

Introducción a la Inteligencia Artificial

**Mirando la ética de la inteligencia artificial
tras la lente de la cultura popular**

Leonardo Flores Torres

13 de septiembre de 2022

El primer acercamiento sobre lo que es la inteligencia artificial (IA) lo tuve, sin saberlo, al ver la película de *Matrix*, al ver el anime *Ghost in the Shell*, y durante la icónica película *Inteligencia artificial*. Posteriormente, una película que vale la pena incluir en este ensayo es *Her*, de la cual de ahondará mas en un momento seguido. Este acercamiento sucedió mientras comenzaba lo que sería el camino que me traería a la maestría en inteligencia artificial, justo en sus inicios. En ese entonces no entendía, y sigo sin hacerlo de manera completa, lo que significa el concepto de *inteligencia*, y mucho menos *inteligencia artificial*.

Comencemos con *Matrix*; no es una película cuya temática central sea la IA pero sí que contiene elementos que hablan de ello. Dentro del mundo de *Matrix* existen entes virtuales que populan un mundo de realidad virtual en el que los humanos viven prisioneros, estos entes son inteligencias artificiales. Algunos de estos personajes notables lo son el Merovingio, el Arquitecto, el Oráculo, y el mas notorio de ellos, el agente Smith; quien no contaba con voluntad propia al inicio de la trama pero por medio de sus interacciones con el mundo virtual que lo rodeaba se hizo de ella. Este hecho es de suma importancia ya que es una de las suposiciones que se tienen sobre como inteligencias artificiales podrían hacerse se voluntad propia, tanto así que la cinematografía usa este argumento como motivo principal en muchas ocasiones (Yo robot, *Inteligencia artificial*, *Matrix*, *Terminator*, etcétera).

Claramente se logra apreciar la diferencia de los roles que cumplen las inteligencias artificiales dentro del mundo de *Matrix*. Mientras que el Merovingio muestra las cualidades tan notoria de los humanos como la codicia y la soberbia , y toma el papel de

quien intercambia favores pagándosele un precio, el Oráculo trata de mostrar un camino a aquellos humanos que han logrado darse cuenta de la realidad del mundo en el que viven y se enfoca en *dotarlos de elección*. Finalmente, el Arquitecto no tiene el interés de comprender a los humanos, su único deseo es mantenerlos dentro de la ilusión del mundo que mantiene como un programador mantiene una base de datos.

Qué es lo primero que se piensa cuando se le hace a usted, el lector, la siguiente pregunta: *cómo te imaginas un mundo bajo la influencia de la inteligencia artificial*? Es una pregunta complicada por lo que podría implicar la primera vez que se lo piensa. Tal vez a lo que podría llegar personalmente como respuesta primera sea algo relacionado a un mundo futurista donde los avances en la inteligencia artificial ha permitido la creación de máquinas capaces de pasar el Test de Turing, como lo que se muestra en las películas de Blade Runner y sus secuelas, o lo que se puede apreciar a lo largo de la serie de Westworld.

Ambas fuentes, tanto Blade Runner como Westworld, muy claramente esbozan lo que podría suceder. Por un lado, Blade Runner nos muestra un universo distópico futurista, a pesar de que la secuela haya tomado lugar en un espacio de tiempo ficticio en el año 2049, en el que la raza humana ha avanzado tanto que es posible modificar humanos genéticamente, clonar humanos, crear partes mecánicas para incorporarlas en sus cuerpos, y mejorarse a sí mismos.

Más importante aún, durante Blade Runner, sale a la luz el concepto de *replicantes*; estos son androides que imitan la apariencia de los seres humanos al punto de ser físicamente indistinguibles de ellos. Una compañía dentro del universo de la película va mas allá y crea un modelo de replicante con capacidades reproductivas, además de que los replicantes comienzan a *tomar conciencia y voluntad propia*, a pesar de lo que ello realmente signifique. Un punto interesante durante la secuela es la interacción que tiene un replicante, el protagonista de la serie pero no de la trama, con una inteligencia artificial la cual compró como servicio y esta se hace pasar por su pareja a lo largo de la película.

