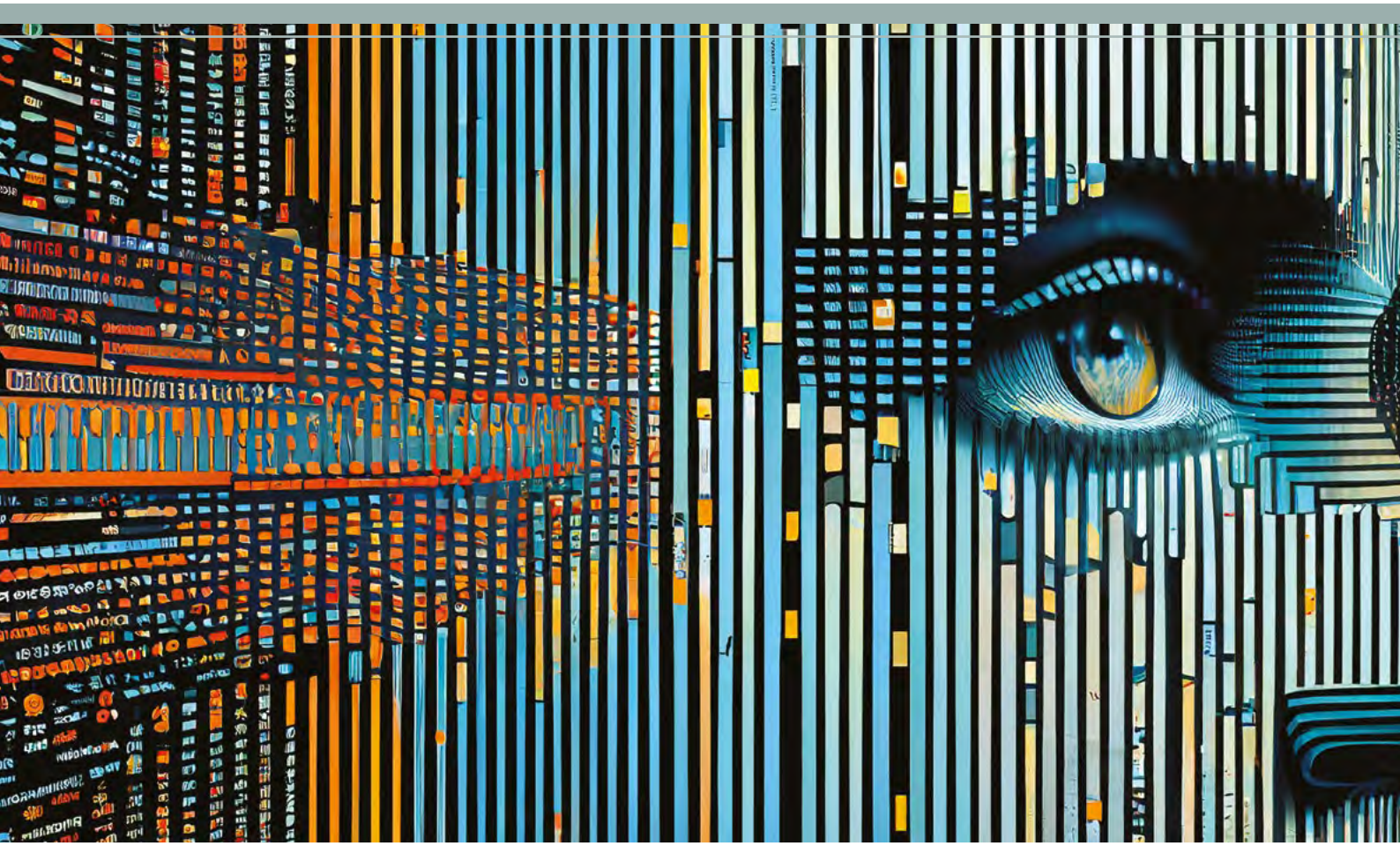

Star Trek y la inteligencia artificial: Apuntes para una política del futuro

Gonzalo Ordóñez Revelo

La inteligencia artificial (IA) recientemente ha cobrado relevancia por la diversidad de cambios que podría suscitar, al igual que por poner en evidencia aspectos de la vida, incluso éticos, acerca de los cuales ciertos científicos estaban reflexionando. Gonzalo Ordóñez Revelo, profesor del Área de Comunicación de la UASB-E, realiza una aproximación a este fenómeno emergente.

En la serie de ciencia ficción *Star Trek*, a mi juicio, una de las mejores del género, un androide conocido como el comandante Data es el tercer miembro en rango de la tripulación del Enterprise.

