«Blush response»: empatía y tolerancia en sistemas de IA

CARLOS GARCÍA-TORRES

Con la popularización de los sistemas de IA cobran nueva actualidad muchas de las expresiones de la literatura de anticipación que tratan de las complicadas relaciones que surgen entre seres humanos y máquinas en el caso eventual -e improbable- de que estas últimas alcancen algún tipo de autoconciencia. En el ámbito de la ciencia ficción son muy conocidas las leyes de la robótica de Isaac Asimov y la peligrosa interacción con Hal 9000 en "2001: odisea en el espacio" de Arthur C. Clarke, así como otras muchas obras que juegan con la idea misma de una "inteligencia artificial" y que se remontan a una larga tradición que proviene del siglo XIX con Samuel Butler. Entre todas ellas tiene un lugar especial la novela "Do Androids Dream of Electric Sheeps?" (¿Sueñan los androides con ovejas eléctricas?) de Philip K. Dick que dio origen al icónico filme "Bladerunner". El título de este artículo proviene de una frase que se utiliza en la novela y en la película, <<blush response>>, respuesta de sonrojo, y que se refiere a un procedimiento útil para diferenciar a las máquinas mimetizadas como seres humanos de los seres humanos auténticos. Conviene aclarar, en primer lugar, que el libro original de 1968 difiere en muchas e importantes formas de la película de 1982. Entre las diferencias principales, de acuerdo con Julián Díez (2015), se encuentran la glorificación de los androides como seres superiores; la inserción de ciudades multitudinarias en la película, en tanto que el libro habla de un planeta despoblado; y, la eliminación de varios personajes como la esposa del protagonista y de un profeta religioso que tiene mucha relevancia en la trama y que, como veremos más adelante, tiene gran importancia en la noción de empatía que presenta la novela. Por mi parte agregaría que en la película se considera un elemento diferenciador muy significativo que se refiere a la angustia existencial de los androides con respecto a la extremada brevedad de sus vidas cuya caducidad está previamente programada.

Estos antecedentes nos llevan a reflexionar sobre los diversos problemas

C. García Torres (✉)
Universidad Técnica Particular de Loja, Ecuador
e-mail: cegarcia@utpl.edu.ec

Disputatio. Philosophical Research Bulletin
Vol.12, No. 26, Dec. 2023, pp. 231-244
ISSN: 2254-0601 | [SP] | **ARTÍCULO**

éticos que dimanan del constante desarrollo de los sistemas de Inteligencia Artificial (en adelante IA). Entre ellos, desde algunos sectores académicos, se plantea la idea de una futura "singularidad" que constituya una forma real de Inteligencia Artificial General, es decir que tenga la capacidad de ser indistinguible de la inteligencia humana. Algunos sistemas de Inteligencia Artificial Generativa han logrado ya dar la impresión de que se está conversando con otros seres humanos. Esta impresión desde luego es errónea y pone en juego a exámenes (como el que ideó Turing) que apelan a diversas formas de establecer la cercanía de un sistema de IA con lo que los seres humanos de forma intuitiva conocemos como inteligencia y, en ocasiones, con lo que conocemos como conciencia. Estos asuntos, aun siendo de la máxima importancia, constituyen caminos de pensamiento iniciales que llevan a otros puntos que se refieren a la reflexión sobre la propia vida humana y a las dimensiones públicas, es decir, ideológicas y políticas de la IA según veremos a continuación.

§ 1. La empatía como característica humana

Advirtamos desde el principio que el tema de la empatía no será tratado en este artículo exhaustivamente. Este tema tiene tantos enfoques y se ha desarrollado tanto que excede largamente las posibilidades de este artículo. Nuestro acercamiento dará preeminencia a determinadas visiones filosóficas y literarias que esperamos den luz sobre los puntos principales a los que nos limitaremos. Estos puntos se refieren a la tolerancia y a los sistemas de IA como expresión política.