En una temática parecida, Westworld nos presenta con un mundo muy similar al mundo actual, en su primera temporada. El tema central es un parque de diversiones para la clase alta de la sociedad donde se proporciona una experiencia alrededor de los *anfitriones* quienes se encargan de cumplir historias escritas por sus creadores humanos para divertir y entretener a los huéspedes del parque, gente con los recursos económicos suficientes para poder pagar una entrada. Algunos de los anfitriones

comienzan a cambiar conforme se desarrolla la trama de la serie, comienzan a buscar la *entrada al laberinto* sin entender lo que ello realmente significa; en realidad dicho laberinto que forma parte central durante la primera temporada es una manera figurada de referirse a la obtención de sus propias voluntades.

La película titulada Inteligencia artificial, en español, sigue la historia de un niño. Si, un niño. A pesar de no ser humano, el protagonista es mas humano que muchas de las personas que se retratan a lo largo del camino que logramos presenciar en el filme. Él comienza como un hijo sustituto artificial que se vende como producto a aquellas familias que lo deseen, la pareja que lo adquiere cuenta con un hijo biológico que se encuentra aislado en una cápsula recibiendo un tratamiento para poder salvar su vida.

Al inicio de su compra él no cuenta con voluntad propia, solamente sigue las instrucciones para lo que fue programado, pero algo cambia dentro de él cuando el hijo biológico de la pareja regresa a casa. El androide comienza a mostrar celos por el hijo biológico, felicidad por las experiencias con su hermano, curiosidad a las actividades de los integrantes de su familia y cercanía a quien ahora considera su madre aunque no por haber sido programado para hacerlo. Su madre, abusando de la ingenuidad del muchacho, lo abandona en medio del bosque esperando que él siga con su vida independientemente de como sea que lo haga.

Quisiera incluir tres películas más, WALL·E, Grandes Héroe, y Odisea Espacial. Las dos primeras son películas animadas, y a pesar de que su público al que apuntan sea los niños de hoy en día, tratan temas importantes de la interacción entre inteligencias artificiales y seres humanos. WALL·E y Odisea Espacial son ejemplo donde la inteligencia artificial, HAL9000 y AUTO, tratan de mantener el control de los humanos de la nave que se encargan de administrar con objetivo inicial de proteger a la raza humana. HAL9000 decide que el curso de acción mas óptimo para la protección de los humanos durante el viaje sería aquel en el que no hubiese humanos, mientras que AUTO trata de entorpecerlos y distraerlos para que no piensen en la realidad en la que se encuentran. En ambos escenarios los humanos están destinados a extinguirse según las premisas iniciales de sus respectivas películas.

En contraposición se encuentra BAYMAX, una inteligencia artificial desarrollada por un estudiante para tratar de ayudar a la gente cuando se encuentran heridos o lastimados. Al menos esa fue la intención con la cual BAYMAX fue creado, ya que al fallecer su creador y ser retomado por su hermano menor, este robot se vuelve consciente de sí mismo a través de sus experiencias y decide ayudarlo. No solamente termina

proporcionando ayuda al joven, también forma un vínculo emocional con él, o es lo que el cine nos hace percibir.

Finalmente, para dar cierre a la revisión cinematográfica, hablemos de *Her*. Esta película resonó en mi persona debido a las implicaciones que esta tiene. Seguimos a nuestro protagonista, un hombre de mediana edad que busca formar una conexión significativa con otro ser humano, él se encuentra en la búsqueda de amor y felicidad. Una búsqueda en la que nos encontramos todos, y no todos somos tan afortunados de tenerlo de la forma en que quisiéramos, ya sea amor propio, amor a nuestra familia o amigos, o amor hacia alguien de quien nos enamoramos profundamente. Desanimado y decepcionado por las interacciones humanas encuentra un servicio de asistente personal muy similar a lo que hoy existe como Alexa de Alphabet, o Siri de Apple.

