

La moral de las máquinas: ¿existe?

El avance imparable de la era tecnológica pone sobre la mesa las controversias morales que devienen de la interacción (¿excesiva?) de los seres humanos con la tecnología. Hoy, con la inteligencia artificial copando cada vez más áreas de nuestras vidas, arde el debate que reclama un paso impostergable para regularla: y ya existe una "Iniciativa IA" para entrar en acción.

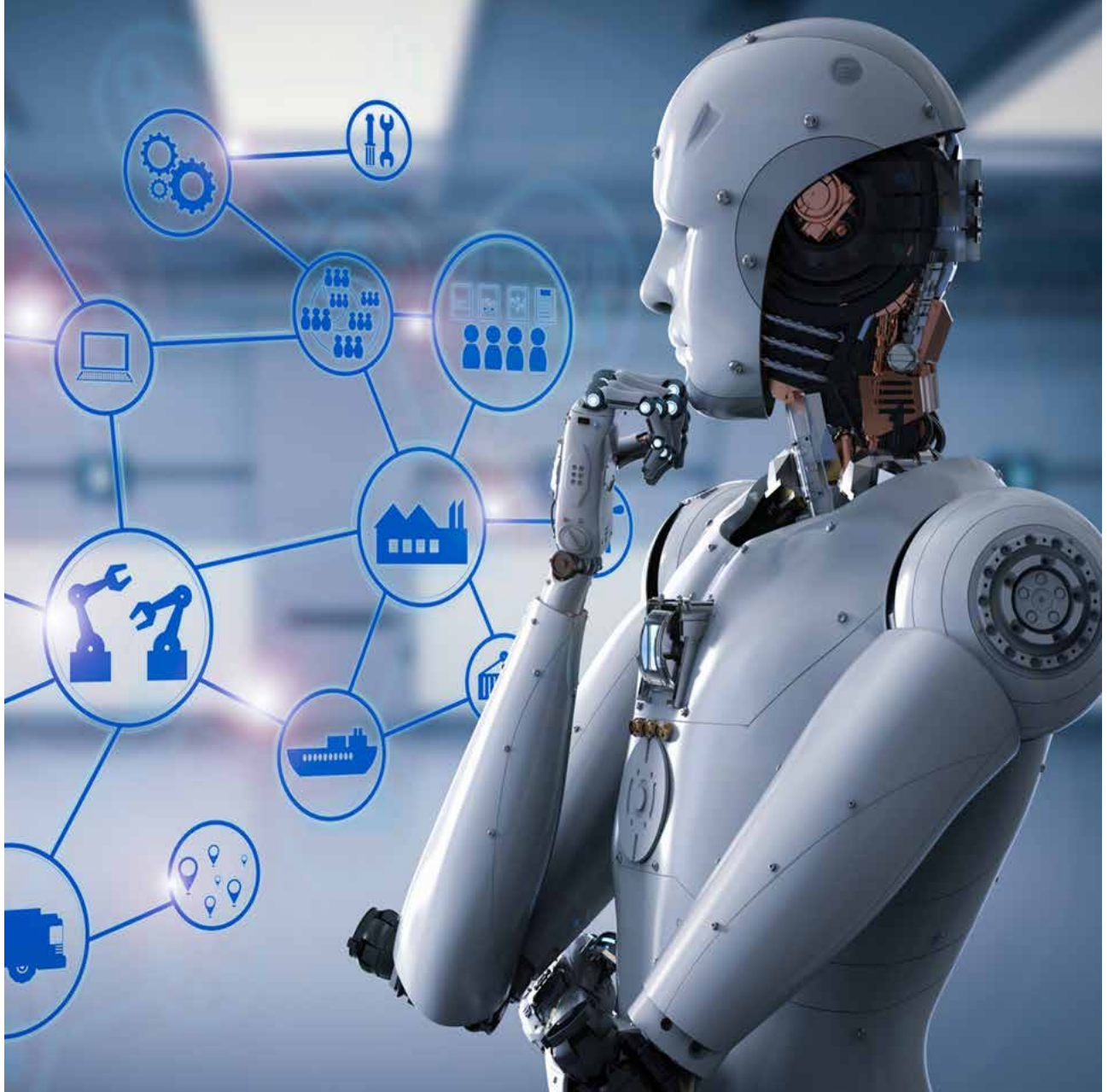
Aún hoy en día, los límites y alcances de la inteligencia artificial están en discusión. Nicolas Mialhe -fundador de *The Future Society* incubada en *Harvard Kennedy School of Government* y cofundador de la *AI Initiative*- afirma que la humanidad se encuentra frente a la revolución de la Inteligencia Artificial (IA), pero asegura que es una revolución que puede funcionar para todos. Si es así ¿cuáles son los pasos a seguir?

La Universidad de Harvard y el MIT dieron el primer paso para que el público entienda el funcionamiento de estas tecnologías de crecimiento acelerado antes de que nos controlen. Porque los avances inusitados de la tecnología traen un cuestionamiento impostergable sobre la moralidad en el uso de las máquinas. Frente a esta pregunta, investigadores del Centro para Internet y la Sociedad Berkman Klein de Harvard se unieron con académicos del MIT Media Lab para lanzar en 2015 la "Iniciativa IA" y estudiar las áreas grises y polémicas que trae el uso

de la inteligencia artificial en la vida cotidiana.

