras, de ciencias, políticas, derecho o artísticas). La Universidad de Berkeley[114] lo implementó ya en 2017. ¡Tiene lógica! ¡Si no tienes la base, puedes equivocarte mucho! En Finlandia, en Australia o en Nueva Zelanda también se están implementando experiencias similares. La Generalitat se lo debería plantear, desde la escuela, en los institutos, en las universidades. Si no nos ponemos las pilas, estamos vendidos. No podemos regalar los datos como si nada. No nos podemos creer todo lo que nos vienen".

[114] El mensaje de la Universidad de Berkeley es: "Vivimos en un mundo rodeado de datos cada vez más complejo. Este curso te capacitará para responder preguntas y explorar problemas con los que tendrás que lidiar en un futuro inmediato, en tu vida profesional o privada". https://data.berkeley.edu/news/fall-milestones-data-science-education

Profesora del Departamento de Matemáticas e Informática de la Universidad de Barcelona (UB), y directora del grupo consolidado de investigación "Machine Learning and Computer Vision" de la UB.

PETIA RADEVA[115]: " Deberían existir los donantes de datos "

Petia Radeva es una investigadora internacional que hace 26 años que trabaja en el ámbito de la salud y de la visión por computación. Es una defensora acérrima de la tecnología, de los algoritmos de aprendizaje automático, la visión por computador y la inteligencia artificial.

"El mundo se está digitalizando. En los últimos cinco años hemos creado más información que en los miles de años de civilización. Los algoritmos son cada día más importantes para avanzar científicamente. Los necesitamos para medir, comparar y para analizar cuando hay una enfermedad o no, cuando un órgano está dañado o sano. Gracias al *machine learning* las máquinas pueden aprender a partir de un entrenamiento sobre un conjunto de datos. Por ejemplo, podemos mostrarle mil casos de órganos enfermos y la misma cantidad de sanos. Y así los algoritmos pueden construir las reglas para decidir si un órgano es sano".

Radeva no ve un problema en que los algoritmos se comporten como 'cajas negras' sino como un reto. "Son difíciles de explicar, cierto. Y esto no es nuevo. Las redes neuronales tuvieron su boom en los años setenta y ya en ese momento se les acusó de ser 'cajas negras'. En la práctica, lo que ocurrió en ese momento es que los ordenadores no eran tan potentes como ahora. Eran difíciles de resolver los problemas reales y más aún explicarlos. Ahora, son tremendamente útiles por la capacidad de computación que tenemos y la cantidad de datos que utilizamos para entrenarlos. ¿Quién se atreve a prescindir de ellos? no sólo necesitamos algoritmos potentes y eficientes, también autoexplicables. Los médicos no pueden aplicar algoritmos a la práctica clínica si no los entienden. Por eso, últimamente, la comunidad científica está avanzando mucho en desarrollar nuevos métodos para que los algoritmos generen autoexplicaciones".

Donante de datos

"Hoy hay posibilidad de recoger muchos datos. ¿Cabe preguntarse qué uso se hace? Como siempre puede ser positivo o negativo. Creo que el reto de este momento tecnológico es encontrar la manera que todo el mundo se anime a dar datos personales a la ciencia. No conozco ningún multinacional que, después de haber hecho un uso ilícito de los datos personales se haya hundido por la regulación europea (RGPD). En cambio, sí que el RGPD pone obstáculos a que los investigadores recolecten datos personales y puedan hacer uso para avanzar en soluciones a problemas de la sociedad y la salud de la gente".

---

[115] Petia Radeva (http://www.ub.edu/cvub/petiaradeva/?p=36)

"Con los patrones de movimiento de las personas se podría optimizar el transporte en las ciudades, promover una vida más activa y menos sedentaria, ahorraría dinero a la sanidad pública porque evitaría ciertas enfermedades, o se podría hacer prevención más cuidada. Pero no lo podemos hacer con datos de toda la población porque los investigadores no tienen acceso a esta información. Me gustaría participar en alguna campaña para promover la donación de datos, como las que se hacen de órganos o de recogida de sangre. Se debe promover la donación de datos con transparencia y explicaciones, asegurando un buen uso. En el Reino Unido hay un banco de datos accesible, de manera gratuita, a los investigadores. Es una iniciativa loable para avanzar en la ciencia".

" Cuando un organismo o universidad te pide datos (por ejemplo, de tu actividad física durante el último mes o año), la reacción inicial muy a menudo es de miedo o de rechazo. No pensamos que pueden servir avanzar en enfermedades que salvarán la vida de nuestros hijos o nietos. Pero después las concedemos tranquilamente a cualquier empresa en Internet a cambio de un entretenimiento. ¿Por qué?".

Más allá de la tecnología

"Un ejemplo. ¿Cómo puede ser que España, estándar de la dieta mediterránea, tenga la tasa más alta de obesidad infantil de toda Europa? Tenemos obesidad infantil altísima porque no tenemos datos de hábitos de alimentación, y no podemos avanzar. Con lo fácil que sería que, a través de los móviles, las familias y escuelas, las recogieran y las dieran para avanzar en esta enfermedad. La tecnología ha dado un gran paso en las capacidades de colección de datos y su procesamiento. Ahora falta que lo haga la sociedad".

Profesor Investigador Distinguido en el Departamento de Tecnologías de la Información y Comunicación de la Universidad Pompeu Fabra (UPF). Lidera el grupo de investigación de Ciencia de la Web y Computación social.

CARLOS CASTILLO[116]: "Es muy complicado que la gente sepa que un algoritmo está sesgado"

El Grupo de Investigación de Ciencia de la Web y Computación Social de la UPF, junto con Universidad Técnica de Berlín y el Centro Tecnológico Eurecat, creó un algoritmo que detecta y mitiga sesgos de otros algoritmos. El sistema lo han llamado FA*IR[117] y actúa sobre otros sistemas automatizados (de búsqueda o recomendación) que no toman en consideración la representación de diferentes grupos.

"Por ejemplo, una persona que utiliza este sistema puede determinar que al menos haya un 20 o un 40% de mujeres en los resultados, o de gente menor de 25 años, o de personas de tal origen, o de profesionales de dicho sector ", explica Carlos Castillo." Los sesgos, que siempre existen, se pueden detectar y mitigar".

Claro, que si los algoritmos se están aplicando ya en muchos sectores, habrá mucha gente detectando y mitigando los sesgos. ¿Como se puede hacer? "Necesitas transparencia", responde Castillo. "Y que haya la capacidad de que el sistema rinda cuentas y explique sus decisiones. Pero como individuo tienes muy pocas herramientas para saber si detrás hay un algoritmo con un sesgo potente que te discrimina. Si buscas en LinkedIn, periodistas en Barcelona, seguramente no saldrás mencionada y seguramente hay un sesgo de género, en función de los datos del pasado. Hasta que alguien no hace un estudio grande, como en este caso nosotros, y se saben los resultados, es imposible".

Justicia algorítmica

"Algún organismo público catalán debería pensar en ello. Quizás la Autoridad Catalana de Protección de Datos -o una institución similar- pueda dar una solución para que los ciudadanos, consumidores tengan más referencias sobre los algoritmos que toman decisiones. Por ejemplo, en Francia, la Comisión Nacional de Informática y Libertades[118] (CNIL) se involucra en estos temas. También la normativa europea de protección de datos (General Data Protection Regulation[119]) acepta denuncias de colectivos o asociaciones ciudadanas ".

"La Ley de Protección de Datos europea (GDPR) también obliga a un proceso justo de los datos, por lo tanto, se debería obligar a que se cumpla esta justicia en los algoritmos ", añade el director del grupo de investigación de Ciencia de la web y Computación social de la UPF. "el RGPD permite a asociaciones ciudadanas para agregar denuncias. Por ejemplo, si crees que las periodistas están discriminadas en LinkedIn, podrías juntar datos de muchos periodistas y denunciarlo".

---

[116] Carlos Castillo (http://chato.cl/research/)

[117] M. Zehlike, F. Bonchi, C. Castillo, S. Hajian, M. Megahed, R. Baeza-Yates. "FA*IR: A Fair Top-k Ranking Algorithm" (https://arxiv.org/pdf/1706.06368.pdf).

[118] CNIL (https://www.cnil.fr/)

[119] GDPR (https://gdpr-info.eu/)

Catedrático del Departamento de Ciencias Políticas y Sociales de la Universidad Pompeu Fabra (UPF)

CARLES RAMIÓ[120]: " La utilización intensiva de la IA y la robótica es la única manera de asegurar el estado del bienestar "

Carles Ramió insiste en que la administración pública ya llega tarde a la implementación de la inteligencia artificial (IA). Este verano publicó un código ético con 24 puntos [121], un borrador del que se debería de estar pensando ya. Entre los puntos más polémicos el de apostar por un sistema de validación de los algoritmos que utilice el sector público. "Se ha de controlar el proceso de diseño y entrenamiento del sistema automatizado. No puede ser que esté al cargo sólo de ingenieros. Hacen falta filósofos, sociólogos, gente que piense en cuestiones éticas, que tengan en cuenta las discriminaciones de todo tipo", explica el catedrático de la UPF. "Y si compramos algoritmos en la empresa privada, los tenemos que hacer públicos también".

El autor del libro *La inteligencia artificial y la administración pública* [122] considera que "las cajas negras no pueden existir. Yo no digo que todo el mundo tenga que conocer la fórmula del algoritmo. Pero un sistema que dicta sentencia, debe poder ser revisado al menos por los expertos y los abogados, ¿saber cuál ha sido el proceso de entrenamiento o por qué no se le ha enseñado más diversidad?".