El término "empatía" es una traducción de la palabra inglesa "empathy" que a su vez es una traducción de la palabra alemana "Einfühlung" acuñada en el siglo XIX por Theodor Lipps y que originalmente tenía una connotación estética referida a la habilidad humana para apreciar las virtudes de belleza de objetos inanimados. Lipps consideraba que un encuentro empático con objetos de arte detona procesos humanos que traen reminiscencias de movimiento corporal que luego llevan a proyectar las propias experiencias individuales sobre el objeto en cuestión. A partir de este punto se crean vínculos entre la percepción estética propia y la percepción de otras personas con mente propia, es decir, la experiencia de otro ser humano (Stueber, 2019). La palabra alemana literalmente significa "sentirse dentro". Esta visión original de la empatía tiene ciertamente relación con aproximación de Dick (2015) que en su novela describe un proceso empático bastante similar al propuesto por Lipps como comprobaremos más adelante. Algunos ejemplos de empatía que señala

Lipps tienen que ver con emociones expresadas a través de los gestos o de la posición del cuerpo y en la propia novela tienen relación con reacciones corporales inconscientes.

La aproximación inicial de Lipps ha dado origen a una serie de consideraciones psicológicas y filosóficas relacionadas con la importancia epistémica de la empatía, así como con la conciencia de la existencia de los otros (lo cual, como veremos más adelante, es más cercano a la idea de alteridad). Entre las corrientes que realizan algunos reparos a esta visión se encuentran principalmente los seguidores de la fenomenología que, pese a las objeciones que tienen, caracterizan la empatía como un concepto irreducible. La neurociencia, por su parte, se acerca a la explicación de Lipps y postula la existencia de ciertas "neuronas espejo" con lo cual pretende dar a la empatía un carácter biológico. Se ha postulado, además, la idea de que la empatía puede ser una aproximación metodológica a las ciencias sociales por su cercanía con la noción de "comprensión" lo cual, como es natural, ha levantado la enérgica oposición de la corriente hermenéutica que considera indispensable la consideración del contexto cultural (Stueber, 2019).

1.1. La empatía y las visiones sobre la fragilidad y finitud de la vida humana en la ficción.

El tema de la brevedad de la vida es una constante literaria y filosófica que, basada en el común e inevitable destino de todo ser viviente, bien puede considerarse como un punto de partida hacia una amplia comprensión humana. Desde los textos de las culturas más antiguas encontramos continuas referencias a la perplejidad que causa la oposición entre la riqueza y exuberancia de la vida frente a la constante presencia de la muerte. La misma aparición de la filosofía, en tanto reflexión organizada sobre la propia vida humana, tiene una gran deuda con la noción de finitud. Por supuesto resulta imposible hacer referencia a todos los textos que tratan este tema esencial, por ello, nos limitaremos a algunos textos fundamentales en el marco propuesto en el presente artículo.

Comenzaremos por el problema de la finitud y abordaremos la posición opuesta que se refiere a la eternidad. Jorge Luis Borges, en el cuento "El inmortal" realiza varias reflexiones sobre las posibilidades de una vida humana sin término. Comienza señalando que la inmortalidad es una característica que todos los animales comparten con la excepción de los seres humanos, dado que los entes irracionales ignoran la circunstancia de la muerte y, desde su limitado punto de vista, su existencia no tiene fin (Borges, 2011). De forma análoga podríamos pensar que las máquinas ignoran completamente el problema de la

muerte como en efecto ocurre con los actuales sistemas de IA. Esta ignorancia proviene de sus propias limitaciones en cuanto a la comprensión del contexto general de la existencia humana. En otras ficciones literarias este parece ser un tema de gran importancia. Recordemos, por ejemplo, el monólogo final de HAL 9000 en "2001: Odisea en el Espacio" (Clarke, 1970) tanto en la novela como en el filme tiene un papel central la desesperada argumentación que, en favor de su propia existencia, realiza la computadora. La eternidad de la vida humana también es rechazada por Jonathan Swift en sus famosos "Viajes de Gulliver" (Swift, 1967) en uno de cuyos capítulos se encuentran seres que no pueden morir pero que se encuentran en un estado muy lastimoso de deterioro de sus facultades con lo cual la eternidad no parece ser muy deseable. Desde estos puntos de vista podemos extraer dos características referidas al escaso tiempo de la vida humana, en primer lugar, el compartir la idea de un tiempo de vida limitado otorga una perspectiva humana única que constituye un punto importante de diferencia con máquinas y con animales; y, en segundo lugar, la comprensión de esta perspectiva humana puede ser un elemento importante para determinar si un sistema de IA puede alcanzar en algún momento capacidades empáticas cercanas a lo humano. Según se verá más adelante la idea la muerte, en el contexto filosófico, está muy relacionada con la noción de alteridad que tiene muchas coincidencias con los acercamientos al concepto de empatía que hemos reseñado.

