

# Ética en la Inteligencia Artificial

## Edgar Ortiz Loyola Rivera Melo

Actualmente es el director del Departamento de Estudios en Ingeniería para la Innovación de la Universidad Iberoamericana. Egresó de la Licenciatura en Ingeniería

Electrónica y de Comunicaciones de la Ibero en 1989. Realizó la Maestría en Telecomunicaciones y Sistemas de Información por la Universidad de Essex, en el Reino Unido. Fue uno de los pioneros en el diseño e implementación de la red de computadoras de la Ibero entre 1989 y 1991. Trabajó en McKinsey & Co. México como Coordinador de Cómputo siendo asesor en tecnología de los consultores.



## José Antonio Morfín Rojas

Ingeniero en Electrónica y de Comunicaciones por la Ibero y maestro en Ingeniería por la University of Canterbury, NZ. Profesor de los programas de Ingeniería Electrónica

de la Ibero desde 1990. Ha vivido la evolución de la tecnología electrónica desde el esplendor de la etapa analógica hasta hoy en que estamos inmersos en la tecnología digital. Ha estudiado el desarrollo exponencial de la tecnología, la historia de la tecnología y de la innovación. Ha participado en la incubación de empresas de base tecnológica.



## Bernard Roeland van der Mersch Huerta Romo

Egresado de la Universidad Iberoamericana de la carrera de Ingeniería

Electrónica y de Comunicaciones. Realizó la Maestría en Calidad en Sistemas Electrónicos en la Universidad de las Américas. En la Ibero ha sido coordinador del programa de ingeniería electrónica de 1992 a 1997, y jefe de laboratorios de electrónica en varias ocasiones. Actualmente, y desde 2016, es el director de Servicios Escolares de la Ibero Ciudad de México.



**L**a Inteligencia Artificial (IA), al igual que otras herramientas tecnológicas que el ser humano ha desarrollado a lo largo de su historia (como la imprenta,

la máquina de vapor, el cómputo y las telecomunicaciones, por mencionar algunas), emerge como una aptitud creada, con la capacidad de transformar profundamente a la humanidad, integrándose cada vez más en la forma en que vivimos, trabajamos y nos comunicamos. Desde los primeros asistentes de voz y los “chatbots”, hasta los automóviles autónomos, la IA va modificando día a día diversos aspectos de nuestras vidas. Sin embargo, mientras los sistemas basados en IA se vuelven cada vez más sofisticados y comunes en todos lados, comienzan a surgir preocupaciones éticas relacionadas con su uso, alrededor de la dignidad humana, la autonomía, la discriminación, la privacidad y la justicia social, entre otras.

En este artículo exploraremos algunas de las definiciones que nos permitan aclarar los conceptos principales alrededor de la Inteligencia Artificial como una herramienta tecnológica. Analizaremos las implicaciones éticas en su uso, estableciendo posteriormente la necesidad de definir marcos de referencia y reglas para su desarrollo e implementación, que permitan promover al ser humano, el bien común y la sustentabilidad ambiental.

La definición de “inteligencia” es un tema en continuo debate y existen diversas aproximaciones para establecer el concepto. Sin embargo, la gran mayoría de las definiciones que enlistan una serie de funciones cognitivas, incluyen la “capacidad de aprender”. La tecnología se manifiesta cuando aplicamos nuestros conocimientos y habilidades usando herramientas, que permiten mejorar nuestras



Fotografía: Adobe Stock.

aptitudes naturales. Sumando la actitud que tengamos sobre la herramienta, su uso se podrá considerar bueno o malo, dependiendo de los marcos éticos que tengamos. [Cf. referencia 1]

La Inteligencia Artificial, sin considerar las historias distópicas de las películas y la ciencia ficción, se define en el diccionario Oxford como: “la teoría y el desarrollo de sistemas de cómputo, capaces de desempeñar tareas que normalmente requieren inteligencia humana, como la percepción visual, el reconocimiento de voz, la toma de decisiones y la traducción entre idiomas”. [Cf. ref. 2]

La diferencia principal entre un sistema desarrollado utilizando IA y un sistema de cómputo tradicional radica en la palabra “aprendizaje”. Mientras los sistemas de cómputo tradicionales se sustentan en secuencias de reglas y decisiones sobre las que el programador tiene un control absoluto del flujo del código, los sistemas que utilizan IA se basan en procesos de “aprendizaje” por parte de alguna estructura de cómputo. Las Redes Neuronales Artificiales (RNA) se entrenan a partir de un número grande de ejemplos. El resultado es una matriz de números reales determinados durante dicho entrenamiento.