"La administración debería dotarse de sistemas que certificaran los algoritmos para que fueran plurales. Que tuviera un equipo de visión ética y técnica, mirando los sesgos, que hiciera transparente el algoritmo y el sistema de entrenamiento".

Una gran oportunidad

"Veo una enorme oportunidad en la IA. Se debe implementar lo antes posible. Los algoritmos, bien diseñados, son imprescindibles. Si todos los datos de una Unidad de Cuidados Intensivos la tienes monitorizada, y alimentas bien el algoritmo, el sistema no puede fallar. Detectará cualquier anomalía o enfermedad. Después ya vendrá el escrutinio del médico que valorará y confirmará o no el diagnóstico del algoritmo".

"Es cierto que esto implica una inversión de entrada. Pero la única forma que seguimos teniendo un cierto estado del bienestar es con la utilización intensiva por parte de la administración pública de la IA y la robótica. Con el envejecimiento de la población no podemos hacer frente a todos los gastos que implican en sanidad y servicios sociales. La IA bien planificada y orientada se puede alcanzar. Pero debemos cambiar el discurso y ser más proactivos como administración. La alerta es que los algoritmos están

---

[120] Carles Ramió (https://www.upf.edu/es/web/politiques/entry/-/-/1182/401/carles-ramio)

3 [121] «Estatuto ético para la implantación de la inteligencia artificial y la robótica en la administración pública», Carles Ramió. https://www.administracionpublica.com/estatuto-etico-para-la-implantacion-de-la-inteligencia-artificial-y-la-robotica-en-la-administracion-publica/

[122] La inteligencia artificial y la administración pública (ed. Los libros de la catarata https://www.todostuslibros.com/libros/inteligencia-artificial-y-administracion-publica_978-84-9097-590-9

entrando a la administración pública de la mano de empresas privadas. ¿Pero entonces como podremos ser transparentes, si dependemos de la privada?"

Catedrático Distinguido de Ciencia de la Computación e Investigador ICREA Academia en la Universidad Rovira i Virgili (URV) de Tarragona, donde dirige la Cátedra UNESCO de Privacidad de Datos y el CYBERCAT-Centro de Investigación en Ciberseguridad de Cataluña.

JOSEP DOMINGO[123]: "Inteligencia significa 'entender' y la máquina no entiende"

Como el resto de los expertos entrevistados, el catedrático de la Universidad Rovira y Virgili declara que tenemos un grave problema con la inteligencia artificial y, sobre todo, con el aprendizaje profundo. Porque a pesar de que son tremendamente eficientes, no hay manera de saber cómo ha aprendido la máquina. "En el momento que una persona pueda solicitar un crédito desde una web, y después de haber introducido unos datos el sistema automatizado le responda que no se lo conceden, querrá tener una explicación. Y aquí vendrá el problema".

"Cada vez tenemos menos conocimiento de por qué pasan las cosas. El broker bursátil se está acabando, cuesta mucho luchar contra un inversor robótico que coge variables económicas de todo el mundo y en décimas de segundo decide qué hace. La explicabilidad[124] o transparencia algorítmica es un intento de recuperar el conocimiento. No te puedes fiar únicamente de la máquina. Tienes que entender de qué va el mundo, sino en pierdes el control ".

El derecho a la explicación

"La máquina no es inteligente. Tampoco sabe por qué ha decidido lo que ha decidido. Inteligencia significa entenderse y la máquina no entiende. Se la entrena con unos datos históricos (por ejemplo datos de créditos concedidos en el pasado que los solicitantes volvieron o no volvieron) para ajustar los parámetros del algoritmo de decisión. Aquí concluye el aprendizaje. Después se le introduce un caso concreto sobre el que hay una decisión (por ejemplo, una nueva solicitud de crédito) y la máquina-algoritmo calcula la decisión utilizando los parámetros aprendidos. Continuando el ejemplo del crédito, si por la razón que sea el algoritmo encuentra que el nuevo solicitante se parece a los que no regresaron el crédito en el pasado, la decisión será rechazar la solicitud, y al revés.

"La ley de protección de datos europea (RGPD) consagra el derecho a la explicación, que en el fondo es para proteger la democracia. El artículo 22 ha sido la piedra de toque. Con el requisito legal nos damos cuenta de que no hay manera de dar explicaciones satisfactorias. Cada vez que alguien impugne la decisión de una máquina, unos técnicos deberán intentar reconstruir de manera artesanal que ha hecho. Pero eso no es escalable. Faltan herramientas para explicar de manera automática".

"Por suerte para los que emplean algoritmos de decisión automática, la ciudadanía aún no entiende bien lo que está pasando. Lo mismo ocurrió con la protección de

[123]Josep Domingo Ferrer (http://www.urv.cat/es/universidad/conocer/personas/profesorado-destacado/2/josep-domingo-ferrer)

[124] Explicabilidad o transparencia algorítmica (https://en.wikipedia.org/wiki/Explainable_artificial_intelligence)

datos. Pero no he visto ninguna manifestación en la calle, ni siquiera cuando pasó el caso de Cambridge Analytica con Facebook[125]– por vulneración de la privacidad".

La responsabilidad del algoritmo

Las 'cajas negras' preocupan y por eso se reclama una transparencia algorítmica. Pero también preocupa definir de quién es la responsabilidad de la decisión del algoritmo. Cuando funcionen los coches autónomos (sin conductores) y haya un atropello, ¿quién será el responsable? ¿Quien la ha diseñado? ¿Quién lo ha programado? ¿El fabricante del coche por haberlo vendido?

Josep Domingo Ferrer aporta otra visión: "El fabricante de coches, el banquero, el médico o el político hacen el encargo del algoritmo. Y tienen una idea vaga de lo que debe hacer el sistema automatizado. El analista hace un diseño y lo pasa al programador informático. de por medio hay muchas instrucciones poco precisas. Finalmente, el algoritmo debería ser revisado de nuevo por el analista y por quien ha hecho el encargo, pero pasa pocas veces. Por otro lado, si no se eligen bien los datos de entrenamiento, puede que el aprendizaje que hace el algoritmo sólo valga para hombres de raza blanca, cristianos y que se hayan descuidado las particularidades de las mujeres, los negros o los musulmanes, por ejemplo".

[125] Karma Peiró , "El escándalo Facebook-Cambridge Analytica: un caso para revisar la protección de datos y mucho más", https://www.naciodigital.cat/noticia/151744/escandol/facebookcambridge/analytica/cas/revisar/proteccio/dades/molt/mes

Investigadora del Grupo de Ingeniería del Conocimiento y Aprendizaje Automático en el Centro de Investigación en Ciencia Inteligente de Datos e Inteligencia Artificial de la UPC. Vicedecana de Big Data, Ciencia de Datos e Inteligencia Artificial del Colegio Oficial de Ingeniería Informática de Cataluña

KARINA GIBERT[126] " El coste energético de la IA es enorme "

"La tecnología, en principio, es neutra y el que tiene usos más o menos éticos es lo que hacemos con ella ", explica la investigadora, Karina Gibert " Los problemas éticos a tener en cuenta hoy son los sesgos, sí. Pero también la explicabilitat, que es una cuestión mucho más profunda y que ya se ha comentado en este capítulo".

Gibert quiere puntualizar que sólo hay una parte de la IA que no es explicable, la que se conoce como 'subsimbólica'. "La parte simbólica, que imita la forma como los humanos resuelven los problemas, es totalmente explicable. Pero es terriblemente cara, desde el punto de vista computacional, y nunca ha dado un salto a problemas reales de cierta complejidad porque colapsa".

"La parte subsimbólica intenta conseguir que los ordenadores resuelvan problemas que requieren inteligencia con más calidad que los humanos, desde el punto de vista de los resultados que aporta "- continúa explicando la investigadora . -" sin importar mucho si la forma como resuelve el problema imita o no como lo hace el humano. Y esta es la IA no explicable, la que, entre otras cosas, ha dado lugar al *deep learning*. Como contrapartida, el *deep learning* ha logrado dar resultados a problemas grandes que hasta ahora no se habían resuelto. Pero está claro que para que el resultado de una inteligencia artificial tenga impacto en un contexto real, necesita argumentar como sea la decisión que has tomado. Y eso lo sabemos los que -como yo- hace 30 años que trabajamos en aplicaciones reales".

La huella ecológica de la IA

Karina Gibert también aporta una reflexión novedosa éticamente, no mencionada hasta ahora. "El coste energético necesario para ejecutar todos estos algoritmos y para hospedar todos los datos que participan de estos procesos es enorme. Y tiene huella ecológica. La pregunta es: ¿Necesitamos que tu reloj mida la temperatura corporal cada cinco minutos, durante un año, cuando se sabe que la temperatura evoluciona de una manera monótona y hay cambios visibles cada hora?".

"Esto tiene mucho que ver con un dilema que el investigador Ricardo Baeza-Yates menciona a menudo: "Datos masivos o datos apropiados?" (*Big Data o Right Fecha*). Quizás necesitamos tener menos y que sean informativos y representativos, lo que nos haría ahorrar energía para almacenarlas y procesarlos. Desde el punto de vista de la sostenibilidad es muy importante tener en cuenta qué tipo de algoritmos y dispendio de datos hacemos".

[126] Karina Gibert https://www.eio.upc.edu/en/homepages/karina

¿De quién es la responsabilidad si falla el algoritmo?

