Reseña del libro "Ética para máquinas" de José Ignacio Latorre

doi: 10.52749/fh.v2i4.7

ANGEL CROVETTO

<https://orcid.org/0000-0001-6358-0354>

Licenciado en Sociología por la Universidad Nacional Federico Villarreal. Egresado de la Maestría en Educación por la Universidad Antonio Ruiz de Montoya. Director de la revista Futuro Hoy del Instituto de Extrapolítica y Transhumanismo de la Sociedad Secular Humanista del Perú. Actualmente se desempeña como Docente en CENFOTUR e investigador del grupo de estudios IET.

angelcrovetto1@gmail.com angelcrovetto.com

Resumen. "Ética para máquinas" de José Ignacio Latorre es un recorrido profundo del desarrollo y la complejidad de la ética aplicada a la Inteligencia Artificial. Con un lenguaje amigable, que, a mi modo de ver, propone una pregunta simple, semánticamente hablando, pero muy difícil de responder: ¿qué vamos hacer cuando las máquinas nos superen?

Palabras clave: Inteligencia Artificial, ética, José Ignacio Latorre, Ética para máquinas

Aproximaciones

A lo largo de la existencia humana hemos intentado siempre superar nuestras capacidades; desde la invención de la rueda, hasta el cohete que nos llevó a la luna. Hemos creado ciudades gigantescas que actualmente albergan a millones de personas. Creamos un sistema de vida basado en el hiperconsumo, bajo el enfoque de la obsolescencia programada. Hemos creado una de las mega autopistas del conocimiento que revoluciona la vida humana, el Internet... Y pienso que producto de todos estos cambios que orientan el progreso humano es que se ha evolucionado al transhumanismo, el cual aborda tres complejos campos del saber y del desarrollo tecnológico [1]: extender sin límite la vida de los seres humanos, elevar el bienestar y potenciar la inteligencia. En este último campo puede ubicarse el análisis que realiza Latorre en su libro Ética para máquinas.

En general el autor describe de forma clara los hitos de la innovación tecnológica que marcaron la evolución de la computación que llevaron a la creación de la Inteligencia Artificial. Latorre plantea una mirada reservada sobre el futuro de la humanidad ya que pone énfasis en que este cambio depende directamente de nosotros – los seres humanos – quienes no hemos tomado las mejores decisiones en el siglo veinte. Y no creo que estemos mejorando nuestro puntaje en los inicios del pre-

sente siglo.

Latorre divide el libro en 5 partes: 1. Máquinas sin alma; 2. Máquinas que parecen inteligentes; 3. Ética para máquinas; 4. Máquinas que nos superarán y 5. Imitar el alma. Para efectos de este artículo de opinión o reseña de libro, deseo enfocarme en las dos últimas partes que – a mi parecer – centran la polémica actual sobre la relación del ser humano y la Inteligencia Artificial: 4. Máquinas que nos superarán y 5. Imitar el alma

Con gran perspicacia, Latorre inicia citando a Nick Bostrom:

"Definamos una máquina ultrainteligente como aquella que puede superar en todas las actividades intelectuales a cualquier humano, por listo que este sea. Dado que el diseño de máquinas es una de estas actividades, una máquina ultrainteligente podría diseñar máquinas aún mejores. Incuestionablemente, habría una explosión de inteligencia, y la inteligencia humana quedaría totalmente rezagada. En consecuencia, la primera máquina ultrainteligente será la última invención que los hombres podrán hacer, asumiendo que esa máquina sea suficientemente dócil como para permitir que mantengamos su control."

Al núcleo de esta reflexión puede agregarse la convergencia de dos tecnologías, una especie de sumatoria sinérgica de esfuerzos potenciados

siempre por nuestros deseos de crear: la inteligencia artificial fuerte y la computación cuántica. Es cierto que ambas tecnologías aún son experimentales, y Latorre aclara que los límites entre ciencia y ciencia ficción pueden ser poco claros, por lo que afirma: "Hay que ser muy cautos al especular sobre lo que es posible o lo que sucederá". Estas dos tecnologías – inteligencia artificial fuerte y computación cuántica – definen una frontera, están justo en el límite donde la ciencia ficción se convierte en ciencia.