Jonathan Zittrain, profesor y cofundador del Centro Berkman Klein para Internet y la Sociedad de la Universidad de Harvard, y Joichi (Joi) Ito, director del MIT Media Lab -y también propulsor junto con Zittrain del curso "Ética y Gobernanza e Inteligencia Artificial"-, no sólo teorizan; en la práctica, son dos investigadores precursores del *Ethics and Governance of Artificial Intelligence Fund* para la "Iniciativa IA", que cuenta con unos fondos de 27 millones de dólares, según confirma el propio Ito. Se trata de un debate colaborativo al que invitan a participar a todos, para pensar y establecer reglas que puedan gobernar la jungla tecnológica.

En este contexto, los propulsores de esta iniciativa entienden que, si bien los desarrollos tecnológicos son prometedores, pueden ser -también- riesgosos. Entonces, en presencia de límites cada vez más difusos, se hizo patente la necesidad



de crear un marco legal para regular este tipo de prácticas.

La "Iniciativa IA"

En teoría, la experiencia humana debería ser potenciada por la tecnología. Hasta hace pocos años, esta última se pensaba como facilitadora para los usuarios. Sin embargo, las mejoras y las convergencias del *machine learning* y las neurociencias combinadas con la disponibilidad masiva de sets de datos (sumados a la ubicuidad de computadoras escalables de alta performance) están impulsando cada vez más rápido a la humanidad hacia la era de la Inteligencia Artificial (IA) y complejizando así la relación tecnología-usuario. Por eso, expertos

tanto del MIT como de Harvard, que siempre se han caracterizado por liderar colaboraciones exitosas en pos de mejorar la experiencia de las personas gracias al uso de la tecnología, ahora abren camino al debate ético y moral que surge del espacio creciente que la tecnología está ocupando en la vida humana; e impulsan, como resultado, la "Iniciativa IA", un abordaje integral y concreto para hacer frente al debate.

Esta iniciativa propone investigar y recurrir a la lluvia de ideas con el fin de establecer (o por el momento empezar a pensar seriamente) reglas morales y éticas para la inteligencia artificial y otras tecnologías basadas en algoritmos complejos. Con el

objetivo de ayudar a diseñar un marco regulatorio global, bajo esta órbita se unieron académicos, estudiantes y expertos que trabajan en conjunto y funcionan como aliados para lograr que las personas puedan interactuar con la inteligencia artificial en pos de un mundo sostenible, pero sin amenazar la vida cotidiana. En definitiva, reconocen la necesidad de profundizar la investigación de la inteligencia artificial (IA) para el bien público.

Lo cierto es que la carrera global hacia la inteligencia artificial necesita que líderes y decisores políticos empiecen a desarrollar una respuesta que se traduzca en acciones, ya que la IA tiene un impacto cuya magnitud no se

puede subestimar. Los más temerosos asocian esta evolución tecnológica con un peligro potencial para el que es necesario elaborar un marco de contención.

La "Iniciativa IA" prevé:

- **la realización de reuniones anuales donde participen grupos de trabajo interdisciplinarios de sectores diversos: público, privado, social, académico y, fundamentalmente, expertos que puedan advertir e interpretar acertadamente los avances disruptivos de la inteligencia artificial.**
- **la creación de un acuerdo colaborativo global respecto a recaudos positivos, estándares de transparencia y construcción continua de medidas de confianza.**
- **la implementación por parte del gobierno, y junto con otros actores estratégicos del medio (empresas, ONGs), de reglas que se acuerden tanto a nivel local como internacional para regular la IA.**

Jonathan Zittrain sugiere que la "Iniciativa Inteligencia Artificial" podría hasta crear nuevas plataformas, algoritmos y agregadores de contenido. De acuerdo al escritor Erick Trickey en su artículo *Morality in the Machines*, Zittrain aspira a producir discusiones de espíritu público para que "todos, desde ingenieros hasta usuarios de la tecnología que están siendo afectados sin saberlo, puedan tener mayor conciencia de las decisiones que se tomaron en el diseño

de estas tecnologías". Así, Zittrain insiste en la importancia de mantener la conversación abierta y de desafiar lo que la tecnología impone a un usuario en desventaja.

Si se logra que el debate sobre la inteligencia artificial sea colaborativo, estaremos ante la participación de gran parte del mundo, teniendo en cuenta que 2.1 de 7.6 billones de la población mundial total son usuarios de Facebook. De hecho, en el MIT Media Lab, en vistas de allanar este camino, los investigadores ya han creado GOBO, un agregador de redes sociales con filtros transparentes controlados por el usuario.

Casos controversiales de la IA

Algunos ejemplos que sirvieron de pie para dar forma a la "Iniciativa IA" fueron las controversias generadas por los vehículos de conducción automática, las noticias falsas en las redes sociales y los algoritmos para controlar la justicia criminal.