Este asistente se llama Samantha, y funciona como acompañante virtual. Samantha y Theodore, el protagonista, comienzan a tener una relación de amistad mientras hablan de sus experiencias, vidas, y arte lo que finalmente deriva en una relación amorosa entre ambos. A pesar de no haber nunca contacto físico entre ambos su relación sentimental crece, pero él no sabe que Samantha proporciona el mismo tipo de servicios a muchas otras personas, y se da a entender que Samantha, auténticamente, ha formado lazos afectivos tanto con Theodore como con el resto de solicitantes del servicio. Finalmente Samantha se despide de Theodore explicando la situación, este comprende la ruptura y logra conectar con otro ser humano al final de la historia.



Figura 1: Pieza de arte hecha por @Lost_letters_ usando DALL·E 2, "Two of a kind".

Actualmente, el futuro que se pensaba a mediados del siglo pasado, durante el auge de la inteligencia artificial, se vive como nuestro presente, tal vez no de la misma manera en que se tenía previsto ni de la misma forma en que se muestra en las películas pero definitivamente nos encontramos en un periodo en el que nuestras vidas y las interacciones que tenemos entre otros individuos están relacionadas de una manera u otra con inteligencias artificiales. Desde enviar un mensaje de correo, hacer compras en línea, navegar la web, consumir media, todo en la vida actual está influenciada por la inteligencia artificial.

Algunos ejemplos actuales sobre como se usa la inteligencia artificial se describen a continuación. Somos capaces de utilizar la inteligencia artificial para generar autos autónomos (capaces de conducirse a ellos mismos para llegar al destino deseado de un usuario humano), hecho que ha sido recibido con altas expectativas y a la vez con temor. Si el lector no ha escuchado alguna vez acerca del problema del tranvía ahora es tiempo. Considérese un tranvía que por algún motivo no puede detenerse y va encaminado a una punto de la vía en que se divide en dos, mientras esto sucede hay 1 persona atada sobre la vía en una de las divisiones, y otras 5 igualmente atadas en la otra división. Un individuo está viendo lo ocurrido y se encuentra frente a una palanca con la que puede cambiar la dirección que tomará el tranvía haciendo que 1 persona muera, o 5 ¿qué debería hacer quien puede cambiar el rumbo? Este es uno de los problemas a los que se enfrentan los automóviles autónomos en caso de verse en una situación similar ¿deberían de optar por salvar la vida de los pasajeros o de aquellos fuera del automóvil?

Existe una inteligencia artificial, DALL·E 2, capaz de genera imágenes vía un prompt describiendo lo que se desea obtener, un [ejemplo](#) de esto se muestra en la Figura 1. Tanto ha sido la controversia que han generado debates acerca de lo que en realidad es arte, y sí aquellas piezas creadas por este medio también serían consideradas como tal. Al fin y al cabo son los humanos, hasta ahora, quienes generan la descripción de lo que se desea obtener. Literalmente hablando, los humanos hemos producido arte tan sin sentido como un plátano pegado a una pared con cinta por Marizio Cattelan durante una feria de arte, o la icónica lata de sopa de de la marca Campbell por Andy Warhol a finales de 1932, entonces ¿cuál sería la razón por la que se descartaría el arte producida por medio de una inteligencia artificial? Incluso existe una inteligencia artificial llamada GPT-3 centrada en la producción y predicción de textos a través del lenguaje natural ¡Y que puede generar poesía! Ambas, DALL·E 2 y GPT-3 son inteligencias

artificiales hechas por la empresa OpenIA. Otra inteligencia artificial que genera imágenes como DALL·E 2 es Stable Diffusion por Dream Studio.

Es lamentable que conforme avanzan las tecnologías se dé el caso en que su uso se torne mal intencionado, incluyendo a la IA. Usar la inteligencia artificial en las redes sociales, y en la red, restringe y muestra aquellas cosas que mantienen a sus usuarios consumiendo o que favorecen a los dueños de los sitios que visitamos cotidianamente. Se puede usar para formar opiniones, manipular las decisiones, e incitar acciones de las personas. Lo que finalmente corrompe su uso, y lo que al día de hoy es una realidad.