Data tiene un cerebro positrónico que le permite aprender y desarrollar una personalidad. El toque de humor filosófico proviene de su permanente intento por comprender el comportamiento humano hasta el punto de que en un capítulo planea su muerte bajo el entendido de que esto quizá sea la característica que nos hace más humanos.

El capítulo 11 de la cuarta temporada, «Un día en la vida de Data», me dejó la sensación de que el esfuerzo por descifrar la naturaleza humana, a pesar de los guionistas, se había invertido: éramos los espectadores los que estábamos intentando descifrar la humanidad de Data, quien afirmaba no comprender las emociones negativas como el odio, mientras que la amistad ahora ya formaba parte de su programación.

A partir de estas ideas desarrollo tres argumentos:

1. La racionalidad sin emociones no existe, por lo que el personaje en la vida real no podría evaluar los datos de la realidad social y tampoco tendría ninguna motivación para hacerlo.
2. La forma de entender, de sentido común, en la IA se conoce como *pensamiento por analogía* e impide pensar en un humanismo artificial paralelo.
3. Frente al pensamiento apocalíptico, que avizoraba el fin de los tiempos por la televisión, el internet, el *smartphone*, el algoritmo y ahora la IA, caben dos posibilidades: pensar en una nueva humanidad o dejar el futuro en manos de los populistas autoritarios.

Consideremos el argumento inicial. En este contexto, en primer lugar, «sin metas, el concepto mismo de inteligencia carece de sentido», nos dice Steven Pinker en *Cómo funciona la mente*. Y la meta más importante de Data es comprender las emociones humanas. En segundo lugar, somos conscientes, en diferentes niveles, de nuestros intereses, las decisiones a las que nos conducen y la información a la que prestamos atención para lograr nuestros objetivos. En tercer lugar, las sensaciones y los pensamientos van acompañados por un aroma emocional: son agradables o desagradables, interesantes o repelentes, excitantes o calmantes. Por último, un ejecutivo, el «Yo», aparece para efectuar las elecciones y mover las plantas del comportamiento, según Pinker.

Antonio Damasio, en su libro *En busca de Spinoza: Neurobiología de la emoción y los sentimientos*, demostró que las personas con daño en la corteza prefrontal son emocionalmente átonas, es decir, no pueden tomar decisiones racionales pues, aunque

> “Personas con daño en la corteza prefrontal son emocionalmente átonas, es decir, no pueden tomar decisiones racionales pues, aunque las capacidades cognitivas como la planificación, la abstracción, el lenguaje, entre otras, no sufran daño, no pueden generar opciones para solucionar problemas.”

las capacidades cognitivas como la planificación, la abstracción y el lenguaje, entre otras, no sufran daño, no pueden generar opciones para solucionar problemas. No son conscientes de las consecuencias de sus actos y se les hace imposible discernir en torno a los fines y la relación entre los acontecimientos y sus efectos sociales, o no cuentan con el razonamiento moral necesario para interpretarlos.

En este contexto, al ser emocionalmente átono, el comandante Data, básicamente sería irracional. Sin embargo, lo vemos tomar decisiones reuniendo una gran cantidad de información, pero en función de proteger a sus amigos o salvar una vida. Esto nos deja en la disyuntiva de si un androide, que podría valorar el entorno social y emocional, sería humano.

Ahora podemos continuar con el segundo argumento. Evolutivamente, el cerebro humano fue diseñado para comer, cortejar, combatir y correr; de aquí surgen nuestras motivaciones fundamentales. Esto lo ha reflexionado Vilayanur S. Ramachandran en su libro *Lo que el cerebro nos dice: Los misterios de la mente humana al descubierto*. Que tengamos sexo sin necesidad de reproducirnos no descarta el peso enorme que la sexualidad tiene en nuestras vidas. Nada peor que mantener una discusión con la pareja antes de comer, seguro termina mal; un estómago satisfecho es una persona feliz. La capacidad de sacrificio de los padres por sus hijos es evidente; la razón fundamental es que compartimos con ellos el 50 % de nuestros genes, es decir, protegemos la herencia que nos pertenece. Finalmente, el espíritu de sobrevivencia humano fue suficientemente documentado con el Holocausto nazi: seguimos vivos en las peores circunstancias.

¿Qué motiva a Data? Comprender la naturaleza humana, por supuesto, pero eso nos deja al principio, es decir, entender para qué. No le interesa la reproducción y la réplica de otros androides de su tipo no cuentan como hijos; algo así como que todos los automóviles de Ford fueran descendientes del primer modelo T de 1908. Puede defenderse con habilidad, huir de un peligro, pero no siente angustia si deja de alimentarse. En suma, le sobran motivos para no ser humano. Sin el sistema de recompensa del cerebro no existe la violencia, pero tampoco el amor, nos dice Pinker en otro de sus libros, *La tabla rasa: La negación moderna de la naturaleza humana*; y sin la felicidad, es difícil entender la tristeza. Una paradoja, sin emociones no hay pensamiento y con solo el pensamiento la vida no tiene sentido.