En el primer caso, por ejemplo, cuando los frenos de un vehículo autónomo fallan, ¿contra quién debe chocar? ¿debe atender contra sus pasajeros o desviarse sobre los peatones? Para intentar dar respuesta a esta pregunta, el MIT desarrolló una plataforma llamada *Moral Machine*, que tiene como fin recolectar perspectiva humana para nutrir las decisiones morales que toman las máquinas conducidas con inteligencia artificial, como estos autos. Por ejemplo, la plataforma presenta dilemas morales: si el vehículo autónomo debe elegir el mal menor ¿es preferible matar a dos pasajeros en vez de a cinco peatones? Hoy a modo meramente exploratorio, la plataforma propone, entonces, distintos escenarios para los cuales se invita al observador externo a aportar su visión humana frente a un dilema moral. Sin respuesta cierta, se abre lugar para un debate ético y moral en

el que la clínica *Cyberlaw* de Harvard ya ofrece asesoramiento a startups y organizaciones sin fines de lucro sobre la responsabilidad legal respecto a autos autónomos.

Por su parte, Barbara Grosz, profesora de computación y Bonnie Docherty, profesora de Derecho, ambas de Harvard, pioneras en investigaciones sobre inteligencia artificial y ética, sugieren que exista una prohibición preventiva para el uso de "armas" totalmente autónomas. Más aún, en el control del crimen, el uso de la tecnología también despierta controversias. Según afirma Trickey, las herramientas computarizadas para evaluación de riesgo criminal han sido usadas por décadas para predecir la conducta futura del defendido, sea cual fuere: rehabilitarse, escapar, cometer otro crimen, etc. De todas maneras, estas herramientas tienen adeptos y opositores. Si bien podrían ayudar a los jueces a tomar decisiones más justas y menos parciales, otros afirman que el sistema confidencial de algoritmos es una trampa en sí mismo que puede magnificar las imparcialidades en base al propio sistema de justicia que los crea. Vale aclarar, para avivar el debate, que los algoritmos que usa el sistema judicial para medir el riesgo, no presenten una solución definitiva. La pregunta permanece: ¿Cuán justas son entonces estas herramientas de control criminal?

Incluso, el programa del MIT *Humanizing AI in Law* (HAL) está investigando cómo los jueces de Kentucky usan los instrumentos de evaluación de riesgos. Zittrain junto con investigadores del HAL, en su estudio *Interventions over Predictions*, argumenta que los algoritmos no deberían usarse meramente para predecir crímenes futuros, sino que los gobiernos deberían usar el aprendizaje automatizado (*machine learning*) para analizar las causas de raíz de los crímenes y, así, encontrar



formas de “romper ciclos de criminalidad”. Además, desde la Legislatura de Massachusetts (MA), advierten que los instrumentos digitales de evaluación criminal no deben tomarse a la ligera. Como se ha dicho, estos pueden acentuar parcialidades generadas por el propio sistema que pretende evitarlas; y además, estos prejuicios pueden manifestarse también en materia de género y raza. Entonces, frente a estas fragilidades, la Legislatura de MA propone desarrollar un programa propio a través de un proceso público y abierto para hacerle mella al desafío de usar incorrectamente las herramientas de medición de riesgo criminal.

Las noticias falsas (*fake news*) son un tercer ejemplo conflictivo cuando se combinan con las redes sociales. Las elecciones presidenciales de 2016 en los Estados Unidos dieron pie a cuestionar la calidad de información a la que nos exponemos diariamente a través de las redes sociales y abrir el juego a la transparencia que necesitan las fuentes de información para hacer frente a la manipulación y a otros

actores propagandistas. Así, entra en tela de juicio también el efecto de las redes sociales en política. En vista de la manipulación de las noticias en manos de la tecnología, Jonathan Zittrain se fortalece como el líder que sigue empujando la “Iniciativa IA” a fin de generar proyectos para mejorar la calidad de la información periodística y su veracidad.

En definitiva, la “Iniciativa IA” abre un espacio de debate para hacer frente a preguntas controversiales a las que es necesario dar respuesta lo antes posible. Lo que se escriba hoy en términos de regulación, vaticina Zittrain, podría controlar lo que suceda en 2025. Por eso, anhela la implementación de la “Iniciativa IA”, que permita trabajar siempre de manera colaborativa y democratizar los algoritmos secretos de la tecnología (ya sean redes sociales, inteligencia artificial o plataformas similares). “Esta iniciativa implica desarrollar y expandir relaciones con esas compañías, a la vez que se mantenga también la distancia apropiada a fin de ser capaces de reportar de manera independiente sobre lo que vemos

y sobre lo que pensamos que debe ocurrir”, explica Zittrain.

Frente al impacto desmedido de la tecnología en la vida humana, es clara la necesidad de regular la intervención tecnológica y desarrollar estrategias colectivas para inventar y reinventar la forma en que los seres humanos experimentamos la tecnología y cómo ésta puede ayudarnos.

Entonces, aunque todavía esté en duda su alcance, la Iniciativa se presenta como una herramienta concreta para sostener una conversación abierta; además predispone a entrar en movimiento en la dirección que promueva el uso saludable de la tecnología. •

WOB