Las máquinas también necesitan principios éticos

La rápida evolución reciente de la inteligencia artificial ha dejado un vacío moral que evalúe los efectos negativos que puede llegar a tener sobre las personas. Sin embargo, urge crear códigos éticos que puedan limitar los perjuicios derivados de la tecnología, ya sean por mal funcionamiento de la misma o por llevar a cabo un mal uso.

"Nunca envíes a un humano a realizar el trabajo de una máquina"
The Matrix (Wachowski, 1999)

Cada vez vivimos en una sociedad más automatizada. Progresivamente, tendemos a delegar funciones que realizan los humanos en unas máquinas más y más inteligentes. La inteligencia artificial está en todas partes; ya ha escapado de las fábricas y de los centros de tecnología punta, y se aparece en nuestros teléfonos móviles o en los servicios de vídeo bajo demanda que tenemos contratados.

El pasado año desembarcaron en España los altavoces inteligentes y, probablemente, en 2019 conozcan una explosión de las ventas espectacular. Los coches que conducimos tienen una capa de informática, sobre la mecánica y la electrónica de su estructura, que les permite tomar decisiones –todavía menores, como encender y apagar los faros automáticamente cuando resulte necesario- para facilitar y simplificar la conducción humana. Todo a nuestro alrededor se tecnifica.

Las ventajas para la sociedad del desarrollo y la aplicación de los sistemas inteligentes son numerosas, como apunta el responsable del departamento digital de la Comisión Europea Roberto Viola en un reciente artículo: *"la inteligencia artificial esta ya haciéndonos más sanos, dándonos un aire y una energía más limpios, manteniéndonos seguros en la red o viajando y mejorando la calidad de nuestro trabajo"* [1].

Viola destaca la aplicación de estas tecnologías en diversos campos, como, por ejemplo, el cuidado de la salud, donde existen proyectos europeos que aplican robots a la detección del cáncer de mama o para asistir a cirujanos de forma remota mediante realidad virtual, por poner dos ejemplos. También menciona iniciativas relacionadas con la conducción de vehículos, mediante sistemas que permiten cambiar de un piloto humano a uno automático.

Otros campos en los que la inteligencia artificial tiene un papel que jugar son la predicción meteorológica y la prevención de desastres naturales, algo de lo que se ocupará un superordenador, cuya entrada en funcionamiento tendrá lugar en Bolonia en 2020. Una de sus aplicaciones reales en Europa es determinar la mejor localización, en los montes que rodean Barcelona, para garantizar el funcionamiento más eficiente de los aerogeneradores. Y podríamos seguir enumerando experiencias y proyectos…

Sin embargo, a pesar de las bondades que promete traer consigo la innovación tecnológica, siempre existe un temor -a menudo inconsciente, pero cada vez más basado en realidades palpables- de que las máquinas inteligentes pueden acabar produciendo más mal que bien a las personas. Son temas relacionados con la destrucción de empleo y el empobrecimiento de amplios estratos de la sociedad, con el poder que tienen para vigilar y controlar, transgrediendo los límites de la privacidad y la libertad individual, o con los sesgos discriminatorios que pueden desarrollar en los procesos automatizados de toma de decisiones, cuyos efectos pueden afectar negativamente a colectivos y a personas concretas.

Cada vez está más extendido el convencimiento de que la inteligencia artificial, tanto su desarrollo como su ejecución y aplicación, debe estar sujeta un código ético, que garantice que sus decisiones y sus acciones no van a resultar perjudiciales ni dañinas para los humanos. Resulta apremiante dotar a las máquinas de una moral.

La inteligencia artificial en los medios

A pesar de que se publican todos los días artículos ciertamente catastrofistas sobre el impacto de las nuevas tecnologías en nuestras vidas, lo cierto es que la visión que exponen los medios de comunicación de las máquinas inteligentes ha mejorado en los últimos años.