A esto le llamamos “aprendizaje”. Como ejemplo, consideremos el ChatGPT de OpenAI<sup>1</sup>, que utiliza un modelo para la generación de texto llamado GPT-3. Este sistema contiene 175 mil millones de valores, escritos en forma de números reales, en sus matrices. Para tener una idea de lo que es entrenar un sistema de este tipo, el número de FLOP<sup>2</sup> que se requirieron para entrenar el modelo GPT-3 fue de 314 millones de petaFLOP<sup>3</sup>.

Existen 3 niveles de Inteligencia Artificial [cf. ref. 3]:

- **Inteligencia Artificial Especializada o Anagosta: Diseñada para ejecutar una función específica.** Es la IA con la que contamos hoy

<sup>1</sup> <https://www.openai.com>

<sup>2</sup> La cantidad de operaciones aritméticas de números reales se conoce como FLOP (Floating Point Operation).

<sup>3</sup> “The rise of artificial intelligence over the last 8 decades”. <https://ourworldindata.org/brief-history-of-ai> : 314,000,000,000,000,000,000 FLOP. (1 petaFLOP=1x10<sup>15</sup> FLOP)



Fotografía: Adobe Stock.

**Mientras los sistemas basados en Inteligencia Artificial se vuelven cada vez más sofisticados y comunes en todos lados, comienzan a surgir preocupaciones éticas relacionadas con su uso, alrededor de la dignidad humana, la autonomía, la discriminación, la privacidad y la justicia social, entre otras.**

en día. Como ejemplo, un sistema puede entrenarse para jugar un juego de mesa como el Go<sup>4</sup>. Su capacidad de procesamiento y almacenamiento le permite jugar y vencer incluso a los mejores humanos en la disciplina, sin embargo, no es capaz de distinguir en una fotografía, un perro o un gato, pues está diseñado para llevar a cabo una tarea específica diferente. Este nivel de IA también es usado para reconocimiento

de imágenes, rostros, voz y lenguaje natural, traducción entre distintos idiomas y generadores de texto, por mencionar algunos.

- **Inteligencia Artificial General (IAG).** Este tipo de IA puede llevar a cabo diversas tareas como lo haría un ser humano. Emulando nuestro comportamiento, reconociendo rostros, entendiendo el lenguaje natural, hablando varios idiomas, manteniendo una conversación, cantando, contando chistes y hasta resolviendo problemas. Este nivel de inteligencia es capaz de aprender nuevas tareas conforme se le presentan, aunque nunca las hubiera conocido antes. Este nivel de IA no existe aún y se podría considerar que falta mucho para lograrlo. Se dificulta hacer un pronóstico de cuánto es “mucho tiempo” ya que el desarrollo de la tecnología es exponencial. [Cf. ref. 4]
- **Superinteligencia.** Nick Bostrom<sup>5</sup>: define superinteligencia como “un intelecto que es mucho más listo que los mejores cerebros humanos en todos los campos,



Fotografía: Adobe Stock.

## La diferencia principal entre un sistema desarrollado utilizando Inteligencia Artificial y un sistema de cómputo tradicional radica en la palabra “aprendizaje”.

incluyendo creatividad, sabiduría y habilidades sociales”. La superinteligencia supone un sistema que es ligeramente más inteligente que un humano, pero podría llegar a ser más inteligente que toda la raza humana. Algunos expertos teorizan que, si esto ocurre, alcanzaremos lo que se le conoce como la Singularidad. Este tipo de IA genera una gran inquietud sobre las posibles consecuencias para el ser humano, ya sea hacia su destrucción o hacia el desarrollo de un mundo utópicamente perfecto. El debate se da en el sentido del comportamiento de esta inteligencia. ¿Será buena o mala? Para llegar a ese debate, primero es importante entender el “comportamiento” de un sistema de

IA Específico. Los errores o faltas éticas que puede cometer, ya sea por cómo fue diseñada o por el propósito con que se le utiliza. Veamos algunos casos bajo la óptica de los valores que promueve la Ibero. [Cf. ref. 6]

### Sobre la dignidad de las personas

En el corazón de los valores de la Ibero se encuentra el compromiso inherente con la dignidad de la persona sin importar la raza, el género, la religión o el nivel socioeconómico. El pensamiento jesuita enfatiza siempre la importancia de tratar a todas las personas con respeto, compasión y empatía. Desde esta perspectiva, la IA debe ser diseñada e implementada de forma que favorezca la dignidad de cada ser humano.