La creación de las ultrainteligencias

Hal 9000 en Odisea del Espacio o Data de Star Trek - Nueva Generación, personajes que son parte de la ciencia ficción, nos llevan a la pregunta: ¿hasta cuándo seguirán siendo parte del reino de la imaginación? La respuesta a esta pregunta tiene varios puntos que deben ser revisados, entre ellos, el "alma" y la "conciencia" de las máquinas.

Cuando Latorre explora el "alma" de las máquinas la asocia a una capacidad de ser conscientes de sí mismas, de saberse parte de las variables que debe considerar para crear escenarios predecibles. Así pasa de ser una cosa a un actor que su sola acción modifica la realidad. Esta capacidad de ser consciente de su influencia y actuar en la realidad es lo que Latorre llama "alma".

Latorre deja entrever que es una posibilidad que las máquinas cobren algún tipo de conciencia de sí mismas, y que no es del todo un simple sueño de ciencia ficción. Cada vez más las máquinas pueden "simular" nuestra conducta, pero la capacidad de distinguir lo bueno y malo está asociada a una combinación compleja de emocionalidad y pensamiento ético que las máquinas no han logrado desarrollar aún; con todo, las máquinas podrían imitar esta capacidad próximamente. Ahora, cuanto más antropomorfizamos las conductas de las máquinas, éstas generan sensaciones extrañas en los seres humanos, lo cual responde al concepto del valle inquietante, propuesto por Masashiro Mori.

Latorre más bien explora la programación de la comprensión ética de las máquinas usando los dos enfoques morales más conocidos: el imperativo categórico kantiano y el utilitarismo. Sin embargo, siempre queda el problema de quién supervisa éticamente al programador de ética, así como la cuestión de quién garantiza que las máquinas opten por una aplicación ética de estos enfoques, en caso de que logren aprender solas.

Latorre propone algunos caminos para acercar la conducta casi humana de las máquinas y su inter-

acción con los seres humanos. Parafraseando, propone el control ético, programación ética o aprendizaje ético: en el primer caso, es necesario que los seres humanos no pierdan la capacidad de apagar el sistema; en el segundo camino, aparece la necesidad de crear un algoritmo que permita a las máquinas diferenciar lo bueno y lo malo en una situación concreta; por último, nos queda dejar que las máquinas puedan tomar sus propias decisiones a partir de la revisión de las mejores interpretaciones éticas de la realidad logradas por el ser humano, logrando esto a través de complejas redes neuronales y *deep learning*. En cualquiera de los casos, Latorre propone de forma explícita que debemos crear mecanismos y/o sistemas que acerquen a las máquinas a una conducta ética que nos ayude a seguir mejorando nuestra vida. El punto aquí es: ¿tenemos los seres humanos, la capacidad de enseñar a las máquinas a conducirse de forma ética? Si nosotros mismos no podemos autorregular nuestras conductas bajo la reflexión ética, ¿cómo podemos enseñar a las máquinas a autorregularse?

A manera de conclusión

Finalmente, el libro *Ética para máquinas*, de José Ignacio Latorre, es un recorrido profundo, con un lenguaje amigable, que, a mi modo de ver, propone una pregunta simple, semánticamente hablando, pero muy difícil de responder: ¿qué vamos hacer cuando las máquinas nos superen? Por ahora es solo una pregunta retórica, pero, ¿cuánto tiempo nos queda para enfrentar la situación y tener una respuesta?

[1] Los defensores de esta corriente del pensamiento afirman que es una consecuencia lógica del desarrollo de la humanidad.

Cómo citar este artículo:

Crovetto, A. (2021). Reseña del libro "éticas para máquinas" de José Ignacio Latorre. *Futuro Hoy*, 2(4), 36-37. <https://doi.org/10.52749/fh.v2i4.7>

Esta obra está bajo licencia internacional Creative Commons 4.0 Reconocimiento 4.0.