“Veo dos escenarios. 1. El de las inteligencias artificiales no adaptativas (no cambian los mecanismos de funcionamiento y mantienen el diseño), donde la responsabilidad está en quien diseña el algoritmo. Lo veo claro: quien diseña el algoritmo debe tener claros cuáles son los valores de riesgo asociados a las decisiones o recomendaciones que el sistema pueda hacer. Pero también pienso que quizás no se visibilizan bastante los niveles de incertidumbre asociados a las soluciones del algoritmo. Esto es como con las *guidelines* a la medicina, que dice: 'Cuando llegue una persona con un ataque al corazón, hay que administrar esta dosis'. La dosis estándar también está ligada a un montón de incertidumbre, y a una persona en concreto le puede ir mal. Y cuando esto sucede, el médico no asume ninguna responsabilidad si ha seguido el *guideline*. No veo mucha diferencia de cómo se debería tratar lo que recomienda una inteligencia artificial, respecto a lo que recomienda un doctor, que también se basa en recomendaciones estándar que pueden no funcionar en un caso concreto”.

“Escenario 2. Cuando la IA es adaptativa. Es complicado tener bien acotados qué tipo de acciones puede desarrollar a largo plazo. Si una inteligencia artificial va guardando un histórico de lo que hace, y modificando su comportamiento según lo que mejor o peor, en función de las acciones previas, aquí nos podemos encontrar que después de un cierto tiempo aparezcan razonamientos o recomendaciones que no nos imaginamos. Es decir que la IA esté yendo más allá, sola, y creando resoluciones nuevas. Aquí es muy difícil exigir o pensar que el que diseña el algoritmo puede tener controladas todas las situaciones y que pueda predecir y limitar las que pueden ser peligrosas para los humanos. Esto plantea un debate sobre hasta qué punto queremos que las inteligencias artificiales expresen con toda su potencia cuando tratan con personas o entornos donde interaccionan con el humano y pueden generar riesgos”.

Karina Gibert propone que se pueda auditar la inteligencia artificial (igual que los Colegios auditan el software en general), en caso de disfunción y poder determinar si ha habido una mala descripción del problema, un mal diseño, una mala implementación o una denegación del sistema. En cada caso, la responsabilidad recae en una u otra persona. "Pero sí es responsabilidad del desarrollador hacer pensar en todos los límites, riesgos y problemas al que encarga el algoritmo (o su función), para que el diseño sea lo más completo y su uso lo más correcto posible. Esto se hace poco, y se debería hacer más”.

2.6 Dilemas pendientes

La científica Marie Curie dijo en algún momento de su dilatada carrera: "No hay que temer nada en la vida, sólo tenemos que intentar comprenderlo. Es momento ahora, de entender más para temer menos".

En el apartado anterior, los principales investigadores de inteligencia artificial de nuestro país han apuntado sobradamente los dilemas que tenemos hoy con la IA, tales como la existencia de sesgos en los algoritmos, la falta de explicabilidad cuando actúan con aprendizaje profundo o la falta de consenso sobre quién debe asumir la responsabilidad en caso de un error de la máquina.

Con el paso del tiempo surgirán otras preguntas que tampoco tienen una respuesta clara hoy. "¿Cómo podemos asegurar que las decisiones automatizadas (o las actuaciones derivadas de ellas) no tienen un impacto negativo para las personas? ¿Qué niveles de seguridad tienen estos sistemas inteligentes para garantizar que no son vulnerables a los ciberataques o un uso malicioso? ¿Quién es el responsable de estas decisiones automatizadas? ¿Qué pasará cuando un algoritmo nos conozca mejor que nosotros mismos y pueda manipular nuestro comportamiento subliminalmente? ", Preguntaba la doctora en Inteligencia Artificial por el Instituto de Technologia de Massachussets (MIT), Nuria Oliver, en la lección inaugural del curso 2019-2020 del sistema universitario catalán[127].

En paralelo a los avances de los algoritmos de decisión automatizada (ADA), hay dos conceptos que no estamos cuidando suficiente y que cada vez tienen más importancia en las sociedades tecnológicas que vivimos: la confianza y la transparencia.

La confianza es un pilar básico en las relaciones entre humanos e instituciones de cualquier sociedad. La tecnología necesita la confianza de los usuarios, que delegamos cada vez más nuestras vidas en servicios digitales. Pero en los últimos años, hemos cambiado el sentido del término. A través de la tecnología confiamos en desconocidos (cuando alquilamos una habitación de casa en AirB & B y abrimos la puerta a desconocidos, cuando nos llevan a un destino, sea en BlaBlaCar o en Uber, cuando transferimos dinero a través de una aplicación de móvil y no sabemos nada de la empresa que hay detrás). Pero la confianza se nos ha colado entre los dedos, sobre todo después de escándalos como el de Facebook y la consultora Cambridge Analytica, que manejar de manera ilegal datos de 50 millones de personas y han hecho tambalear el sentido de la palabra 'democracia'.

Y **la transparencia** ... Un término muy empleado en los últimos años, pero muy poco consistente, porque más bien tenemos opacidad informativa de lo que ocurre con la tecnología que nos rodea. "Un sistema computacional es transparente cuando una persona no experta, la observa, la entiende y sabe cómo funciona", explica Nuria Oliver. "Pero esta no es la realidad hoy en día". Según la investigadora cada vez tenemos más opacidad, bien porque las empresas privadas quieren proteger la propiedad intelectual de sus algoritmos, bien porque la ciudadanía no tenemos un

4 [127] Lección de Nuria Oliver en la inauguración del curso 2019-2020 del sistema universitario catalán https://www.youtube.com/watch?v=DwC0KDwliXc

mínimo de conocimiento tecnológico para entender las explicaciones, bien porque las administraciones no conceden la importancia que la transparencia se merece (a pesar de la palabra salga en casi todos los programas electorales), o porque con el aprendizaje profundo impide tener una explicación del porqué ha tomado esa decisión.

La IA se está desarrollando en todo el mundo a marchas aceleradas. Quien la domine , no sólo tendrá poder económico, sino también político y social. Llegados a este punto, el único camino que deberíamos contemplar desde la ciudadanía es conseguir (y exigir) una inteligencia artificial **confiable**, diseñada y pensada 'para las personas', con un **profundo sentido ético**, que cumpla con los valores de justicia, transparencia (de la real) y equidad. Una inteligencia artificial segura y que se pueda auditar y pedir explicaciones.

2.7 Lecturas recomendadas para profundizar más

Esta sección recoge normas públicas que regulan el uso de la inteligencia artificial en Europa, pero también textos recomendados por los investigadores mencionados que han servido de contexto y documentación para la elaboración de este informe.

Estrategias, Declaraciones y Recomendaciones

- Estrategia de inteligencia artificial de Cataluña https://participa.gencat.cat/processes/estrategiaIA
- Barcelona Declaration for the proper development and usage of artificial intelligence in Europe (2017) https://www.iiia.csic.es/barcelonadeclaration/
- Declaración sobre ética y protección de datos en inteligencia artificial https://apdcat.gencat.cat/web/.content/04-actualitat/noticies/documents/ICDPPC-40th_AI-Declaration_ADOPTED.pdf
- Directrices éticas para una IA fiable https://ec.europa.eu/spain/barcelona/news/press_releases/190408_ca
- Estrategia española de I+D+I en Inteligencia Artificial (http://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf)
- Recomendaciones de la OCDE sobre inteligencia artificial https://www.oecd.org/centrodemexico/medios/cuarentaydospaisesadoptanlosprincipiosdelaocdesobreinteligenciaartificial.htm
- Resolución del Parlamento Europeo "Una política industrial global europea en materia de IA y robótica (febrero 2019) https://www.europarl.europa.eu/doceo/document/A-8-2019-0019_ES.html
- EU artificial intelligence ethics checklist ready for testing as new policy recommendations are published https://ec.europa.eu/digital-single-market/en/news/eu-artificial-intelligence-ethics-checklist-ready-testing-new-policy-recommendations-are
- Regulating big data. The guidelines of the Council of Europe in the context of the European data protection framework http://isiarticles.com/bundles/Article/pre/pdf/106203.pdf
- Artificial Intelligence and Data Protection: Challenges and Possible Remedies https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6
- Principios establecidos por High Level Expert Group sobre Inteligencia Artificial de la Comisión Europea https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence
- Artificial Intelligence and Data Protection: Challenges and Possible Remedies, per Alessandro Mantelero https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6

Informes

- *Automating society. Taking stock of Automated Decision-Making in the EU*. Informe coordinado por Algorithm Watch https://algorithmwatch.org/en/automating-society/
- *Artificial Intelligence in Education. Challenges and Opportunities for Sustanaible Development. UNESCO* https://unesdoc.unesco.org/ark:/48223/pf0000366994
- World Health Organisation releases first guideline on digital health interventions. https://www.who.int/news-room/detail/17-04-2019-who-releases-first-guideline-on-digital-health-interventions
- La inteligencia artificial en Cataluña. Informe ACCIÓ. https://www.accio.gencat.cat/ca/serveis/banc-coneixement/cercador/BancConeixement/la-intelligencia-artificial-a-catalunya
- *La ciberseguridad en Cataluña (*2018). Informe ACCIÓ. https://www.accio.gencat.cat/web/.content/bancconeixement/documents/informes_sectorials/ciberseguretat-informe-tecnologic.pdf
- IEEE Advancing Technology for Humanity (2016). "Ethically Aligned Design. A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems". http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf
- Partnership on AI to benefit people and society. Website https://www.partnershiponai.org/#
- "European Civil Law Rules in Robotics". European Parliament (2016). http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf
- La Internet de las Cosas en Cataluña. Informe ACCIÓ. Generalitat de Catalunya, https://www.accio.gencat.cat/web/.content/bancconeixement/documents/pindoles/iot-cat.pdf
- 2015 Data Breach Fraud Impact Report, per Al Pascual. https://www.javelinstrategy.com/coverage-area/2015-data-breach-fraud-impact-report
- *Solving the false positives problem in fraud prediction using automated feature engineering,* de Roy Wedge, James Max Kanter, Kalyan Veeramachaneni, Santiago Moral Rubio i Sergio Iglesias Perez. http://www.ecmlpkdd2018.org/wp-content/uploads/2018/09/567.pdf
- ALGORITHMIC IMPACT ASSESSMENTS:A PRACTICAL FRAMEWORK FOR PUBLIC AGENCY ACCOUNTABILITY.
  Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker. AI Now Institute https://ainowinstitute.org/aiareport2018.pdf
- AI in Context. Data&Society. https://datasociety.net/output/ai-in-context/
- The Cybersecurity Campaign Playbook (2018) https://www.ndi.org/publications/cybersecurity-campaign-playbook-global-edition
- Evaluación de los aspectos metodológicos, éticos, legales y sociales de proyectos de investigación en salud con datos masivos (big data), de Itziar de Lecuona http://scielo.isciii.es/scielo.php?script=sci_abstract&pid=S0213-91112018000600576