El problema más espinoso de la filosofía y la neurociencia del cerebro es que seamos autoconscientes. En la saga de las películas *Terminator*, la IA Skynet toma conciencia de sí misma cuando los humanos intentan apagarla. Es el momento en que las máquinas deciden que los humanos son una amenaza para su supervivencia y provocan una guerra nuclear, para exterminar a la mayor cantidad de la población.

El final de Data es épico: le pide al capitán Piccard que elimine su conciencia, para sentir la muerte, que es lo que otorga sentido a la vida.

“El problema más espinoso de la filosofía y la neurociencia del cerebro es que seamos autoconscientes.”

El problema de fondo de estos razonamientos es el supuesto, de sentido común, de que la conciencia está separada de las sensaciones, las emociones y del entorno; algo que ocurre solamente en el cerebro y no en el cuerpo.

La conciencia de la sed surgió como un fenómeno impulsado por un «interoceptor» (es decir, un sensor interno). Los mecanismos hipotalámicos y del cerebro medio para sentir los cambios de la concentración de solutos se hicieron más pertinentes para la supervivencia, y de manera espectacular, como lo fueron las intenciones adecuadas que promovían este objetivo. Esto ha sido estudiado y sintetizado por Derek Denton en su libro *El despertar de la consciencia: La neurociencia de las emociones primarias*.

El comandante Data evidentemente está dotado de infinidad de sensores, pero que no afectan su supervivencia fundamental, únicamente proveen de información que requiere

para cumplir con su programación. Incluso si sufriera un daño masivo, puede transferir su memoria con todas sus experiencias, aunque esto puede ser un problema porque lo humano no está localizado meramente en el cerebro, sino en el cuerpo en su totalidad, lo cual incluye su experiencia social y cultural.

Todo un espectro de efectos viscerales, cardiovasculares, respiratorios y endocrinos evocados por la emoción producen una avalancha de sensaciones que retroalimentan los procesos corticales y del cerebro basal originales que iniciaron la emoción. Denton dice que darse cuenta de las propias sensaciones se convierte en un elemento dominante en la amplificación del estado emocional.

Así que el asunto de la humanidad no se resuelve, para un androide, con incorporar sensores para registrar procesos internos, como un daño en la batería, o externos, como que una fuente de calor pueda hacer daño a la piel artificial; es la manera de ser humano la que también está en juego.

El problema de fondo es el pensamiento por analogía. Según Carl Sagan en *El mundo y sus demonios: La ciencia como una luz en la oscuridad*, suponemos que la IA, al ser nuestra creación, anhela lo que nosotros y creemos que se comportará como los humanos, ya sea para el bien o para el mal. Hay algo de cierto en este argumento, pero tiene un límite que explicamos a continuación.

> **Suponemos que la IA, al ser nuestra creación, anhela lo que nosotros y creemos que se comportará como los humanos, ya sea para el bien o para el mal.**

Antonio Quintana Carrandi, en una reseña que escribe para el blog *Sitio de ciencia ficción*, señala que Brent Spiner, el actor que interpreta al androide, comentó que Data sí tenía emociones, pero que él no lo sabía. Ello nos lleva a la pregunta: ¿Qué son emociones para el androide? Si pensamos en las tesis de Yuval Noah Harari de *Homo Deus: Breve historia del mañana*, el sesgo algorítmico vendría a ser la tendencia de un sistema de aprendizaje automático cuando refleja un aspecto a varios de la perspectiva cultural, política y social de sus creadores. En este contexto, es correcto pensar que el comandante Data tenga este interés humano, en la medida en que es un sesgo que proviene de su creador.

Pero el pensamiento por analogía también puede conducir al error: una IA autoconsciente, que aprende por sí misma y que posee la capacidad de sentir, bien puede desarrollar otro tipo de principios, así como sus emociones pueden corresponder a motivos radicalmente distintos. ¿Cómo amaría un androide sin los límites de la reproducción, el placer, el egoísmo, el machismo, los celos, el odio, el resentimiento? ¿Cuál sería la pasión de una IA sin el deseo?

La mayoría de los investigadores en este campo, como Max Tegmark, «estiman que la vía más rápida hacia la superinteligencia consiste en dejar de lado la emulación del cerebro y construirla de alguna otra manera (tras lo cual la emulación del cerebro podría seguir interesando o quizá no)». Esta es la apreciación de este cosmólogo en su libro *Vida 3.0*.

