La apuesta ética por la inteligencia artificial. Una perspectiva europea

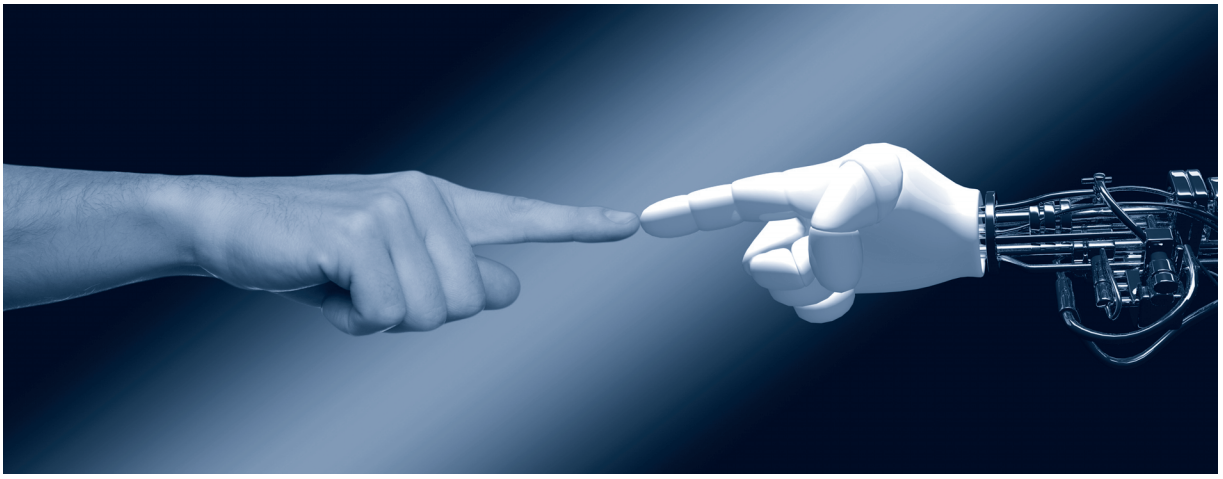
LAS NOTICIAS SOBRE INTELIGENCIA ARTIFICIAL SIEMPRE GENERAN ALARMA, SOBRE TODO ACERCA DE CREAR ENTES AUTÓNOMOS QUE EMULEN A LOS HUMANOS. EL ESTADO ACTUAL DE LA TECNOLOGÍA HACE PENSAR QUE ESO ES TODAVÍA IRREALIZABLE. SIN EMBARGO, EXISTEN RIESGOS EN EL USO DE ESAS TECNOLOGÍAS QUE DEMANDAN UN CONTROL SOCIAL

JOSLAY POLANCO MEDINA

En los últimos años nos han inundado noticias y avances relacionados con la inteligencia artificial. No pocas de ellas alarmantes, y algunas incluso haciendo referencia a la posibilidad de que en pocos años perderemos el control sobre estas tecnologías. Aunque desde el inicio de esta ciencia se ha especulado con el alcance de la inteligencia artificial general, es decir, aquella que pueda tener como objetivo final la creación de entes autónomos, que además sean capaces de aprender por si mismos y tomar decisiones en entornos ambientes cambiantes, emulando a los humanos, la gran mayoría de expertos concluye que –al menos de momento– este tipo de hitos es difícilmente realizable. Por eso, teniendo en cuenta el estado actual de la técnica, es importante ser conscientes que los riesgos que el desarrollo de esta tecnología puede acarrear, no son los que derivan de una superinteligencia, sino los que proceden de IA's específicas. Esto es, las diseñadas e implementadas para resolver problemas concretos, pero que pueden derivar en vulneraciones de los derechos de los ciudadanos y, por tanto, deben ser objeto de control social. Piénsese, por ejemplo, en la utilización de estas tecnologías para controlar el acceso a servicios públicos o beneficios sociales; para determinar quiénes deben acceder a las fronteras europeas; si se debe otorgar o no un beneficio procesal a una persona privada de libertad. Las posibilidades de permeación de la IA y el potencial impacto sobre la vida social es infinito. Por lo que no es extraño que las instituciones comunitarias estén resueltas a crear un marco ético para orientar no sólo la respuesta jurídica, sino también para brindar criterios de orientación al resto de «stakeholders» de la cadena de vida de la IA, donde jugarán un papel fundamental las empresas del sector. No obstante, antes de pasar a comentar –con la brevedad del caso– el mencionado marco ético, conviene aclarar antes que la IA que es objeto de preocupación más inmediata es la IA blanca. Esto es, el *machine learning* en sus diversas variantes.