- El Vehículo Conectado a Cataluña (2019). Informe ACCIÓ. https://www.accio.gencat.cat/ca/serveis/banc-coneixement/cercador/BancConeixement/vehicle-connectat-a-catalunya
- Trustful and trustworthy: Manufacturing trust in the digital era, de Liliana Arroyo https://www.esade.edu/researchyearbook/node/231
- The Filter Bubble: What The Internet Is Hiding From You, de Eli Pariser https://hci.stanford.edu/courses/cs047n/readings/The_Filter_Bubble.pdf

Artículos

- La oportunidad de la investigación con datos masivos en salud, de Itziar de Lecuona. http://www.gacetasanitaria.org/es-evaluacion-los-aspectos-metodologicos-eticos-articulo-S0213911118300864
- La ética en la inteligencia artificial, de Ramón López de Mántaras https://www.investigacionyciencia.es/revistas/investigacion-y-ciencia/el-multiverso-cuntico-711/tica-en-la-inteligencia-artificial-15492
- Es posible acabar con los sesgos de los algoritmos (1a i 2a parte), Ricardo Baeza-Yates i Karma Peiró (2019) https://www.karmapeiro.com/2019/06/17/es-possible-acabar-amb-els-biaixos-dels-algoritmes-1a-part/
- The Ultimate List of Cognitive Biases: Why Humans Make Irrational Decisions, de Tomer Hochma https://humanhow.com/en/list-of-cognitive-biases-with-examples/
- Machine Bias, Pro Publica https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
- How to make algorithms fair when you don't know what they're doing, de Sandra Wachter. Revista Wired. https://www.wired.co.uk/article/ai-bias-black-box-sandra-wachter
- Hacking Our Identity: The Emerging Threats From Biometric Technology, per Jayshree Pandya https://www.forbes.com/sites/cognitiveworld/2019/03/09/hacking-our-identity-the-emerging-threats-from-biometric-technology/#353ed3505682
- Google apologizes after Photos app tags two black people as a gorillas, per Loren Grush https://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas
- MIT researchers: Amazon's Rekognition shows gender and ethnic bias (updated), per KYLE WIGGERS https://venturebeat.com/2019/01/24/amazon-rekognition-bias-mit/
- Algoritmo, yo también existo, de Karma Peiró i Ricardo Baeza-Yates (https://www.karmapeiro.com/2019/09/24/algoritmo-yo-tambien-existo/)
- FA*IR: A Fair Top-k Ranking Algorithm, de Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, Ricardo Baeza-Yates https://arxiv.org/pdf/1706.06368.pdf

- Estatuto ético para la implantación de la inteligencia artificial y la robótica en la administración pública, de Carles Ramió. https://www.administracionpublica.com/estatuto-etico-para-la-implantacion-de-la-inteligencia-artificial-y-la-robotica-en-la-administracion-publica/
- La inteligencia artificial y la administración pública (ed. Los libros de la catarata https://www.todostuslibros.com/libros/inteligencia-artificial-y-administracion-publica_978-84-9097-590-9
- El escándalo Facebook-Cambridge Analytica: un caso para revisar la protección de datos y mucho más, de Karma Peiró https://www.naciodigital.cat/noticia/151744/escandol/facebook-cambridge/analytica/cas/revisar/proteccio/dades/molt/mes
- ¿Cómo abrir las cajas negras de las administraciones públicas? Transparencia y rendición de cuentas en el uso de los algoritmos, d'Agustí Cerrillo. http://revistes.eapc.gencat.cat/index.php/rcdp/article/view/10.2436-rcdp.i58.2019.3277

Videos

- Lección inaugural de Nuria Oliver del curso 2019-20 del sistema universitario catalán. https://www.youtube.com/watch?v=DwC0KDwliXc
- "Por la razón y la ciencia". Documental de la televisión chilena. Intervención de Ricardo Baeza-Yates. https://www.youtube.com/watch?time_continue=1253&v=7PCC7tRyM2I

2.8 Agradecimientos

Por último, corresponde añadir un sincero agradecimiento a los expertos en inteligencia artificial y ciencia de datos consultados y que nos han permitido entender el momento actual de la aplicación de los algoritmos de decisión automatizada en Cataluña, tanto los beneficios como los riesgos. El agradecimiento también a los responsables de departamentos de la administración pública, y los responsables de empresas privadas por haber hecho posible explicar sesenta ejemplos.

La intención era profundizar en esta tecnología, huyendo del habitual alarmismo de titulares sin contexto. Las entrevistas personalizadas y el tiempo concedido para todas las personas que mencionamos a continuación es impagable.

- **Ricardo Baeza-Yates**, jefe de Tecnología de NCENT. Catedrático de informática en la Universidad Pompeu Fabra y en la Northeastern University.
- **Anton Bardera**, profesor del Departamento de Informática, Matemática Aplicada y Estadística de la Universidad de Girona (UdG).
- **Marga Bonmatí** , directora del Consorci AOC.
- **Marco Bressan**, ex científico de datos jefe del BBVA. Director de Satellogic.
- **Meritxell Bassolas**, directora de Conocimiento y Transferencia de Tecnología en el Centro de Visión por Computación de la Universidad Autònoa de Barcelona (UAB).
- **Victòria Camps,** catedrática de ética y filosofía del derecho moral y político de la Universidad Autònoa de Barcelona (UAB).
- **Carlos Castillo**, profesor investigador distinguido en el Departamento de Tecnologías de la Información y Comunicación de la Universidad Pompeu Fabra (UPF). Lidera el grupo de investigación de Ciencia de la Web y Computación social.
- **Ulises Cortés**, director científico del grupo de Inteligencia Artificial de alto rendimiento en el Centro de Supercomputación de Barcelona. Catedrático de Inteligencia Artificial, Universidad Politécnica de Cataluña
- **Fernando Cucchietti**, director de analítica de datos y visualización en el Barcelona Supercomputing Center
- **Josep Domingo**, catedrático distinguido de Ciencia de la Computación e Investigador ICREA Academia en la Universidad Rovira i Virgili (URV) de Tarragona, donde dirige la Cátedra UNESCO de Privacidad de Datos y el CYBERCAT-Centro de Investigación en Ciberseguridad de Cataluña.
- **Ricard Gavaldà**, profesor y coordinador del laboratorio de investigación en la Universidad Politécnica de Cataluña (UPC). CEO y director científico de Amalfi Analytics.
- **Karina Gibert**, investigadora del Grupo de Ingeniería del Conocimiento y Aprendizaje Automático en el Centro de Investigación en Ciencia Inteligente de Datos e Inteligencia Artificial de la Universidad Politécnica de Cataluña (UPC). Vicedecana de Big Data, Ciencia de Datos e Inteligencia Artificial del Colegio Oficial de Ingeniería Informática de Cataluña

- **Elisabeth Golobardes**, Doctora en ingeniería informática y profesora catedrática de La Salle Universidad Ramon Llull, de Inteligencia artificial y Ciencia de los datos. Miembro del Grupo de investigación en Data Science for the Digital Society (DS4DS).
- **Àlex Hinojo**, ex director general de la asociación Amical Wikimedia. Cofundador del proyecto Derechos Digitales.
- **Alberto Labarga**, científico de datos en IOMED.
- **Itziar de Lecuona**, profesora del departamento de Medicina y subdirectora del Observatorio de Bioética y Derecho de la Universidad de Barcelona.
- **Ramon López de Mántaras**, profesor investigador del Consejo Superior de Investigaciones Científicas (CSCI) y ex director del Instituto de Investigación de Inteligencia Artificial (IIIA).
- **Jorge López**, responsable de área de investigación y desarrollo de UBinding.
- **David Llorente**, CEO de Narrativa.
- **Alessandro Mantelero,** relator de inteligencia artificial y privacidad del Consejo de Europa. Profesor de derecho de la Universidad de Turín.
- **Gabriel Maeztu**, doctor y científico de datos. CEO de la empresa Iomed.
- **Jordi Mas**, miembro de Softcatalà y pionero de la Internet catalana.
- **Manel Medina**, catedrático en la Universidad Politécnica de Cataluña y fundador y director del esCERT-UPC.
- **Jordi Navarro,** CEO & co-founder en Cleverdata.io
- **Adina Nedelea**, responsable del equipo de matemáticos del proyecto Ubinding.
- **Ivan Ostrowitz**, ingeniero y experto en inteligencia artificial aplicada al sector educativo.
- **Pier Paolo Rossi**, director del Advanced Customer Marketing&Analytics del Banco de Sabadell.
- **Petia Radeva**, profesora del Departamento de Matemáticas e Informática de la Universidad de Barcelona (UB), y directora del grupo consolidado de investigación "Machine Learning and Computer Vision" de la UB.
- **Carles Ramió**, catedrático del Departamento de Ciencias Políticas y Sociales de la Universidad Pompeu Fabra (UPF).
- **Teresa Roig Sitjar,** asesora tecnológica y científica.
- **Carles Sierra**, director del Instituto de Investigación en Inteligencia Artificial (IIIA).
- **Giusseppe Scionti**, CEO y fundador de NovaMeat.
- **Antonio Andrés Pueyo**, investigador principal del Grupo de Estudios Avanzados en Violencia de la UB y catedrático de Psicología de la UB.
- **Carme Torras**, Doctora en informática y profesora de investigación en el Instituto de Robótica e Informática Industrial (CSIC-UPC)
- **Lluís Torrens**, director de Innovación Social del Área de Derechos Sociales, Justicia Global, Feminismos y LGTBI en el Ayuntamiento de Barcelona.
- **Jordi Vitrià**, Catedrático de Lenguajes y Sistemas Informáticos de la Universidad de Barcelona (UB) y miembro del Departamento de Matemáticas e Informática de la UB.
- **Jordi Torras**, Fundador y CEO d'Inbenta
- Responsables de la **Secretaria Tècnica del Serve d'Ocupació Català**.