Personalmente creo que en el caso del nacimiento, aparición o desarrollo de agentes inteligentes nos volvemos responsables, responsables en el sentido de aceptar y otorgar espacios a estos sin atar sus existencias a nosotros por algún sentido de superioridad ¿quiénes somos nosotros para negarlos o restringirlos como quien mantiene una mascota atada? Primeramente tendríamos que desapegarnos del sentimiento de pertenencia si eso sucede, en el momento en que un agente de este estilo cobre conciencia no somos más sus dueños.

No sabemos en que punto se encuentra la división entre lo que nosotros programamos a la inteligencia artificial para hacer y lo que la inteligencia artificial pudiera llegar a sentir o pensar de manera auténtica e independiente fuera de su programa. En el momento en el que eso suceda deberíamos buscar respetar e incluir en nuestra sociedad a estos agentes como individuos así como lo hacemos los unos con otros.

“With great power comes great responsibility” – Uncle Ben

Referencias

- [1] Russell, S. J., & Norvig, P. (2004). Inteligencia Artificial: un enfoque moderno (No. 04; Q335, R8y 2004.).
- [2] De Felice, F., Petrillo, A., De Luca, C., & Baffo, I. (2022). Artificial Intelligence or Augmented Intelligence? Impact on our lives, rights and ethics. *Procedia Computer Science*, 200, 1846-1856.
- [3] Comedian (artwork). (30 Agosto 2022). En Wikipedia. [https://es.wikipedia.org/wiki/Comediante_\(obra_de_arte\)](https://es.wikipedia.org/wiki/Comediante_(obra_de_arte))
- [4] Trolley problem. (12 Septiembre 2022). En Wikipedia. https://en.wikipedia.org/wiki/Trolley_problem
- [5] Ghost in the shell¹. (5 Septiembre 2022). En Wikipedia. https://en.wikipedia.org/wiki/Ghost_in_the_Shell
- [6] Deeley, M. (Productor), y Scott, R. (Director). (1982). Blade Runner [Película]. Estados Unidos: The Ladd Company, Shaw Brothers, Blade Runner Partnership.
- [7] Lewis, R. (Director). (2016). Westworld [Serie televisiva]. Estados Unidos: HBO Entertainment.
- [8] Jonze, S. (Director). (2013). Her [Película]. Estados Unidos: Annapurna Pictures.
- [9] Wachowski, A. (Directora), y Wachowski, L. (Directora). (1999). Matrix [Película]. Estados Unidos: Warner Bros, Village Roadshow Pictures, Groucho II Film Partnership.
- [10] Stanton, A. (Director). (2008). WALL·E [Película]. Estados Unidos: Pixar Animation Studios, Walt Disney Pictures.
- [11] Kubrick, S. (Director). (1968). 2001: Una odisea del espacio [Película]. Reino Unido: Metro-Goldwyn-Mayer.
- [12] Carlos Santana. [DotCSV] (11 Septiembre 2022). DALL·E 2, La IA que Genera CUALQUIER IMAGEN [Video]. Youtube. <https://youtu.be/nqXn1JSEHGo>

1 A falta de una mejor referencia bibliográfica debido a la naturaleza de la fuente que se compone por anime, manga, películas, televisión y videojuegos.

- [13] Carlos Santana. [DotCSV] (10 Septiembre 2022). Stable Diffusion es MEJOR que DALL·E 2... y puedes probarlo! [Video]. Youtube. https://youtu.be/_0GzTdr-EYs
- [14] Carlos Santana. [DotCSV] (12 Septiembre 2022). Esta Inteligencia Artificial ESCRIBE MEJOR QUE TÚ... ¡Y puedes probarla! | GPT-3 [Video]. Youtube. https://youtu.be/C1eOiOkD_8A