En cambio, de manera general, la duración de la vida humana no tiene gran importancia para los personajes de "¿Sueñan los Androides con ovejas eléctricas?" dado que todos ellos han pasado por una reciente y perturbadora "guerra final" que ha dejado contaminada toda la tierra y ha matado a la mayor parte de su población. Con este antecedente la inminencia de la muerte o de la destrucción de la propia mente (lo cual pude considerarse como un equivalente) es un hecho aceptado y prontamente olvidado. Por el contrario, en la película "Bladerunner", los androides saben que han nacido con una fecha de caducidad determinada y tienen gran ansiedad por saber esa fecha, por alargar su vida y por aprovechar los breves años que les han sido concedidos. El dilema de la fragilidad y de la finitud de la vida humana pasa de esta manera desde los seres humanos auténticos a los entes artificiales. Este es uno de los principales puntos de ruptura entre la novela y el film. Para los androides de "Bladerunner" su motivación es la duración de la vida, para los de la novela el principal incentivo es la búsqueda de la empatía. En la novela existe una religión ampliamente difundida cuyo principal personaje es Wilbur Mercer. Mediante una "cajas de empatía" los seres humanos son capaces de sentir lo que Mercer siente al subir interminablemente una colina y recibir

golpes de piedra que lo lastiman cada vez más hasta que cae derrotado y vuelve a comenzar; al mismo tiempo los seres humanos son capaces de sentir lo que los otros individuos que se encuentran conectados a la máquina sienten. <<Y recuerdo que pensé cuanto mejores somos, cuánto mejoramos cuando estamos con Mercer. A pesar del dolor. Sufriendo dolor físico, pero espiritualmente unidos; he sentido a todos los demás, todos los que se fundían a la vez a lo largo de todo el mundo>> (Dick, 2015: 287). Se trata de una especie de empatía mediada por un amplificador tecnológico. Como se ha dicho los androides ambicionan sentir la empatía que los seres humanos sienten lo cual les está vedado.

1.2. ¿Decrece la empatía humana?

Según habíamos dicho en el acápite anterior uno de los puntos principales de la idea de singularidad y en las pruebas que se proponen para determinar si una máquina determinada es consciente (Lenharo, 2023) puede ser la noción de empatía. Es claro que un sistema de IA en determinadas condiciones puede hacerse pasar por un ser humano y aun vencer en el juego propuesto por Turing (Turing, 1950). De hecho, el ChatGPT o cualquier sistema de IA generativa puede dar una respuesta que parezca empática solo por el medio de aproximación estadística y en realidad es lo que estos sistemas hacen todo el tiempo. También sistemas de voz como Alexa o Siri simulan algunas formas de empatía. Pero puede decirse que lo propio sucede con los seres humanos según ha demostrado Peter Singer en una charla de Ted Talks del año 2013. Es un hecho no controvertido que emitimos respuestas de falsa empatía en muchas de las situaciones de la vida diaria. Surge entonces una pregunta, ¿se puede hablar de una decreciente capacidad empática en el propio género humano? La historia política de la humanidad, desde Hobbes, busca creer que los seres humanos mejoramos nuestra capacidad de empatía que, en el sentido político, podría asimilarse a la comprensión de las actitudes individuales y a la tolerancia de sus creencias. Es corriente suponer que desde un estado de barbarie hemos llegado a la civilización gracias a esas características empáticas que nos diferencian de los animales (al menos de algunos de ellos). Pero esta afirmación que parecía cierta entre la edad media y el siglo XIX, resultó muy difícil de aceptar a partir del siglo XX cuando aparecieron alarmantes brotes de lo que se llama intolerancia pero que bien se podría llamar “antipatía humana” y cuyo ejemplo más extremo es el nazismo y la persecución judía de la Segunda Guerra Mundial. A partir de entonces numerosas muestras de “antipatía” por nuestros congéneres han aparecido en tiempos y lugares varios. El esclavismo que se pensaba extinto regresa en peores formas y lo propio sucede con el

racismo y con el odio a los migrantes, o a las personas de diferentes orientaciones sexuales. Ciertamente el aumento de la intolerancia demuestra un decrecimiento de la empatía humana. Cada vez más, cuando vemos atrocidades en todo el mundo, somos menos capaces de sonrojarnos o de sentir vergüenza por nuestra propia especie. De esta forma bien podríamos decir que nos acercamos más a las máquinas y nos alejamos de lo humano. Los sistemas de IA, por su parte, tampoco se acercan a lo auténticamente humano.