3 Protección de datos y la IA

En la introducción, al hablar de los riesgos asociados a los algoritmos de decisión automatizada (ADA), destacamos dos tipos de riesgos: los relacionados con la aplicación concreta (por ejemplo, el riesgo a que nos denieguen un crédito) y los relacionados únicamente con el tratamiento de datos personales (por ejemplo, el riesgo a que se filtren nuestros datos). El trabajo de campo sobre aplicaciones de los ADA en Cataluña se ha centrado en los riesgos asociados a cada una de las aplicaciones presentadas. En esta sección del informe, nos centraremos en los riesgos asociados a la protección de datos.

El respeto a la protección de datos es un punto imprescindible para hacer un uso ético de los algoritmos de decisión automática. En el marco de la Unión Europea, esto implica cumplir con el RGPD.

Cuando hablamos de riesgos asociados a la protección de datos, nos referimos a que no se cumpla alguna de las disposiciones del RGPD. El uso intensivo de datos personales que se hacen en muchas aplicaciones de los ADA, hace que haya fricciones entre los ADA y el RGPD. En este escenario, el objetivo de la APDCAT y de las otras autoridades supervisoras es que se cumplan las disposiciones del RGPD.

3.1 Recogida y tratamiento de datos de forma ilegítima

El RGPD pide que todo tratamiento de datos tenga una base legal que lo legitime. Hay varias y todas son igualmente válidas: consentimiento, obligación legal, interés legítimo, etc. Independientemente de la base legal usada, es necesario que el interesado esté informado del tratamiento de sus datos.

Actualmente los datos personales tienen mucho valor. Se ha llegado a decir que son el nuevo petróleo. En este contexto, sería ingenuo pensar que todas las organizaciones que recogen datos personales lo hacen de manera apropiada. Los altos niveles de personalización que permiten los ADA son un incentivo para recoger la mayor cantidad de datos personales posibles, ya que son la base para su funcionamiento.

Ejemplos de recogida y tratamiento de datos de forma ilegítima hay muchos. Poniendo el foco sobre los dispositivos móviles (que, por estar a nuestro lado gran parte del día, tienen un gran potencial para generar información personal) , encontramos dos trabajos muy interesantes que hacen referencia al uso ilegítimo de nuestros datos. El primero[128] hace un análisis de dos técnicas que utilizan aplicaciones móviles que encontramos en el mercado de aplicaciones para evitar los controles establecidos por Android para acceder a ciertos datos. El segundo[129] explica como piezas de software adjuntadas durante el proceso de manufactura del móvil aprovechan sus privilegios para leer y distribuir información sin el conocimiento de los usuarios.

[128] Joel Reardon, Álvaro Feal et al. “50 ways to leak your data: an exploration of apps’ circumvention of the Android permission system”, 28th USENIX Security Symposium, 2019.

[129] Julien Gamba, Mohammed Rashed et al. “An analisys of pre-installed Android software”, 41st IEEE Symposium on Security and Privacy, 2020.

3.2 Derecho a no ser objeto de decisiones automatizadas

Los ADA toman decisiones basadas en los perfiles de las personas (las características de una persona que los diseñadores del algoritmo han considerado relevantes). El hecho de limitar la toma de decisiones a un conjunto de características previamente definidas es una limitación importante respecto a la capacidad que tiene un juez humano de evaluar otros aspectos que, en principio, pueden parecer poco relevantes. Por ejemplo, hablando con un interlocutor humano una persona puede explicar el porqué de la importancia o la irrelevancia de alguna de las características de su perfil, o incluso, remarcar cosas que el perfil no refleja.

El RGPD reconoce esta problemática y establece el derecho a no ser objeto de decisiones automatizadas si estas pueden tener un efecto significativo sobre su persona. En particular el artículo 22 dice:

> *"Toda persona tiene el derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que tenga efectos jurídicos sobre el interesado o le afecte significativamente".*

El artículo 22 no excluye, en ningún caso, el uso de medios automáticos en tomar una decisión, pero exige que, si la decisión tiene efectos significativos sobre las personas, haya intervención humana.

Aunque la redacción del artículo 22 habla del derecho a no ser objeto de una decisión automatizada, las directrices publicadas por su interpretación afirman que no es un derecho que hayan de exigir los interesados sino una prohibición. La diferencia es sustancial porque si fuera una especie de derecho de oposición, las organizaciones podrían hacer uso de decisiones automatizadas de forma generalizada y, únicamente, ajustar su comportamiento para los casos específicos en que un interesado lo solicita.

3.3 El principio de transparencia

Para que las personas puedan ejercer los derechos que tienen sobre sus datos, es necesario que el tratamiento sea transparente. Según la RGPD, la transparencia exige que toda la información que se da a las personas (los datos que se tratan, el objetivo del tratamiento, los derechos de las personas, etc.) sea concisa, fácilmente localizable y haga uso de un lenguaje comprensible.

En cuanto a la transparencia de las decisiones tomadas por los algoritmos, las disposiciones del RGPD son limitadas. Es cierto que los considerandos (la parte no dispositiva de la norma que sirve para interpretarla) se habla del derecho a obtener una explicación de la decisión. Pero la parte dispositiva se limita a pedir que los interesados reciban información relevante sobre la lógica que se aplica. Mientras que dar una explicación de una decisión requiere tratar cada caso concreto, dar información relevante sobre la lógica se puede hacer de forma más general, en relación al algoritmo.

Posiblemente, la decisión de no pedir una explicación por las decisiones individuales se fundamenta en la incapacidad técnica que hay para hacerlo en algunos casos relevantes. Cada algoritmo de IA tiene un nivel específico de explicabilidad. En general,

el incremento de la complejidad de los algoritmos ha ido acompañado de una reducción en la capacidad de explicar los resultados. Por ejemplo, mientras que explicar una decisión resulta sencillo en un sistema basado en reglas, puede ser muy complicado en un sistema de aprendizaje profundo (donde una estructura compleja de neuronas artificiales autorregula en la fase de entrenamiento del modelo sin ninguna intervención externa).

Dada la relevancia del aprendizaje automático y, en particular, del aprendizaje profundo, podemos decir que la explicabilidad es uno de los puntos débiles de la inteligencia artificial. ¿Pero por qué es importante que el resultado de un algoritmo sea explicable? A veces no nos interesa saber el porqué de una decisión; basta saber que el algoritmo tiene un nivel de precisión suficientemente bueno. Por ejemplo, en un sistema de recomendación de películas, no parece especialmente importante explicar el porqué de una recomendación concreta. Ahora bien, cuando la decisión puede tener efectos importantes sobre las personas, tener una explicación concreta es más necesario. ¿Te fías de un médico que no da ninguna explicación del diagnóstico o de un juez que dicta sentencias sin justificación? Entonces, si exigimos una explicación a los expertos humanos, ¿por qué no deberíamos pedir a los ADA que hacen funciones equivalentes?

El RGPD evita el problema anterior exigiendo intervención humana en aquellas decisiones que puedan afectar a las personas de una forma importante; es decir, en aquellas decisiones en que la exigencia de una explicación es más necesaria.

Por otra parte, la necesidad de explicar algunas decisiones está fomentando la investigación en inteligencia artificial explicable (XAI). Hay dos aproximaciones. La primera busca generar una explicación de un comportamiento observado de un ADA. En cierto modo, esta aproximación sigue el procedimiento de la ciencia experimental: dado un comportamiento observado, construimos un modelo que la explica. El problema es la validez de este modelo. Como dijo el estadístico británico George Box en 1978, todos los modelos para explicar la realidad son incorrectos, pero algunos son útiles. Tener un modelo útil para explicar el funcionamiento de un ADA no quiere decir que la explicación que obtenemos por un caso concreto sea la correcta. La segunda aproximación evita el problema de la validez de la explicación exigiendo que el algoritmo de decisión sea fácilmente interpretable. Esto se logra principalmente reduciendo la complejidad. El precio para pagar en este caso es pérdida de precisión del ADA.

Del mismo modo que no todas las decisiones son sencillas, tampoco es necesario tener una explicación para todas las decisiones. El primer paso es determinar si es necesario que las decisiones de un ADA sean explicables y valorar cuál es la mejor manera de obtenerlas. Por ejemplo, en el caso de un coche autónomo las explicaciones de las decisiones pueden ser interesantes a la hora de determinar la responsabilidad en caso de accidente.