Una de las consideraciones éticas más importantes en la IA es el peligro potencial de que ésta viole la dignidad humana. Un sistema de IA puede reforzar y amplificar sesgos existentes e inequidades en nuestra sociedad, perpetuando la discriminación y la injusticia. Por ejemplo, en los sistemas de reconocimiento facial con fotografías de la población de los EEUU<sup>6</sup> se ha demostrado que existe menos precisión para identificar gente con tonos de piel más oscuros y como resultado los falsos positivos han derivado en arrestos erróneos e injustificados. Del mismo modo, algunos sistemas de IA utilizados en los procesos de contratación de trabajadores pueden perpetuar los sesgos raciales o de género si éstos son entrenados con datos igualmente sesgados.

Para evitar estos problemas, los ingenieros que desarrollan sistemas de IA deben priorizar la equidad e inclusión en sus conjuntos de datos y modelos de entrenamiento. Deben, además, construir el sistema con transparencia y asu-

<sup>6</sup> <https://www.france24.com/es/20200624-denuncian-arresto-erróneo-de-hombre-negro-por-sistema-de-reconocimiento-facial>, <https://es.wired.com/articulos/software-de-reconocimiento-facial-puede-conducir-a-detenciones-erroneas-de-personas-inocentes>

**El pensamiento jesuita enfatiza siempre la importancia de tratar a todas las personas con respeto, compasión y empatía. Desde esta perspectiva, la Inteligencia Artificial debe ser diseñada e implementada de forma que respete la dignidad de cada ser humano.**

miendo la responsabilidad sobre los sistemas desarrollados, a fin de que los usuarios puedan entender la manera en la que dicho sistema de IA toma las decisiones, y poder corregir los sesgos o errores que puedan surgir. De esta forma, la IA puede convertirse en una herramienta que promueva la dignidad humana y la justicia social, en lugar de ser una fuente de discriminación e injusticia.

### **Sobre la autonomía y libertad del ser humano**

Otro aspecto importante de los valores en la Ibero es el de la autonomía y la libertad de las personas, la habilidad de cada una para tomar sus propias decisiones y llevar a cabo acciones que reflejen sus valores y metas. La Inteligencia Artificial puede enriquecer la autonomía de las personas si nos provee de nuevas herramientas e información, pero puede también convertirse en un sistema manipulador y coercitivo a través del uso de tecnología persuasiva. Por ejemplo, los algoritmos de redes sociales pueden mostrarnos contenidos que refuercen nuestros valores y creencias, e incluso sesgos, mientras nos oculta puntos de vista opuestos o diversos. Esto puede crear visiones muy cortas y sesgadas que limiten nuestra exposición a ideas y perspectivas diversas, haciendo más difícil para nosotros tomar decisiones informadas.

Para promover la autonomía de las personas en la era de la IA, los desarrolladores de estos sistemas deben construirlos de forma que prioricen el conocimiento informado, la transparencia y el control del usuario. Los usuarios deben tener la facultad de enten-

der cómo el sistema de IA toma decisiones y decidir si aceptan o modifican las decisiones conforme éste lo requiera. Esto exige un entendimiento profundo de las implicaciones éticas de la IA, así como un compromiso para diseñar sistemas de IA que promuevan el desarrollo del ser humano en lugar de buscar maximizar el rendimiento o la eficiencia.

### **Sobre el bien común**

La Inteligencia Artificial debe ser diseñada y desplegada de forma que promueva el bien común y la sustentabilidad ambiental. El bien común, de la mano de la sustentabilidad ambiental, es la idea de que todos estamos conectados y nuestras acciones deben ser guiadas por una preocupación hacia el bienestar de todos y del planeta. Esto significa que la IA debe ser utilizada para buscar el bien mayor, el que alcance a más personas y no sólo a los intereses de unos cuantos.