Un informe de la Universidad de Standford ha medido el tono de los artículos que hablan sobre inteligencia artificial en medios no especializados, clasificándolos como neutrales, positivos y negativos. Como se puede observar en el siguiente gráfico, el discurso mediático sobre el tema va abandonando la neutralidad a partir de 2016 y se transforma en abiertamente positivo. Por el contrario, el porcentaje de textos que dan una imagen negativa de esta tecnología se mantiene muy bajo a lo largo de todo el periodo considerado.

Fuente: Stanford University. *The AI Index 2018 Annual Report*

Y, sin embargo, de forma inconsciente los medios nos ofrecen otra visión de la innovación más preocupante, a través de las noticias que publican con frecuencia que levantan problemas de carácter ético –y en ocasiones legal- relacionados con la tecnología.

La imagen siguiente fue presentada al comienzo del AI Now 2018 Symposium, celebrado en Nueva York en octubre, y representa una línea de tiempo de eventos acaecidos (se puede hablar de titulares) relacionados con el sector tecnológico que han levantado dudas éticas.

Podemos encontrar desde los distintos problemas de seguridad de la privacidad de los usuarios de Facebook, hasta los accidentes fatales sufridos por coches autónomos. Desde el vacío legal que experimentan actualmente los sistemas de reconocimiento facial, a la entrada en vigor del Reglamento General de Protección de Datos de la UE.

En suma, se trata de una forma gráfica de demostrar que el desarrollo tecnológico es cualquier cosa menos neutral para las personas, y que la aplicación de unos principios éticos que nos protejan a todos y que limiten los posibles efectos negativos del maquinismo, es una prioridad.

Los grandes riesgos que presenta la tecnología actual

Son numerosos los expertos del sector tecnológico y de actividades asociadas a él que expresan sus miedos y preocupaciones acerca de la inteligencia artificial. Elon Musk, el fundador de empresas de alta tecnología como Tesla, ha expresado en numerosas ocasiones el peligro que entrañan para la humanidad los sistemas inteligentes[2]. El propio Stephen Hawking[3], en sus últimos años de vida, también se manifestó en esa dirección.

En este sentido, el Pew Research Center recabó en el verano de 2018 la opinión sobre el tema de un colectivo formado 979 pioneros tecnológicos, innovadores, desarrolladores, líderes políticos y empresariales, investigadores y activistas. Las principales preocupaciones que éstos expresaron sobre el impacto de la

inteligencia artificial fueron las siguientes:

La pérdida de control de los individuos sobre sus vidas. Acabamos cediendo a las herramientas herméticas basadas en el código aspectos clave de nuestras vidas. Perdemos el control de los procesos al desconocer el funcionamiento de estas complejas herramientas y sacrificamos la independencia, la privacidad y la capacidad de elegir.

Uso descontrolado de los datos por las empresas y los gobiernos. La inteligencia artificial deposita en manos de las compañías y de los poderes públicos poderosos medios de control y manipulación de la gente. Se trata de un aspecto difícil de regular dada la globalidad de las redes y la dispersión de la información que circula a través de ellas.

Pérdida de puestos de trabajo. La disrupción que provoca la llegada de los sistemas inteligentes al entorno laboral derivará en la sustitución de trabajadores por máquinas, creando grandes cifras de desempleo.

Dependencia y pérdida de habilidades. Frente a los que pregonan que la inteligencia artificial aumentará la capacidad del ser humano, existe una opinión contraria que postula que la dependencia de la tecnología erosionará nuestras habilidades y nos hará dependientes de las máquinas, minando nuestra iniciativa.

Desorden mundial. Los más catastrofistas predicen que las estructuras sociopolíticas tradicionales sufrirán daños por el cambio tecnológico y que, en conjunto, el planeta irá cayendo en el caos. El uso militar descontrolado de armas inteligentes, la propaganda y las noticias falsas o el cibercrimen cada vez más sofisticado, son elementos que tienden a desestabilizar el equilibrio del sistema.