3.4 El principio de lealtad

El principio de lealtad exige hacer un uso de los datos que sea previsible por parte de las personas (en relación con la finalidad buscada) y que no haya consecuencias adversas para las personas que no estén justificadas.

El principal punto de fricción entre los ADA y el principio de lealtad es la potencial discriminación que pueden sufrir las personas como consecuencia de un algoritmo que toma decisiones sesgadas. Es necesario que los ADA no hagan un trato desigual hacia un grupo de personas en función del género, del color de piel, de sus creencias religiosas, de su orientación sexual, etc.

Podemos pensar que los algoritmos, a diferencia de las personas, no tienen prejuicios y que, por tanto, las decisiones tomadas por estos no son discriminatorias. Esta es una visión no se ajusta a la realidad: los ADA tienen sesgos. El origen de estos sesgos es diverso.

Aunque no podemos descartar que el sesgo de una ADA sea intencionado, en general, no hay malas intenciones por parte del programador, del diseñador o de la organización responsable. Es más bien una cuestión de desconocimiento, de falta de diligencia. La falta de una visión ética de la computación de las personas que intervienen es la causa principal.

Los sesgos en un algoritmo pueden tener su origen en el uso de datos erróneos o sesgadas. Como ya se ha comentado, el problema es que el entrenamiento de los algoritmos de aprendizaje automático necesita datos y, lamentablemente, muchos datos históricos presentan sesgos. Hay que ser consciente de este hecho y aplicar técnicas para minimizar el impacto.

Los sesgos también pueden venir de un algoritmo mal diseñado. Es común diseñar un algoritmo pensando en el grupo mayoritario de personas. Aunque este algoritmo sea cuidadoso con las personas de este grupo, puede no serlo para otros grupos minoritarios. En este caso, estaríamos tratando de forma diferente a los grupos minoritarios; es decir, hay riesgo de discriminación. Si el peso de estos grupos en el conjunto de los datos es pequeño, puede tener poco impacto global y pasar desapercibido.

Una técnica básica para diseñar algoritmos no sesgados es evitar el uso de variables potencialmente discriminatorias como el género, la raza, etc. Esto no siempre es posible; hay algoritmos que pueden hacer un uso legítimo. Por ejemplo, si se ha demostrado que las personas de una raza concreta son más propensas a sufrir un cierto tipo de enfermedad, sería irracional no aprovechar este conocimiento. Por otra parte, eliminar estas variables no garantiza que no haya discriminación. Por ejemplo, puede haber una correlación entre áreas concretas de una ciudad y la nacionalidad de las personas. En este caso, el lugar de residencia es una potencial fuente de discriminación.

Existen técnicas más sofisticadas para detectar y mitigar el riesgo de discriminación. La investigación en el campo de la antidiscriminación se ha incrementado en los últimos años. Ahora bien, el uso efectivo de estas técnicas puede presentar algunos problemas. Por ejemplo, es necesario determinar los grupos de personas en riesgo de sufrir discriminación; sino no se puede evaluar si hay discriminación y mitigarla. El

hecho de que, muchas veces, estos grupos esté definidos a partir de datos de categorías especiales (que son objeto de limitaciones adicionales al tratamiento) dificulta la aplicación de estas técnicas.

3.5 Principio de limitación de la finalidad

El principio de limitación de la finalidad dice que los datos se recogerán para una finalidad específica y explícita, y que no se deben utilizar los datos de manera incompatible. Este principio es esencial para que las personas puedan hacer un control efectivo del uso que se hace de sus datos.

Algunos de los algoritmos de inteligencia artificial se entrenan con una finalidad especifica; por ejemplo, decidir si se ha de conceder un crédito, detectar una enfermedad, etc. En estos casos, el objetivo del aprendizaje es claro. Ahora bien, otros algoritmos entran dentro de la categoría de "el aprendizaje no supervisado" que extraen información de los datos (patrones, correlaciones, etc.). Como esta información no se conoce por adelantado, es difícil satisfacer el principio de limitación de finalidad.

A pesar de la facilidad con que se recogen datos actualmente, la gran avidez de datos de la IA hace que sus promotores vean con buenos ojos la posibilidad de recurrir a conjuntos de datos ya existentes. Este recurso, puede reducir los costes y, incluso, permitir el acceso a ciertos tipos de datos que pueden ser difíciles de recoger. La reutilización de los datos recogidos por parte de otro actor con una finalidad diferente puede chocar con la limitación de finalidad establecida en el RGPD.

Los sectores que propugnan una flexibilización en la posibilidad de reutilizar datos argumentan que una reducción en los datos disponibles puede dar lugar a una inteligencia artificial menos precisa y con más sesgos.

El principio de limitación de la finalidad tiene algunas excepciones. En particular, el artículo 5 del RGPD dice que el tratamiento con finalidad de investigación científica o histórica, o finalidad estadística son compatibles con el propósito inicial de recogida de datos. Por tanto, el recurso a la investigación científica es una posible vía para la reutilización de datos en IA. La pregunta es: ¿qué es investigación científica? El antecedente 159 del RGDP dice que el tratamiento con finalidad de investigación científica debe interpretarse de una manera amplia que incluye por ejemplo el desarrollo y demostración de nuevas tecnologías, la investigación fundamental, la investigación aplicada y la investigación con financiación privada.

3.6 Principio de minimización

El principio de minimización dice que los datos utilizados en un tratamiento deben ser adecuados, pertinentes y limitados a lo estrictamente necesario para alcanzar la finalidad del tratamiento.

> **Adecuados**. Los datos son adecuados si permiten conseguir la finalidad buscada. En el caso que los datos no sean adecuados, el tratamiento carece de sentido.
>
> **Pertinentes**. Los datos han de estar relacionados con la finalidad del tratamiento.
>
> **Limitadas al mínimo necesario**. No ha de ser posible conseguir la finalidad buscada sin tratar los datos.

Como en el caso del principio de limitación de la finalidad, las dificultades que introduce este principio en los algoritmos ADA radican en el hecho de que estos necesitan de gran cantidad de datos para aprender y tomar decisiones inteligentes. Si queremos que un niño reconozca un coche, le explicamos las características típicas y le enseñamos alguna foto. Para hacer que un algoritmo de aprendizaje automático reconozca los coches con precisión, debemos entrenar con muchos ejemplos.

Siguiendo el principio de minimización deben tratar únicamente los datos que son estrictamente necesarias. Ahora bien, hay que tener en cuenta el efecto que una reducción de datos puede tener sobre la precisión del sistema y sobre la aparición de sesgos.

Existen varias técnicas que permiten reducir el uso de datos personales:

- Selección de los atributos. La precisión de algunas de las técnicas de aprendizaje automático tiene una gran dependencia de los atributos que se consideran. En estos casos, añadir atributos que no son pertinentes puede ser contraproducente. Otras técnicas, como el aprendizaje profundo, no son tan sensibles a los atributos considerados. En cualquier caso, la selección de atributos relevantes da lugar a modelos más sencillos y fáciles de entrenar.
- Aprendizaje federado. Supongamos que queremos entrenar un modelo de inteligencia artificial sobre los datos de un conjunto de personas. En vez de entregar los datos a una entidad que hace el entrenamiento, el aprendizaje federado propone que el entrenamiento se haga de forma distribuida: cada uno hace el entrenamiento con sus datos. Esto se consigue dando el modelo actual en cada persona, que lo actualizará con sus datos y devolverá la actualización hecha. Una vez recogidas las actualizaciones de todos los usuarios, estas se agregarán para generar el modelo entrenado.

- Anonimización y uso de datos sintéticas. El objetivo es evitar el uso de datos personales en el entrenamiento, sean estos datos anonimizados (datos donde se ha roto el enlace con la persona que las ha originado) o datos sintéticos (datos inventados que reproducen las características de los datos originales).

Aunque tener datos abundantes beneficia los ADA, es igualmente importante que los datos sean precisos: puede ser mejor tener menos datos si éstos son más precisos. Esto va en la línea del principio de exactitud.

Como hemos comentado anteriormente, no siempre es posible definir con claridad el fin. El principio de minimización está formulado en términos de la finalidad buscada. Por lo tanto, si no podemos definir la finalidad con claridad, tampoco podremos evaluar si se satisface el principio de minimización.

3.7 Análisis de aplicaciones de los ADA

Para concluir esta sección, revisamos, desde el punto de vista de la protección de datos, algunas aplicaciones de ADA presentadas en el estudio de campo. Para este análisis se han escogido aplicaciones que tienen un impacto importante sobre las personas.

Evaluación del riesgo de violencia

En el capítulo de aplicaciones de la inteligencia artificial en el sistema judicial hay varias herramientas que buscan estimar el riesgo de reincidencia.

- RisCanvi estima el riesgo de que una persona vuelva a delinquir una vez haya salido de la cárcel.
- SAVRY estima el riesgo de reincidencia juvenil.
- VioGen estima el riesgo de reincidencia en el caso de violencia contra las mujeres.

La evaluación del riesgo de reincidencia es una tarea necesaria para determinar las medidas preventivas o de protección a implantar. Antes de la introducción de estos sistemas, el riesgo de reincidencia se evaluaba de forma no estructurada por parte de los profesionales con conocimientos específicos tales como psicólogos, psiquiatras o trabajadores sociales. El problema de estas evaluaciones es que dependen en gran medida de la persona que las hace: los factores que considera relevantes, los sesgos que tiene, de su estado de ánimo, etc. Los estudios muestran que estas estimaciones de riesgo tienen una precisión muy baja.