Las instituciones comunitarias están decididas a crear un marco ético con criterios que orienten a las empresas del sector

En el machine learning la intervención humana es más reducida. Los algoritmos aprenden de los datos proporcionados y extraen patrones o realizan predicciones para resolver problemas de todo tipo



LOS DISTINTOS TIPOS DE INTELIGENCIA ARTIFICIAL

El concepto de machine learning fue planteado por Arthur Samuel en 1959 en la revista IBM Journal. En ese trabajo hacía referencia a un programa de ordenador creado por él, y que aprendía por si mismo a jugar al ajedrez previo «período de entrenamiento». Curiosamente, Samuel reconocía que el programa terminó derrotándole, avizorándose quizás una de las fórmulas para medir los hitos de esta tecnología, que se repetiría en 1996 cuando un algoritmo de la propia IBM venciera al entonces campeón del mundo de este deporte Garry Kasparok.

El machine learning, que forma parte un campo especialmente relevante de la Ciencia de la Inteligencia Artificial, se identifica con el aprendizaje automático por una serie de algoritmos. Cuando nos referimos a automático hacemos referencia a la capacidad de emular una de las características del ser humano: Aprender a partir de los ejemplos, de la experiencia. En este caso, esa experiencia se transmite mediante «datasets» que recogen miles e incluso millones de ejemplos que los algoritmos analizan, y a partir de allí pueden detectar patrones realizar asociaciones que puedan aplicar para nueva información.

Así, por ejemplo, una base de datos con información detallada de miles de pisos (ubicación, número de habitaciones, baños, precios y otras características) puede permitir que un algoritmo de machine learning pueda determinar el precio de otros pisos que no se encuentren en esa base de datos inicial si lo acompañamos de la información apropiada.

Quizás la clave para entender la diferencia entre el machine learning y otro tipo de programas basados en reglas, es que la intervención humana es mucho más reducida, y como se dijo, estos algoritmos aprendan de los datos proporcionados para extraer patrones o correlaciones, o también a realizar predicciones que puedan aplicarse a la resolución de problemas de cualquier tipo.

Como puede adivinarse la elaboración de estos modelos dista mucho de ser sencilla, y exige la conjunción no sólo de programación, sino un avanzado conocimiento matemático y una certera valoración de la capacidad para resolver los problemas en cuestión. Si bien en cierto, que ya se ofertan plataformas de «IA As A Service», y en pocos años es previsible una importante generalización de su uso, se requerirá en cualquier caso de personal cualificado no sólo desde el punto de vista técnico, sino con visión comercial y capacidad de realizar valoraciones éticas.

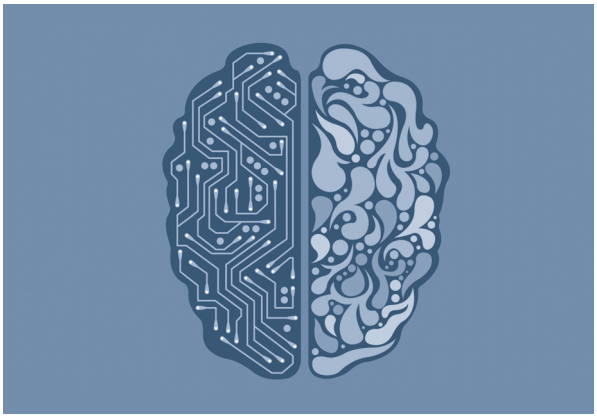
Aunque se suele hablar de machine learning para agrupar distintas soluciones, es importante diferenciar entre modelos de aprendizaje supervisado y no supervisados. La diferencia radica en el tipo de datos utilizados para el «entrenamiento» de los modelos. Así, en el supervisado se «enseña» utilizando datos de entrenamiento que contienen «etiquetas» con información clave para resolver los futuros problema o realizar predicciones. Por tanto, en este tipo de modelos es fundamental contar con bases de datos que recojan la información acertada a partir de la cual el algoritmo aprenderá.