Las amenazas a corto plazo

Por otro lado, se nos presentan riesgos mucho más concretos y más inmediatos, que ha destacado el MIT a principios de este año, basándose en sucesos ocurridos recientemente. En concreto habla de:

Los accidentes de los vehículos autónomos. El coche de Uber que atropelló a un peatón en marzo de 2018 en Arizona pone en relieve que, o bien la tecnología de estos automóviles no está lo suficientemente madura para que circulen solos, o que la interacción con el conductor humano es insuficiente. En cualquier caso, actualmente son un factor de riesgo.

Los bots que manipulan la opinión pública y la intención de voto. El caso Cambridge Analytica, que saltó a los medios en marzo del pasado año, demostró cómo se puede manipular la intención de voto del electorado haciendo uso de la información de la gente (en ese caso de los usuarios de Facebook), es decir, explotando adecuadamente el big data.

Armas inteligentes. Los empleados de Google se rebelaron ante la intención de la empresa de vender tecnología a la fuerza aérea de Estados Unidos y consiguieron abortar el acuerdo para participar en el proyecto Maven. Pero el peligro de que los gobiernos –o, peor aún, terroristas- desarrollen armas autónomas sigue allí, y, de hecho, Microsoft y Amazon parecen no tener problemas morales para trabajar en iniciativas en ese campo.

La identificación facial como herramienta de control. Otra de las aplicaciones de la inteligencia artificial que está en boga en la actualidad es el reconocimiento facial. Se trata de una tecnología que puede invadir en derecho a la privacidad de las personas y que puede acumular sesgos que lleven a discriminar a determinados colectivos.

Falsificación de vídeos. Se trata de un tema en el que los algoritmos han demostrado su destreza: crear vídeos falsos de personalidades para desacreditarlas o campañas agresivas de desprestigio para manipular la opinión pública.

La discriminación de los algoritmos. Los sesgos que desarrollan o que llevan en su concepción los programas basados en la inteligencia artificial pueden llevarles a discriminar a colectivos, por ejemplo, por motivos raciales o de género.

Las leyes de la robótica

El escritor de ciencia ficción Isaac Asimov se aventuró, tan pronto como en 1942, a postular tres leyes que el funcionamiento de las máquinas inteligentes debería siempre respetar, destinadas a proteger a los humanos de cualquier perjuicio. Son las siguientes:

1. Un robot no puede hacer daño a una persona o, por omisión, permitir que la persona se haga daño.
2. Las máquinas deben obedecer las órdenes dadas por seres humanos a no ser que infrinjan la primera ley.
3. El robot debe proteger su existencia si dicha existencia no infringe las leyes anteriores.

Asimov propuso este breve código ético en un momento en el que la inteligencia artificial no era más que una fantasía tecnológica, pero, en fechas más recientes, tanto Google como Microsoft han ampliado y actualizado esas tres leyes, que quedan como en la figura siguiente.

Inteligencia artificial fiable: la visión de la Comisión Europea

La preocupación que despiertan los problemas éticos asociados a los sistemas inteligentes ya ha provocado reacciones en organismos e instituciones, orientadas en la dirección de crear un marco normativo que regule su actividad.

Una de las iniciativas más destacadas es la de la Comisión Europea, que en diciembre de 2018 publico un borrador sobre los principios éticos que deben guiar el desarrollo de una inteligencia artificial fiable. De acuerdo con el informe, el concepto “fiable” tiene dos componentes:

1. La inteligencia artificial debe respetar los derechos fundamentales, la regulación aplicable y los principios y valores básicos, asegurando que cumple un “propósito ético”.
2. Debe ser técnicamente robusta y confiable, dado que, aún con buenas intenciones, la falta de maestría tecnológica puede causar un daño no intencionado.

De esta manera, las aplicaciones de una inteligencia artificial fiable deben cumplir una serie de requerimientos:

Responsabilidad. Tienen que existir mecanismos que establezcan la responsabilidad en el caso de que los sistemas causen algún perjuicio y medios previstos para compensar a los damnificados.

Gobernanza de los datos. Dado que la inteligencia artificial hace un uso intensivo de datos, se debe garantizar la calidad de los mismos, evitando que puedan contener sesgos discriminatorios o que puedan ser falsos o maliciosos, y asegurar que son los adecuados para utilizar en cada caso.