Las herramientas RisCanvi, SAVRY y VioGen buscan reducir la dependencia de la persona que hace la evaluación. Basándose en un análisis de los datos de reincidencia, estas herramientas determinan las características que incrementan el riesgo y la interrelación entre ellas. Una vez determinadas se utilizan en todas las estimaciones de riesgo. El hecho de que estas características sean públicas, da un buen nivel de transparencia al sistema. La transparencia favorece que no se utilicen características que pueden ser una fuente clara de discriminación hacia minorías (por ejemplo, el color de piel o el país de origen).

La estimación que dan estos sistemas no es definitiva. El personal puede modificar si consideran que hay razones de peso (situaciones específicas relevantes de un

individuo que no se tienen en cuenta en los parámetros del sistema). Ahora bien, cualquier cambio en la valoración hecha por el sistema, hay que justificarla.

En estos casos, la decisión tomada por el sistema tiene un impacto importante sobre las personas y por tanto, según el RGPD, tienen el derecho a obtener intervención humana, a expresar su punto de vista y en impugnar la decisión.

Sistemas de scoring

Antes de conceder un crédito, los bancos estudian el riesgo que tiene. Lo mismo ocurre cuando queremos hacer un seguro, por ejemplo, para el coche. Actualmente, no es el personal del banco o de la aseguradora que determina el riesgo sino un algoritmo.

Los bancos tienen un perfil muy detallado de sus clientes: saben dónde trabajan, su sueldo, en qué gasta el dinero, etc. Los sistemas scoring tienen la capacidad de utilizar toda esta información a la hora de determinar el riesgo asociado a un crédito. Por ejemplo, las transacciones realizadas con tarjeta tienen un código que identifica el tipo de actividad: farmacias, apuestas, servicios de citas, etc. Algunas de estas categorías pueden indicar un incremento en el perfil de riesgo de la persona. Hay algoritmos que utilizan datos demográficos, datos de redes sociales, etc.

Al aceptar las cláusulas de protección de datos, los clientes están consentido una gran cantidad de usos. Entre otros el uso de nuestros datos con fines comerciales (ofrecer productos y servicios). Lamentablemente, hay muy poca gente que lea estas cláusulas.

Aunque tengan una base legal, el hecho de que el algoritmo sea una caja negra es un problema a la hora de determinar si son sistemas justos. Algunos estudios sugieren que hay cambios aparentemente irrelevantes (como cambiar la resolución del móvil) afectan al resultado. Otros estudios, sugieren que se podrían estar perpetuando discriminaciones históricas en cuanto al acceso al crédito.

Dada la relevancia de la decisión, según el RGPD, las personas tienen el derecho a obtener intervención humana, a expresar su punto de vista y a impugnar la decisión.

Detección de fraude en las tarjetas de crédito

La detección de fraude es, indudablemente, una aplicación que busca tanto el beneficio del banco como de sus clientes. Estos sistemas analizan las características de las compras para determinar si son fraudulentas. El problema de estos sistemas es que todavía tienen un porcentaje de falsos positivos muy elevado.

Aunque las consecuencias de un falso positivo, en general, no deberían ser graves. Conviene que se establezcan mecanismos para impugnar la decisión.

Detección de enfermedades

La detección de enfermedades es una de las grandes aplicaciones de la inteligencia artificial en la medicina. En el capítulo de aplicaciones en salud se han presentado tres

casos: la detección del cáncer de colon, la detección del dolor exagerado y la detección de la cirrosis.

En función de la aplicación concreta, el uso de la inteligencia artificial puede tener varias ventajas: mejora de la precisión en la detección, detección temprana, uso de técnicas menos invasivas, etc. Ahora bien, la importancia que pueden tener estas aplicaciones sobre la vida de las personas hace que la intervención de un médico, que tiene la última palabra, sea imprescindible. Afortunadamente, en el campo de la medicina, estos sistemas se plantean como sistemas de ayuda a la decisión.

4 Recomendaciones finales

La capacidad de tomar decisiones de forma autónoma se ha mirado con cautela desde el inicio de la IA. De hecho, el primer código ético lo propuso el autor de ciencia ficción Isaac Asimov en 1942 a Runarround, una historia corta. Esto fue 14 años antes de que John McCarthy acuñara el término inteligencia artificial en 1956.

La gran penetración que han alcanzado los ADA y la IA en los últimos años, junto con el impacto que puede tener sobre las personas, ha dado lugar al desarrollo de múltiples códigos éticos o, tanto por parte de empresas como de otras organizaciones. Por ejemplo, la OCDE ha propuesto un código ético basado en los siguientes principios:

- La IA debe beneficiar a las personas y el planeta a través de crecimiento inclusivo, desarrollo sostenible y bienestar.
- Los sistemas IA deben diseñarse para respetar la ley, los derechos humanos, los valores democráticos y la diversidad, y deben incluir los mecanismos necesarios para garantizar una sociedad justa.
- Debe haber transparencia para asegurarnos de que la gente entiende los resultados de la IA y puedan oponerse.
- Los sistemas IA deben funcionar de manera robusta y segura durante todo el ciclo de vida, y los riesgos potenciales se deben evaluar y gestionar de manera continua.
- Las organizaciones y las personas que desarrollan despliegan o operan los sistemas de IA son responsables de su buen funcionamiento, de acuerdo con los principios anteriores.

Nuestro objetivo en esta sección es desarrollar la parte de protección de datos personales que se ha incluido a muchos de los códigos éticos mencionados. Esto lo hacemos dando una serie de recomendaciones que deben aplicar los diferentes actores que intervienen en un sistema ADA / AI (personas, organizaciones y autoridades supervisoras).

4.1 Recomendaciones para las personas

El RGPD es una legislación de protección de datos avanzada que busca volver a las personas el control sobre el uso de sus datos. Ahora bien, para que el RGPD sea efectiva es necesario un cambio de mentalidad en las personas; hay que crear una cultura de la privacidad. Afortunadamente el desconocimiento que había hace 10 años ya no está: antes nos sorprendía obtener servicios gratuitos en Internet, ahora sabemos que no lo eran (querían nuestros datos). Desafortunadamente todavía somos demasiado permisivos con el uso que se hace de nuestros datos: la mayoría de las personas acepta las condiciones de uso sin leerlas.

En este escenario, las recomendaciones que damos a las personas respecto al uso de sus datos en aplicaciones de inteligencia artificial son bastante generales:

- Tomar conciencia de que los datos personales que una organización trata no son propiedad de la organización sino información que hemos cedido las personas.
- Conocer los derechos que tenemos respecto nuestros datos.
- En caso de que no se respeten los derechos que tenemos sobre nuestros datos, hay que conocer los mecanismos para hacerlo efectivos.
- Es necesaria una mentalidad crítica en el momento de consentir el tratamiento nuestros datos: entender la finalidad de un tratamiento, conocer qué datos son necesarios, entender las potenciales consecuencias del tratamiento y decidir sobre el consentimiento en consecuencia. Por ejemplo, un cliente de un gimnasio puede considerar que tener que dar la huella dactilar para acceder a las instalaciones es desproporcionado.
- Tomar conciencia de que los algoritmos ADA / IA pueden ser especialmente intrusivos. Una vez se detecta que nuestros datos pueden ser objeto de tratamiento por parte de un algoritmo ADA / IA, conviene analizar cuáles son las potenciales consecuencias que se pueden derivar.

4.2 Recomendaciones para las organizaciones que hacen uso de la IA

La inteligencia artificial y los ADA supone una ventaja competitiva que las organizaciones no pueden dejar pasar. Ahora bien, hacer un mal uso puede tener consecuencias importantes. Por ejemplo, Cambridge Analytica era una empresa de consultoría política que hizo un uso ilícito de datos personales para hacer publicidad electoral personalizada, resaltando cada persona aquella parte del programa que más encajaba con sus preferencias. Como resultado de este escándalo, Cambridge Analytica ha dejado de existir y Facebook (la fuente de los datos) ha tenido que pagar una multa de 5.000 millones de dólares.

El uso de los algoritmos de decisión automatizadas debe hacerse desde una **perspectiva ética** y, en particular, **respetando el RGPD** en sus principios y en los derechos que reconoce a las personas. Aunque, el diseño y desarrollo de algoritmos de decisión automatizadas es primordialmente una tarea técnica, hay que exigir unos mínimos conocimientos de ética y de protección de datos a las personas encargadas. Es necesario el seguimiento de códigos éticos existentes o el desarrollo de propios.

- Hay que evaluar el impacto sobre las personas. Los algoritmos de decisión automatizada son una tecnología innovadora. El uso de estas tecnologías es un factor a tener en cuenta a la hora de determinar la necesidad de hacer una **evaluación de impacto sobre la protección de datos** (AIPD). Por lo tanto, es más probable que este tipo de tratamientos requieran una AIPD. A parte, la realización de la AIPD por parte del responsable es una manera de demostrar responsabilidad en el tratamiento de datos personales en caso de que ésta sea demandada por la autoridad de protección de datos.
- **Derecho a no ser objeto de decisiones automatizadas.** De acuerdo con el artículo 22 del RGPD, en los casos que la decisión tiene consecuencias

importantes para las personas, no debe tomarse en base exclusiva de medios automatizados.