§2. La tolerancia como expresión política de la empatía

Como ya habíamos adelantado en el acápite anterior, las diversas nociones de empatía de las que se ha hablado en el principio de este artículo y las diversas vertientes de conocimiento desde las cuales se abordan, bien pueden condensarse, en el aspecto político, en la noción de tolerancia. La idea de tolerancia ha tenido un largo desarrollo a lo largo de los siglos hasta el punto de que es posible afirmar que la historia política de occidente es una historia del desarrollo de la tolerancia. En efecto se trata de un concepto esencialmente occidental, aunque se encuentre latente tras las prescripciones de benevolencia hacia el prójimo de todas las grandes religiones. Como es bien sabido la tolerancia comenzó como un concepto aplicado a la religión. Las guerras religiosas europeas hicieron comprender la necesidad de tolerancia religiosa como forma mantener la paz y la prosperidad de los estados. Las grandes herejías de la edad media ocasionaron numerosas persecuciones y la reforma protestante llevó a la escisión definitiva del catolicismo con lo cual se crearon minorías religiosas en varios países, así, por ejemplo, en Inglaterra permaneció una minoría católica, en tanto que en España y Portugal la fe protestante se tornó minoritaria. Estos hechos llevaron a numerosos actos discriminatorios y a la marginación social y política de amplios grupos humanos. Diversos doctrinarios, entre los que se destacan Baruch Spinoza, Erasmo de Rotterdam, Pierre Bayle, Johannes Altusio y Cristian Thomasius, establecieron un cuerpo de prescripciones que demostró la necesidad de la tolerancia religiosa y que dio origen a la idea de tolerancia política. Rousseau (1901), por su parte, afirmaba en su "Contrato Social" que cada Estado debe tener una religión única que sirva como elemento cohesionador de todos los ciudadanos.

La idea general de la tolerancia alude a una actitud general ante la vida que considera que existe una característica esencial en los seres humanos y en sus creencias que los torna merecedores de respeto. En este sentido tiene directa relación con esas visiones sobre la empatía que hemos revisado en la primera parte de este artículo. Por todo esto reiteramos la idea de que la tolerancia es la expresión pública (y por tanto expresión política) de la empatía. Esta

afirmación, como se verá más adelante, es medular para el desarrollo de la idea de la IA como una oportunidad que concierne a toda la humanidad.

2.1. Las nociones de inteligencia y la tolerancia

La comprensión de nuestros conceptos sobre la inteligencia ha resultado un tema central para la definición de que funciones de la máquina pueden entenderse como "inteligencia artificial". En este artículo no revisaremos las diversas formas de aproximación a una definición de inteligencia. Nos limitaremos a proponer la idea de que la palabra "inteligencia" tiene una raíz latina en el verbo "intellegere" que libremente traducido significa "comprender". A partir de esta afirmación podemos retomar esa tradición filosófica que buscaba establecer la empatía como un método de conocimiento en las ciencias humanas y que la asociaba con la comprensión (Verstehen, understanding). Adhirieron a este método filósofos como Wilhelm Dilthey que señalaba que se puede explicar a la naturaleza, pero solo se puede comprender la vida del alma (Stueber, 2019) esto desde luego en concordancia con los postulados originales de la "Filosofía de la vida" cercanos a la importancia de la vivencia individual en la percepción e interpretación de la realidad. Según la interpretación que hace Lukács <<El gran descubrimiento de Dilthey reside, por tanto, en sostener que nuestra fe en la realidad del mundo exterior brota de la vivencia de la resistencia y de los obstáculos con que tropezamos en nuestras relaciones volitivas con las personas y las cosas del mundo exterior>> (Lukács, 1983: 339-340) Como se ve se trata de una posición que, pese a sus orígenes lógicos y psicológicos se encuentra aún muy cercana a cuestiones metafísicas pero que nos sirve para ilustrar el punto de la cercanía de los conceptos de inteligencia y de empatía. Establecida esta cercanía y habiendo ya determinado en un acápite anterior que la tolerancia es la expresión política de la empatía bien podemos inferir un nexo entre inteligencia y tolerancia. Y existiendo este nexo entre la inteligencia humana y la tolerancia podemos afirmar que es posible concebir la existencia de un nexo similar entre IA y tolerancia. Tanto es así que aparece implícitamente en los párrafos 66 y 67 de la Declaración de la UNESCO sobre Ética de la Inteligencia Artificial (UNESCO, 2021).