Una de las aplicaciones más promisorias de la IA recientemente surge en los sistemas de salud, ya que tiene el potencial de mejorar la salud de los pacientes, reducir costos e incrementar el acceso a más gente a servicios médicos. La IA aplicada a la salud debe diseñarse con las necesidades de los pacientes en mente y debe además priorizar su privacidad y confidencialidad.

Algunos organismos internacionales como la Unión Europea [cf. ref. 7], la Unesco [cf. ref. 8] o la NASA [cf. ref. 9], han propuesto marcos de referencia éticos para la IA. A continuación, se enlistan algunos principios y valores propuestos:

- § Respeto, protección y promoción de los derechos humanos, libertades fundamentales y dignidad humana.
- § Cuidado del medioambiente y el ser humano.
- § Asegurar la diversidad e inclusión.
- § Promover sociedades justas, pacíficas e interconectadas.
- § Autonomía restringida que promueve la seguridad de las personas.
- § Responsabilidad compartida y colegiada con supervisión y determinación por seres humanos.

- § Seguridad y protección.
- § Evitar sesgos discriminatorios y asegurar accesibilidad y justicia.
- § Sostenibilidad.
- § Derecho a la privacidad y protección de datos.
- § Transparencia y explicabilidad.
- § Concientización y educación sobre la Inteligencia Artificial.

La tecnología se desarrolla tan rápidamente que es indispensable diseñar y probar los marcos de referencia éticos de forma que, cuando evolucionemos a las siguientes etapas, se tenga una Inteligencia Artificial acotada por dichos marcos.


Para finalizar, queremos retomar la idea del escritor científico Isaac Asimov en la que se considera su última obra, *La receta del tiranosaurio*. Podemos vislumbrar dos escenarios opuestos hacia el futuro de nuestra sociedad: no entender nuestras propias capacidades, continuar con los *egosistemas* y por lo tanto llevar acciones que empeoren nuestra situación a grados extremos o desarrollar verdaderos ecosistemas que nos permitan florecer como humanidad. Es más interesante hablar de cómo sería la vida en el futuro si logramos crecer y continuar nuestro desarrollo como sociedad.

Para este segundo escenario, la tecnología evolucionará de tal forma que abastecerá las necesidades básicas del ser humano, permitiéndonos un tiempo mayor de ocio. Las personas podrán buscar, a través de la educación y el aprendizaje, nuevas actividades que las motiven, dándoles un mayor sentido y valor, sin tener que preocuparse por realizar actividades tediosas y repetitivas. Es decir, existirá la posibilidad de continuar con un gusto por su propia instrucción a lo largo de su vida y no únicamente por un período corto de tiempo, para alcanzar un medio “inmediato” de subsistencia.

Siguiendo los lineamientos descritos y participando activamente sobre las implicaciones éticas de la Inteligencia Artificial podemos asegurar que su desarrollo sea utilizado para promover el bienestar de las personas, la justicia social y un futuro sustentable para todos.



Fotografía: Adobe Stock.

La Ibero, consciente de la importancia del estudio y desarrollo de personas que sean capaces de diseñar, implementar y sostener estas tecnologías, desde un comportamiento ético y sustentado en los valores ignacianos, iniciará en agosto la Licenciatura en Ingeniería en Ciencia de Datos. 

## Referencias

- [1] W. Brian, Arthur, *The Nature of Technology: What it is and How it Evolves*, Free Press, USA, 2009.
- [2] <https://www.oxfordlearnersdictionaries.com/us/definition/english/intelligence?q=intelligence>
- [3] <https://codebots.com/artificial-intelligence/the-3-types-of-ai-is-the-third-even-possible>
- [4] Kurzweil, Ray, *The Singularity is Near*, Viking Press, 2005.
- [5] <http://www.nickbostrom.com/superintelligence.html>
- [6] <https://ibero.mx/sites/all/themes/ibero/descargables/corpus/co588.pdf>
- [7] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- [8] UNESCO, *The Ethics of Artificial Intelligence*, 2022.
- [9] NASA/TM-20210012886, NASA Framework for the Ethical Use of Artificial Intelligence (AI).