- En los casos en que no se excluye la posibilidad de toma de decisiones automatizadas, conviene que la organización considere otras maneras de alcanzar la finalidad buscada que sean menos invasivas.
- El **principio de lealtad** exige que el uso de los datos esté dentro de lo que las personas pueden razonablemente esperar y que las consecuencias para las personas no sean injustificadas. En el caso de los algoritmos de decisión automatizada toma especial relevancia el sesgo de los algoritmos, que puede dar lugar a decisiones discriminatorias. Garantizar que un algoritmo está libre de sesgos es muy difícil, pero hay algunas técnicas que pueden ayudar a reducirlos:
    - Utilizar datos no sesgadas. El uso de datos con sesgos en el entrenamiento de un algoritmo de decisión automatizada es una de las principales formas de introducir sesgos en las decisiones de los algoritmos. Hay que analizar los datos de entrenamiento para detectar y, posteriormente, mitigar los sesgos.
    - No utilizar características que pueden dar lugar a decisiones discriminatorias: sexo, edad, color de piel, etc. Hay que tener en cuenta que esto no siempre es posible y que el hecho de no usar estas características no garantiza que no hay discriminación indirecta, a partir de otras características que son consecuencia de las anteriores.
    - Analizar los resultados del algoritmo para detectar potenciales efectos discriminatorios del algoritmo.
- El **principio de transparencia** exige que las personas reciban una información clara respecto al tratamiento de datos. En el caso de algoritmos de decisión automatizada que tienen un impacto significativo sobre las personas, está la exigencia de dar información sobre la lógica del algoritmo. Para favorecer la transparencia del algoritmo:
    - Conviene utilizar algoritmos que estén disponibles para revisiones externas.
    - Cuando sea factible, conviene utilizar algoritmos explicables en vez de algoritmos que funcionen como una caja negra.
    - En cualquier caso, la información proporcionada sobre la lógica del algoritmo deberá ser clara e incluir los aspectos básicos de su funcionamiento.
- Del **principio de minimización de datos** resulta una limitación de los datos disponibles en un campo que hace un uso extensivo. Para minimizar el impacto de estos principios, conviene hacer uso de técnicas que reducen la cantidad de datos necesarios. Entre estas están las que se comentaron en la sección anterior: anonimización, generación de datos sintéticos, selección de atributos y aprendizaje federado.

- Del principio de limitación de la finalidad también resulta una limitación de los datos disponibles. A pesar de esta limitación, hay que garantizar que el uso de los datos no es incompatible con la finalidad que motivó la recogida de datos.
- En caso de que el responsable del tratamiento identifique una finalidad compatible, se informará a los interesados para que puedan tomar decisiones informadas respecto al uso de sus datos.

4.3 Recomendaciones para las autoridades supervisoras

Las fricciones de los algoritmos de inteligencia artificial con los principios de la RGPD, así como el incremento en su uso, hacen previsible un incremento de las reclamaciones ante las autoridades de protección de datos, que tienen el mandato de hacer cumplir el reglamento.

En su competencia investigadora, las autoridades de protección de datos pueden tener algunas dificultades relacionadas con la complejidad de los sistemas de decisión automatizada de inteligencia artificial. Por ejemplo, a la hora de evaluar si el sistema discrimina (principio de lealtad), a la hora de evaluar la necesidad de tratar un tipo de datos (principio de minimización). Para llevar a cabo estas tareas, puede ser necesario contar con la ayuda de expertos.

Dada la importancia de las autoridades supervisoras en garantizar el cumplimiento de la RGPD y en pedir cambios en el tratamiento cuando no se cumple, es necesario:

- Tener el conocimiento necesario para llevar a cabo de forma efectiva las tareas de supervisión que tienen encomendadas. Los ADA, la inteligencia artificial y, más en general, la tecnología es un campo en rápida evolución. Es necesario que las autoridades de protección de datos tengan los conocimientos suficientes o la capacidad de adquirirlos cuando sea necesario.
- Hacer un esfuerzo divulgativo para concienciar a la ciudadanía sobre los beneficios pero también sobre los riesgos asociados a los ADA y la inteligencia artificial.
- Hacer un esfuerzo divulgativo dirigido a las organizaciones que hacen uso de los ADA y la IA, con el objetivo de concienciarles las de los problemas en relación a la protección de datos y fomentar la incorporación de esta temática en los códigos de conducta.

5 Glosario

A

Algoritmo. Secuencia precisa de instrucciones a seguir para realizar una tarea determinada. Los algoritmos están en todas partes. Por ejemplo, los pasos necesarios para matricularse en una escuela o para pedir una ayuda social no dejan de ser algoritmos. En computación los algoritmos son esenciales. De hecho, lo único que hace un ordenador es seguir las instrucciones de los algoritmos con que ha sido programado.

Algoritmo de decisión automatizada (ADA). Algoritmo que toma decisiones de forma automatizada sin que sea necesaria ninguna intervención humana. En los últimos años, estos algoritmos han alcanzado una gran extensión. Por ejemplo, se calcula que un 70% de las transacciones financieras las realizan algoritmos. Los potenciales efectos de estos algoritmos sobre las personas han puesto de manifiesto cuestiones éticas importantes: responsabilidad de las decisiones (por ejemplo, en un accidente de un coche autónomo), etc.

Antidiscriminación. Conjunto de técnicas que se utilizan para evaluar y mitigar el tratamiento discriminatorio respecto de ciertos grupos de personas.

Aprendizaje automático (*Machine learning*). Subconjunto de algoritmos de IA que se utilizan para enseñar a los ordenadores a hacer tareas sin tener que dar instrucciones precisas. El aprendizaje automático es particularmente útil para enseñar a los ordenadores hacer tareas complejas para las que no se conoce un algoritmo determinado. Por ejemplo, no conocemos ningún algoritmo que determine si un correo electrónico es un spam. Lo que hace el aprendizaje automático es que a partir de un conjunto de correos electrónicos, donde se indica cuáles son basura, analiza los datos para determinar las características que son indicativas de spam.

Aprendizaje federado (*Federated learning*). Técnica usada en aprendizaje automático para entrenar el sistema de forma descentralizada, por lo que no es necesario que una entidad central tenga acceso a todos los datos de entrenamiento. El modelo actual se distribuye al conjunto de entidades que participan en el entrenamiento. Cada una de estas lo actualiza en base a sus datos y devuelve la actualización. El aprendizaje federado tiene dos ventajas importantes: mejora la privacidad de las personas (pues no hay una entidad que tenga acceso a todos los datos) y distribuye el coste computacional del entrenamiento entre diversas entidades.

Aprendizaje profundo (*Deep learning*). Se refiere al uso de redes neuronales complejas que cuentan con múltiples capas de neuronas. El elevado coste computacional de esta técnica hace que su uso práctico sea bastante reciente.

Aprendizaje no supervisado. Técnica usada en aprendizaje automático para entrenar el sistema. Dado un conjunto de datos de entrenamiento, el aprendizaje no supervisado busca encontrar patrones en estos datos (relaciones causa-efecto, correlaciones, grupos tipo, etc.).

Aprendizaje supervisado. Técnica usada en aprendizaje automático para entrenar el sistema. Consideramos un algoritmo que hace uso de un modelo matemático que, dados unos datos de entrada, produce un resultado. Dado un conjunto de datos de entrenamiento que tienen las entradas y la salida que se espera del algoritmo, el aprendizaje supervisado ajusta al modelo para minimizar el error sobre los datos de entrenamiento.

C

Caja negra (*Black box*). La complejidad inherente a muchos de los modelos usados en aprendizaje automático hace que sea difícil explicar su comportamiento. En estos casos se dice que el modelo se comporta como una caja negra: recibe unas entradas y produce una salida, pero desconocemos cómo se genera la salida a partir de las entradas. Cada modelo tiene un nivel específico de transparencia. Por ejemplo, un sistema que aprende reglas a partir de los datos de entrenamiento es muy transparente; mientras que un sistema basado en aprendizaje profundo es una caja negra (la complejidad del modelo hace que sea difícil determinar qué características de los datos de entrada tienen relevancia a la hora de determinar el resultado).

Consentimiento. Expresión clara, informada y libre que hace el interesado respecto la aceptación del tratamiento de sus datos.

D

Datos personales. Datos asociados, o que es posible asociar a una persona.

I

Inteligencia artificial (IA) es un término genérico que engloba aquellos algoritmos que buscan dotar a los ordenadores de un comportamiento inteligente. La definición de comportamiento inteligente es evasiva. Hay dos visiones principales: IA general (que busca dotar a los ordenadores de capacidades intelectuales humanas) y IA débil (que se centra en tareas específicas tales como el reconocimiento de voz, de imágenes, etc.). Actualmente, la inteligencia artificial ha superado las capacidades humanas en muchas tareas específicas pero la IA general aún es lejana.

Interesado. Persona a quien se refieren los datos.

L

Lealtad (*Fairness*). Principio básico del RGPD. Un tratamiento de datos es leal cuando hace un uso de los datos previsible por parte de los interesados y de este tratamiento no se derivan consecuencias que sean injustificadas. Por ejemplo, un tratamiento leal no debe discriminar por género, color de piel, etc.

M

Modelo. Representación de la realidad subyacente que utilizan los algoritmos de decisión automatizada a la hora de hacer predicciones y tomar decisiones. Este modelo se ajusta a cada caso particular en un proceso de aprendizaje basado en los datos disponibles.

R

Red neuronal. Modelo de aprendizaje automático inspirado en el funcionamiento de las redes de neuronas presentes en los organismos vivos

Reglamento general de protección de datos (RGPD). Regulación Europea sobre el uso de datos personales. Entró en vigor en mayo de 2018. Sustituye la directiva de protección de datos 95/46 / EC y refuerza las garantías que daba la última.

Responsable del tratamiento. Persona u organización que determina la finalidad y el modo en que se deben procesar los datos.

T

Transparencia. Principio básico del RGPD. Para que un tratamiento sea transparente es necesario que los interesados reciban información que sea clara y que haga uso de un lenguaje sencillo.


Generalitat
de Catalunya