Una vez abordado lo hasta aquí discutido, podemos desarrollar el tercer argumento. Puede parecer extraño que el razonamiento incluya al populismo, pero recordemos que el desarrollo para una vacuna tarda años o incluso décadas. Sin embargo, la primera vacuna contra el COVID-19 estuvo lista el 11 de diciembre de 2021 y fue desarrollada por Pfizer-BioNTech, con la tecnología de ARN mensajero. Fue menospreciada por Donald Trump, Jair Bolsonaro, entre otros populistas autoritarios, a pesar de que el número de vidas que salvaron las vacunas fue muy superior al de las personas que murieron.

La innovación tecnológica constituye un hecho político por estar asociada con la reconstrucción social. Tegmark afirma que la continuación natural de esta tendencia pasa por usar nanorrobots, sistemas de biorretroalimentación inteligentes y otras tecnologías para sustituir, a principios de la década de 2030, los sistemas digestivo y endocrino, la sangre

y el corazón, para a continuación reemplazar el esqueleto, la piel y el cerebro durante las dos décadas siguientes.

En primer lugar, se encuentran los casos en los que la invención, el diseño o la disposición de un dispositivo o sistema técnico específico se convierte en una manera de resolver un tema relacionado con una comunidad en particular. Bien enfocados, estos ejemplos son bastante directos y fáciles de entender. En segundo lugar, se encuentran los casos que pueden denominarse de «tecnologías inherentemente políticas», sistemas hechos por el ser humano que parecen requerir o ser fuertemente compatibles con tipos particulares de relaciones políticas. Langdon Winner, en *La ballena y el reactor, una búsqueda de los límites en la era de la alta tecnología*, lo ha discutido ampliamente.

En este momento, dado este análisis, es posible una vía distinta, la del futuro surgimiento de una humanidad paralela de máquinas sentipensantes que puedan transformar la existencia humana, permitirnos viajar a Marte, resolver el problema de la contaminación, mejorar la gestión de las ciudades a niveles jamás pensados, extender la vida humana, ampliar la socialización y el intercambio de ideas, fuente fundamental de la paz humana, civilizatoria.

Por lo tanto, la tarea más importante no es estudiar los «efectos» e «impactos» del cambio técnico, sino evaluar las infraestructuras materiales y sociales que crean las tecnologías específicas para la actividad de nuestras vidas. Tomando las palabras de Winner, debemos tratar de imaginar y procurar construir regímenes técnicos que sean compatibles con la libertad, la justicia social y otros fines políticos claves.

Las posibilidades que Tegmark señala para la relación de la IA con la sociedad humana pueden ser la de un dios esclavizado; utopía libertaria, algo como Neo, el elegido, en la película *Matrix* (1999) de las hermanas Lana y Lilly Wachowski; dios protector, dictador benévolo, algo muy a tono con la tendencia humana hacia el pensamiento de grupo y cuidador de zoológico, que me recuerda al filme *El planeta de los simios* (1968), de Franklin Schaffner. En cualquier caso, la decisión tendría que estar en nuestras manos, pero antes deberíamos pensar en qué tipo de humanidad queremos, ahora que la tendencia es hacia una vida cíborg.

El problema de fondo entonces no es si la IA será tan humana como nosotros y, por lo tanto, con la misma capacidad de Skynet de destruirnos o, incluso, más humana, como el caso del comandante Data, sino si estaremos a la altura de los androides que sean mejores humanos que nosotros. La respuesta quizá provenga de lo que Stephen Hawking denominó *evolución autodiseñada*, que permitirá intervenir en la agresividad humana y mejorar la inteligencia y la resistencia a enfermedades, pero con la posibilidad de que se generen problemas políticos, irresolubles frente a los humanos no mejorados. Isaac Stanley-Becker lo manifiesta en su artículo «Stephen Hawking temía por una raza de "superhumanos" capaces de manipular su propio ADN».

Para el futuro, las sociedades tendrán que elegir entre continuar aliándose con los líderes populistas dispuestos a asegurar la tradición y valores de tiempos pasados o, por el contrario, tomar la decisión de afrontar las posibilidades y riesgos de lo que sigue (pero que ya está en curso): la fusión de las máquinas con las personas y la posibilidad de que la IA sea la que cree mejores IA.

A mi modo de ver las cosas, el futuro será el de dos humanidades: una nueva que nacerá de la Inteligencia Artificial General y la nuestra, pequeñas humanidades enfrentadas por diferencias artificiales, creadas por infinitas versiones de Hitler, siempre actualizadas, en varias versiones del populismo autoritario.