Se trata, por tanto, de modelos en los que la intervención humana guía de manera un poco más intensa al algoritmo al proporcionarle las respuestas correctas. Siguiendo con el ejemplo del modelo de predicción del precio de un piso. Las «etiquetas» serían las variables comentadas y que

se refieren a las características de los pisos (número de habitaciones, metros cuadrados, ubicación, etc.).

En el aprendizaje no supervisado, como puede deducirse, no se transmiten etiquetas en la información de entrenamiento. En este caso, a partir del «dataset» de entrenamiento el propio algoritmo detectaría ciertos patrones que utilizaría, por ejemplo, para clasificar ejemplos futuros. El aprendizaje no supervisado, dada su flexibilidad, está generando avances importantes en muchos campos. Incluso está permitiendo que los modelos puedan aprender a partir de la competencia entre ellas, por ejemplo, mediante redes generativas antagónicas o GAN, que están detrás de otros hitos significativos, y también problemáticos, como los «deep fakes» o ultrafalsificaciones que se han usado para suplantar la identidad de personas reconocidas.

Además del aprendizaje supervisado y no supervisado, también existen modelos intermedios denominados aprendizaje semi–supervisado o aprendizaje por refuerzo. En estos casos, la información de entrenamiento carece de etiquetas y se permite al algoritmo que realice inferencias iniciales, a partir de las cuales recibe una «recompensa» o *feedback* que le orienta sobre las respuestas objetivas que debería alcanzar. Este tipo de algoritmos se ha utilizado para entrenar IA que jueguen de manera autónoma a videojuegos, como se demostró con el experimento de Dota 2, pero más interesante aún, está llamada a contribuir de manera determinante en procesos de automatización industrial, donde ya encontramos ejemplos especialmente significativos como los que ha desarrollado GoogleX con sus «granjas» de brazos robóticos.



Un caso que merece mención aparte, dada su especial problemática, es el aprendizaje profundo o «deep learning». Son los modelos que se han desarrollado tomando como referencia el funcionamiento del cerebro humano y, por tanto, construidos mediante redes neuronales artificiales. Estas redes organizan la información por capas, tanto como sean necesarias, lo que les permite procesar información compleja como imágenes o sonido. Por lo tanto, es la tecnología que está detrás de los asistentes virtuales como Alexa o Siri, pero también de muchas aplicaciones que requieren el procesamiento de imágenes complejas como las que se utilizan en el sector sanitario, y también se usan en tecnologías muy complejas, como las necesarias para permitir la conducción autónoma.

Los modelos desarrollados tomando como referencia el funcionamiento del cerebro humano tienen muchas aplicaciones (asistentes virtuales, procesamiento de imágenes complejas en el sector sanitario o tecnologías más complejas, como la conducción autónoma)

Pese a las bondades que pueden derivar de este tipo de algoritmos, su evolución lleva aparejada una problemática peculiar: son algoritmos de caja negra. Esto supone que en no pocas ocasiones resulta problemático interpretar sus resultados. Por tanto, son poco transparentes, y su uso puede no ser aconsejable en casos donde puedan derivar en la afectación de derechos fundamentales de las personas o se trate de acceso a bienes o servicios esenciales, cuestión sobre la que la normativa ético–jurídico de la Unión Europea ha puesto el foco de su programa normativo.

LA PROPUESTA ÉTICA DE LA UNIÓN EUROPEA PARA LA INTELIGENCIA ARTIFICIAL: UNA TECNOLOGÍA HUMANO–CÉNTRICA

Con miras a responder a los retos de la IA, la Unión Europea creó un Grupo de Expertos de Alto Nivel sobre inteligencia artificial en junio 2018. Este comité de expertos, publicó unas directrices éticas, y se acogió un enfoque novedoso: la IA como tecnología fiable. Es decir, la propuesta promovida por la UE se basa no sólo en un programa ético que responda a los riesgos, para lo cual se recurre al principio de precaución, sino que además pretende extender la apuesta comunitaria por el desarrollo de una industria competitiva fundamentada en el prestigio y la fiabilidad de la industria comunitaria, así como el respeto de los derechos y los valores fundamentales de la UE.