2.2. Los sistemas de IA como expresión política e ideológica

El carácter político de los sistemas de IA ya ha sido demostrado en la literatura científica y de manera especial en la corriente de crítica de la IA por sus relaciones, bastante evidentes, con los grandes capitales, con las tendencias imperialistas y con numerosos sesgos y discriminaciones que se acercan a la idea

de la "supremacía blanca" (Alí, 2022). La discusión sobre estos puntos tiene, como es natural, un importante trasfondo ideológico. Sobre estos aspectos cabe decir que el propio origen de los sistemas de IA en el seno de la gran industria informática norteamericana, así como su desarrollo desde los años 50 al amparo de grandes corporaciones y en algún caso de universidades, demuestra que no se trata de una tecnología inocua y carente de cualquier finalidad de dominación, por el contrario, sus inicios tienen connotaciones claramente militares y se encontraron confinados al contexto cultural y étnico anglosajón. Su desarrollo actual tiene directa relación con la inyección de enormes capitales que han hecho posible la ampliación de las capacidades físicas de computación que permiten la recogida y el manejo de datos con los cuales se alimentan los algoritmos. Es decir, sin un gran capital no hubieran sido posibles los sistemas de IA generativa que tienen tanta popularidad en este momento. Por tanto, resulta claro que los sistemas de IA actuales tienen vinculación con las estructuras de capital de los países del norte desarrollado. Resulta claro, así mismo, que sus desarrollos futuros estarán también ligados a los intereses de esas estructuras y que estos intereses están mediados por las circunstancias geopolíticas globales y por la ideología dominante entre las clases industriales de Estados Unidos y de algunos países europeos.

Hecha esta primera aproximación conviene entonces reflexionar sobre la importancia de la regulación de los sistemas de IA en un marco internacional comprometido con los derechos humanos y con los valores democráticos. Los instrumentos internacionales regulatorios emitidos por la Unión Europea y la declaración de la UNESCO señalan los derroteros deseables para el futuro de las aplicaciones de IA. Cosa diferente es que los desarrolladores adopten conscientemente esas normas y esas recomendaciones. Un hecho alentador es el reciente regreso de los Estados Unidos a la UNESCO y la consecuente voluntad política de apoyar las actividades de este organismo en cuanto a la regulación y direccionamiento de los desarrollos de la IA.

En la novela "¿Sueñan los Androides con Ovejas Eléctricas? Se prefiguran algunas de las implicaciones ideológicas de la IA. Cuando el héroe de la novela interpela al gran capitalista sobre los peligros del desarrollo de androides progresivamente más humanos surge el siguiente diálogo: <<señor Rosen. Nadie obliga a su organización a hacer evolucionar los robots humanoides hasta el punto en el que….

> Producimos lo que quieren los colonos, nuestros clientes -dijo Rosen- Seguimos el viejo principio subyacente a cualquier aventura comercial. Si nuestra firma no hubiera fabricado estos modelos progresivamente más humanos, otros lo habrían hecho>> (Dick, 2015: 176-177).

El desarrollo de los sistemas de la IA a la luz de las consideraciones de las empresas que los crean tiene finalidades comerciales que se insertan en la visión liberal de la economía, en la preeminencia de la oferta y la demanda y en la confianza en la regulación del mercado. Este último punto, sin embargo, constituye tal vez uno de los peligros mayores para el futuro de la IA y para las propias consecuencias que estas tecnologías pueden tener para la humanidad. Especialmente a la vista de un mundo que mantiene seculares desigualdades y niveles de pobreza y de marginación que pueden ser mejorados o empeorados por el uso de los sistemas de IA. De forma que se puede decir que el verdadero peligro no está en una futura "singularidad" que sojuzgue a los seres humanos sino en las enormes injusticias que sin necesidad de la existencia de la IA ya existen en todo el planeta. Es precisamente en este punto que la tolerancia, como un factor esencial de la democracia juega un papel principal en las consideraciones éticas sobre la IA como lo veremos en los siguientes acápites.