Diseño para todos. Los sistemas y aplicaciones deben ser concebidos para ser utilizados por cualquier persona, independientemente de su edad, nivel socioeconómico o situación de discapacidad.

Gobernanza del grado de autonomía de la inteligencia artificial. Las máquinas inteligentes funcionan con un grado de autonomía que depende del grado de sofisticación requerido por la tarea a realizar. Cuanto mayor es el grado de autonomía, mayor debe ser la gobernanza a la que esté sometido el sistema y el número de pruebas que garanticen su óptimo funcionamiento.

No discriminación. Los sesgos discriminatorios, directos o indirectos, presentes en los algoritmos pueden perjudicar seriamente y marginar a colectivos sociales. Es importante eliminar los posibles sesgos existentes, tanto en los conjuntos de datos seleccionados, como en el propio mecanismo de toma de decisiones del sistema.

Respetar la autonomía individual. La inteligencia artificial debe ser capaz de proteger a los ciudadanos de los abusos de poder que los gobiernos puedan llevar a cabo mediante sistemas inteligentes. El objeto es distribuir los beneficios derivados de la tecnología, así como defender la pluralidad de los valores humanos y la autodeterminación y autonomía individuales.

Respeto por la privacidad. La privacidad y la protección de datos deben estar garantizadas en todo el proceso asociado a la inteligencia artificial.

Robustez. Los algoritmos deben ser seguros y fiables, además de lo suficientemente robustos como para gestionar errores o inconsistencias que puedan aparecer durante las fases de diseño, desarrollo, ejecución, difusión y uso.

Seguridad. Un principio que implica que los sistemas harán lo que se espera que hagan, sin dañar a las personas, los recursos o el medio ambiente.

Transparencia. La inteligencia artificial no puede funcionar como una caja negra que solamente los técnicos entienden. Es necesario explicar los mecanismos que utilizan las máquinas para tomar decisiones y para adaptarse a su entorno, así como el origen y la dinámica de los datos que son utilizados y creados por los sistemas.

Un futuro digital inclusivo, fiable y sostenible

El Foro Económico Mundial también ha presentado su propia propuesta para paliar los posibles efectos nocivos de la tecnología. En un informe, publicado a finales del pasado año, apuesta por una sociedad digital inclusiva, fiable y sostenible.

Para alcanzar este objetivo, ha definido seis grandes objetivos basados en esos tres principios que, si bien hablan de la innovación tecnológica en general, se pueden aplicar específicamente en el caso particular de la inteligencia artificial. Son los siguientes:

1. *No dejar a nadir atrás*: asegurar el acceso a un internet de alta calidad para todos.
2. *Empoderar a los usuarios a través de buenas identidades digitales*: asegurar que todo el mundo puede participar de la sociedad digital a través de la identidad y mecanismos de acceso.
3. *Hacer que los negocios trabajen para la gente*: ayudar a las empresas a navegar la disrupción digital y a evolucionar hacia nuevos modelos y prácticas de negocio responsables.
4. *Mantener seguro a todo el mundo*: dar forma a normas y prácticas que construyan un entorno tecnológico seguro y resiliente.
5. *Crear reglas nuevas para un juego nuevo*: crear nuevos mecanismos de gobernanza flexibles, participativos y basados en la experiencia, para complementar la política tradicional y la regulación.
6. *Romper la barrera de los datos*: desarrollar innovaciones que nos permitan beneficiarnos de los datos, a la vez que protegen los intereses legítimos de todas las partes implicadas.

Humanizar la inteligencia artificial

Una última iniciativa en este campo es la que ha lanzado Telefónica, que la ha convertido en una de las primeras empresas del mundo en crear unas pautas éticas sobre inteligencia artificial. El gran objetivo es que esta tecnología pueda garantizar su impacto positivo en la sociedad.

De esta forma, este compromiso de la compañía, presentado en octubre de 2018, implica que todos sus proyectos que incorporen inteligencia artificial deberán ser evaluados con una serie de principios, que garanticen la humanización de la tecnología en beneficio de todos.