Aunque cabría preguntarse si esto supondrá, de partida, una desventaja en la carrera por el liderazgo de la IA, lo cierto es que la UE apuesta por la IA ética, sea una forma de distinción que a la larga garantice la competitividad de las empresas internas e, incluso, la creación de verdaderos campeones europeos en la materia. Se prevé que el conocido «efecto Bruselas» sea suficiente para convertir la estrategia comunitaria en la referencia ineludible para la implementación de esta tecnología. De ahí que ya se empiece a hablar de dejar atrás fórmulas y mantras que han inundado la economía digital, como el archiconocido «muévete

rápido y rompe cosas», que derivó en el predominio de modelos basados en el MVP (*Mínimum Viable Product*), y se propongan sustituirlos por otros donde destaque el «Minimum Virtuous Product» reflejando la necesidad de potenciar la responsabilidad social empresarial.

No se trata de una experiencia inédita, y basta con recordar el impacto que ha tenido el Reglamento General de Protección de Datos u otras apuestas normativas comunitarias totalmente disruptivas, como el marco comunitario aplicable a la industria química (Reglamento Reach), que se han convertido en una suerte de «Golden Rules» y estándares mínimos ineludibles si se pretende acceder al enorme mercado europeo. Por tanto, no es aventurado sostener que la UE desplegará su potente poder blando para convertir las directrices éticas en los estándares globales de la IA. Cabe recordar que, a diferencia de otras jurisdicciones, la UE tiene estructurados mecanismos de cooperación, y ha reticulado una intrincada red de colaboración que garantiza su presencia inmediata en los principales foros globales y regionales. En estos se apoya, nuevamente, en el prestigio que le otorga fomentar buena parte de sus propuestas en la promoción de los derechos y valores fundamentales.

La IA comunitaria debe tener como orientación fundamental un enfoque centrado en el hombre. Eso salvaguardará la dignidad humana y evitará su instrumentalización

En cuanto a la propuesta específica, se construye sobre la base de principios de alto nivel, pero todos ellos derivan en uno estructural. La IA comunitaria ha de tener como orientación fundamental un enfoque centrado en el hombre. Aunque es previsible que esto supondrá un cuestionamiento por algunos sectores, parece razonable que es un punto de consenso que debe estructurar toda relación del hombre con la tecnología. De esta forma, de partida, colocamos la tecnología al servicio de la humanidad. Esto no es meramente teórico, sino que sirve de sustento para proscribir cualquier tipo de modelo de IA que atente contra la dignidad del hombre o su instrumentalización, cuestión más que evidente a nada que se profundice en el denominado «capitalismo de la atención». Lo mismo puede sostenerse respecto a modelos que pueden derivar en la anulación del individuo y su reducción a una mera estadística, desconociendo el valor intrínseco del ser humano, como el que proponen los modelos de «scoring social» que ya se prueban en algunos países •

ACTIVIDADES Y FOROS

JORNADAS

Son reuniones o mesas de diálogo breves de carácter nacional que se proponen fomentar la relación entre hombres de empresa, académicos y profesionales en general, sobre temas relacionados con los objetivos del Instituto. A lo largo del curso 2021–2022 estarán centradas en el propósito de la empresa.

EL TRABAJO EN BUSCA DE SENTIDO

Universidad de Navarra. Pamplona. Jornada online. Pamplona, 22 de junio de 2021. Ponentes: Michael Pirson (International Humanistic Management), Andrés Sendagorta (Presidente Grupo SENER), Iñaki Vélaz (Instituto Empresa y Humanismo), Manuel Guillén (Director Cátedra de Ética Empresarial IECO–UV).

EL PROPÓSITO CORPORATIVO COMO IMPULSOR DE LA COMPETITIVIDAD

Universidad de Navarra. Jornada online. Pamplona, 18 de noviembre de 2021. Ponentes: José Antonio Alfaro (Universidad de Navarra), José Luis Aldaba (ISS Facility Services) y Álvaro Lleo(Universidad de Navarra).

El propósito corporativo mejora de la competitividad a través de las persona