§3. Empatía humana y alteridad

La noción misma de la empatía se encuentra directamente relacionada con nuestra percepción de la existencia de los otros, así como de sus perspectivas y de sus necesidades, lo que muchas veces se ha llamado "alteridad".

La idea de alteridad tiene una muy respetable trayectoria filosófica en la cual destaca Emmanuel Lévinas. Este pensador, sobre este tópico, tiene formulaciones muy amplias y complejas que se relacionan más con aspectos de la trascendencia humana, a veces con la idea de una revelación de carácter metafísico y, a la vez, como fundamento inicial del propio andamiaje con el que construye sus formulaciones filosóficas (Navia, 2023). Como decíamos la noción de alteridad tiene antecedentes ilustres en Husserl y en Heidegger. Resulta claro que para los efectos de este artículo bastaría tomar la idea de alteridad como la noción básica de la "conciencia del otro" pero, aún dentro de esta formulación muy limitada la palabra "conciencia" puede llevar a toda una serie de reflexiones problemáticas sobre los propios sistemas de IA. Bástenos decir, por ahora, que de acuerdo con cierta formulación de Lévinas, tomada de Heidegger y citada por Navia <<en Ser y tiempo se encontraba ya la preocupación-de-ser del estar-ahí humano y la preocupación por el otro hombre; que uno a otro, se solicitaban; que el estar ahí en su autenticidad era un estar-junto-a-otro y un ser-para-otro>> (Navia, 2023: 106). Esta cita aclara un concepto de alteridad que me parece muy importante y que sobrepasa nuestra simple formulación inicial. También convendría resaltar que aparte de esta formulación existe en el mismo autor una aproximación a la alteridad a través de la conciencia de la muerte y a la apropiación de la muerte de los otros. Esta

última idea también parte de Heidegger (Navia, 2023) y resulta relevante para este artículo por su cercanía con las ideas que presenta Dick en "¿Sueñan los androides con ovejas eléctricas?". Lamentablemente por las propias limitaciones de este artículo debemos omitir esta y otras relaciones de la idea de alteridad entre ellas la que se refiere a la conciencia del otro y el tiempo.

Dicho todo esto es importante aclarar las relaciones de los conceptos que hasta ahora se han expuesto. Diremos entonces que podemos considerar la alteridad como el paso inicial que pude llevar a la empatía (respetando las grandes diferencias de contexto que median entre estas dos nociones); que, como ya habíamos señalado, se puede considerar a la tolerancia como la forma política de la empatía; y que, la tolerancia resulta ser una condición inevitable de cualquier concepción de democracia. En resumen, afirmamos que existe un camino directo entre la alteridad, la empatía, la tolerancia y la democracia. Un camino que podríamos entender como de puertas sucesivamente más estrechas (según la conocida metáfora bíblica) que conducen a una concepción general de democracia. Esta digresión cobrará mayor sentido en los siguientes acápites en donde se aclara su relación con el tema general del artículo.

3.1. La intolerancia como característica de la máquina

Es claro que siendo la tolerancia una característica esencialmente humana su ausencia, es decir la intolerancia, puede atribuirse fácilmente a una máquina cualquiera. Si buscamos razones para esta atribución encontramos enseguida que la principal es la inexistencia de algo que pueda llamarse "conciencia" en cualquier tipo de entidad no humana. Con esto entramos de nuevo en el delicado problema de determinar posibilidades conscientes en sistemas de IA. Por lo pronto los caminos de investigación que se han tomado no parecen estar cerca de lograr algo parecido a una entidad consciente, con lo cual queda cerrado, por ahora, el debate sobre posibilidades de empatía y tolerancia en sistemas de IA. De tal forma que se puede pensar que la máquina en tanto cumplidor eficiente de su programación u operador inconsciente de algoritmos probabilísticos carece por completo de la capacidad de entender los conceptos de tolerancia esenciales a los sistemas democráticos y a los derechos humanos. Tolerar, en último término, significa poder comprender las diferencias humanas, y esta capacidad para los sistemas de IA generativa, por ejemplo, sólo se desarrolla de manera muy superficial dado que las condiciones generales del entrenamiento de sus algoritmos obedecen a un contexto cultural y político claramente inserto en formas de capitalismo muy cercanas a los principios egoístas del estado de naturaleza de Hobbes o a los intereses humanos que postuló Locke y más cercanos todavía al poder del mercado de Adam Smith. Es