Los principios establecidos por Telefónica son los siguientes:

Las aplicaciones de la inteligencia artificial deben arrojar resultados justos, sin sesgos discriminatorios en relación con la raza, el origen étnico, la religión, el género, la orientación sexual, la discapacidad o cualquier otra condición personal.

Igualmente, los sistemas deben ser transparentes y explicables, de forma que los usuarios sepan que están interactuando con inteligencia artificial, qué datos suyos se usan y para qué.

La inteligencia artificial debe estar al servicio de la sociedad y generar beneficios tangibles para las personas, cuyos derechos humanos no pueden verse vulnerados. En este sentido, la empresa se ha propuesto ayudar a cumplir los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas.

Privacidad y seguridad desde el diseño: las políticas de privacidad y seguridad de la compañía cobran en estos principios especial relevancia para preservar los datos tanto personales como anónimos y agregados.

Finalmente, Telefónica se compromete a verificar la lógica y los datos utilizados por los proveedores, para asegurarse de que son ciertos.

[Photo by Negative Space from Pexels]

Anderson, J. Rainie, L. y Luchsinger, A. (2018) “Artificial Intelligence and the Future of Humans”. Pew Research Center. Disponible en: http://www.pewinternet.org/2018/12/10/artificial-intelligence-and-the-future-of-humans/

European Commission (2018) “Draft ethics guidelines for trustworthy AI. Working Document for stakeholders’ consultation”. Disponible en: https://ec.europa.eu/knowledge4policy/publication/draft-ethics-guidelines-trustworthy-ai_en

Hao, K. (2018) “Establishing an AI code of ethics will be harder than people think” en *MIT Technology Review*. Disponible en: https://www.technologyreview.com/s/612318/establishing-an-ai-code-of-ethics-will-be-harder-than-people-think/?utm_campaign=the_algorithm.unpaid.engagement&utm_source=hs_email&utm_medium=email&utm_content=68751142&_hsenc=p2ANqtz-9PpuFG5ee_s_AVa0yS7-Fz7WtRJbfGG2pdbJDlllKyg7agaJuYoU8ab1yoq6ZWPzGC1uUkezd6s8FzSEuPEj8WqhydXrm_T7Sl1MuGU0RRArC8-bE&_hsmi=68751142

Hao, K. y Knight, W. (2019) “Never mind killer robots—here are six real AI dangers to watch out for in 2019” en *MIT Technology Review*. Disponible en:

https://www.technologyreview.com/s/612689/never-mind-killer-robotshere-are-six-real-ai-dangers-to-watch-out-for-in-2019/?utm_campaign=the_download.unpaid.engagement&utm_source=hs_email&utm_medium=email&utm_content=68817960&_hsenc=p2ANqtz-za180woDOZ6_Zuc8J6_yf2uEe5LOq-c6mWyI7_DubiBTzgkT8Q5A2qnaDOPZJvG4aJ8WKAriWIgJMC5PPN7A8jmu0fblaQpguY-gsMMFYL76wGn8&_hsmi=68817960

Rodríguez, P. (2017) "Lo + Visto 7. Inteligencia artificial. Las máquinas que aprenden solas". Fundación Telefónica. Disponible en: https://www.fundaciontelefonica.com/arte_cultura/publicaciones-listado/pagina-item-publicaciones/itempubli/622/

Smith, J., St Amour, L. y O´Halloran, D. (2018) "Our Shared Digital Future. Building an Inclusive, Trustworthy and Sustainable Digital Society". World Economic Forum. Disponible en: http://www3.weforum.org/docs/WEF_Our_Shared_Digital_Future_Report_2018.pdf

Stanford University (2018) "The AI Index 2018 Annual Report" Disponible en: http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf

Telefónica (2018) "Principios de IA de Telefónica". Disponible en: https://www.telefonica.com/es/web/negocio-responsable/nuestros-compromisos/principios-ia

Viola, R. (2018) Artificial intelligence, real benefits. Disponible en: https://ec.europa.eu/digital-single-market/en/news/artificial-intelligence-real-benefits