decir, un contexto que ignora los desarrollos políticos de los últimos tres siglos y entre estos desarrollos ignora, sobre todo, el principal producto político humano y la principal expresión de la tolerancia que es la Declaración Universal de los Derechos Humanos. Por todas estas razones la Declaración de la UNESCO sobre Ética de la Inteligencia Artificial hace énfasis en la presencia de un responsable humano de cualquier posible daño que pueda ser imputado a un sistema de IA (UNESCO, 2021).

3.2 La intolerancia como valor político actual

La tolerancia política y religiosa ha sido uno de los valores y actitudes fundamentales que, a lo largo de cinco siglos han sostenido las nociones de democracia y el propio orden geopolítico mundial. En el siglo XX se daba por sentado que la tolerancia mantendría ese papel central y que no se podría regresar a las etapas de la historia humana en donde era notoria su ausencia. Las razones para esta confianza compartida se apoyaban en las terribles experiencias históricas que la intolerancia trajo a lo largo de los siglos especialmente los genocidios de la segunda guerra mundial y los posteriores.

Sorprendentemente con el avance de las comunicaciones y la irrupción de las redes sociales las posiciones políticas intolerantes (relacionadas especialmente con el racismo, el nacionalismo, la homofobia, el machismo, etc.) que durante muchos años fueron claramente minoritarias encontraron nuevos canales de expresión y numerosos seguidores. A partir de aquí los partidos políticos que fomentan estas tendencias progresivamente logran mayores ventajas políticas habiendo alcanzado el poder en algunos países y estando muy cerca de alcanzarlo en otros. Por eso se puede afirmar que, paradójicamente, la intolerancia ha alcanzado el carácter de valor político en el siglo XXI. Una parte importante de este auge está dado por los desarrollos tecnológicos que han permitido la aparición de las redes sociales y la consiguiente difusión masiva de información que, en el caso de estos grupos, es mayoritariamente falsa. Este tipo de información construye una realidad ilusoria que eufemísticamente se ha llamado "posverdad". Para los constructos de esta "posverdad" son muy útiles las capacidades de muchos de los sistemas de IA generadores de texto, imagen y voz.

A la par se ha dado un resurgimiento de la intolerancia religiosa desde y hacia los países que profesan las formas más ortodoxas del islam. Lo cual constituye un punto más de alarma y de posibilidad de violencia generalizada.

Habíamos dicho antes que parece haber un decrecimiento general de la empatía humana y resulta que a ese decrecimiento se agrega el difundido desprestigio político de la tolerancia y de su consecuencia principal que son los

derechos humanos. La difusión de estas posiciones constituye un peligro general para la humanidad. Este peligro se potencia si es que el entrenamiento de los algoritmos sufre una contaminación en este sentido, como parece ya haber sucedido.

En el fondo de nuestro mismo concepto de civilización se encuentra la noción de empatía humana, que de todo lo dicho puede resumirse como la posibilidad de comprender al otro. Esta comprensión, según se ha dicho da origen a la palabra "inteligencia" a través del verbo latino "intelligere". Sin esta capacidad de comprensión cualquier tipo de inteligencia, incluyendo la humana, no deja de ser una inteligencia artificial.

REFERENCIAS

Alí S. M. (2022) "Yarden Katz. Artificial Whiteness: Politics and Ideology in Artificial Intelligence" Book Review. Kalfou. Vol 9. No. 2.

Borges, J. L. (2011). Cuentos completos. España: Penguin Random House Grupo Editorial España.

Clarke A. (1970) 2001 Una Odisea en el Espacio. Biblioteca Básica Salvat. Barcelona.

Díez J. (2015) "Introducción" en Dick P. ¿Sueñan los Androides con Ovejas Eléctricas? Cátedra. Madrid.

Dick P.K. (1982) Do the Androids Dream on Electric Sheep? Ballantine Books.

Dick P.K. (2015) ¿Sueñan los Androides con Ovejas Eléctricas? Cátedra. Madrid.

Lenharo M. (2023) If AI becomes conscious: here´s how researchers will know. Nature 2023 Aug 24. doi: 10.1038/d41586-023-02684-5. Online ahead of print.

Lukács G. (1983) El Asalto a la Razón. Grijalbo. México.

Navia M. (2023) La alteridad absoluta de la muerte en Emmanuel Lévinas. Anuario Filosófico. Vol. 56. No. 1. 101-123.

Rousseau J.J. (1901) "The Social Contract" en Ideal Empires and Republics. Universal Classical Library. M. Walter Dunne Publisher. Washington & London.

Stueber K. (2019) "Empathy" en Stanford Enciclopedia of Philosophy disponible en https://plato.stanford.edu/entries/empathy/

Swift J. (1967) Los Viajes de Gulliver. John W, Clute Editor. México.

Turing A. M. (1950) Computing Machinery and Intelligence. Mind, New Series. Vol 59. No. 236. 433-460.

UNESCO (2021). Recomendación sobre la ética de la inteligencia artificial. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa.

Blush response: empathy and tolerance in AI systems

Starting from the idea of using empathy as a useful way to identify an AI system indistinguishable from human beings, postulated in the novel "Do Androids Dream of Electric Sheep?" and in the film

Bladerunner, some of the philosophical and ethical problems related to empathy and otherness raised by generative AI systems and those that a possible general AI system may raise, as well as their relationship with notions of tolerance, and democracy are reviewed. It is also postulated that the development of AI systems entails certain ideological and political aspects that constitute ethical problems, including the gradual decrease in empathy and the global political rise of intolerance.
Keywords: Empathy · Tolerance · Otherness · Artificial Intelligence.

Blush response: empatía y tolerancia en sistemas de IA

A partir de la idea del uso de la empatía como un modo útil para identificar un sistema de IA indistinguible de los seres humanos, postulada en la novela "¿Sueñan los androides con ovejas eléctricas?" y en la película Bladerunner, se revisan algunos de los problemas filosóficos y éticos relacionados con la empatía y la alteridad que plantean los sistemas de IA generativa y los que puede plantear un posible sistema de IA general, así como su relación con las nociones de tolerancia y de democracia. Se postula además que el desarrollo de los sistemas de IA conlleva determinados aspectos ideológicos y políticos que constituyen problemas éticos, entre ellos el gradual decrecimiento de la empatía y el auge político global de la intolerancia.
Palabras Clave: Empatía · Tolerancia · Alteridad · Inteligencia Artificial.

CARLOS GARCÍA TORRES es Profesor Titular en el Departamento de Derecho y Coordinador de la Cátedra UNESCO de Ética y Sociedad en la Educación Superior de la Universidad Técnica Particular de Loja, Ecuador. Es Doctor en Derecho y Ciencias Sociales por la Universidad Nacional de Educación a Distancia, España. Sus intereses de investigación se concentran en la bioética, la ética, la filosofía del derecho y el derecho romano. Es autor de obras como: Derecho romano: una revisión sumaria (Dykinson, 2011); Derecho Romano (UTPL, 2020) o Sociología Jurídica (UTPL, 2020). **CONTACTO**: Cátedra UNESCO de Ética y Sociedad en la Educación Superior, Universidad Técnica Particular de Loja, Calle Marcelino Champagnat s/n, 110107 San Cayetano Alto, Loja, Ecuador. e–mail (✉): cegarcia@utpl.edu.ec — iD: https://orcid.org/0000-0003-1170-6765.

HISTORIA DEL ARTÍCULO | ARTICLE HISTORY
Received: *8–March–2023*; Accepted: *17–November–2023*; Published Online: *30–December–2023*

COMO CITAR ESTE ARTÍCULO | HOW TO CITE THIS ARTICLE
García Torres, Carlos (2023). «"Blush response": empatía y tolerancia en sistemas de IA». *Disputatio. Philosophical Research Bulletin* 12, no. 26: pp. 231–244.